Local Convexity Inspired Low-complexity Noncoherent Signal Detector for Nano-scale Molecular Communications

Bin Li, Mengwei Sun, Siyi Wang, Weisi Guo, Chenglin Zhao

Abstract

Molecular communications via Diffusion (MCvD) represents a relatively new area of wireless data transfer with especially attractive characteristics for nano-scale applications. Due to the nature of diffusive propagation, one of the key challenges is to mitigate inter-symbol interference (ISI) that results from the long tail of channel response. Traditional coherent detectors rely on accurate channel estimations and incur a high computational complexity. Both of these constraints make coherent detection unrealistic for MCvD systems. In this paper, we propose a low-complexity and non-coherent signal detector, which exploits essentially the local convexity of the diffusive channel response. A threshold estimation mechanism is proposed to detect signals blindly, which can also adapt to channel variations. Compared to other non-coherent detectors, the proposed algorithm is capable of operating at high data rates and suppressing ISI from a large number of previous symbols. Numerical results demonstrate that not only is the ISI effectively suppressed, but the complexity is also reduced by only requiring summation operations. As a result, the proposed non-coherent scheme will provide the necessary potential to low-complexity molecular communications, especially for nano-scale applications with a limited computation and energy budget.

Index Terms

Molecular communications; non-coherent detector; low-complexity; inter-symbol-interference; local convexity

Bin Li, Mengwei Sun and Chenglin Zhao are with the School of Information and Communication Engineering (SICE), Beijing University of Posts and Telecommunications (BUPT), Beijing, 100876 China. (Email: stonebupt@gmail.com)

Siyi Wang is with the Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province 215123, China. (E-mail: siyi.wang@xjtlu.edu.cn)

Weisi Guo is with the School of Engineering, The University of Warwick, Coventry, CV4 7AL, UK. (E-mail: weisi.guo@warwick.ac.uk)

Manuscript received April 5th, 2015.

I. INTRODUCTION

FOR more than a century, humanity has relied on wave-based communications as the primary means of carrying information without a tether. As we move forward, we have found new operating environments and new dimensions that require wireless communications. For example, in nano-medicine, we may need to communicate across micro- and nano-scale distances, using low-energy and -complexity nano-machines. In this and many other challenging environments, wave-based communications may not always be best suited to transferring data.

A. Development Background

Molecular communications, a biologically-inspired concept, involves the transfer of information via a population of molecular messengers [1]–[6]. The data can be embedded onto the various chemical and physical properties of the molecular messengers [7], [8]. For example, in nature, molecular communications is used across macro- (e.g. communication between moths can be up to a few kilo-metres) and micro-distance scales (e.g. quorum sensing). Generally, MCvD relies on the diffusion process [9], and in some cases, this can be accelerated by ambient or assisted flow (e.g., convection flow) [10]. In terms of application, a variety of envisaged scenarios have been proposed in the IEEE standardization document (IEEE P1906.1-2015), one of which is in the promising area of nano-medicine: in-body targeted drug delivery or surgery [11].

Recently, one of the first experimental test-bed of molecular communications has been developed and reported in [12]. It consists of a transmitter that emits short pulses of molecules bearing the data using the binary concentration shift keying (B-CSK) modulation. The channel is a random-walk based diffusion channel with optional induced drift. The receiver is a chemical detector that will covert chemical concentration into an electrical signal. It was demonstrated that generic text information can be reliably conveyed via the MCvD process, and the communication distance may even approach several meters in the unbounded free-space.

B. Coherent and Non-Coherent Receiver Design

Molecular communication channels are significantly more stochastic than alternative wireless channels [13], [14]. The expected characteristic at the receiver (i.e., magnitude of the peak response) of even a large number of molecules undergoing random walk can be unpredictable for the following reasons. In

a fixed distance molecular channel, the diffusivity parameter (rate of diffusion) may vary as a function of time due to changes in the channel temperature and external disturbances (i.e., convection current). In a mobile molecular channel, small changes in the distance may also result in significant variations in the expected characteristic at the receiver. This has been demonstrated in a recent paper by the authors where small physical perturbations were added to the transmitter [15]. The observed channel frequency response was significantly shifted, indicating that micro-disturbances in the channel or the transmitter / receiver can cause significant changes to the channel response. Therefore, whilst traditional detection techniques can achieve a good performance with channel information, in reality estimating the channel response accurately may prove to be extremely difficult.

In order to mitigate the inter-symbol interference (ISI), a number of coherent detection schemes, which utilize the knowledge of the channel response, have been widely studied [16]–[19], e.g., the maximum *a posteriori* (MAP) detector and the minimise mean square error (MMSE) methods. After accurately acquiring the channel response, a decision-feedback equaliser (DFE) has been proposed [16]. In addition to the aforementioned post-equalisation schemes, a pre-equalisation scheme has also been investigated [20]. Whilst the existing coherent methods can mitigate ISI, they require the accurate channel state information (CSI) (i.e. the equivalent diffusion channel response), as well as complex digital signal processing (DSP) implementations.

In contrast to coherent detection, a non-coherent scheme can avoid the challenge of acquiring or estimating the channel response and the complex DSP operations associated with coherent methods. In our previous work, a simple non-coherent detection scheme is proposed for molecular communications [21]. The non-coherent method essentially exploits the long-tail property of diffusion channel and utilizes the *concentration difference* among two adjacent slots to detect the unknown signal. When the symbol duration is greatly larger than the diffusion time of peak concentration, the suggested difference-based non-coherent detector could obtain the promising performance with a simple implementation. However, for other applications where the symbol duration is relatively small, then its performance will degrade significantly, as the difference operation can not effectively eliminate the self-interference from the previous slots. So, this difference-based scheme is vulnerable to self-interference and its application is thereby limited only to some low-rate applications.

Inspired by the *time-domain* local convexity feature of molecular concentration in each slot, in this study we propose a new non-coherent signal detection scheme, which will be more robust to the self-

interference and hence has broader applications. Irrespective of fluctuations in the channel conditions (i.e. diffusivity), the CSI has certain features that can be exploited at the receiver side. By exploiting the local time-domain convex shape of the channel response, a simple yet effective metric is constructed, which is then utilised to judge whether a new molecule transmission has arrived in the current interval. In order to detect signals blindly, an adaptive threshold is also designed, which can adapt to channel changes based on the received signals. The major innovation of our new scheme is that, by excluding the complicated matrix multiplications in MAP and MMSE methods, the proposed scheme's DSP implementation involves only a number of summations. As shown by the numerical analysis, even without a channel estimator or accurate CSI, the proposed non-coherent scheme may achieve a promising detection performance and is robust to channel variations.

The rest of the article is structured as follows. In Section II, the systems model is introduced and existing coherent detection algorithms are briefly reviewed. A new non-coherent detection algorithm is then proposed in Section III, including the construction of convex metrics and the designing of blind decision threshold. Simulation results are provided in Section IV, and we conclude the whole investigation in Section V.

II. System Model & Review of Coherent Detection

In the section, we first elaborate on the signal model of MCvD systems studied in the literature [16]. This is followed by a short review of existing MAP and MMSE coherent detection schemes.

A. Channel and Noise Model

A model of the physical-link MCvD channel is illustrated by Fig.1. At the beginning time of each symbol T_b, the molecular transmitter emits one short pulse with the duration of T_p. The on-off keying (OOK) line code is used [21], [22], i.e., the binary information symbol is $\alpha_k \in \mathcal{A} = \{0, 1\}$, $k = 0, 1, \cdots, K$. Each emitted pulse with a duration T_p consists of a number of molecules A. After propagating through the diffusion channel $h(t)$, the number of molecules received at the receiver is:

$$\begin{aligned} r(t) &= A \times \sum_{k=0}^{K} \alpha_k \texttt{rect}\left(\frac{t - T_p/2}{T_p} - kT_b\right) \otimes h(t) + w(t), \\ &= A \times \sum_{k=0}^{K} \alpha_k y(t - kT_b) + w(t). \end{aligned} \tag{1}$$

Here, T_p is the duration each emitted pulse and A is the release rate. $K = \lfloor t/T_b \rfloor$ denotes the total number of symbols until time t; a rectangular pulse shaper is adopted by the transmitter, which has amplitude 1. The notation $\otimes$ denotes the convolution, and the channel impulse response at the receiver is $y(t) \triangleq \texttt{rect}\left[(t - T_p/2)/T_p - kT_b\right] \otimes h(t)$ [21]. Usually, the symbol period is much larger than the emission duration, i.e. $T_p \ll T_b$. E.g., in [12] T_b is 3s whilst T_p is configured only to 100 ms. This is because a large T_p will further expand the channel response (i.e. $y(t)$) and result in more serious self-interference. The parameter $w(t)$ represents the additive noise introduced by the imperfect counting process, e.g. the receiver may have a low-sensitivity chemical-to-electrical converter (i.e. leading to an unknown fluctuation in the number of molecules received), especially when the nano-scale low-complexity sensors are required.

In practice, there are two kinds of noise sources in MCvD, both of which are modeled as memoryless additive noise. It is claimed in [23] that the additive noise will be aroused either by the discreteness of molecule numbers (*physical-sampling* noise), or the imperfection of a receptor (e.g. when counting the number of molecules, known as *physical-counting* noise). As the physical-sampling noise in the context of the discrete binary concentration modulation would be negligible, here we focus only on the counting noise, which is assumed to be independent and identically distributed (i.i.d) zero-mean white Gaussian processes with a variance of σ_w^2, i.e. $w(t) \sim \mathcal{N}(0, \sigma_w^2)$. It is noted that, although the Gaussian approximation is used for the convenience of analysis [16], the designed scheme is independent of the statistical distributions of counting noises, as the noise-dependent likelihood function will not be required by the proposed non-coherent detector.

In general, the channel impulse response of a 3-D unbounded diffusion channel can be characterised by:

$$h(t) = \frac{1}{(4\pi D t)^{3/2}} \exp\left(-\frac{|d|^2}{4Dt}\right), \tag{2}$$

where $D \sim 10^{-9}\text{m}^2/\text{s}$ is the diffusion coefficient; d is the Euclidean distance between a transmitter and a receiver [9], [24], [25].

B. Previous Works

As in previous works, one normally assumes that the synchronisation of the molecular communication link has already been accomplished [26]. In existing coherent detection methods, the receiver samples the

molecular concentration at a specific rate T_b [1], and the resulting discrete signal is:

$$r_k = \sum_{l=0}^{K} \alpha_l y_{k-l} + w_k, \tag{3}$$

where $r_k \triangleq r(kT_b)$, $y_k \triangleq y[(k+1)T_b]$ and $w_k \triangleq w(kT_b)$. For convenience, the received signal will be further expressed as $\mathbf{r}_{K\times 1} = \mathbf{Y}\boldsymbol{\alpha} + \mathbf{w}$. If we define the number of previous symbols that contribute towards ISI as I, then $\mathbf{r}_{K\times 1} \triangleq [r_k, r_{k-1}, \ldots, r_{k-K+1}]^T$, where K is the block length of information bits; $\boldsymbol{\alpha}_{(K+I)\times 1} \triangleq [\alpha_k, \alpha_{k-1}, \ldots, \alpha_{k-K-I+1}]^T$; $\mathbf{Y}_{K\times(K+I)}$ is the *circulant* channel matrix constructed from the equivalent channel response $\mathbf{y}_{I\times 1} \triangleq [y_0, y_1, \cdots, y_I]^T$, and $\mathbf{w}_{K\times 1}$ is additive noise vector. The channel matrix $\mathbf{Y}_{K\times(K+I)}$ is expressed as [21]:

$$\mathbf{Y} = \begin{bmatrix} y_0 & y_1 & \cdots & y_I & 0 & \cdots & 0 \\ 0 & y_0 & y_1 & \cdots & y_I & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & y_0 & y_1 & \cdots & y_I \end{bmatrix}.$$

Here, we assume the channel response y_k will keep invariant in a transmission block of the length K. Taking the kth received sample for example, it will be expressed as $r_k = a_k y_0 + \underbrace{a_{k-1}y_1 + \cdots + a_{k-I}y_I}_{\text{ISI}}$, which contains the ISI from the previous I intervals [2].

1) MAP: For the coherent MAP schemes, the accurate estimation of channel response should be acquired first. Thus, relying on the posterior distribution, the unknown symbols are recovered via:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{MAP} &= \arg\max_{\boldsymbol{\alpha}\in\mathcal{A}^K} \Pr(\boldsymbol{\alpha}|\mathbf{r}, \mathbf{Y}), \\ &= \arg\max_{\boldsymbol{\alpha}\in\mathcal{A}^K} \prod_{k=0}^{K} \Pr(\alpha_k|\alpha_{k-1}) \prod_{k=0}^{K} \Pr(r_k|r_{0:k-1}, \alpha_{0:k}), \end{aligned} \tag{4}$$

where the likelihood densities $\Pr(r_k|r_{0:k-1}, \alpha_{0:k})$ follow the Gaussian distribution, as suggested by [16]. Based on the trellis search [19], a sequential estimation strategy will be used to solve Eq. (4). For the i.i.d information source $\alpha_k \in \mathcal{A}$ with the equal priors for "0" and "1", the MAP scheme is equivalent to a maximum likelihood (ML) method. Note that, for the non-Gaussian noises [23], the corresponding likelihood density will be required when performing the Bayesian estimation.

[1]The sampling time may be smaller than the symbol duration for coherent detection [17]. As will be seen later, for the considered non-coherent detector, an over-sampling technique would also be adopted where $T_s < T_b$.

[2]Note that, the above vector notation does not apply if $k < I$.

2) MMSE: Another widely used sub-optimal detector is based on the MMSE criterion [27], which will minimise the covariance matrix of residual errors, i.e.,

$$\hat{\boldsymbol{\alpha}}_{MMSE} = \arg \min_{\boldsymbol{\alpha} \in \mathcal{A}^K} \mathbb{E}\left[(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T\right]. \tag{5}$$

Based on the linear and Gaussian model in Eq. (3), the MMSE estimation of estimated symbols is derived from:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{MMSE} &= \mathbb{E}(\boldsymbol{\alpha}|\mathbf{r}), \\ &= \mathbb{E}(\boldsymbol{\alpha}) + \boldsymbol{\Lambda}_w \mathbf{Y}^T (\mathbf{Y}\boldsymbol{\Lambda}_w \mathbf{Y}^T + \boldsymbol{\Lambda}_w)^{-1} (\mathbf{r} - \mathbf{Y}\boldsymbol{\alpha}), \end{aligned} \tag{6}$$

where $\mathbb{E}(\cdot)$ represents the statistical expectation, $\boldsymbol{\Lambda}_w$ denotes a $K \times K$ diagonal matrix with its elements are all equal to σ_w^2. From Eq. (5), an accurate estimation of CSI $\mathbf{y}$ is required when constructing $\mathbf{Y}$.

Note that, a finite memory I is necessary in both the sequential MAP and MMSE schemes [16], [27]. That is, even though the molecular concentration from previous transmissions will be sustained for a very long time, coherent detectors have to truncate the memory length to a finite number $I < \infty$ in order to alleviate the implementation complexity. Other threshold-based methods, e.g. in [28], utilize the knowledge of the channel response, which may be essentially treated as a coherent detector.

Coherent signal detectors will no doubt obtain an appealing detection performance [16], [18], [26], [27], but still face challenges in operating as part of a realistic molecular communication system. For one thing, due to the nature of the diffusion process and its sensitivity to channel disturbances, the response $\mathbf{y}_{I\times 1}$ can be difficult to estimate. Even if the *a priori* pilot sequence is available, the accurate acquisition of CSI is still resource intensive and has a short coherence time. For another, traditional coherent schemes, such as MAP or MMSE, will inevitably involve computational matrix inversions or polynomial operations. The acquisition of CSI and the resulting complex DSP implementation, as a consequence, will be the main stumbling-blocks of applying such coherent schemes to low-complexity and low-power scenarios.

III. Non-Coherent Signal Detector

The design of non-coherent detection will be of significant benefit to molecular communications, especially for nano-scale systems that have a lower energy and computation budget. A one-step interference mitigation scheme has been suggested by [29] to mitigate ISI from the previously time-slot. The information at time k is demodulated via a threshold, provided the interference from previous symbols $k - k_0$ ($k_0 \geq 2$) could be ignored. Thus, this non-coherent scheme may be suitable for very low data-rate molecular

communications, i.e., the symbol duration is larger than the arrival time of the peak concentration, t_m. When we want to achieve a relatively high data-rate, the ISI from multiple previous transmission slots will be accumulated (see Eq. (3)), which will cause severe ISI in the one-step SIC detection scheme [29]. The differential detector reported in our previous work [21] assumes the accumulated ISI in two adjacent slots will be comparable, which is only applicable to some low-rate applications. Other non-coherent detection schemes include the one proposed in [18], which describes a family of weighted sum detectors.

In this work, we propose a simple yet efficient non-coherent detector premised on the local convexity of the diffusion channel response, which may differ dramatically from the common coherent concepts [16], [19], [20]. Note that, rather than a spatial concept, the local convexity in fact refers to the time domain. As seen later, the diffusive channel response will be only be convex for a short duration. Essentially, we exploit certain stable features of the diffusion channel that exist irrespective of the changes of the channel. The new detection scheme, therefore, may lead to a class of low-complexity and robust techniques that can aid the design of nano-scale molecular communication systems.

A. Local Convexity

For the diffusion-based molecule propagation channel, we will have the following two remarks.

Remark 1: There exists two time deviations t_1 and t_2 from the start point $t = 0$, such that for the interval, i.e., $t \in [t_1, t_2]$, the channel response $h(t)$ is locally convex; while for the remaining two regions, i.e., $t \in (0, t_1) \cup (t_2, \infty)$, the channel response is locally concave.

We may easily have the 2nd order derivative of the channel response $h(t)$, i.e.,

$$\frac{\partial^2 h(t)}{\partial t^2} = \exp\left(-\frac{d^2}{4Dt}\right) \times \frac{(d^4 - 20d^2Dt + 60D^2t^2)}{128\pi^{3/2}D^{7/2}t^{11/2}}. \tag{7}$$

Letting the 2nd partial derivative be zero, i.e. $\frac{\partial^2 h(t)}{\partial t^2} = 0$, we will equivalently have $60D^2t^2 - 20d^2Dt + d^4 = 0$ since the term $\exp\left(-\frac{d^2}{4Dt}\right)$ is always larger than 0. Thus, we can obtain two inflection points on the channel response $h(t)$, i.e.,

$$t_1 = \frac{(5 - \sqrt{10}) \times d^2}{30D}, \; t_2 = \frac{(5 + \sqrt{10}) \times d^2}{30D}. \tag{8}$$

It is easily shown that for the region $t \in [t_{I_1}, t_{I_2}]$, we will have:

$$\left.\frac{\partial^2 h(t)}{\partial t^2}\right|_{\frac{(5-\sqrt{10})\times d^2}{30D} < t < \frac{(5+\sqrt{10})\times d^2}{30D}} \leq 0. \tag{9}$$

For the other regions $t \in (0, t_1) \cup (t_2, \infty)$, we similarly obtain:

$$\left.\frac{\partial^2 h(t)}{\partial t^2}\right|_{t \le \frac{(5-\sqrt{10})\times d^2}{30D} \cup t \ge \frac{(5+\sqrt{10})\times d^2}{30D}} > 0. \tag{10}$$

Remark 2: The maximum value of the channel response $h(t)$ will be achieved when $t = t_m$, and we have $t_1 < t_m < t_2$.

Similarly, let the 1st partial derivative to be zero, i.e.,

$$\frac{\partial h(t)}{\partial t} = \frac{d^2}{2Dt^2 \times (4\pi D)^{3/2}} \times \exp\left(-\frac{d^2}{4Dt}\right) - \frac{6\pi D}{(4\pi f t)^{5/2}} \times \exp\left(-\frac{d^2}{4Dt}\right) = 0. \tag{11}$$

Then, we may obtain:

$$t_m = d^2/6D, \ t_{I_1} < t_m < t_{I_2}. \tag{12}$$

Note that from Eq. (12), the arrival time of peak concentration t_m will be proportional to the square of the distance d. For some macro-scale molecular communications, the symbol duration T_b may be smaller than t_m in the case that a relatively high data rate is required.

Since the duration of emission pulse (i.e. T_p) is significantly smaller than both t_m and t_2 (or t_1)[3], the locally convex property of the equivalent response $y(t)$ will remain the same as $h(t)$, i.e.,

$$\frac{\partial y(t)}{\partial t} \simeq \frac{\partial h(t)}{\partial t}, \ \frac{\partial^2 y(t)}{\partial t^2} \simeq \frac{\partial^2 h(t)}{\partial t^2}. \tag{13}$$

For a more general case, $p(t) \triangleq \texttt{rect}\left[(t - T_p/2)/T_p - kT_b\right]$ accounts for a common gate (or indicator) function, i.e., $p(t) = 1$ in the time region $(-T_p/2, T_p/2]$ and, otherwise, $p(t) = 0$. For clarity, we consider the equivalent discrete signal model, and the received response $y(k)$ is then expressed as:

$$y(k) = \sum_{k_0=0}^{+\infty} h(k - k_0) \times p(k_0) = \sum_{k_0=0}^{K_0-1} h(k - k_0). \tag{14}$$

Here, $K_0 = \lfloor T_p/T_s \rfloor$ is a discrete length, and T_s is the time duration of the over-sampling process. Denote two inflection indexes by $K_1 = \lfloor t_1/T_s \rfloor$ and $K_2 = \lfloor t_2/T_s \rfloor$. Given an arbitrary integer k_1 satisfying

[3]See [12] where the emission time is anout $T_p = 100$ ms, while the measured inflection times are about $t_1 = 4$ s and $t_2 = 12$ s at the distance of 4 meters.

$k_1 \in [0, K_2 - K_1]$, in the sub-region $k \in [K_1 + k_1 + k_0, K_2 - k_1 + k_0]$ we will have the following relation:

$$\begin{aligned} y(k+k_1) + y(k-k_1) &= \sum_{k_0=0}^{K_0-1} h(k+k_1-k_0) + \sum_{k_0=0}^{K_0-1} h(k-k_1-k_0), \\ &\leq \sum_{k_0=0}^{K_0-1} 2 \times h(k-k_0) = 2 \times y(k). \end{aligned} \tag{15}$$

The above Eq. (15) indicates that, even if the shaping function $p(t)$ is not an ideal *Dirac-delta* function (i.e. $K_0 > 1$), the equivalent response $y(t)$ will remain convex around the peak time, for $K_0 < K_{I_2}$. As mentioned above, this condition can be easily satisfied in practice, as in a common communication system we usually use a very small K_0 in order to control ISI.

Based on the above two remarks, it is concluded that the shape of received molecular numbers is ***locally convex*** around the diffusion peak. In practice, the time deviation of the peak concentration within each duration T_b, which is denoted by $T_\Delta \in [0, T_b]$, will remain unknown. Thus, a pre-timing process is necessary to estimate T_Δ. For simplicity, in the analysis we assume the peak time (i.e. $kT_b + T_\Delta$) has been acquired.

In order to fully exploit the aforementioned convex property and design a non-coherent detection scheme, an over-sampling technique should be adopted, which will remain relatively different from existing coherent detection methods where a sampling period T_s can be configured directly to T_b as in eq. (3). In the case of over-sampling, we assume $L_0 T_s = T_b/2$ for convenience, where $L_0 = \lfloor T_b/T_s \rfloor/2$ is an integer larger than 1. Thus, the over-sampled signal in the kth duration is:

$$r_{k,l} \triangleq \sum_{n=0}^{K} \alpha_k \times y(t)|_{t=(k-n)T_b+T_\Delta+lT_s} + w_{k,l}, \quad l \in [-T_b/T_s, T_b/T_s - 1], \tag{16}$$

where the i.i.d noises $w_{k,l}$ follow a Gaussian distribution, i.e. $w_{k,l} \sim \mathcal{N}(0, \sigma_w^2)$.

B. Local Convexity Detector

The designed non-coherent detector mainly contains three steps. Firstly, a pre-smoothing process is used to suppress noise. Then, the convex metric will be constructed. Finally, the obtained convex metric will be compared with the threshold, and the unknown information symbol will be estimated.

*1) **Pre-Smoothing**:* Firstly, L zeros will be padded to the front of sequence, and the zero-padding signal is written as:

$$\mathbf{s} = [0, \cdots, 0, r_{1,-L}, r_{1,-L+1}, \cdots, r_{1,L}, \cdots, r_{K,-L}, \cdots, r_{K,L}]^T. \tag{17}$$

where $r_{k,l}$ $(-L \leqslant l \leqslant L)$ denote the l-th over-sampled signal of the k-th symbol duration.

A moving average (MA) process with a length of $2L+1$ ($L \leqslant \lfloor T_b/T_s \rfloor/2 \rfloor$) will be applied then. For $n > L$, the smoothed signal $\breve{r}(n)$ is:

$$\breve{r}(n) = \frac{1}{2L+1} \times \sum_{q=n-L}^{n+L} s_q. \tag{18}$$

*2) **Local Convexity**:* In the proposed non-coherent scheme, the local convexity of received concentrations will be exploited to recover unknown information. Recall that a convex function $f(x)$ satisfies $1/2 \times [f(x+\Delta x) + f(x-\Delta x)] < f(x)$, or $f(x+\Delta x) + f(x-\Delta x) < f(x+\delta x) + f(x-\delta x)$ when $\Delta x > \delta x$. Thus, we should evaluate the convexity around a diffusion peak time of each symbol duration. Based on the above elaboration, we define a group of sub-metrics as:

$$c_k^{(1)} \triangleq \sum_{l=1}^{L} \left[\breve{r}(kT_b + T_\Delta) - \frac{\breve{r}(kT_b + T_\Delta + lT_s) + \breve{r}(kT_b + T_\Delta - lT_s)}{2} \right], \tag{19a}$$

$$c_k^{(2)} \triangleq \frac{1}{2} \times \sum_{l=1}^{L-1} \{ \breve{r}(kT_b + T_\Delta + lT_s) + \breve{r}(kT_b + T_\Delta + LT_s - lT_s) - \breve{r}[kT_b + T_\Delta + (l+1) \times T_s] - \breve{r}[kT_b + T_\Delta + LT_s - (l+1) \times T_s] \}, \tag{19b}$$

$$c_k^{(3)} \triangleq \frac{1}{2} \times \sum_{l=1}^{L+1} \{ \breve{r}(kT_b + T_\Delta + lT_s) + \breve{r}(kT_b + T_\Delta + LT_s + lT_s) - \breve{r}[kT_b + T_\Delta + (l+2) \times T_s] - \breve{r}[kT_b + T_\Delta + LT_s + (l-2) \times T_s] \}. \tag{19c}$$

Here, $kT_b + T_\Delta$ accounts for the peak concentration time of the kth symbol, which may be redefined as $T_\Delta(k) \triangleq kT_b + T_\Delta$. It should be noted that, after the over-sampling process, the discrete signal should be indexed via k and l. For clarity, in the above definition and following analysis, we focus on the convexity sub-metrics constructed from the $2L+1$ discrete samples that are centered at the peak concentration, which may be directly indexed by k as in Eq. (19). This is because such convexity sub-metrics $c_k^{(i)}$ $(i = 1, 2, 3)$ may be of more significance to the estimation of unknown symbols α_k. However, for more general cases where the analyzed samples are not centered at peak concentrations (e.g. $kT_b + T_\Delta + l'T_s$), the constructed convexity sub-metrics should be indexed by k and l' jointly, e.g. $c_{k,l'}$. Since the construction of general convex sub-metrics is the same as Eq. (19), the expressions will be omitted in the following analysis.

Taking the first metric in Eq. (19) for example, when the information bit is $\alpha_k = 1$, this sub-metric will be larger than 0, which hence can be used as a convex metric. Then, the convexity metric will be

specified by:

$$c_k = c_k^{(1)} + c_k^{(2)} + c_k^{(3)}. \tag{20}$$

Remark 3: The convexity metric c_k will effectively suppress the ISI generated by the previous $(k-k_0)$ symbols.

Due to the heavy tail and the slow decay of concentration from previous symbols, the residual interference e_{k-k_0} at the kth interval, generated from the $(k-k_0)$th previous symbol, may be linearly approximated by Eq. (21), as demonstrated in Fig. 2.

$$\begin{aligned} e_{k-k_0}(k,l) &= \alpha_{k-k_0} \times y[T_\Delta(k-k_0) + lT_s], \\ &\backsimeq \alpha_{k-k_0} \times y[T_\Delta(k-k_0) + LT_s] + b_{k-k_0} \times (L+l) + o\{(L+l)^2\}, \\ &\backsimeq C_{k-k_0} + b_{k-k_0} \times (L+l), \quad -L \leqslant l \leq L. \end{aligned} \tag{21}$$

where $C_{k-k_0} = \alpha_{k-k_0} \times y[t - T_\Delta(k-k_0) + LT_s] > 0$ is the intercept of an approximated line, and $b_{k-k_0} \simeq \alpha_{k-k_0} \times \frac{dy(t)}{dt}|_{t=(k-k_0)T_b+LT_s} < 0$ is its negative slope.

Relying on Eq. (21), in the presence of ISI from the $(k-k_0)$-th time, the expectation of the first sub-metric of time bin $[T_\Delta(k) - LT_s, T_\Delta(k) + LT_s]$ can be simplified to:

$$\begin{aligned} \mathbb{E}\{c_k^{(1)}\} \simeq & \frac{-1}{2}\sum_{l=1}^{L} \alpha_k y[T_\Delta(k) + lT_s] - \frac{L}{2}\sum_{k_0=1}^{I}\left(C_{k-k_0} + \frac{3(L-1)}{2} b_{k-k_0}\right) \\ & - \frac{1}{2}\sum_{l=1}^{L} \alpha_k y[T_\Delta(k) - lT_s] - \frac{L}{2}\sum_{k_0=1}^{I}\left(C_{k-k_0} + \frac{L-1}{2} b_{k-k_0}\right) \\ & + L \times \left[\alpha_k y(T_\Delta(k)) + \sum_{k_0=1}^{I}(C_{k-k_0} + b_{k-k_0} L)\right], \\ = & \alpha_k \times \left\{ L y[T_\Delta(k)] - \frac{\sum_{l=1}^{L-1}[y(T_\Delta(k) + lT_s) + y(T_\Delta(k) - lT_s)]}{2} \right\} \\ & + \sum_{k_0=1}^{I}(2Lb_{k-k_0} - 2L^2 b_{k-k_0}). \end{aligned} \tag{22}$$

Similarly, the expectations of the second and third sub-metrics will be simplified, i.e.,

$$\begin{aligned} \mathbb{E}\{c_k^{(2)}\} = \frac{\alpha_k}{2} \times \sum_{l=1}^{L-1} \{ & y[T_\Delta(k) + lT_s] + y[T_\Delta(k) + LT_s - lT_s] \\ & - y[T_\Delta(k) + (l+1) \times T_s] - y[T_\Delta(k) + LT_s + (l+1) \times T_s] \}, \end{aligned}$$

$$\mathbb{E}\{c_k^{(3)}\} = \frac{\alpha_k}{2} \times \sum_{l=1}^{L-1} \{y[T_\Delta(k) + lT_s] + y[T_\Delta(k) + LT_s - lT_s] \\ -y[T_\Delta(k) + (l+2) \times T_s] - y[T_\Delta(k) + LT_s + (l+2) \times T_s]\}.$$

It is noted from Eq. (22) that the expectation of the first sub-metric $c_k^{(i)}$ will be independent of the additive noise. This is because the constructed convex sub-metrics involve only the linear summation or subtraction operations of the zero-mean noise samples. In Eq. (22), the first term $\alpha_k \times \mu_1$ accounts for the useful signal with $\mu_1 \triangleq Ly(T_\Delta(k)) - \frac{1}{2} \times \sum_{l=1}^{L-1}[y(T_\Delta(k) + lT_s) + y(T_\Delta(k) - lT_s)] > 0$, the second term $\sum_{k_0=1}^{I}(2Lb_{k-k_0} - 2L^2 b_{k-k_0}) \to 0$ may be dropped, as the decay slope usually is very small for a large k_0, i.e. $b_{k-k_0} \to 0$. So, the convexity metric c_k will suppress the ISI from the previous $(k-k_0)$-th time interval. Similar conclusions can be drawn to the other two convexity sub-metrics.

From the simplified expression of $c_k^{(1)}$, an explanation will be further given on how the convexity sub-metrics could suppress ISI. In the case of $\alpha_k = 1$, the convex shape will occur around the peak concentration, so we have c_k significantly larger than zero. Note that, even though the ISI signals from the previous $(k-k_0)$-th slot may probably exhibit the concave shape, the convexity of c_k will almost not be affected. This is because the diffusion response usually exhibits the long tail, and the ISI signal from the $(k-k_0)$-th slot would be approximated by a linear function, which of course will not change the convex shape of useful signals in the k-th slot. To sum up, the convex shape is determined by the useful signals related with α_k, while the second term as in Eq. (21) will not affect the convexity.

Note from Eq. (22) that, the channel response $y(t)$ will appear in the expectation of convex metric $c_k^{(1)}$. This is just to illustrate the composition of $c_k^{(1)}$ under different symbols α_k. In practice, the calculation of the three convexity metrics will rely only on the received concentration $r(t)$, as in Eq. (19). By directly employing the underlying received signal $r(t)$ instead of the channel response $y(t)$, the designed convexity-based scheme can be regarded as a non-coherent detector.

With the above derivations, the expectation of the local convexity metric becomes:

$$\mathbb{E}\{c_k\} \backsimeq \alpha_k \times (\mu_1 + \mu_2 + \mu_3), \quad \mu_1, \mu_2, \mu_3 > 0, \tag{23}$$

where the constants μ_1, μ_2 and μ_3 are three positive constants that are related with the channel convexity

and the MA length L, i.e.,

$$\mu_2 \triangleq \frac{1}{2} \times \sum_{l=1}^{L-1} \{y[T_\Delta(k) + lT_s] + y[T_\Delta(k) + LT_s - lT_s)] \\ -y[T_\Delta(k) + (l+1) \times T_s] - y[T_\Delta(k) + LT_s + (l+1) \times T_s]\},$$

$$\mu_3 \triangleq \frac{1}{2} \times \sum_{l=1}^{L-1} \{y[T_\Delta(k) + lT_s] + y[T_\Delta(k) + LT_s - lT_s] \\ -y[T_\Delta(k) + (l+2) \times T_s] - y[T_\Delta(k) + LT_s + (l+2) \times T_s]\}.$$

In order to design a decision threshold, we should investigate the expected values of the constructed convexity metric in the presence of different information bits. Provided that we have the zero-mean counting noise, then the expectation of c_k in the case of $\alpha_k = 1$ is approximated by:

$$\mathbb{E}\{c_k|\alpha_k = 1\} = \lambda NL \times \left\{ \frac{y\left[T_\Delta(k)\right] - y\left[T_\Delta(k) - \frac{L}{2}T_s\right]}{4} + \frac{y\left[T_\Delta(k)\right] - y\left[T_\Delta(k) + \frac{L}{2}T_s\right]}{4} + \right. \\ \left. \frac{y\left[T_\Delta(k) - \frac{L}{2}T_s\right] - y\left[T_\Delta(k) - LT_s\right]}{4} + \frac{y\left[T_\Delta(k) + \frac{L}{2}T_s\right] - y\left[T_\Delta(k) + LT_s\right]}{4} \right\}, \tag{24}$$

where λ is a weighting factor ranged from 0 to 1, and N denotes the total number of sub-metrics. Here, we have $N = 3$. When the information bit of the kth duration is $\alpha_k = 0$, then the expected value will become:

$$\mathbb{E}\{c_k|\alpha_k = 0\} \to 0.$$

This is due to the fact the convexity metric c_k involves two components in Eq. (22), i.e., the linear combination of zero-mean noises and the other negligible self-interference term that is related with the decay slope $b_{k-k_0} \to 0$.

3) ***Post-smoothing****:* A post-smoothing or post-filtering procedure is further suggested to reduce the noise. In practice, a similar MA process will be used, i.e., $\breve{c}(k) = \frac{1}{2L_0+1} \times \sum_{l=k-L_0}^{k+L_0} c_{k,l}$, where the averaging length $L_0 < L$ will be set to 2∼3.

Finally, a threshold γ will be used to determine whether the current convexity metric is sufficiently dominant, and then the information bit of each interval will be estimated as in Eq. (25).

$$\hat{\alpha}_k = \begin{cases} 1, & \breve{c}_k > \gamma, \\ 0, & \breve{c}_k \leq \gamma. \end{cases} \tag{25}$$

Here, the threshold γ is of importance to the design of the proposed non-coherent detector, as it serves as a measure of the strength of local convexity.

*4) **Threshold design**:* It can be easily seen that, from the definition of convexity metrics in Eqs. (19)–(20), the threshold γ will be related to three factors:

1) the length of each convexity sub-metric, namely L (recall that each sub-metric is the linear combination of L terms);
2) the total number of sub-metrics N;
3) the specific convexity of the received concentration.

For the constructed metric, we have the following constraint relationships:

$$\gamma < \frac{NL}{2} \times \left\{ y[T_\Delta(k)] - \frac{y[T_\Delta(k) + LT_s] + y[T_\Delta(k) - LT_s]}{2} \right\}, \tag{26}$$

$$\gamma > \frac{NL}{4} \times \min_{\substack{L_1,L_2,L_3,L_4 \in [-L,L], \\ L_1 < L_2 \le L_3 < L_4}} \left\{ y[T_\Delta(k) + L_2T_s] + y[T_\Delta(k) + L_3T_s] - y[T_\Delta(k) + L_1T_s] - y[T_\Delta(k) + L_4T_s] \right\}. \tag{27}$$

Here, the upper bound of the threshold γ is determined by the maximum convex value among all combinations of received samples when constructing convex sub-metrics, as in Eq. (19). Recall the molecular number will reach the local maximum at the time $T_\Delta(k)$ of the k-th symbol (see Fig. 2), while one of the molecular number of times $T_\Delta(k) \pm LT_s$ may achieve the minimum value. Thus, the maximum value is specified by $\frac{NL}{2} \times \left\{ y[T_\Delta(k)] - \frac{1}{2} \times \left\{ y[T_\Delta(k) + LT_s] + y[T_\Delta(k) - LT_s] \right\} \right\}$. Similarly, the lower bound of the threshold will be determined by the minimum convex value among all combinations, as in Eq. (27).

With the equal priors of $\alpha_k = 1$ and $\alpha_k = 0$, the threshold is determined by:

$$\begin{aligned} \gamma &\triangleq \frac{1}{2} \times \left(\mathbb{E}\{c_k | \alpha_k = 1\} + \mathbb{E}\{c_k | \alpha_k = 0\} \right) \\ &\simeq \frac{1}{2} \times \lambda NL \times \left\{ y[T_\Delta(k)]/2 - \frac{y\left[T_\Delta(k) + LT_s\right] + y\left[T_\Delta(k) - LT_s\right]}{4} \right\}. \end{aligned} \tag{28}$$

It is found from Eqs. (24) and (28) that the threshold γ will be viewed as a linearly equal combination of the median values in four secant lines, which can roughly reflect the convexity shape of the underlying diffusion channel. Recall that, the convexity metric c_k is also a linear combination of multiple median values, so the designed threshold may produce the promising detection performance, as demonstrated by subsequent simulations. However, the design of convexity metrics and the corresponding threshold

remains an open problem. For example, the construction of the convex sub-metrics in Eq. (19) and the combination strategy in Eq. (20) as well as the threshold may be optimised in future work.

Since the accurate channel response $y(t)$ is unavailable for the non-coherent detection scenario, the threshold γ can be in practice configured approximately via Eq. (29). In this simplest case, we use the received sample $r(k)$ as the instantaneous estimation of $y(k)$ within a single interval, i.e.,

$$\hat{\gamma} = \frac{1}{2} \times \lambda NL \times \left[r(T_\Delta(k))/2 - \frac{r\left(T_\Delta(k) + LT_s\right) + r\left(T_\Delta(k) - LT_s\right)}{4} \right]. \tag{29}$$

Alternatively, an adaptive threshold $\hat{\gamma}(k)$ can be suggested to reduce the influence of noise and further improve the detection performance. For the static channel with unknown response, we will have:

$$\hat{\gamma}(k) \simeq (1-\beta) \times \hat{\gamma}(k-1) + \beta \times \frac{\lambda NL}{2|\mathcal{K}_s|} \sum_{m \in \mathcal{K}_s} \left[\frac{r(T_\Delta(m))}{2} - \frac{r\left(T_\Delta(m) + LT_s\right) + r\left(T_\Delta(m) - LT_s\right)}{4} \right], \tag{30}$$

where β is a forgotten parameter which is ranged in $[0.9\ 0.99]$; the initial threshold estimation can be set to $\hat{\gamma}(0) = 0$; $\mathcal{K}_s \subset \{1, 2, \cdots, k\}$ accounts for a sub-set of the time slots where its convexity metric is sufficiently large and $|\mathcal{K}_s|$ represents the number of elements, i.e.,

$$\mathcal{K}_s = \left\{ m \middle| 1 \le m \le k,\ r(T_\Delta(m))/2 - \frac{r\left(T_\Delta(m) + LT_s\right) + r\left(T_\Delta(m) - LT_s\right)}{4} > 0.5 \times \sigma_w \right\}. \tag{31}$$

It is seen that, as k increases, the effects of additive noise will vanish gradually and the adaptive threshold $\hat{\gamma}(k)$ will converge to a stable value, i.e. $\gamma(k) = \gamma(k-1)$, and then we have

$$\begin{aligned} \lim_{k\to\infty} \hat{\gamma}(k) &= \lim_{k\to\infty} \frac{\lambda NL}{2} \times \frac{1}{|\mathcal{K}_s|} \sum_{m \in \mathcal{K}_s} \left[r(T_\Delta(k))/2 - \frac{r\left(T_\Delta(k) + LT_s\right) + r\left(T_\Delta(k) - LT_s\right)}{4} \right] \\ &= \frac{\lambda NL}{2} \times \left[y(T_\Delta(k))/2 - \frac{y\left(T_\Delta(k) + LT_s\right) + y\left(T_\Delta(k) - LT_s\right)}{4} \right] \\ &= \gamma. \end{aligned} \tag{32}$$

In the designed adaptive threshold, there are three points need to be noteworthy in practice. (1) In each symbol interval, the threshold $\gamma(k)$ will be re-calculated, which is then used to detect the signal. (2) For the static channel with an unknown response, the received signals of multiple intervals have been utilized to improve the accuracy of the estimated threshold, as in Eq. (30). (3) For the variant channel, the threshold could also adapt to channel changes based on the received signals. In the case, the sub-set $\mathcal{K}_s$ will be redefined to $\{k - K_c + 1, k - K_c + 2, \cdots, k\}$, where K_c accounts for the static length in which

the channel is temporarily unchanged. The forgotten parameter β will reflect the correlation of varying channels. Specifically, the higher the channel correlation, the smaller the forgotten parameter is. Consider the fast varying channel with a low correlation for example, we may have $K_c = 1$ and $\beta = 0.99$, and consequently, the threshold $\gamma(k)$ will be estimated only based on the signal of the kth interval.

C. Theoretical Analysis

Since the convex sub-metrics $c_k^{(i)}$ $(i = 1, 2, 3)$ are all based on the linear combination of the noisy observation $r(n)$, thus the constructed convexity metric c_k will be Gaussian distributed under the i.i.d noises. Thus, the likelihood distributions of the convexity metric c_k in the case of two information bits are respectively given by:

$$p\{c_k = x|\alpha_k = 1\} \sim \mathcal{N}\{x; \mathbb{E}\{c_k|\alpha_k = 1\}, \mathbb{V}\{c_k\}\}, \tag{33}$$

$$p\{c_k = x|\alpha_k = 0\} \sim \mathcal{N}\{x; 0, \mathbb{V}\{c_k\}\}. \tag{34}$$

Here, $\mathbb{V}\{c_k\}$ accounts for the variance of the convexity metric c_k. It is seen that $\mathbb{V}\{c_k\}$ will be related with the specific construction method of each sub-metric $c_k^{(i)}$ $(i = 1, 2, 3)$, the averaging length $2L + 1$ as well as the noise variance σ_w^2. For simplicity, we assume the analysis length L to be an odd integer. For the i.i.d noise samples, we can further derive the variance of the convexity metric c_k which is constructed from Eqs. (19) and (20), i.e.,

$$\begin{aligned}\mathbb{V}\{c_k\} &= \sigma_w^2 \times \left[L^2 + 1 + \left(\frac{1}{2}\right)^2 \times (2L - 4) + 2 \times \left(\frac{3}{2}\right)^2\right], \\ &= \sigma_w^2 \times (L^2 - 0.5L + 4.5).\end{aligned} \tag{35}$$

Given the threshold in Eq. (28) and the equal priors for "0" and "1", i.e., $p(\alpha_k = 1) = p(\alpha_k = 0) = 0.5$, the theoretical error probability of the designed detection method can be calculated via:

$$\begin{aligned}P_E &= p(\alpha_k = 1) \times \int_0^{\gamma} p\{x|\alpha_k = 1\}dx + p(\alpha_k = 0) \times \int_{\gamma}^{\infty} p\{x|\alpha_k = 0\}dx, \\ &= \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\mathbb{E}^2\{c_k|\alpha_k = 1\}}{8 \times \sigma_w^2 \times (L^2 - 0.5L + 4.5)}}\right).\end{aligned} \tag{36}$$

Note from Eq. (36) that the derived BER performance, which is similar to the theoretical result of the OOK-based communications system [30], is related closely with the expectation of a convexity metric c_k

in the case of $\alpha_k = 1$. For the proposed non-coherent detector, the analytical BER curve will be unavailable as the term $\mathbb{E}^2\{c_k|\alpha_k = 1\}$ cannot be determined due to the unknown channel response. However, the theoretical BER curve demonstrates the influence of related parameters (e.g. L and σ_w^2), which can be used to validate the practical detection performance if $\mathbb{E}\{c_k|\alpha_k = 1\} = 2\gamma$ can be estimated, e.g. based on Eq. (30).

D. Complexity Analysis

It is found that the complexity of the new non-coherent scheme is approximately on the order of $\mathcal{O}(L^2)+\mathcal{O}(L)$ summation operations, i.e., requiring no multiplication. The first term $\mathcal{O}(L^2)$ comes from the moving average operations, while the second term $\mathcal{O}(L)$ is attributed to the calculations of convexity metrics and the threshold. For the previously mentioned coherent MAP scheme, the number of multiplications are measured by $\mathcal{O}(2^I\vartheta)$, where ϑ accounts for the complexity of evaluating the involved likelihood function. For a linear MMSE method, the computational complexity is $\mathcal{O}(I^3)$. Thus, the required multiplications of coherent detectors (e.g. $\mathcal{O}(I^3)$) increases rapidly with the increasing of I, whilst the required summations of the new non-coherent scheme increases with $\mathcal{O}(L^2)$.

IV. Numerical Simulations

In the simulation studies, we focus on the processing of discrete time samples. While the sample duration T_s may vary with different applications, the proposed non-coherent detection scheme is basically independent of any specific T_s. Without loss of generality, the maximum concentration of $y(t)$ is located at $t_m = 14 \times T_s$. The information interval is set to $T_b = 10 \times T_s$, i.e. $T_b < t_m$. In order to simulate the discrete channel response, the diffusion coefficient is $D = 4.75 \times 10^{-8}\text{m}^2/\text{s}$, the distance is d=65nm and the sampling time is $T_s = 1.05 \times 10^{-9}$s. It is easily noted that the received concentrations will be always positive, as the ISI has usually a long-tail as demonstrated by measurement [12], [13], which will be dominant to the additive counting noise. Given a high symbol rate, i.e. $T_b < t_m$, severe ISI will occur at the receiver after the transmitted signal have propagated via the diffusion channel. Each bit error ratio (BER) curve is derived numerically based on 10^6 binary symbols. In the analysis, the average length is set to $L = 5$ and a long ISI of $I = 30 \times T_b$ is assumed. The signal to noise ratio (SNR) is defined as:

$$\text{SNR} \triangleq 10 \times \log_{10}\mathbb{E}\left(\frac{\sum_{k=1}^{K} r_k^2}{K\sigma_w^2}\right). \tag{37}$$

In the first experiment, we will investigate the effect from different configurations of the decision threshold γ. It is shown by Fig. 3 that, for the suggested threshold in Eq. (30), the weighting factor λ will have some impact on the detection performance. Taking SNR $= 29$ dB for example, the best BER value of 2.25×10^{-3} will be achieved when $\lambda = 0.5$, while the worst BER is about 0.35 if $\lambda = 0.01$. Meanwhile, we have noted that, in the high SNR regions (e.g., > 25 dB), the optimal λ is independent of SNR. Thus, we may conclude that the weighting factor λ may be set to around 0.5 in practice.

For the non-coherent detection performance, we study the effect of different symbol durations T_b, for a fixed peak concentration time of $t_m = 14 \times T_s$. In order to construct the significant and effective convexity metrics, the symbol duration T_b is expected to be larger than the peak diffusion time, i.e. $T_b > t_m$. It is easily understood that, the larger the symbol duration T_b, the weaker the ISI from the adjacent durations, which also results in a more dominant convex shape. The conclusion has been verified by the simulation results. As demonstrated in Fig. 4, if we increase the symbol duration from $T_b = 20T_s \backsimeq 1.43t_m$ to $T_b = 30T_s \backsimeq 2.15t_m$, then a rough detection gain of 2dB will be achieved by the non-coherent scheme. Although the ISI can be reduced with a sufficiently large T_b, the resulting data rate will be restricted. Fortunately, we can configure the symbol duration slightly smaller than a peak diffusion time when pursuing the high data rate, $T_b = 0.7 \times t_m$. As shown by simulation results, in the serious ISI case the designed non-coherent detection scheme can also acquire good BER performance. However, this is achieved at the expense of low energy efficiency (e.g. requiring high SNR). In conclusion, the designed scheme is relatively robust in the case that the symbol duration is smaller than t_m, which is of promise to promote the data rate of molecular communications.

In Fig. 5, we further plot the theoretical BER curves of our new detection scheme, in which the full response is assumed to be known. Thus, the expectation on the convexity metric c_k can be accurately evaluated based on Eq. (28). We can see that, for the non-coherent detector, the theoretical BER curve in Eq. (36) can be interpreted as the low bound, as the threshold γ (or the expectoration on the convexity metric c_k) can only be approximated in practice. From the comparative results in Fig. 5, it is found that the simulation result agrees with the theoretical BER curves.

In Fig. 6, the BER performance of the proposed adaptive threshold in Eq. (30) as well as the threshold calculated from Eq. (28) with a known channel response are plotted together. In the simulation, we configure the forgotten parameter to $\beta = 0.915$. We can see that the adaptive threshold will basically achieve the promising BER performance. It is noted that, in high SNRs (e.g > 26 dB), the calculated

threshold based on a known channel response may lead to better BER performance. However, in lower SNRs, the calculated threshold seems to be not accurate. That is mainly because, in the low SNRs (e.g. SNR $\leqslant$ 25 dB), the optimal weighting factor λ would be deviated slightly from the configured value 0.5 (see Fig. 3), which may cause the slight degradation on the BER performance. In general, it is observed that the adaptive threshold can be applied in practice.

A. Comparative Performance

In the second set of experiments, the exact channel response is assumed to be known by the two previously mentioned coherent detection schemes i.e., the MAP or MMSE schemes, which, of course, have to be obtained at the expense of sophisticated estimation techniques. In contrast, our proposed scheme assumes that the channel response is *blind* to the non-coherent detection process. It is observed that, from Fig. 7, the coherent MAP scheme can obtain the optimum detection performance, by maximizing the *a posteriori* probability and mitigating the ISI to the minimum. The designed non-coherent detection scheme, despite the dramatically reduced implementation compared with MAP or MMSE, is slightly inferior to such coherent schemes. Taking the high SNR regime as an example (e.g., SNR $>$ 22 dB), a detection SNR gap between our proposed scheme and the MMSE scheme is approximately 2dB.

What is clear from our previous analysis is that, although classical coherent schemes may acquire the more attractive detection performance, their implementation complexity $\mathcal{O}(I)^3$ will become easily unaffordable as far as a large I is concerned. In order to reduce the computational complexity, as a compromise, in practice the ISI length I would be truncated to a smaller value. E.g., in some energy-constrained implementation scenarios, the processing length I could be restricted to 10. We observed from Fig. 8 that, in this case, the MAP detector that changes and becomes comparable to the non-coherent scheme which, yet, has a much lower complexity than both two coherent schemes. The proposed non-coherent scheme will achieve a comparable performance with the MMSE method when $I = 10$.

B. Channel Estimation Error

Although channel parameter estimation is investigated in [31], to the best of our knowledge, the existing work has not shown how the diffusion channel can be realistically estimated when it is time varying. The challenge may probably lie in two aspects, i.e., the lack of a chemical pilot signal (due to the limited chemical bandwidth) and the *non-reciprocal* nature of a diffusion channel. The reality is that the molecule channel is not static and there are frequent disturbances aroused by both ambient flow (temperature

differences) and active air flow (movement of bodies). Even when a pilot channel is available, the non-reciprocal diffusion channel is still difficult to obtain accurately, due to its short coherence time. Typically, the channel may change faster than the travel time of molecular signals.

We assume that, for the coherent detectors (i.e. MAP and MMSE), the channel estimation can be available at a specific time, by resorting to the dedicated chemical pilot. However, the diffusion channel will change randomly after a period of time. For example, the previous channel response CH1 may have changed to CH2 due to unknown disturbance, as shown in Fig. 9(a). In most cases, coherent detectors will be unaware of the time-variation of channel response. This is because, in practice, it is very expensive to employ the bandwidth-demanding pilot to estimate the frequently changed channels for any times. Thus, they would probably use the previously acquired CSI (with errors) until the dedicated chemical pilot can be attainable again. For the non-coherent method, however we assume it will have no *a priori* information on the CSI, and the threshold will be determined based on the received signals as in Eq. (30). Numerical results in the presence of the varying diffusion channel are demonstrated by Fig. 9(b). As expected, since the detection performances of the coherent detectors hinges on the channel estimation and, provided an inaccurate CSI, their BER performances degrade significantly. From the simulation results in Fig. 9(b), the SNR loss of 5dB may be observed in the MAP scheme. For the MMSE scheme, the BER floor may even be demonstrated. It is noted that the proposed non-coherent detector relies on the local convexity of diffusion channel response, which will also update its threshold based on the received signals after the change in channel response. So, it is essentially independent of the channel estimation and is thereby robust to unknown CSI. As demonstrated by Fig. 9(b), its BER performance will be affected slightly. This provides great promise to the realistic application of the non-coherent detector.

V. Conclusions

A novel non-coherent signal detector is proposed for the emerging field of molecular communications. The data rate of molecular communications is typically limited by severe ISI, which is difficult to avoid due to the diffusion process. Whilst up-to-date accurate CSI is difficult to acquire and requires complex signal processing operations, there exists common features in the diffusion channel that can be exploited at the receiver side to suppress ISI. By utilizing the local convex features of the channel, the proposed non-coherent detector can operate in the absence of channel estimations and mitigate ISI from multiple symbols in high data rate transmissions. In certain channel scenarios, simulation results have shown that the performance of the proposed blind detector can approach those achieved by high complexity

coherent detectors. Yet, the signal processing implementation of the proposed non-coherent detector is simple and involves only a number of summation operations. More importantly, it is found that the BER performance of the designed scheme, which requires no explicit CSI, is relatively robust to time varying channels. In general, the proposed low-complexity and non-coherent signal detector will be vital to the realistic implementation of molecular communications, especially in the application of mobile molecular communications between nano-robots in the area of nano-medicine, where the low-complexity implementation is critical.

ACKNOWLEDGMENT

This work of B. Li was supported by Natural Science Foundation of China (NSFC) under Grants 61471061 and the Fundamental Research Funds for the Central Universities under Grant 2014RC0101. The work of S. Wang was in part supported by the Research Development Fund (RDF-14-01-29) of Xi'an Jiaotong-Liverpool University.

REFERENCES

[1] I. F. Akyildiz, J. M. Jornet, and M. Pierobon, "Nanonetworks: A new frontier in communications," *Communications of the ACM*, vol. 54, no. 11, pp. 84–89, Nov. 2011.

[2] I. F. Akylidiz, F. Fekri, R. Sivakumar, C. Forest, and B. Hammer, "Monaco: fundamentals of molecular nano-communication networks," *IEEE Wireless Communications*, vol. 19, no. 5, pp. 12–18, May 2012.

[3] T. Nakano, A. Eckford, and T. Haraguchi, *Molecular Communications*. Cambridge University Press, Oct. 2013.

[4] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano, "Molecular communication," in *Proceedings of the NSTI Nanotechnology Conference and Trade Show*, 2005, pp. 391–394.

[5] L. Parcerisa Giné and I. F. Akyildiz, "Molecular communication options for long range nanonetworks," *Computer Networks*, vol. 53, no. 16, pp. 2753–2766, Nov. 2009.

[6] T. Nakano, T. Suda, Y. Okaie, M. J. Moore, and A. V. Vasilakos, "Molecular communication among biological nanomachines: A layered architecture and research issues," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3, pp. 169–197, Sep. 2014.

[7] A. W. Eckford, N. Farsad, S. Hiyama, and Y. Moritani, "Microchannel molecular communication with nanoscale carriers: Brownian motion versus active transport," in *10^{th} IEEE Conference on Nanotechnology (IEEE-NANO)*, 2010, pp. 854–858.

[8] A. Mafra-Neto and R. T. Cardé, "Fine-scale structure of pheromone plumes modulates upwind orientation of flying moths," *Nature*, vol. 369, no. 6476, pp. 142–144, May 1994.

[9] H. B. Yilmaz, A. C. Heren, T. Tugcu, and C.-B. Chae, "Three-dimensional channel characteristics for molecular communications with an absorbing receiver," *IEEE Communications Letters*, vol. 18, no. 6, pp. 929–932, Jun. 2014.

[10] T. D. Wyatt, *Pheromones and animal behaviour: communication by smell and taste*. Cambridge University Press, Mar. 2003.

[11] B. Atakan, O. B. Akan, and S. Balasubramaniam, "Body area nanonetworks with molecular communications in nanomedicine," *IEEE Communications Magazine*, vol. 50, no. 1, pp. 28–34, Jan. 2012.

[12] N. Farsad, W. Guo, and A. W. Eckford, "Tabletop molecular communication: Text messages through chemical signals," *PloS one*, vol. 8, no. 12, p. e82935, Dec. 2013.

[13] N. Farsad, N. Kim, A. Eckford, and C. Chae, "Channel and noise models for nonlinear molecular communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 12, pp. 2392–2401, Dec. 2014.

[14] W. Guo, S. Wang, A. Eckford, and J. Wu, "Reliable communication envelopes of molecular diffusion channels," *Electronics Letters*, vol. 49, no. 19, pp. 1248–1249, Sep. 2013.

[15] W. Guo, B. Li, S. Wang, and W. Liu, "Molecular communications with longitudinal carrier waves: Baseband to passband modulation," *IEEE Communication Letters*, vol. 19, no. 9, pp. 1512–1515, Jun. 2015.

[16] D. Kilinc and O. B. Akan, "Receiver design for molecular communication," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 705–714, Dec. 2013.

[17] W.-A. Lin, Y.-C. Lee, P.-C. Yeh, and C.-h. Lee, "Signal detection and ISI cancellation for quantity-based amplitude modulation in diffusion-based molecular communications," in *IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 4362–4367.

[18] A. Noel, K. C. Cheung, and R. Schober, "Optimal receiver design for diffusive molecular communication with flow and additive noise," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3, pp. 350–362, Sep. 2014.

[19] L.-S. Meng, P.-C. Yeh, K.-C. Chen, and I. F. Akyildiz, "On receiver design for diffusion-based molecular communication," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6032–6044, Nov. 2014.

[20] B. Tepekule, A. Pusane, M. Kuran, and T. Tugcu, "A novel pre-equalization method for molecular communication via diffusion in nanonetworks," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1311–1314, Aug. 2015.

[21] B. Li, M. Sun, S. Wang, W. Guo, and C. Zhao, "Low-complexity non-coherent signal detection for nano-scale molecular communications," *IEEE Transactions on NanoBioscience*, 2015, doi: 10.1109/TNB.2015.2504542.

[22] M. Pierobon and I. F. Akyildiz, "Information capacity of diffusion-based molecular communication in nanonetworks," in *Proceedings IEEE INFOCOM*, 2011, pp. 506–510.

[23] ——, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2532–2547, Jun. 2011.

[24] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*. Springer, 2003.

[25] M. Pierobon and I. F. Akyildiz, "A physical end-to-end model for molecular communication in nanonetworks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 4, pp. 602–611, May 2010.

[26] H. ShahMohammadian, G. Messier, and S. Magierowski, "Blind synchronization in diffusion-based molecular communication channels," *IEEE communications letters*, vol. 17, no. 11, pp. 2156–2159, Nov. 2013.

[27] S. Kay, *Fundamentals of statistical signal processing, volume 1: Estimation theory*. Prentice Hall, 1998.

[28] L. S. Meng, P. C. Yeh, K. C. Chen, and I. F. Akyildiz, "On Receiver Design for Diffusion-Based Molecular Communication," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6032–6044, Nov. 2014.

[29] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks," in *IEEE International Conference on Communications (ICC)*, 2011, pp. 1–5.

[30] J. G. Prokias, *Digital Communication*, 3rd ed. McGraw Hill, Oct. 1995.

[31] A. Noel, K. C. Cheung, and R. Schober, "Joint channel parameter estimation via diffusive molecular communication," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 1, pp. 4–17, Mar. 2015.

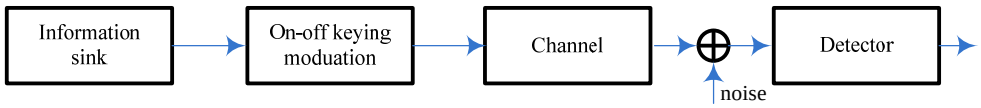


Fig. 1. A physical-layer schematic structure of molecular communications.

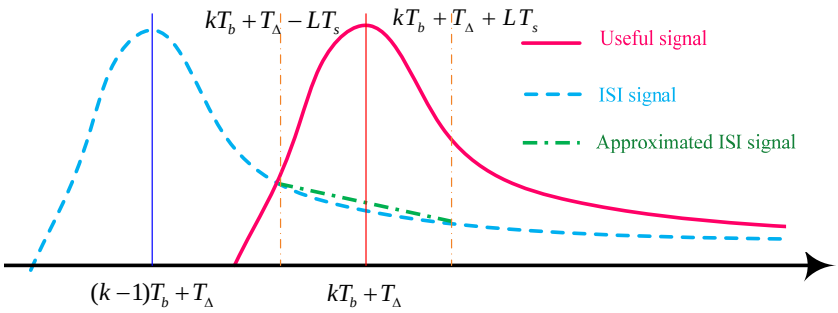


Fig. 2. The linear approximation of ISI generated from the previous concentration.

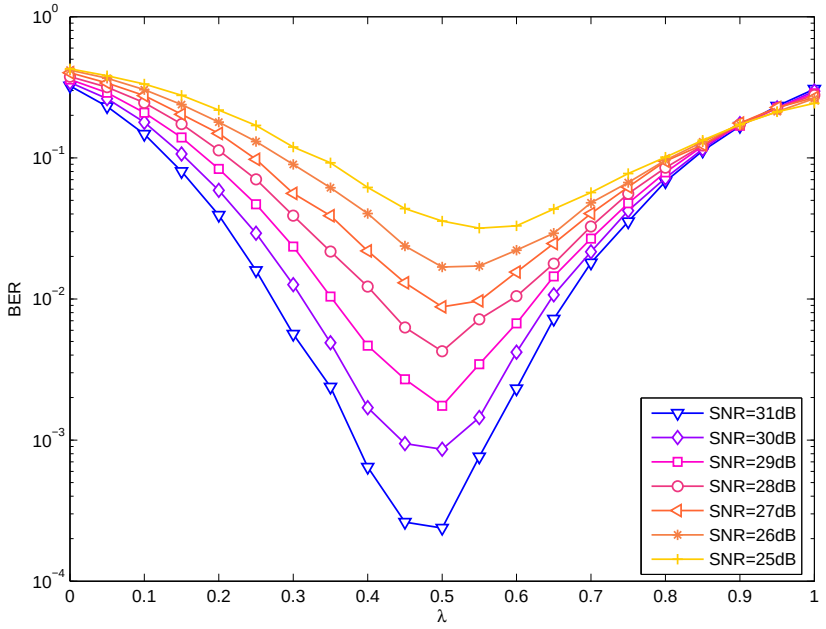


Fig. 3. BER performance of the non-coherent detector with various threshold configurations or different weighting factors.

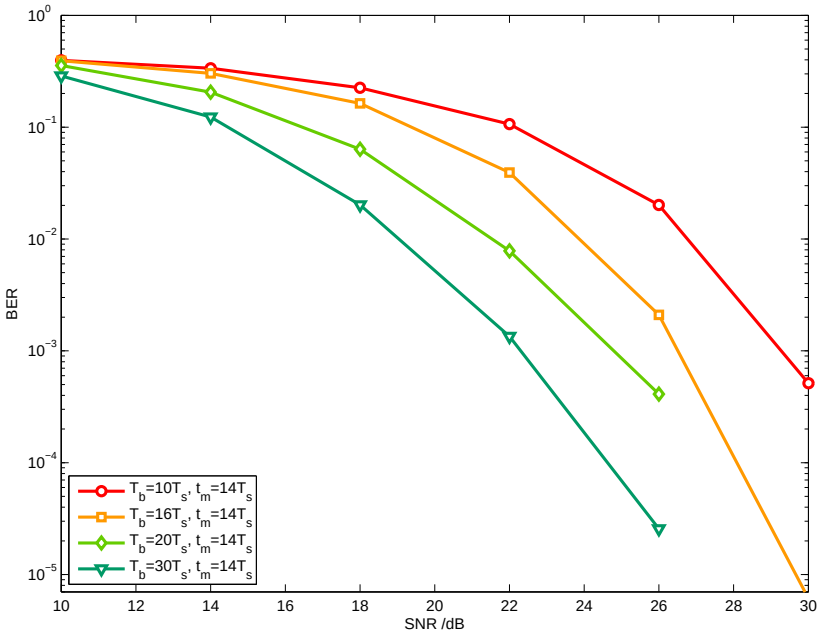


Fig. 4. BER performance of the non-coherent detector with different symbol durations.

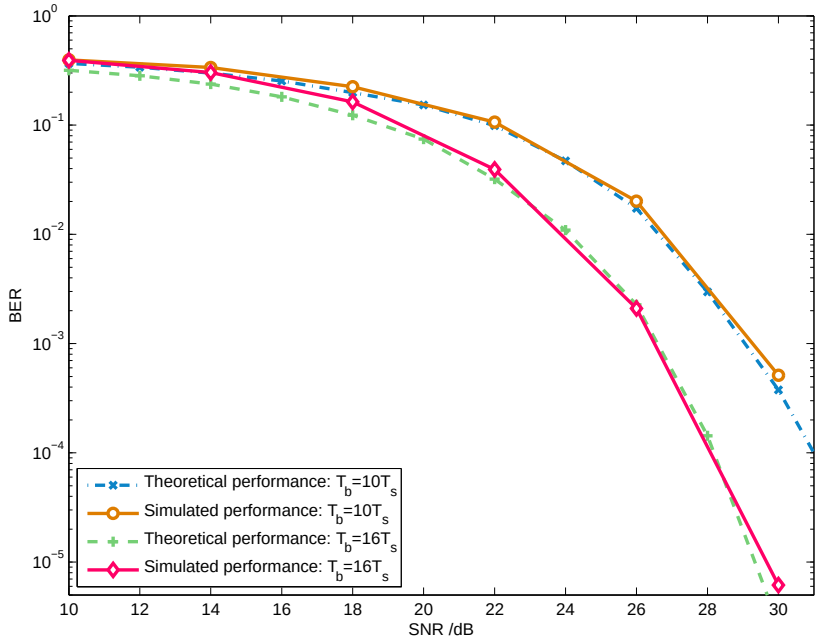


Fig. 5. Theoretical BER performance vs simulation results.

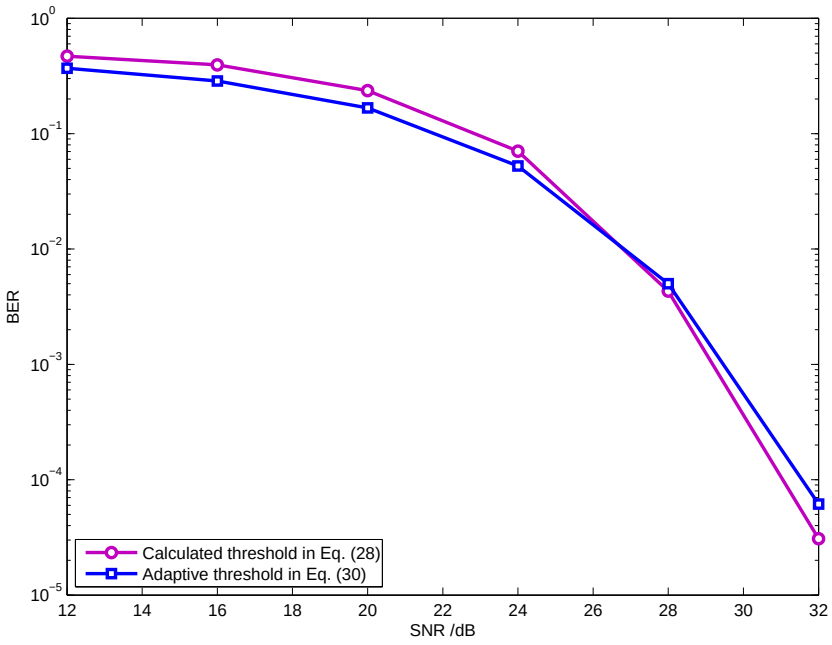


Fig. 6. BER performance of the calculated threshold vs the adaptive threshold.

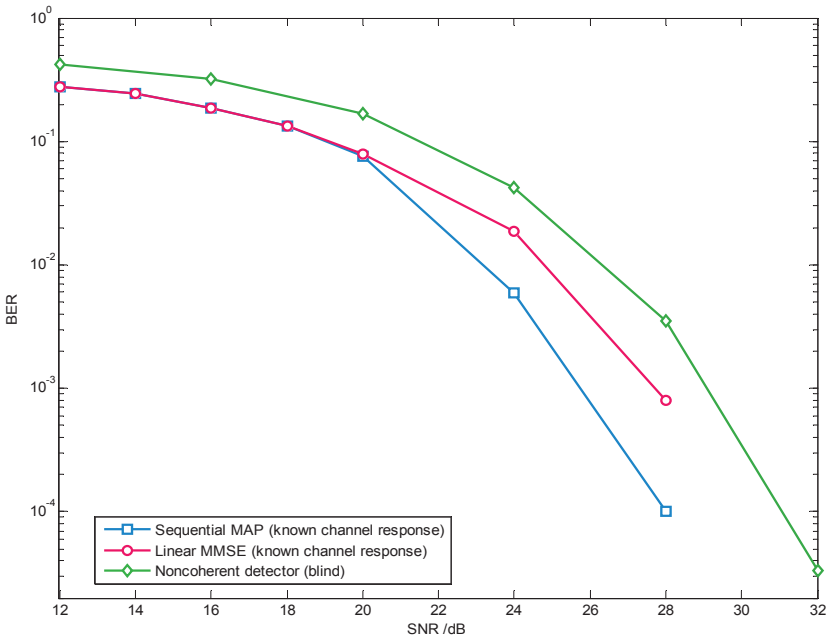


Fig. 7. BER performance of existing coherent detection schemes (MAP and MMSE) and the proposed non-coherent detection technique.

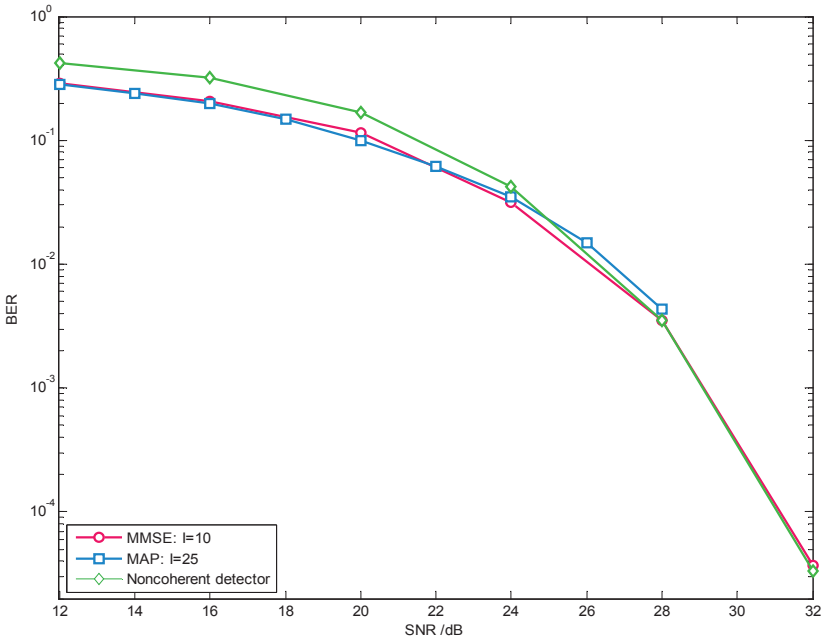
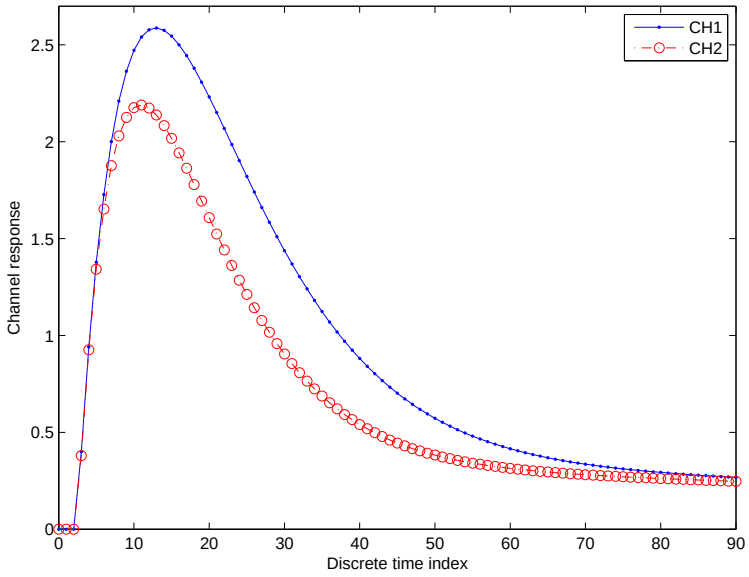
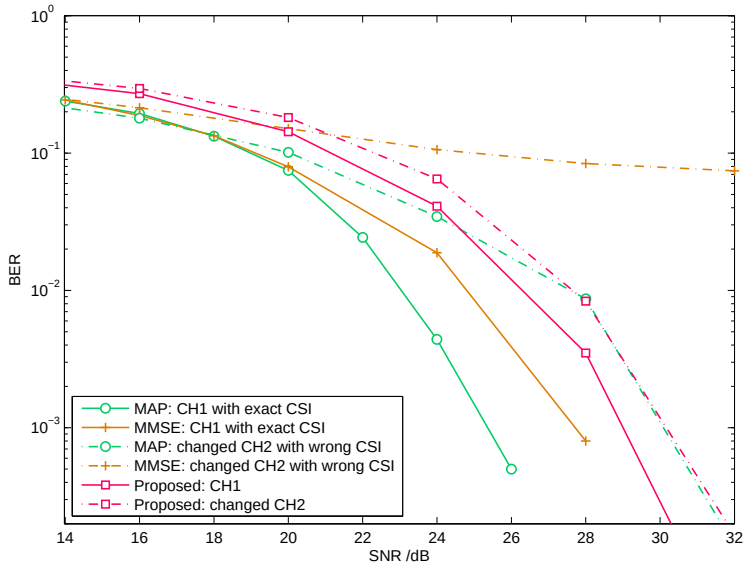


Fig. 8. BER performance of both coherent detection schemes (MAP and MMSE) and the non-coherent scheme with different ISI truncation length.



(a) Time-varying channel response. CH2 denotes the varied channel response after a period of time.



(b) BER performance of both coherent detection schemes (MAP and MMSE) and the non-coherent scheme with varied channel response

Fig. 9. BER performance with channel estimation errors or channel variations.