

Higher Order Derivatives: Improved Pre-Processing and Receivers for Molecular Communications

Mustafa Can Gursoy and Urbashi Mitra

Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA

E-mail: {mgursoy, ubli}@usc.edu

Abstract—Significant inter-symbol interference (ISI) challenges the achievement of reliable high data-rate molecular communication (MC) links. Inspired by recent results showing the ISI mitigation capability of pre-processing received signals by differentiation, the impact of using higher order derivatives is studied herein. The trade-off between ISI mitigation and noise amplification with higher order derivatives is characterized. Optimal maximum-likelihood sequence detection (MLSD) is investigated as well as low complexity banded MLSD to exploit the pulse narrowing induced by differentiation. Furthermore, analysis suggests the existence of an optimal derivative order. The bit error ratio (BER) for a fixed threshold detector is tightly approximated and employed to find the optimal derivative order. Numerical results confirm the aforementioned trade-off and show that reliable communication can be established using symbol durations considerably smaller than the peak time.

Index Terms—Diffusive molecular communications, receiver design, MLSD, higher order derivatives, pre-processing, detector design.

I. INTRODUCTION

Molecular communication via diffusion (MCvD) employs the diffusive nature of chemical signals to communicate [1]. After their release by the transmitter, the emitted molecules randomly propagate and arrive at the receiver in a probabilistic fashion, which results in signal attenuation and inter-symbol interference (ISI). Higher data rates further exacerbate ISI in MCvD channels [2].

Numerous transmitter and receiver strategies have been proposed to mitigate the effects of ISI. Different modulation strategies employing single or multiple molecule types have been considered [3]–[6], as well as pre-equalization methods [7], [8]. Motivated by the finite-impulse response type behavior of the MCvD channel, maximum *a posteriori* (MAP) and maximum likelihood (ML) sequence detectors are proposed in [9]. Due to the high computational complexity of these detectors, lower complexity equalizers (e.g. decision feedback and minimum mean squared error) are also proposed in [9].

A recent work provides a promising alternative strategy: applying a discrete-time derivative on the over-sampled received signal to provide strong ISI reduction at the cost of noise enhancement. As the peak time of the derivative signal precedes the peak of the arrival density function, [10] shows that given the MCvD system can operate using high transmission power, high data rates can be achieved. Exploiting

this phenomenon, [11] proposes a rising-edge-based detection algorithm and shows the beneficial properties of the derivative signal on a macro-scale molecular MIMO testbed. We see from these prior works that the derivative signal induces a less diffusive signal: the signal peak occurs earlier and the decay rate of the differentiated signal is faster after the peak.

This paper extends these earlier works. In particular,

- 1) We generalize the initial findings of [10] and suggest a pre-processing framework based on $m > 0$ derivatives.
- 2) We derive the maximum-likelihood sequence detection (MLSD) [9] for an arbitrary derivative order and propose a reduced complexity banded MLSD.
- 3) The trade-off between the memory in the banded MLSD and derivative order is investigated.
- 4) The BER of a fixed threshold detector is tightly approximated and employed to determine the optimal derivative order to be paired with a threshold detector.
- 5) Numerical analysis coupled with simulation suggests an inherent trade-off between noise, ISI and the applied derivative-order validating the notion of an optimal derivative order.
- 6) Finally, strong performance of derivative pre-processed communication necessitates transmitters with a large molecule pool, which is verified via numerical simulation.

The rest of the paper is organized as follows: Section II presents the channel model. Section III elaborates upon the benefits of higher order derivative pre-processing. Section IV presents the optimal detector as well as proposing a lower-complexity alternative. Section V derives an approximate error probability expression to be used as a cost function in derivative order optimization. Section VI presents the numerical results and Section VII concludes the paper.

II. CHANNEL MODEL

We assume a point transmitter and a synchronized spherical absorbing receiver residing in a three-dimensional unbounded environment. The spherical receiver's radius is denoted by r_r , and the distance between the transmitter and the center of the spherical receiver is denoted by r_0 . The overall topology of interest is presented in Fig 1. Denoting the diffusion coefficient of the utilized molecule as D and $d = r_0 - r_r$, [12] computes the arrival density function with respect to time as

$$f_{hit}(t) = \frac{r_r}{d + r_r} \frac{1}{\sqrt{4\pi Dt}} \frac{d}{t} e^{-\frac{d^2}{4Dt}}, \quad t \in (0, \infty). \quad (1)$$

This research has been funded in part by one or more of the following grants: ONR N00014-15-1-2550, NSF CCF-1718560, NSF CCF-1410009, NSF CPS-1446901, NSF CCF-1817200, and ARO W911NF1910269.

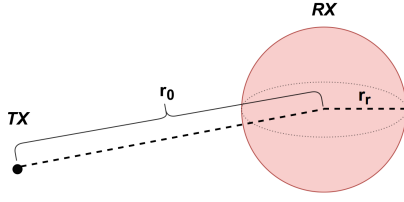


Fig. 1. The considered transmission/reception environment.

In sequential bit transmission scenarios, the time-slotted MCvD system is characterized by the channel coefficient vector $\mathbf{h}$, whose elements can be calculated by

$$h[n] = \int_{(n-1)t_s}^{nt_s} f_{hit}(t) dt, \quad n = 1, 2, \dots \quad (2)$$

where t_s is the sample duration. Note that denoting the duration of a binary concentration shift keying (BCSK) symbol as t_b and the samples per bit parameter as N , one can write $t_s = \frac{t_b}{N}$.

The number of molecules counted by the receiver can be modeled by the linear time-invariant (LTI)-Poisson channel [13], where each sample is a Poisson distributed random variable with rate parameter according to the convolution of the channel with the transmitted sequence. Throughout the paper, we also consider an external Poisson noise process with a rate of λ_s molecules per sample. Overall, using the Gaussian approximation of the LTI-Poisson arrival signal [14], the noisy received signal is approximately distributed as

$$y[n] \sim \mathcal{N}(\mu[n], \mu[n]), \quad (3)$$

where,

$$\mu[n] = \left(\sum_{k=1}^{LN} h[k] x[n-k+1] \right) + \lambda_s. \quad (4)$$

Here, L denotes the channel memory in bits, and N is the number of samples per bit duration. Furthermore, $\mathbf{x}$ denotes the overall transmission sequence. We assume an idealized transmitter in this paper, which implies that $\mathbf{x}$ has a one-to-one deterministic relationship with the transmitted bit sequence $\mathbf{s}$. Specifically, using BCSK [3], $x[n]$ can only be non-zero at the transmission instant $n = (i-1)N + 1$ and if $s[i] = 1$. We employ BCSK throughout the paper and denote the number of emitted molecules to transmit a bit-1 by M .

In discrete time, convolution with the channel can be represented by a Toeplitz matrix $\mathbf{H}$, thus our discrete-time received signal is given by

$$\mathbf{y} = (\mathbf{H}\mathbf{x} + \lambda_s \mathbf{j}) + \boldsymbol{\eta}, \quad (5)$$

where all vectors are of dimension $SN \times 1$ and $\mathbf{H}$ is square. Furthermore, S is the length of the transmission block¹, $\mathbf{j}$ denotes the SN -vector of ones, and $\boldsymbol{\eta}$ is the noise vector. Note that due to the channel characterization presented in (4), the variance of the n^{th} element of $\boldsymbol{\eta}$ is $\mu[n]$. In other words, $\boldsymbol{\eta} \sim$

¹We assume guard bands of no-transmission between the transmission blocks of length S .

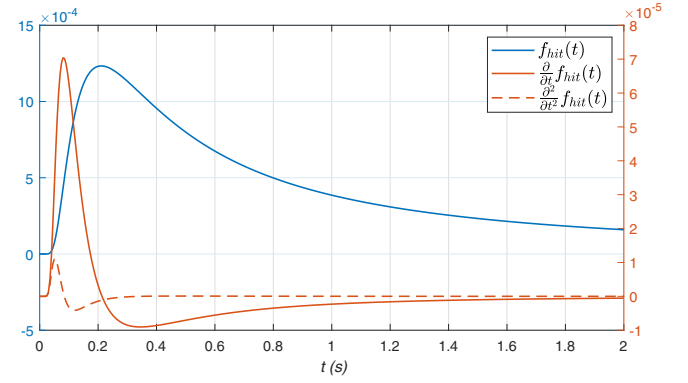


Fig. 2. $f_{hit}(t)$, $\frac{\partial f_{hit}(t)}{\partial t}$, and $\frac{\partial^2 f_{hit}(t)}{\partial t^2}$ curves with respect to time (correspond to blue, orange, and orange (dashed) curved respectively). $d = 10 \mu\text{m}$, $r_r = 5 \mu\text{m}$, $D = 80 \mu\text{m}^2 \text{s}^{-1}$. Note that the magnitudes of the first and second order derivatives are an order of magnitude smaller than the original signal.

$\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \text{diag}\{\mathbf{H}\mathbf{x}\} + \lambda_s \mathbf{I}$. The first summand in the covariance matrix is widely referred to as the signal-dependent noise of molecular communications [9], [14].

III. HIGHER ORDER DERIVATIVE-BASED PRE-PROCESSING

A. Analysis of Derivative Pre-processing

We first investigate properties of the pulse shape after applying an arbitrary number of differentiations on $f_{hit}(t)$. We define $t_{p,(m)}$ as the leftmost local maximum of $\frac{\partial^m f_{hit}}{\partial t^m}$ (i.e. the peak time of $\frac{\partial^m f_{hit}}{\partial t^m}$).

Proposition 1. $t_{p,(m)}$ is a monotonically decreasing function of m .

Sketch of Proof. The proof follows from directly calculating the leftmost vanishing points of the derivatives of (1). It should be noted that at the leftmost vanishing point of $\frac{\partial^{m+1} f_{hit}}{\partial t^{m+1}}$, $\frac{\partial^{m+2} f_{hit}}{\partial t^{m+2}}$ is always negative, ensuring the first local extremum of $\frac{\partial^m f_{hit}}{\partial t^m}$ is indeed a local maximum. $\square$

For visualization purposes, Figure 2 presents $f_{hit}(t)$ and its first and second derivatives. As proven, with increasing derivative order, the peak shifts leftwards. Furthermore, one interesting observation from Figure 2 is that the right tail is decaying considerably faster as the derivative order increases, suggesting a narrowing of the effective pulse shape duration. These two phenomena suggest that using higher order derivatives in pre-processing can allow communicating at data rates that are considerably faster than conventional MCvD. However, we note that this decay of the right tail provides better ISI mitigation at the cost of a reduction in signal amplitude. We next develop the discrete-time model for m^{th} derivative pre-processing.

B. Generalized Derivative-Based Pre-Processing

In this subsection, we generalize the initial idea of [10] and present a generalized framework on using discrete-time forward derivatives before detection in MCvD receivers. First,

we introduce the $SN \times SN$ discrete-time forward derivative operator $\mathbf{D}$

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & -1 & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}. \quad (6)$$

Note that using such a notation, $\mathbf{D}^m$ corresponds to the discrete-time m^{th} order derivative operator. After the application of the m^{th} order derivative, the pre-processed arrival vector can be represented as:

$$\mathbf{y}_{(m)} = \mathbf{D}^m \mathbf{y} = \mathbf{D}^m (\mathbf{H}\mathbf{x} + \lambda_s \mathbf{j} + \boldsymbol{\eta}). \quad (7)$$

It should be noted that the curves in Figure 2 characterize the *expected* continuous time arrival signal for a single-shot transmission. In a time slotted communication system with consequent bit transmissions, the convolutional nature of the channel and the variance of the arrival signal are also in effect, as described by (5). The effect of the m^{th} order derivative operator on the signal-dependent variance sequence is represented in the $\boldsymbol{\eta}_{(m)} = \mathbf{D}^m \boldsymbol{\eta}$ term in (7) resulting in the noise covariance: $\boldsymbol{\Sigma}_{(m)} = \mathbf{D}^m \boldsymbol{\Sigma} (\mathbf{D}^T)^m$. The structure of the transformed noise covariance suggests noise enhancement and subsequent noise correlation due to the differentiation process. In fact, as m increases, the banded diagonal of $\mathbf{D}^m$ increases in width, further increasing coloration and noise enhancement.

IV. DETECTOR DESIGN

A. Optimal Detector

We develop the maximum-likelihood sequence detector [9] after differentiation. Given knowledge of the channel impulse response and λ_s at the receiver, the resultant detector is

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \max_{\mathbf{x}} P(\mathbf{y}_{(m)} | \mathbf{x}) \\ &= \arg \min_{\mathbf{x}} \left\{ \ln(|\boldsymbol{\Sigma}_{(m)}|) + \right. \\ &\quad \left. (\mathbf{y}_{(m)} - \mathbf{w}_{(m)})^T \boldsymbol{\Sigma}_{(m)}^{-1} (\mathbf{y}_{(m)} - \mathbf{w}_{(m)}) \right\}, \end{aligned} \quad (8)$$

where $\mathbf{w}_{(m)} = \mathbf{D}^m (\mathbf{H}\mathbf{x} + \lambda_s \mathbf{j})$ is the processed mean vector. Given properties of maximum likelihood detection and one-to-one functions, the detected transmission vector sequence $\hat{\mathbf{x}}$ directly yields $\hat{s}$.

B. Banded MLSD

Given the pulse narrowing effect of differentiation, we can reduce the complexity of MLSD (2^S for S bits) by approximating the duration of the pulse shape to be less than L' , [15], [16]. Our goal is to design a receiver amenable to implementation in a micro- or nano-scale device. Thus, we propose a reduced memory detector based on the MLSD presented in (8). The proposed detector works in an online fashion rather than acting on the block as a whole and considers a much smaller memory window of $L' < L$, including the intended bit.

At each time instant i , based on the window between $(i - L' + 1)^{\text{th}}$ and i^{th} bits (both inclusive), the detector performs $2^{L'}$ likelihood computations. We first define $\mathbf{H}^{i,L'}$, $\mathbf{x}'$, and $(\mathbf{D}^m)^{i,L'}$ as the corresponding channel, candidate transmission sequence, and derivative operator matrices/vectors for the shorter memory (L') consideration, respectively. We then collect the received signal vector for this observation window, denoted by $\mathbf{y}^{i,L'}$ (i.e. $\mathbf{y}^{i,L'} = [y[(i - L')N + 1] \dots y[iN]]^T$), and apply the $(\mathbf{D}^m)^{i,L'}$ operator on it. We then extract the *useful* components and their statistics from these vectors/matrices. Note that we are only interested in the samples corresponding to the last bit in the L' -bit candidate string. Furthermore, we remove the last m samples, and only consider the $(N - m) \times 1$ vectors and $(N - m) \times (N - m)$ matrices corresponding to the first $(N - m)$ samples of the intended bit. This operation is due to the very nature of $\mathbf{D}^m$ and is done to avoid adding ISI into the system by mixing the intended bit's samples with samples from the next transmitted bit. Using the τ subscript to denote these truncated vectors and matrices, the proposed detector detects the i^{th} transmitted bit by

$$\begin{aligned} \hat{s}[i] &= \arg \min_{\mathbf{x}'} \left\{ (\mathbf{y}_{(m),\tau}^{i,L'} - (\mathbf{D}^m)^{i,L'}_{\tau} \mathbf{H}_{\tau}^{i,L'} \mathbf{x}'_{\tau})^T \boldsymbol{\Sigma}_{(m),\tau}^{i,L'}{}^{-1} \right. \\ &\quad \left. (\mathbf{y}_{(m),\tau}^{i,L'} - (\mathbf{D}^m)^{i,L'}_{\tau} \mathbf{H}_{\tau}^{i,L'} \mathbf{x}'_{\tau}) + \ln(|\boldsymbol{\Sigma}_{(m),\tau}^{i,L'}|) \right\}. \end{aligned} \quad (9)$$

Per one block, the proposed detector is of complexity $\mathcal{O}(2^{L'}S)$, compared to the $\mathcal{O}(2^S)$ of (8) and $\mathcal{O}(2^L S)$ of the optimal Viterbi decoder.

C. Fixed Threshold Detector

Since many micro- or nano-scale applications of MCvD may require simplistic transceiver structures, using even simpler detectors may be desirable. Motivated by this goal, as proposed in [10], a fixed threshold detector described as

$$\hat{s}[i] = \max(y_{(m)}[(i-1)N+1], \dots, y_{(m)}[iN-m]) \stackrel{1}{\geq} \gamma. \quad (10)$$

can be employed, where γ is the designed threshold. Using arguments similar to that in Section IV-B, the last m samples of the pre-processed arrival vector are discarded.

V. ERROR ANALYSIS AND THE OPTIMIZATION OF m

As noted in Section III, the order of the discrete-time forward derivative operator m reveals a fundamental trade-off between the ISI mitigation capability of the receiver and the incurred noise amplification. This leads to m being an optimization parameter in terms of error performance. It is this optimization which we will address herein. Motivated by the widespread use of fixed threshold detectors in the literature [2], [10], [17], we provide an approximation on the bit error ratio of a receiver employing the $\mathbf{D}^m$ pre-processing and utilizing the detector described in (10). We will then use this approximation as a cost function over which we optimize m for a fixed threshold detector-based receiver.

In order to fully observe the effects of a memory of L symbols, we consider a bit stream of length L . In this hypothetical bit stream, the approximate expected arrival counts to the last bit's samples can be written as

$$\begin{aligned}\boldsymbol{\mu}_L &= E(\mathbf{y}_L | \mathbf{x}_L) = E\left(\begin{bmatrix} y[(L-1)N+1] \\ \vdots \\ y[LN] \end{bmatrix} \middle| \mathbf{x}_L\right) \\ &= \mathbf{H}_L \mathbf{x}_L + \lambda_s \mathbf{j}_N,\end{aligned}\quad (11)$$

where $\mathbf{j}_N$ is an N -vector of ones, $\mathbf{x}_L$ is the LN samples long transmission sequence, corresponding to a certain L bit-long stream $\mathbf{s}_L$, and

$$\mathbf{H}_L = \begin{bmatrix} h[(L-1)N+1] & \cdots & h[1] & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \ddots & \vdots \\ h[LN-1] & \cdots & h[2] & h[1] & 0 & \\ h[LN] & \cdots & h[3] & h[2] & h[1] & \end{bmatrix}. \quad (12)$$

Note that fully capturing a window of length L requires evaluating 2^L combinations of $\mathbf{s}_L$. Motivated by the immense computational complexity of this operation, we consider a shorter memory, denoted by L' . Upon this point, the subscript L' denotes the vectors/matrices generated with L' bits into consideration. That being said, the associated covariance matrix is $\boldsymbol{\Sigma}_{L'} = \text{diag}\{\boldsymbol{\mu}_{L'}\}$, due to the LTI-Poisson channel behavior and its Gaussian approximation described in Equations (3)-(4). Furthermore, we note that using an m^{th} order derivative operator $\mathbf{D}^m$ of size $N \times N$, the processed vector is distributed as $\mathbf{y}_{L', (m)} \sim \mathcal{N}(\boldsymbol{\mu}_{L', (m)}, \boldsymbol{\Sigma}_{L', (m)})$, where $\boldsymbol{\mu}_{L', (m)} = \mathbf{D}^m \boldsymbol{\mu}_{L'}$ and $\boldsymbol{\Sigma}_{L', (m)} = \mathbf{D}^m \boldsymbol{\Sigma}_{L'} (\mathbf{D}^T)^m$.

After obtaining the arrival statistics, we express the bit error ratio expression as an average of conditional error probabilities. Conditioning on the first $L' - 1$ elements of $\mathbf{s}_{L'}$, denoted by $\mathbf{s}_{L', \text{ISI}}$, we write

$$P_{e, L'} = \frac{1}{2^{L'-1}} \left(\sum_{\forall \mathbf{s}_{L', \text{ISI}}} P_{e | \mathbf{s}_{L', \text{ISI}}} \right), \quad (13)$$

where

$$P_{e | \mathbf{s}_{L', \text{ISI}}} = \frac{1}{2} \left(P(R_{(m)} < \gamma | \mathbf{s}_{L'}[L'] = 1) + P(R_{(m)} > \gamma | \mathbf{s}_{L'}[L'] = 0) \right). \quad (14)$$

Here, $R_{(m)} = \max_j y_{L', (m)}[j]$. Note that $R_{(m)}$ denotes the maximum of $N - m$ correlated and differently distributed Gaussian random variables². At this point, we invoke Clark's approximation to approximate $R_{(m)}$ as a normal random variable in a recursive manner [18]. The recursion is motivated by the fact that $\max(A, B, C) = \max[\max(A, B), C]$ and operates by approximating the PDF of the maximum as a Gaussian in each iteration. Herein, we exemplify the first iteration of the recursion that finds the moments of $\max(y_{L', (m)}[1], y_{L', (m)}[2])$. Note that the following approach

²Due to the strategy employed in (10), we remove the last m samples in the calculations.

can be further continued until $y_{L', (m)}[N - m]$ to approximate $R_{(m)}$. We first define two parameters, a and α as

$$\begin{aligned}a &= \sqrt{\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2 - 2\tilde{\sigma}_{1,2}} \\ \alpha &= \frac{\tilde{\mu}_1 - \tilde{\mu}_2}{a}\end{aligned}\quad (15)$$

where $\tilde{\mu}_p$ and $\tilde{\sigma}_{p,q}$ values are their corresponding entries in $\boldsymbol{\mu}_{L', (m)}$ and $\boldsymbol{\Sigma}_{L', (m)}$, respectively. Denoting $\psi(\cdot)$ as the standard normal PDF, [18] finds the first and second moments of $\max(y_{L', (m)}[1], y_{L', (m)}[2])$, $\tilde{\nu}_1$ and $\tilde{\nu}_2$, as

$$\begin{aligned}\tilde{\nu}_1 &= \tilde{\mu}_1 Q(-\alpha) + \tilde{\mu}_2 Q(\alpha) + a\psi(\alpha) \\ \tilde{\nu}_2 &= (\tilde{\mu}_1^2 + \tilde{\sigma}_1^2)Q(-\alpha) + (\tilde{\mu}_2^2 + \tilde{\sigma}_2^2)Q(\alpha) \\ &\quad + (\tilde{\mu}_1 + \tilde{\mu}_2)a\psi(\alpha),\end{aligned}\quad (16)$$

where $Q(\cdot)$ is the Gaussian Q -function. After obtaining $\tilde{\nu}_1$ and $\tilde{\nu}_2$, we approximate $\max(y_{L', (m)}[1], y_{L', (m)}[2]) = \tilde{z}_{1,2} \sim \mathcal{N}(\tilde{\nu}_1, \tilde{\nu}_2 - \tilde{\nu}_1^2)$, and continue with the iteration. Note that in reality, maximum of two Gaussians is *not* a Gaussian itself [18], [19], but we approximate as such to come up with a tractable and recursive approach.

Upon approximating the first and second order statistics of $R_{(m)}$ conditioned on $\mathbf{s}_{L'}[L'] = i$ as $\mu_{R, (m) | i}$ and $\sigma_{R, (m) | i}^2$, respectively, we recall to treat $R_{(m)}$ as a normal random variable, and obtain

$$P(R_{(m)} < \gamma | \mathbf{s}_{L'}[L'] = 1) \approx Q\left(\frac{\mu_{R, (m) | 1} - \gamma}{\sigma_{R, (m) | 1}}\right) \quad (17a)$$

$$P(R_{(m)} > \gamma | \mathbf{s}_{L'}[L'] = 0) \approx Q\left(\frac{\gamma - \mu_{R, (m) | 0}}{\sigma_{R, (m) | 0}}\right). \quad (17b)$$

Note that the means and variances are different for (17a)-(17b), as they are obtained through different conditionings on the intended bit.

Overall, the presented approach becomes most accurate when one considers all 2^L possible combinations of $\mathbf{s}_L$, rather than $2^{L'}$. However, using Equations (11)-(17) with L' still provides a useful sub-optimal cost function as a proxy of the true error probability, hence can be used to optimize m . Furthermore, the expressions need to be computed only once for each candidate m value (before data transmission), which makes utilizing them with L' a feasible approach. The accuracy of the approach is presented in Section VI.

VI. NUMERICAL RESULTS

This section presents numerical bit error ratio (BER) results regarding different detector strategies, the proposed cost function, and the derivative orders. The BER simulations are performed using the channel model described by Equations (1)-(5). In order to scale the signal transmission power of the transmitter and the external noise power, we define the signal-to-noise ratio (SNR) as

$$\text{SNR} = \frac{M}{N\lambda_s}. \quad (18)$$

Note that the denominator $N\lambda_s$ is equivalent to the Poisson noise rate parameter for the duration of a whole bit, hence our definition of SNR is normalized on a per-bit basis.

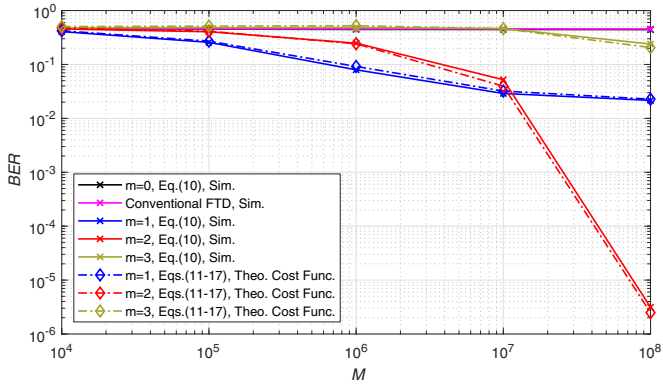


Fig. 3. BER vs. M using Equation (10). $S_r = 0.5$, $\text{SNR} = 10\text{dB}$, $d = 10\text{ }\mu\text{m}$, $r_r = 5\text{ }\mu\text{m}$, $D = 80\text{ }\mu\text{m}^2\text{ s}^{-1}$, $N = 5$, $S = 1000\text{ bits}$, $L = 100$, $L' = 10$. γ values numerically optimized through exhaustive search.

As another key parameter, we define the unitless parameter $S_r = \frac{t_b}{t_{p,(0)}}$, where $t_{p,(0)} = \frac{d^2}{6D}$ denotes the peak time of $f_{hit}(t)$. Note that S_r is a measure of how fast the system is communicating with respect to the channel peak time (i.e. smaller S_r means higher data rate).

A. BER vs. M , and the Effectiveness of the Cost Function

First, we compare the simulated BERs for $m = 0, 1, 2$, and 3 , with the associated approximations obtained by Equations (11)-(17) in Figure 3. The BER of the conventional detector, which is a fixed threshold detector (FTD) on the total number of arrivals within a bit duration [3], is also presented in the figure. Note that the case where $m = 1$ corresponds to the approach considered in [10].

There are several conclusions to be drawn from Figure 3. Firstly, the approximations for the BER are tight, despite considering a greatly reduced memory of $L' = 10$ versus the actual $L = 100$. Note that this is due to the strong pulse narrowing introduced by the derivative operator. The key is that the approximated BER provides the accurate trend for the impact of the differentiation order relative to the simulated performance. Furthermore, we see that there is a correlation between the number of molecules, M and the differentiation order m . For smaller M , a smaller m yields improved BER. This phenomenon is mainly due to the noise amplification incurred at each derivative operation. However, as M increases, this adverse effect is less of an issue, and the system benefits more from the ISI mitigation provided by higher orders (in this case $m = 2$).

B. Equation (9): BER vs. L'

We further investigate the choice of L' in Figure 4 on the performance of banded equalization (9). Here too, there is a trade-off between L' and m which is unsurprising. That is, as m increases, the effective pulse shape narrows and decays more quickly, resulting in less ISI. Thus, a smaller L' suffices. However, as L' increases, $m = 1$ provides smaller noise amplification. Overall, for the system parameters employed in

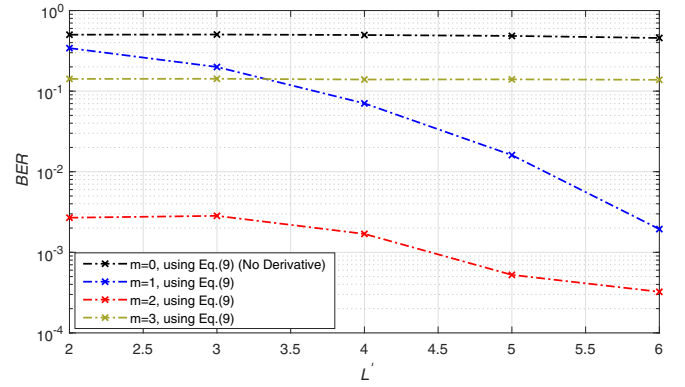


Fig. 4. BER vs. L' using banded MLSD in Equation (9). $S_r = 0.5$, $M = 10^7$ molecules, $\text{SNR} = 20\text{dB}$, $d = 10\text{ }\mu\text{m}$, $r_r = 5\text{ }\mu\text{m}$, $D = 80\text{ }\mu\text{m}^2\text{ s}^{-1}$, $N = 5$, $S = 1000\text{ bits}$, $L = 100$.

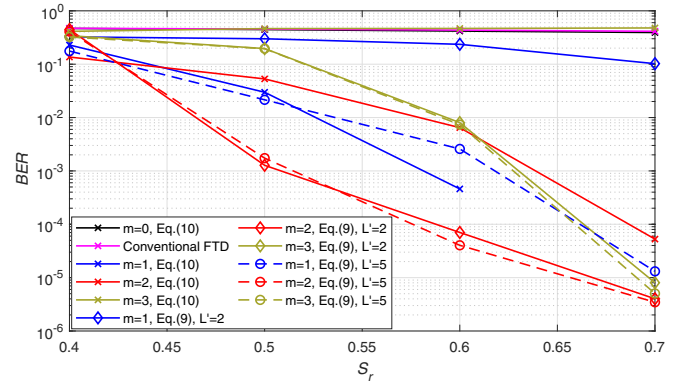


Fig. 5. BER vs. S_r . $M = 10^7$ molecules, $\text{SNR} = 10\text{dB}$, $d = 10\text{ }\mu\text{m}$, $r_r = 5\text{ }\mu\text{m}$, $D = 80\text{ }\mu\text{m}^2\text{ s}^{-1}$, $N = 5$, $S = 1000\text{ bits}$, $L = 100$. γ values numerically optimized through exhaustive search. For $m = 1$ and $S_r = 0.7$, $\text{BER} < 10^{-6}$.

Figure 4, a BER of 2.7×10^{-3} can be obtained with $m = 2$ by using a memory as small as $L' = 2$.

C. Error Performance and Data Rate

In this subsection, we provide BER curves with respect to S_r through presenting Figure 5. Recalling $S_r = \frac{t_b}{t_{p,(0)}}$ as a measure of data rate, going leftward in Figure 5 is equivalent to communicating faster. Note that since $S_r < 1$, all data points in Figure 5 correspond to a bit duration that is smaller than the channel peak time.

The results of Figure 5 suggest that for the system with considered parameters, $m = 2$ provides the best overall error performance. At high data rates, ISI is extremely high and the consideration of $m = 1$ does not sufficiently mitigate it. This phenomenon can also be validated from the large performance gap between banded MLSD results with $L' = 2$ and 5 for $m = 1$. Note that on the contrary, since $m = 2$ and 3 narrow the effective pulse duration more effectively, banded MLSDs' performances do not vary drastically with L' . Secondly, it can be observed from Figure 5 that $m = 2$ also outperforms the selection of $m = 3$. We note that although a larger m results in better ISI mitigation, the performance of $m = 3$ is limited due

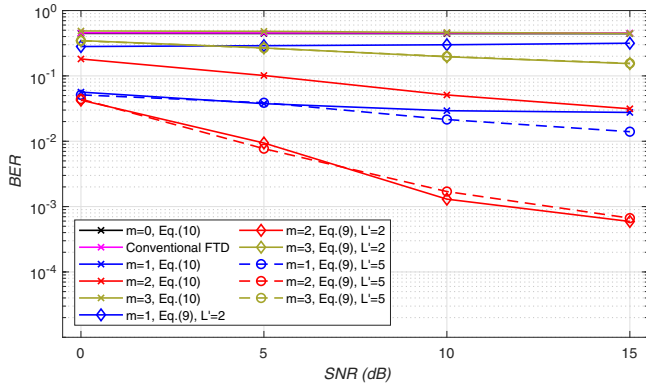


Fig. 6. BER vs. SNR. $S_r = 0.5$, $M = 10^7$ molecules, $SNR = 10\text{dB}$, $d = 10\text{ }\mu\text{m}$, $r_r = 5\text{ }\mu\text{m}$, $D = 80\text{ }\mu\text{m}^2\text{s}^{-1}$, $N = 5$, $S = 1000$ bits, $L = 100$. γ values numerically optimized through exhaustive search.

to its higher noise amplification. Our empirical observations suggest that when M increases even further, there is a point where $m = 3$ outperforms $m = 2$, confirming the inferences made in Subsection VI-A.

Overall, with the proper selection of $m = 2$, a BER of 1.3×10^{-3} can be obtained with a banded MLSD that performs only $2^2 = 4$ likelihood computations per bit, when $S_r = 0.5$. Note that whilst the derivative-based pre-processing can achieve such a performance, extreme ISI causes the conventional methods to yield unreliably high error rates.

D. BER vs. SNR

This subsection presents the error performance with respect to the external noise power, λ_s . Recalling the definition of SNR from (18), Figure 6 presents the relationship with BER and SNR.

As expected, the results of Figure 6 show that error performance improves with increasing SNR. However, it can be observed that derivative orders do not benefit from increasing SNR equally. Similar to the arguments made in Subsection VI-C, the selection of $m = 1$ provides inadequate ISI mitigation and faces undesirable error floors. On the other hand, the selection of $m = 3$ also does not benefit from increasing SNR considerably. Given that $m = 3$ provides strong ISI mitigation unlike $m = 1$, it can be inferred that the amplified data-dependent noise is more detrimental to the system than external noise. Overall, for the system in Figure 6, selecting $m = 2$ yields the best error performance gain with respect to SNR.

VII. CONCLUSIONS

In this paper, a framework for utilizing higher order derivatives to pre-process the arrival signal has been provided. A trade-off between ISI combating and noise amplification is shown, hinting to an optimal derivative order that minimizes BER. Through deriving an approximate error probability expression, a cost function has been presented to find the optimal derivative order for a threshold detector. In terms of further

detector design, MLSD for a derivative pre-processed receiver is investigated. Noting the MLSD's high computational complexity, the effective pulse width narrowing of the derivative operator is exploited to propose a sub-optimal banded MLSD strategy employing less memory. Overall, our results suggest that although the cost of high transmission power must be incurred, derivative based pre-processing allows considerably faster communication, while still preserving reliability.

REFERENCES

- [1] T. Nakano, A. W. Eckford, and T. Haraguchi, *Molecular communication*. Cambridge University Press, 2013.
- [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. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, Feb. 2016.
- [3] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Apr. 2011, pp. 1–5.
- [4] H. Arjmandi, A. Gohari, M. N. Kenari, and F. Bateni, "Diffusion-based nanonetworking: A new modulation technique and performance analysis," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 645–648, Mar. 2013.
- [5] S. Pudasaini, S. Shin, and K. S. Kwak, "Run-length aware hybrid modulation scheme for diffusion-based molecular communication," in *Int. Symp. Commun. Inf. Tech. (ISCIT)*, 2014, pp. 439–442.
- [6] M. C. Gursoy, D. Seo, and U. Mitra, "Concentration and position-based hybrid modulation scheme for molecular communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [7] 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.
- [8] M. Movahednasab, M. Soleimanifar, A. Gohari, M. Nasiri-Kenari, and U. Mitra, "Adaptive transmission rate with a fixed threshold decoder for diffusion-based molecular communication," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 236–248, Jan. 2016.
- [9] 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.
- [10] H. Yan, G. Chang, Z. Ma, and L. Lin, "Derivative-based signal detection for high data rate molecular communication system," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1782–1785, Sep. 2018.
- [11] Y. Huang, X. Chen, M. Wen, L. Yang, C. Chae, and F. Ji, "A rising edge-based detection algorithm for MIMO molecular communication," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 523–527, Apr. 2020.
- [12] 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, Jun. 2014.
- [13] G. Aminian, H. Arjmandi, A. Gohari, M. Nasiri-Kenari, and U. Mitra, "Capacity of diffusion-based molecular communication networks over LTI-Poisson channels," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 2, pp. 188–201, Nov. 2015.
- [14] V. Jamali, A. Ahmadzadeh, and R. Schober, "On the design of matched filters for molecule counting receivers," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1711–1714, May 2017.
- [15] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [16] A. Kavcic and J. M. F. Moura, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Info. Theory*, vol. 46, no. 1, pp. 291–301, 2000.
- [17] B. Koo, C. Lee, H. B. Yilmaz, N. Farsad, A. Eckford, and C. Chae, "Molecular MIMO: From theory to prototype," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 600–614, 2016.
- [18] C. E. Clark, "The greatest of a finite set of random variables," *Oper. Res.*, vol. 9, no. 2, pp. 145–162, 1961.
- [19] W. R. Greer Jr. and G. J. La Cava, "Normal approximations for the greater of two normal random variables," *Omega*, vol. 7, no. 4, pp. 361–363, 1979.