Time-Slotted Transmission over Molecular Timing Channels[☆]

Yonathan Murin[a,*], Nariman Farsad[a], Mainak Chowdhury[a], Andrea Goldsmith[a]

[a] *Department of Electrical Engineering, Stanford University, Stanford, USA*

Abstract

This work studies time-slotted communication over molecular timing (MT) channels. The transmitter, assumed to be perfectly synchronized in time with the receiver, is required to send K bits to the receiver using K information particles. It *releases a single information particle in each time-slot*, where the information is encoded in the *time of release*. The receiver decodes the transmitted information based on the *random* time of arrivals of the information particles during a *finite-time* observation window. The maximum-likelihood (ML) detector is derived in terms of the permanent of a matrix involving the arrival times, and shown to to have an exponential computational complexity, thus, rendering it impractical. Therefore, two additional (practical) detectors are presented: The first is a symbol-by-symbol detector. The second is a sequence detector which is based on the Viterbi algorithm (VA), yet, the VA is used differently than in its common application in electromagnetic communications where the channels are linear. Numerical simulations indicate that the proposed sequence detection algorithm significantly improves the performance compared to the symbol-by-symbol detector. For a short number of transmitted symbols, the numerical results indicate that the performance of the proposed sequence detector closely approaches the performance of the highly complicated ML detector. Finally, the proposed sequence detector is numerically compared with a one-shot transmission scheme that releases all K particles simultaneously to send a single symbol out of a constellation of size 2^K. It is shown that while for a small number of bits the one-shot scheme is better, when the number of bits is medium to large, the sequence detector achieves significantly better performance.

Keywords: Time-slotted communication, Molecular communications, Timing channels, Maximum likelihood detection, Viterbi algorithm, Permanent

[☆] This work was presented in part in the 3[rd] ACM International Conference on Nanoscale Computing and Communication [1].

[*] Corresponding author.

Email addresses: `moriny@stanford.edu` (Yonathan Murin), `nfarsad@stanford.edu` (Nariman Farsad), `mainakch@stanford.edu` (Mainak Chowdhury), `andreag@stanford.edu` (Andrea Goldsmith)

1. Introduction

Molecular communications (MC) is a new communication paradigm in which nano-scale devices communicate with each other using chemical signaling, based on exchanging small *information particles* [2, 3]. Several methods are used in MC systems to encode the transmitted information; this information can be encoded in the type of the released particles [4], in their concentration [5, 6], in their number [7], or in their time of release [8–10]. There are also several mechanisms for transporting the information particles from the transmitter to the receiver: diffusion, active transport, bacteria, and flow, as described in [3, Sec. III.B] and the references therein. While the research in this field is in its early stages, several basic experimental systems already serve as a proof of concept for transmitting short messages at low bit rates [11, 12].

Traditional electromagnetic communications (EM) and MC share several similarities that motivate using tools and algorithms common in EM communications systems in designing MC receivers. This holds in particular for channels that exhibit inter-symbol interference (ISI). On-off transmission based on diffusion of information particles [13] was studied in [14]. In this setup the receiver recovers the transmitted information from the measured concentration of information particles. The work [14] modeled this channel as a linear channel with finite memory and additive Gaussian noise, and derived the optimal, i.e., maximum likelihood (ML), sequence detector using the Viterbi algorithm (VA) [15]. A similar setup was studied in [16, 17]. Similarly to [14] the work [16] assumed known channel state information at the receiver and derived a technique for ISI mitigation along with a reduced-state ML sequence detection algorithm. On the other hand, [17] studied non-coherent multiple-symbol detection, derived the ML detector, and proposed an approximate detection metric and a suboptimal detector to cope with the high complexity of the ML detector. On-off transmission over a diffusive MC channel with flow was studied in [18], which proposed an ML sequence detection algorithm, and designed a family of weighted sums detectors. While the works [14, 16, 17] and [18] proposed to deal with ISI by applying sequence detection techniques, [19] proposed to use a transmission scheme with two types of information particles, thus, reducing the amount of ISI between consecutive transmissions.

In this work, we study receiver design for MC systems in which the information is encoded in the *time of release of indistinguishable information particles*. We refer to this model as the molecular timing (MT) channel. The idea of encoding information in the transmission timing is not new. For instance, [20] used this approach to describe communication in the brain at the synaptic cleft, where two chemical synapses communicate over a chemical channel [20]; [21] used this model to study bacterial communication over a microfluidic chip; and [22] derived the capacity of the single-server queue in which the information is conveyed via the messages timing. We use the common assumption that each particle, which arrives at the receiver during its observation window, is absorbed and removed from the environment [6, 9, 23, 24]. This assumption is accurate for modeling many sensors. Thus, the random delay until a particle arrives at the receiver can be represented as an additive noise term. For an unbounded observation window, and when the transportation mechanism is diffusion *without flow*, this additive noise follows the Lévy distribution [25, 26], whereas with flow the noise follows an inverse Gaussian (IG) distribution [27, 28].

Previous studies on MT channels either considered a *memoryless* channel [9, 26–29], or focused on the information theoretic aspects of transmission over MT channels [8, 30, 31]. In the current study we focus on the more practical setup of *time-slotted communication* in which the receiver has a *bounded observation window*. Time-slotted communication over MT channels with a bounded observation window introduces three main challenges. First, for MCs with indistinguishable particles, an information particle released during the k^{th} time-slot may arrive *after* the information particle released in the $(k+k_0)^{\text{th}}, k_0 > 0$ time-slot. Thus, *order may not be preserved* as was observed in [8, 30, 31]. As the information is encoded in the time of release, the lack of ordering leads to a channel with ISI.[1] Second, as the receiver has a *bounded observation window*, some of the particles *may not arrive within this observation time*, which can be viewed as *deletions or erasures*. This is one of the differences compared to [8, 30, 31] which assumed a non-bounded observation interval. Third, in contrast to the works [14, 16] and [18] which use a linear channel model with additive (and in some cases Gaussian) noise, MT channels are *not linear* and the additive noise is *not Gaussian*. Thus, new decoding techniques must be designed.

The first result of the current study is the derivation of the ML detector for time-slotted communication over MT channels with a bounded observation window. To derive this ML detector we use results from order statistics theory [32]. As the ML detector involves calculating the permanent [33] of a matrix, its complexity is exponential. Currently, no polynomial algorithm is available for calculating the permanent. Thus we conjecture that this detector cannot be implemented (or even approximated) with polynomial complexity. To support transmission of bit sequences of arbitrary length, we first derive a simple symbol-by-symbol detector, which extends the detector presented in [26]. Clearly, as this detector ignores the ISI, it is sub-optimal. Then, to account for the ISI, while using reasonable computational complexity, we develop a sequence detection algorithm that is based on the VA. However, in contrast to the traditional VA algorithm in EM communications which assumes a linear channel model, the proposed VA is based on counting the number of arrivals in each time-slot, which approximates the accurate arrival times. Numerical simulations show that the proposed sequence

[1] This is also the main difference compared to the "bits through queues" paradigm [22] in which order is enforced by the queue.

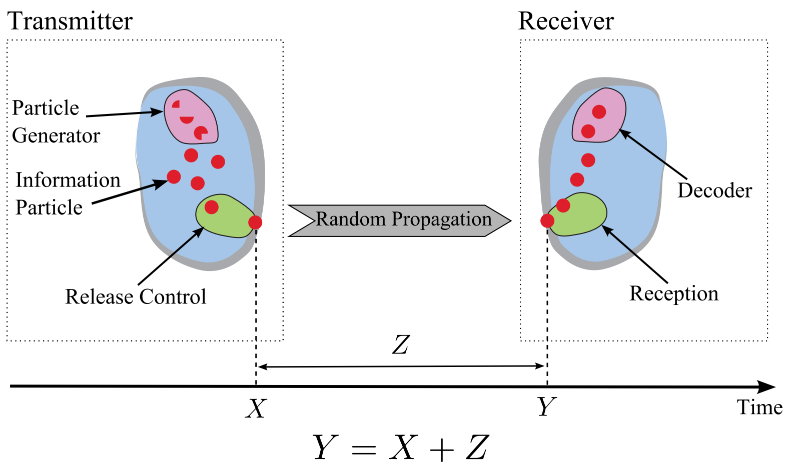

Figure 1: MC timing channel. X denotes the release time, Z denote the *random* propagation time, and Y denotes the arrival time.

detection algorithm significantly improves the performance compared to the symbol-by-symbol detector, and for a short number of transmitted symbols it closely approaches the high-complexity ML detector.

The rest of this paper is organized as follows. The problem formulation is presented in Section 2. Optimal detection in the presence of ISI is studied in Section 3. The symbol-by-symbol detector is presented in Section 4, while the sequence detector is derived in Section 5. Numerical results are presented in Section 6 and concluding remarks are provided in Section 7.

***Notation*:** We denote the set of real numbers by $\mathfrak{R}$, the set of positive real numbers by $\mathfrak{R}^+$, and the set of integers by $\mathfrak{N}$. Other than these sets, we denote sets with calligraphic letters, e.g., $\mathcal{V}$. We denote random variables (RVs) with upper case letters, e.g., X, Y, and their realizations with lower case letters, e.g., x, y. An RV X takes values in the set $\mathcal{X}$, and we use $|\mathcal{X}|$ to denote the cardinality of a finite set. We use $f_Y(y)$ to denote the probability density function (PDF) of a continuous RV Y on $\mathfrak{R}$, $f_{Y|X}(y|x)$ to denote the conditional PDF of Y given X, and $F_{Y|X}(y|x)$ to denote the conditional cumulative distribution function (CDF). Finally, we use $\mathsf{erfc}(\cdot)$ to denote the complementary error function given by $\mathsf{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} du$, and $\log(\cdot)$ to denote the natural logarithm.

2. Problem Formulation

2.1. System Model

We consider an MC channel, in which information is modulated in *the time of release of the information particles*. This channel is illustrated in Figure 1. We assume that as the information particles are *identical and indistinguishable* at the receiver, the receiver can only use the time of arrival to decode the intended message. Furthermore, the information particles propagate from the transmitter to the receiver through some *random* propagation mechanism (e.g. diffusion). We further make the following assumptions about the system:

A1) The transmitter perfectly controls the release time of each information particle, and the receiver perfectly measures the arrival times of the information particles. In addition, the transmitter and the receiver are perfectly synchronized in time.

A2) An information particle that arrives at the receiver is absorbed and removed from the propagation medium.

A3) All information particles propagate independently of each other, and their trajectories are random according to an i.i.d. random process.[2]

Note that these assumptions are consistent with those made in previous works, e.g., [6, 9, 23, 24, 26], in order to make the models tractable.

Let $\mathcal{X}$ be a finite set of constellation points on the real line: $\mathcal{X} \triangleq \{\xi_0, \xi_1, \ldots, \xi_{L-1}\}$, $0 \leq \xi_0 \leq \cdots \leq \xi_{L-1} \leq T_s$, where $T_s < \infty$ denotes the symbol duration. The k^{th} transmission takes place at time $(k-1)T_s + X_k, X_k \in \mathcal{X}, k = 1, 2, \ldots, K$. At this time, a *single* information particle is released into the medium by the transmitter. The transmitted information is encoded in the sequence $\{(k-1)T_s + X_k\}_{k=1}^{K}$, which is assumed to be independent of the random propagation time of *each* of the information particles. Let Y_k denote the time of arrival of the information particle released at time $(k-1)T_s + X_k$. It follows that $Y_k > X_k$. Thus, we obtain the following additive noise channel model:

$$Y_k = (k-1)T_s + X_k + Z_k, \quad k = 1, 2, \ldots, K, \tag{1}$$

where Z_k, is a random noise term representing the propagation time of the particle transmitted at the k^{th} time slot.[3]

The observation window at the receiver is assumed to be bounded, namely, the receiver observes all $Y_k \leq (K + \kappa)T_s, \kappa \geq 0$. This implies that if $Z_k > (K + \kappa - k + 1)T_s - X_k$, then the particle released in the k^{th} time-slot is not observed by the receiver, and is thus considered to be lost. Let $\{t_1 \leq t_2 \leq \cdots \leq t_n\}, 1 \leq n \leq K$ be the sequence of arrival times at the receiver, observed within its observation interval. Thus,

[2]This is a reasonable assumption for many different propagation schemes in MC such as diffusion in dilute solutions, i.e., when the number of particles released is much smaller than the number of molecules of the solutions.

[3]Note that Assumption **A3**) implies that all the RVs Z_k are independent.

$t_n \leq (K+\kappa)T_s$. Note that $t_j \in \{y_1, y_2, \ldots, y_K\}$, yet, due to the fact that the particles are indistinguishable, *the receiver does not know the mapping* $\{t_j\}_{j=1}^n \mapsto \{y_k\}_{k=1}^K$. Thus, the receiver decodes the sequence $\{X_k\}_{k=1}^K$ based on observing $\{t_j\}_{j=1}^n$.

In most of this work we restrict our attention to the case of binary modulations and set $\mathcal{X} = \{0, \Delta\}$. Let $\{S_k\}_{k=1}^K, S_k \in \{0,1\}$, denote a sequence of independent and equiprobable bits to be sent over the channel to the receiver, where the k^{th} bit is sent in the k^{th} slot. Without loss of generality assume that the bit $b \in \{0,1\}$ is mapped to the symbol $b \cdot \Delta$. Let $\hat{S}_k$ denote the estimate of S_k at the receiver, and define $P_{\varepsilon,k} \triangleq \Pr\{\hat{S}_k \neq S_k\}$. Our objective is to design receivers that minimize the following average probability of error:

$$P_\varepsilon = \frac{1}{K}\sum_{k=1}^{K} P_{\varepsilon,k}. \tag{2}$$

We emphasize that the above description of communication over MT channels is fairly general and can be applied to different propagation mechanisms as long as Assumptions **A1**)–**A3**) are not violated. In the rest of the paper we focus on *diffusion-based propagation without flow* which is governed by the Lévy distribution.

2.2. The Diffusion-Based MT Channel

In diffusion-based propagation, the released particles follow a random Brownian path from the transmitter to the receiver. In this case, to specify the random additive noise term Z_k in (1), we define a Lévy-distributed RV as follows:

Definition 1. *Let Z be Lévy-distributed with location parameter μ and scale parameter c [34]. The PDF of Z is given by:*

$$f_Z(z) = \begin{cases} \sqrt{\frac{c}{2\pi(z-\mu)^3}} \exp\left(-\frac{c}{2(z-\mu)}\right), & z > \mu \\ 0, & z \leq \mu \end{cases}, \tag{3}$$

and its CDF is given by:

$$F_Z(z) = \begin{cases} \mathsf{erfc}\left(\sqrt{\frac{c}{2(z-\mu)}}\right), & z > \mu \\ 0, & z \leq \mu \end{cases}. \tag{4}$$

Let d denote the distance between the transmitter and the receiver, and D denote the diffusion coefficient of the information particles in the propagation medium. Following along the lines of the derivations in [27, Sec. II], and using the results of [35, Sec. 2.6.A], it can be shown that for 1-dimensional pure diffusion, the propagation time of each of the information particles follows a Lévy distribution, denoted in this work by $\mathscr{L}(\mu, c)$ with $c = \frac{d^2}{2D}$ and $\mu = 0$. Thus, $Z_k \sim \mathscr{L}(0, c), k = 1, 2, \ldots, K$.

Remark 1. The work [36] showed that a scaled Lévy distribution can also model the first arrival time in the case of an infinite, three-dimensional homogeneous medium without flow. Hence, our results can be extended to 3-dimensional space by simply introducing a scalar factor.

Next, we study ML detection for time-slotted communication over diffusion-based MT channels.

3. ML Detection for Arbitrary T_s

Before studying detection for arbitrary values of T_s, we first briefly discuss the case of $T_s \gg \Delta$, namely, the channel is (approximately) memoryless and (almost) optimal detection can be applied for each symbol separately. The probability of error in this case is clearly a lower bound on the performance in the case of arbitrary T_s, since when $T_s \gg \Delta$ there is (almost) no ISI and (almost) all particles arrive at the receiver within its observation time interval.

3.1. ML Detection When $T_s \gg \Delta$

The optimal decision rule, assuming a memoryless channel, was presented in [26, Prop. 1]. We repeat it here for ease of reference:

Proposition 1. *Let $y = y_k - (k-1)T_s$. The decision rule that minimizes the probability of error for $T_s \gg \Delta$, is given by:*

$$\hat{X}_{\text{no-ISI}}(y) = \begin{cases} 0, & y < \theta \\ \Delta, & y \geq \theta, \end{cases} \tag{5}$$

where θ is the unique solution, in the interval $[\Delta, \Delta + \frac{c}{3}]$, of the following equation in y:

$$y(y-\Delta)\log\left(\frac{y_1}{y_1-\Delta}\right) = \frac{c\Delta}{3}, \quad y > \Delta > 0. \tag{6}$$

Furthermore, the probability of error of this decision rule is given by:

$$P_{\varepsilon,\text{no-ISI}} = 0.5\left(1 - \mathsf{erfc}\left(\sqrt{\frac{c}{2\theta}}\right) + \mathsf{erfc}\left(\sqrt{\frac{c}{2(\theta-\Delta)}}\right)\right). \tag{7}$$

Next, we discuss ML detection for arbitrary values of T_s.

3.2. ML Detection for Arbitrary T_s

Before discussing the ML detector, we first observe that the ISI in MT channels is fundamentally different from the ISI in traditional EM communications. In EM communications the channel is commonly assumed to be linear with finite impulse response (FIR):[4]

$$\upsilon_k = h_0\zeta_k + \sum_{j=1}^{L} h_j\zeta_{k-j} + w_k, \tag{8}$$

where $\{\zeta_k\}$ are the channel inputs, $\{h_j\}_{j=0}^{L}$ are the (usually known or estimated) channel response coefficients, w_k is an additive noise independent of the channel coefficients and of the transmitted signal, and υ_k is the channel output. Thus, a symbol transmitted at time index k *systematically and linearly* affects the current and future L output symbols via the channel response $\{h_j\}_{j=0}^{L}$. Note that in (8) it is assumed that $\zeta_k = 0, k \leq 0$. Therefore, the first symbol experiences no ISI, the second symbol experiences ISI from the first symbol, and so on. Moreover, for a *given* channel input sequence, and assuming $\{h_j\}_{j=0}^{L}$ is known, the ISI in the k^{th} symbol, $k \geq L$, given by $\sum_{j=1}^{L} h_j\zeta_{k-j}$, is deterministic.

On the other hand, in MT channels the ISI is caused by the random propagation (the additive noise) which is never known, and there is a *non-negligible probability* that a particle that was transmitted at time slot $n < k$ will arrive at time slot k. This implies that in MT channels, in contrast to EM channels, the channel outputs are *not ordered*. In addition to this fundamental difference, we note that due to the random propagation, some of the information particles may not arrive at the receiver within its bounded observation interval. Thus, in contrast to EM channels, the considered MT channel imposes deletions and the receiver is required to decode the transmitted data when the number of arrivals is possibly less then the number of released particles. Finally, we note that in MT channels an information particle *arrives only once*, while in EM communications, due to the multipath, ζ_k is assumed to be observed at the receiver both at time instance k and $k + l, l < L$. The above differences render the problem of designing time-slotted communication schemes for MT channels significantly different than designing time-slotted communication scheme for EM channels.

Remark 2. Part of the above observations were stated in [38, Remark 1]. Yet, in [38] the authors considered on-off concentration modulation and tackled the randomness in

[4] Even if the channel has an infinite impulse response (IIR) we can approximate it using an FIR channel [37, Ch. 11.4.1].

the ISI by analyzing the *expected* channel output. In diffusion-based MT channels the mean and variance of the channel output are not defined (since the Lévy noise has no mean and variance). Thus, analysis of the expected channel output is not applicable.

Before presenting the ML detector for arbitrary values of T_s, we recall that $\{t_1 \leq t_2 \leq \cdots \leq t_n\}, 1 \leq n \leq K$, denote the sequence of arrival times at the receiver, observed within its observation interval. We further recall that $t_j \in \{y_1, y_2, \ldots, y_K\}$, yet, due to the fact that the *particles are indistinguishable*, the receiver does not know the mapping $\{t_j\}_{j=1}^n \mapsto \{y_k\}_{k=1}^K$. Let $\mathsf{perm}(\mathbb{M})$ denote the permanent of a square matrix $\mathbb{M}$, see [33], and let $\mathbf{x}$ and $\mathbf{t}$ be shorthand notations for $\{x_k\}_{k=1}^K$ and $\{t_j\}_{j=1}^n$, respectively. Further define $f_k(y, \mathbf{x}) \triangleq f_{Y|X}(y|x_k)$, where $f_{Y|X}(y|x_k) = f_Z(y; \mu = x_k)$. $F_k(y, \mathbf{x})$ is defined similarly. The detector that minimizes the probability of error is stated in the following theorem:

Theorem 1. *The detector that minimizes the probability of error in detecting* $\mathbf{X}$ *based on* $\{t_1, t_2, \ldots t_n\}$ *is given by:*

$$\hat{\mathbf{X}}_{\mathrm{ML}}(\mathbf{t}) = \underset{\mathbf{x}}{\operatorname{argmax}}\, \mathsf{perm}(\mathbb{M}(\mathbf{x}, \mathbf{t})), \tag{9}$$

where $\mathbb{M}(\mathbf{x}, \mathbf{t})$ *is a* $K \times K$ *matrix given by:*

$$\mathbb{M}(\mathbf{x}, \mathbf{t}) = \begin{bmatrix} f_1(t_1, \mathbf{x}) & f_2(t_1, \mathbf{x}) & \cdots & f_K(t_1, \mathbf{x}) \\ f_1(t_2, \mathbf{x}) & f_2(t_2, \mathbf{x}) & \cdots & f_K(t_2, \mathbf{x}) \\ \vdots & & & \vdots \\ f_1(t_n, \mathbf{x}) & f_2(t_n, \mathbf{x}) & \cdots & f_K(t_n, \mathbf{x}) \\ 1 - F_1(t_n, \mathbf{x}) & 1 - F_2(t_n, \mathbf{x}) & \cdots & 1 - F_K(t_n, \mathbf{x}) \\ \vdots & & & \vdots \\ 1 - F_1(t_n, \mathbf{x}) & 1 - F_2(t_n, \mathbf{x}) & \cdots & 1 - F_K(t_n, \mathbf{x}) \end{bmatrix},$$

with the last $K - n$ *rows of* $\mathbb{M}(\mathbf{x}, \mathbf{t})$ *identical and equal to:*

$$\begin{bmatrix} 1 - F_1(t_n, \mathbf{x}) & 1 - F_2(t_n, \mathbf{x}) & \cdots & 1 - F_K(t_n, \mathbf{x}) \end{bmatrix}.$$

Proof (of Theorem 1). As all the symbols are equiprobable, the detector that minimizes the probability of error is the ML detector, given by:

$$\hat{\mathbf{X}}_{\mathrm{ML}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, f_{\{T_j\}_{j=1}^n | \{X_k\}_{k=1}^K}(\{t_j\}_{j=1}^n | \{x_k\}_{k=1}^K). \tag{10}$$

Note that $\{t_j\}_{j=1}^n$ are the first n arrivals, or equivalently, the smallest n values in the sequence $\{y_k\}_{k=1}^K$. Further note that given $\{x_k\}_{k=1}^K$, the RVs $\{Y_k\}_{k=1}^K$ are independent but *non-identically* distributed, since each Y_k has a different offset parameter. Given

x_k, the PDF of Y_k is $f_k(y, x_k)$, and its CDF is $F_k(y, x_k)$. Thus, using results from order-statistics theory [39, pg. 309], see also [40, Eq. (16)],[5] we have:

$$f_{\{T_j\}_{j=1}^n|\{X_k\}_{k=1}^K}(\{t_j\}_{j=1}^n|\{x_k\}_{k=1}^K) = \frac{1}{(K-n)!}\mathsf{perm}(\mathbb{M}(\mathbf{x}, \mathbf{t})). \tag{11}$$

Plugging (11) into (10) concludes the proof.

Remark 3. While the works [8, 31, 40] also considered optimal detection when order is not preserved, they assumed that all the particles arrive. Furthermore, the works [8, 31] expressed the joint density of the sequence $\{t_j\}_{j=1}^K$ in terms of a folded density of $\{y_k\}_{k=1}^K$, see [31, Eq. (4)], and not as in (11).

3.3. On the Computational Complexity of the ML Detector

Analyzing the ML detector in (9), we note that it consists of two parts: First, for each candidate input sequence $\mathbf{x}$, one should calculate $\mathsf{perm}(\mathbb{M}(\mathbf{x}, \mathbf{t}))$. Second, one should find the $\mathbf{x}$ that minimizes this permanent. The applicability of this detector depends on its computational complexity. More precisely, assuming an input bit sequence of reasonable length (more then 10-15 bits), the computational complexity must be at most polynomial for the algorithm to be applicable. Thus, a natural question is *can the ML detector in* (9) *be calculated with polynomial complexity?*

To answer this question we note that even though the evaluation of the permanent attracted the attention of researchers for almost two centuries (see [41] for a comprehensive survey), the *best known* algorithm for calculating the permanent is by Ryser from 1963 [42], see also [43, Ch. 1.3]. The complexity of this algorithm is $\Theta(2^K K^2)$. Therefore, since the permanent should be evaluated for each candidate $\mathbf{x}$, it seems that the ML detector in (9) *cannot* be evaluated using polynomial complexity. In fact, if the $\text{argmax}_{\mathbf{x}}$ operation in (9) is implemented via an exhaustive search the resulting complexity is $\Theta(2^{2K})$.

Since direct evaluation of the permanent is of exponential complexity, one may consider approximating the permanent. Such an approximation with polynomial complexity was presented in [44], yet, it is not clear if and how this approximation can be combined with the search in (9) (recall that the search itself is of exponential complexity). A common approach to calculate the $\text{argmax}_{\mathbf{x}}$ in (9) in polynomial time is using the VA. Yet, the VA requires $n = K$, namely, the number of observations to be equal to the number of transmitted symbols, and the joint density $f_{\{T_j\}_{j=1}^n|\{X_k\}_{k=1}^K}(\{t_j\}_{j=1}^n|\{x_k\}_{k=1}^K)$

[5] Note that we are interested in the density of the *smallest* n RVs out of k independent RVs.

should obey:

$$f_{\{T_j\}_{j=1}^{k}|\{X_k\}_{k=1}^{K}}(\{t_j\}_{j=1}^{n}|\{x_k\}_{k=1}^{K}) = \prod_{k=1}^{K} f_{T_k|X_k}(t_k|x_k).$$

As it does not seem that such a representation holds, we conjecture that the ML detector in (9) cannot be accurately approximated in polynomial time using the results of [44].

Since an exponential detection algorithm cannot be implemented even for moderate values of K, we next consider sub-optimal detectors. In the next section we consider a simple symbol-by-symbol detector which is asymptotically optimal, i.e., for $T_s \gg \Delta$.

4. Symbol-by-Symbol Detection

Before presenting the symbol-by-symbol detector we note that the detector in (5) should be adapted to the setting in which ISI is present. This follows as there might be more than a single arrival within the slot boundaries. This is again fundamentally different than in EM communications where the optimal symbol-by-symbol detector can simply be used in the presence of ISI, at the cost of a higher probability of error. We propose the following (sub-optimal) adaptation to the detector in (5):

***A Symbol-by-symbol decision rule*:** Recall that θ is the optimal decision threshold defined in Prop. 1 for the case of $T_s \gg \Delta$. Let α_k denote the number of arrivals in the time interval $[(k-1)T_s, (k-1)T_s + \theta), k = 1, 2, \ldots, K$, and β_k denote the number of arrivals in the interval $[(k-1)T_s + \theta, kT_s)$. The detector applies the following decision rule:[6]

$$\hat{X}_k(\alpha_k, \beta_k) = \begin{cases} 0, & \alpha_k > 0 \text{ and } \beta_k = 0 \\ \Delta, & \alpha_k = 0 \\ q_k, & \alpha_k > 0 \text{ and } \beta_k > 0, \end{cases} \tag{12}$$

where $q_k \in \{0, \Delta\}$ is a Bernoulli RV with $\Pr\{q_k = 0\} = 0.5$, i.i.d. over time.

The first two events on the right hand side (RHS) of (12) are an extension of (5) to the case of multiple arrivals at the same time slot. On the other hand, for the third event we note that, as we cannot distinguish between different information particles, it is not clear which decision should be taken. Therefore, we simply toss a coin via the RV q_k.

Next, we derive the exact probability of error of the adapted detector (12). We show that for $T_s \gg \Delta$ the performance of the adapted detector approaches (7). An intuitive

[6]Note that this detector is not necessarily the optimal symbol-by-symbol detector in the presence of ISI, yet, asymptotically, i.e., for $T_s \to \infty$, it achieves the performance of the detector in (5).

explanation for this result is the observation that when $T_s \gg \Delta$, the probability that a particle will not arrive in its time slot is negligible, and (12) simplifies to (5). We first define the function $\psi(c, T_s, \Delta, k, \eta_1, \eta_2)$, which is used to calculate the probability of interference in the different error events.

$$\psi(k,\eta_1,\eta_2) \triangleq 0.5^k \cdot \prod_{n=0}^{k-1} \Bigg\{ 2 + \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s+\eta_1)}}\right) + \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s+\eta_1-\Delta)}}\right) \\ - \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s+\eta_2)}}\right) - \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s+\eta_2-\Delta)}}\right) \Bigg\}. \quad (13)$$

We further define $\mathsf{erfcs}(y,x) \triangleq \mathsf{erfc}\left(\sqrt{\frac{c}{2(y-x)}}\right)$. Finally, we define the terms:

$$\rho_{k,1}^{(0)} \triangleq \psi(k,0,T_s) \cdot (1 - \mathsf{erfcs}(T_s,0)), \quad (14a)$$

$$\rho_{k,1}^{(1)} \triangleq \psi(k,0,T_s) \cdot (1 - \mathsf{erfcs}(T_s,\Delta)), \quad (14b)$$

$$\rho_{k,2}^{(0)} \triangleq \psi(k,\theta,T_s) \cdot (1 - \mathsf{erfcs}(T_s,0) + \mathsf{erfcs}(\theta,0)) - \rho_{k,1}^{(0)}, \quad (14c)$$

$$\rho_{k,2}^{(1)} \triangleq \psi(k,\theta,T_s) \cdot (1 - \mathsf{erfcs}(T_s,\Delta) + \mathsf{erfcs}(\theta,\Delta)) - \rho_{k,1}^{(1)}, \quad (14d)$$

$$\rho_{k,3}^{(0)} \triangleq \psi(k,0,\theta) \cdot (1 - \mathsf{erfcs}(\theta,0)) - \rho_{k,1}^{(0)}, \quad (14e)$$

$$\rho_{k,3}^{(1)} \triangleq \psi(k,0,\theta) \cdot (1 - \mathsf{erfcs}(\theta,\Delta)) - \rho_{k,1}^{(1)}, \quad (14f)$$

$$\rho_{k,4}^{(0)} \triangleq 1 - \sum_{j=1}^{3} \rho_{k,j}^{(0)}, \quad (14g)$$

$$\rho_{k,4}^{(1)} \triangleq 1 - \sum_{j=1}^{3} \rho_{k,j}^{(1)}. \quad (14h)$$

The probability of error of the detector (12) is presented in the following theorem:

Theorem 2. *The probability of error of the decision rule* (12) *is given by:*

$$P_\varepsilon = \frac{1}{2K} \cdot \sum_{k=0}^{K-1} \left\{ \rho_{k,1}^{(0)} + \rho_{k,3}^{(0)} + \rho_{k,2}^{(1)} + 0.5\left(\rho_{k,4}^{(0)} + \rho_{k,4}^{(1)}\right) \right\}. \quad (15)$$

PROOF. The proof is provided in Appendix A.

The following corollary characterizes the asymptotic performance of the detector (12), i.e., when $T_s \to \infty$.

Corollary 1. *When $T_s \to \infty$ then P_ε in* (15) *approaches P_ε in* (7).

PROOF. When $T_s \to \infty$ then the arguments of the $\mathsf{erfc}(\cdot)$ functions in (13) go to 0. This implies that the $\mathsf{erfc}(\cdot)$ functions go to 1, and therefore $\psi(k, \eta_1, \eta_2) \to 1$ as well, as long as η_1 and η_2 are positive or finite. In this case $\rho_{k,1}^{(0)}, \rho_{k,1}^{(1)}, \rho_{k,4}^{(0)}$, and $\rho_{k,4}^{(1)}$ all go to zero. It can further be observed that when $T_s \to \infty$ then $P_{\varepsilon,k}$ is independent of k. Therefore, the probability of error in (15) specializes to (7).

While Corollary 1 indicates that the symbol-by-symbol detector in (12) is asymptotically optimal, when T_s is of the order of Δ this detector performs significantly worse compared to the ML detector in (9), as indicated in Section 6. The main reason for this gap is that (12) does not detect the *sequence* $\{X_k\}_{k=1}^{K}$; instead, it detects each X_k individually, thus ignoring the channel memory. In the next section we propose a novel sequence detection algorithm that achieves performance close to the performance of the ML detector.

5. Sequence Detection Algorithm

We first observe that for small values of T_s, or for heavy-tailed noise distributions like the Lévy distribution, the channel (1) resembles an IIR channel in EM communications, in the sense that the current channel input depends on *all* previous transmitted symbols. Yet, in contrast to the linear model common in EM communication, the channel (1) is not linear. Moreover, in some of the slots there are no arrivals, and since the observation interval is bounded the total number of arrivals can be smaller than the number of transmissions. Therefore, *applying traditional sequence detection algorithms is far from straight-forward.* To tackle the challenge of lack of measurements in arbitrary slots, we use a simplified discrete channel model that counts the number of arrivals at each slot. Thus, we obtain valid measurements even for slots with no arrivals. Then, we apply sequence detection based on the simplified model using a modified VA. Clearly, as we use a simplified channel model, the resulting detector is sub-optimal. Yet, our simulations indicate that it can achieve performance close to the ML detector.

In its most general form, the VA solves a MAP estimation problem of the state sequence of a *finite-state* discrete-time Markov process observed in memoryless noise [15]. However, the channel in our problem may have memory of length *equal to the number of transmitted symbols*, which implies that the state space *is not bounded.* We tackle this by truncating the channel response. More precisely, by noting that the density of the noise decays with time, it follows that when the transmission time and observation time are far enough apart, one can approximate this density by zero, thus, truncating the channel and resulting in a channel with finite memory. Yet, it should be noted that in contrast to EM communications, in our setting such truncation requires a

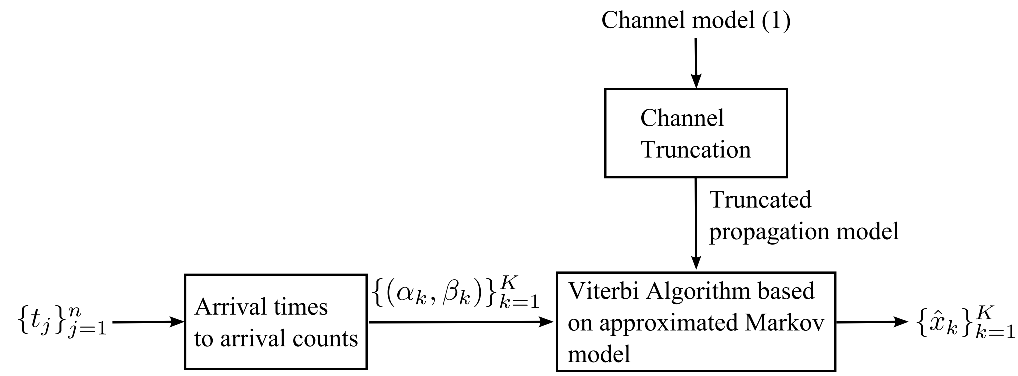

Figure 2: Block diagram of the proposed sequence detector. $\{t_j\}_{j=1}^n$ are the observed arrival times. (α_k, β_k) is the arrival count in the k^{th} time-slot. $\{\hat{x}_k\}_{k=1}^K$ are the decoded symbols.

separate treatment for cases in which the observed channel output is not a valid output of the assumed model. We denote this event as an *out of model event*. A block diagram of the proposed approach is depicted in Figure 2.

5.1. From Arrival Times to Arrival Counts

To overcome the lack of measurements in some of the slots, and to partially cope with the lack of ordering, we use the sequence $\{t_j\}_{j=1}^n$ to calculate the sequence $\{V_k\}_{k=1}^K$, $V_k = (\alpha_k, \beta_k)$, as defined in Section 4. $\{V_k\}_{k=1}^K$ is the input to our algorithm. Note that with this transformation we have a valid measurement for every time slot. It should be noted that the proposed detection algorithm can be used with any partition of the slot. We choose the above partition since it minimizes the probability of error in a single symbol transmission and it achieves the optimal detection performance for $T_s \gg \Delta$.

5.2. Truncating the Channel

To obtain a finite state model we truncate the channel: We assume that in the k^{th} time slot one can observe arrivals of information particles transmitted in time slots $\{k-l\}_{l=0}^L$, thus, obtaining a channel of memory L. Clearly, a larger L captures the channel statistics more accurately at the cost of exponentially increased computational complexity. We propose to choose L as a given fraction of the worst case probability of arrival of past information particles. Let p_k^{ISI} denote the probability that *any past* transmitted information particle arrives at time slot k. We have the following proposition:

Proposition 2. *The probability of arrival of past information particles at time slot k*

increases with k and is given by $p_k^{ISI}=\sum_{i=1}^{k-1} p_{k,i}^{ISI}$, where $p_{k,i}^{ISI}$ is given by:

$$p_{k,i}^{ISI}=0.5\Bigg\{\mathsf{erfc}\left(\sqrt{\frac{c}{2(k-i+1)T_s}}\right)-\mathsf{erfc}\left(\sqrt{\frac{c}{2(k-i)T_s}}\right) \\ +\mathsf{erfc}\left(\sqrt{\frac{c}{2((k-i+1)T_s-\Delta)}}\right)-\mathsf{erfc}\left(\sqrt{\frac{c}{2((k-i)T_s-\Delta)}}\right)\Bigg\}. \tag{16}$$

Proof. The proof is provided in Appendix B.

The above proposition implies that the *last slot experiences the largest ISI*. Thus, L can be chosen to cover a given fraction of p_K^{ISI}. We propose to choose L such that the probability of arrival from the past L symbols will be larger than a pre-defined threshold λ times p_K^{ISI}:

$$L=\operatorname*{argmin}_l\left\{\sum_{n=K-l+1}^{K} p_{K,n}^{\mathrm{ISI}}\geq \lambda p_K^{\mathrm{ISI}}\right\}. \tag{17}$$

Example 1. *Let $c=1,\Delta=1,T_s=1.5$ and $K=100$. Then, $p_K^{ISI}=0.649$, and by setting $\lambda=0.5$ we obtain $L=2$. On the other hand, if we set $\lambda=0.95$ then $L=44$. We further note that, roughly speaking, larger Δ and T_s does not change L for a given λ. Instead, it reduces p_{K-1}^{ISI}, which also reduces the probability of error. Finally, we note that lower values of the parameter c, which controls the width of the Lévy distribution, results in lower L. We elaborate more on the relations between Δ, T_s, c, and λ in Section 6.*

5.3. The Markov Structure

For a given L, we *assume* a finite-memory channel, i.e., $V_k=(\alpha_k,\beta_k)$ depends on the transmissions at time slots $k-L,k-L+1,\dots,k$. Thus, we write:

$$\Pr\{\{V_k\}_{k=1}^K|\{X_k\}_{k=1}^K\}=\prod_{j=1}^{L}\Pr\{V_j|\{X_i\}_{i=1}^j,\{V_i\}_{i=1}^{j-1}\} \\ \times\prod_{k=L+1}^{K}\Pr\{V_k|\{X_i\}_{i=k-L}^k,\{V_i\}_{i=1}^{k-1}\}. \tag{18}$$

Note that in (18), V_k depends on all the previous V's, thus, it does not represent a process with finite memory. This follows from the fact that each particle can arrive at the receiver *only once*. Motivated by the case of $T_s\gg\Delta$, or by the large Δ regime,[7]

[7]These are reminiscent of the high SNR regime in EM communications.

in which the dependence between V_k and $\{V_j\}_{j=1}^{k-1}$ is very weak, we apply the approximation $\Pr\{V_k|\{X_i\}_{k-L}^{k}, \{V_i\}_{i=1}^{k-1}\} \approx \Pr\{V_k|\{X_i\}_{k-L}^{k}\}$, which results in a finite memory. In Section 6 we show that even with this approximation, the performance of the proposed sequence detector can approach the performance of the ML detector. With this approximation in hand, we now propose the following sequence detector:

$$\hat{\mathbf{X}}_{\text{SD}}(\mathbf{v}) = \underset{\{x_k\}_{k=1}^{K}}{\text{argmax}} \sum_{j=1}^{L} \log\Big(\Pr\{v_j|\{x_i\}_{i=1}^{j}\}\Big) + \sum_{k=L+1}^{K} \log\Big(\Pr\{v_k|\{x_i\}_{i=k-L}^{k}\}\Big). \tag{19}$$

The first summation on the RHS of (19) constitutes L steps of initializations, while the second summation can be efficiently calculated using the VA. Next, we elaborate on the initialization of the VA, the calculation of the steady-state branch metric, and the traceback.

5.4. Branch Metric, Initialization, and Traceback

***Branch Metric*:** Equation (19) implies that the trellis in the proposed VA will consist of 2^L states (for each time-slot), characterized by $\{x_i\}_{i=k-L}^{k-1}$, and the transition metrics will be given by $\log\big(\Pr\{V_k = v|\{x_i\}_{i=k-L}^{k}\}\big)$. Note that since we only *assume* a truncated channel model, v *may not* belong to the set of possible output pairs defined by the assumed model. More precisely, the assumption of finite memory of order L implies that $V_k \leq L+1, k > L$. On the other hand, the *measured* V_k is generated by the original channel, and therefore $V_k \leq k$. Note that this phenomena is fundamentally different than IIR channels in EM communications in which even if one assumes an FIR channel instead of an IIR channel, the output sequence is still valid under the model assumptions. To see this note that (8), assuming an IIR channel, can be written as:

$$\upsilon_k = h_0\zeta_k + \sum_{j=1}^{k-1} h_j\zeta_{k-j} + w_k = h_0\zeta_k + \sum_{j=1}^{L} h_j\zeta_{k-j} + \tilde{w}_k. \tag{20}$$

Thus, if the channel impulse response $\{h_j\}_{j=0}^{\infty}$ decays fast enough, then $\tilde{w}_k \approx w_k$. Moreover, the output sequence $\{\upsilon_k\}_{k=1}^{K}$ is a valid sequence even if an FIR is assumed.

To deal with the case of an *out of model* event, we first define the set $\mathcal{V}_{\text{tm}}$, which contains all the possible V's in our truncated channel model:

$$\mathcal{V}_{\text{tm}} \triangleq \{(\alpha,\beta) : \alpha+\beta \leq L+1, \alpha \in \mathfrak{N}, \beta \in \mathfrak{N}\}.$$

Next, we define the set $\mathcal{V}_{\text{res}} = \{(\alpha,\beta) : L+1 < \alpha+\beta \leq K, \alpha \in \mathfrak{N}, \beta \in \mathfrak{N}\}$, namely, $\mathcal{V}_{\text{res}}$ is the set of all V's that cannot be generated by our model. The branch (transition) metric is now given by:

$$\mathrm{TM}(v,\{x_i\}_{i=k-L}^{k}) = \begin{cases} \log\left(\Pr\{v|\{x_i\}_{i=k-L}^{k}\}\right), & v\in\mathcal{V}_{\mathrm{tm}}, \\ \log\left(1-\sum_{(\tilde{v}\in\mathcal{V}_{\mathrm{tm}})}\Pr\{\tilde{v}|\{x_i\}_{i=k-L}^{k}\}\right), & v\in\mathcal{V}_{\mathrm{res}}. \end{cases} \tag{21}$$

Remark 4. Note that in (21) the terms $\log\left(\Pr\{v|\{x_i\}_{i=k-L}^{k}\}\right)$ are calculated *based on the finite memory model*, namely, *without accounting for the tails* of the noise (propagation) density. Therefore, as the support of the Lévy distribution is $\mathfrak{R}^{+}$, the probability of the out of model event, given by $1-\sum_{(\tilde{v}\in\mathcal{V}_{\mathrm{tm}})}\Pr\{\tilde{v}|\{x_i\}_{i=k-L}^{k}\}$, is larger than zero. This implies that all the events in the set $\mathcal{V}_{\mathrm{res}}$ are indistinguishable and have the same metric: $\log\left(1-\sum_{(\tilde{v}\in\mathcal{V}_{\mathrm{tm}})}\Pr\{\tilde{v}|\{x_i\}_{i=k-L}^{k}\}\right)$.

***Initialization and traceback*:** In the initialization phase we use the metric:

$$\log\left(\Pr\{V_j=v|\{x_i\}_{i=1}^{j}\}\right).$$

Here, as $j<L$, the out of model event is not required. Finally, as there is no termination transmission, we trace back starting from the state with the maximal accumulated metric.

In Appendix C we provide an example of the proposed algorithm. Specifically, we consider the case of $L=1$, detail the sets $\mathcal{V}_{\mathrm{tm}}$ and $\mathcal{V}_{\mathrm{res}}$, and describe how the respective transition probabilities are calculated.

6. Numerical Results

Next, we demonstrate our results via numerical simulations. To simplify the presentation we assume that the slot length scales proportionally to Δ, and present the average probability of error as a function of Δ for a given value of c. The parameter c is known to be the dispersion of the noise [26], which is reminiscent of the standard deviation in the case of the Gaussian distribution. This implies that larger values of c correspond to more ISI, while when c is small the distribution is more concentrated around the transmission time. Hence, we also present our results for different values of c.[8]

[8]Recall that $c=\frac{d^2}{2D}$, where d is the distance between the transmitter and the receiver, and D is the diffusion coefficient.

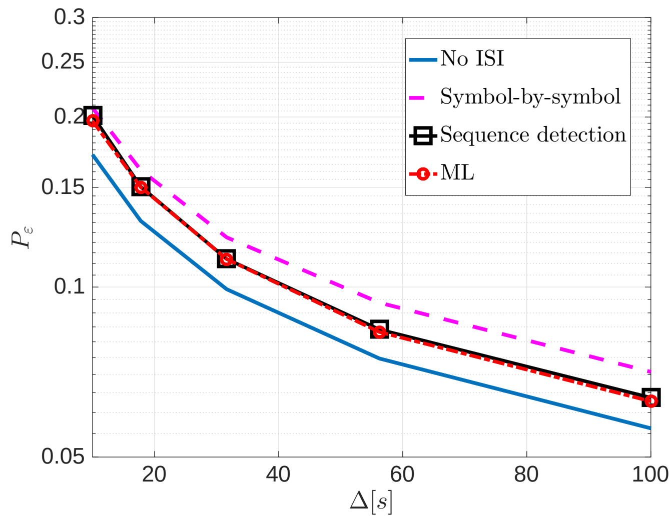

Figure 3: P_ε vs. Δ for $c = 2$ and $T_s = 1.5\Delta$. The sequence detection is implemented assuming $L = 2$. The block length is $K = 6$ and $\kappa = 2$.

6.1. Time-Slotted Communication

We begin the numerical analysis with the study of the different detection approaches. Figure 3 depicts P_ε versus Δ for $c = 2$ and $T_s = 1.5\Delta$. Here, as the ML detector has exponential complexity, we consider a short block of length $K = 6$. The sequence detector is implemented with $L = 2$, $\kappa = 2$, and for each Δ point 10^4 trials were carried out. It can be observed that the sequence detector proposed in Section 5 achieves almost the same probability of error as the ML detector, yet, it entails only a polynomial complexity. It can further be observed, as expected, that both detectors are significantly outperformed by the no-ISI detector; this is due to the severe ISI which results in non-ordered arrivals, or even arrival of only part of the information particles. It should further be noted that 1) We use a truncated channel model which captures less than 60% of the ISI; and 2) We apply sequence detection based on a "hard decision" version of the channel output measurements, namely, we decode based on the pairs (α_k, β_k). Moreover, in contrast to the work [45] which considered finite-length ISI for Gaussian channels, and showed that the VA can obtain average probability of error very close to the no-ISI transmission, it is not clear if commensurate performance can be obtained for our model, in particular due to the heavy tails of the Lévy distribution. Figure 3 also indicates the large gains of the sequence detector over the symbol-by-symbol detector. For example, to achieve $P_\varepsilon = 0.1$ the sequence detector requires $\Delta = 42[s]$, compared to $\Delta = 51[s]$ required by the symbol-by-symbol detector.

Figure 4 depicts P_ε versus Δ for $c = 0.1$ and $T_s = 15\Delta$. Here $K = 100$, and the

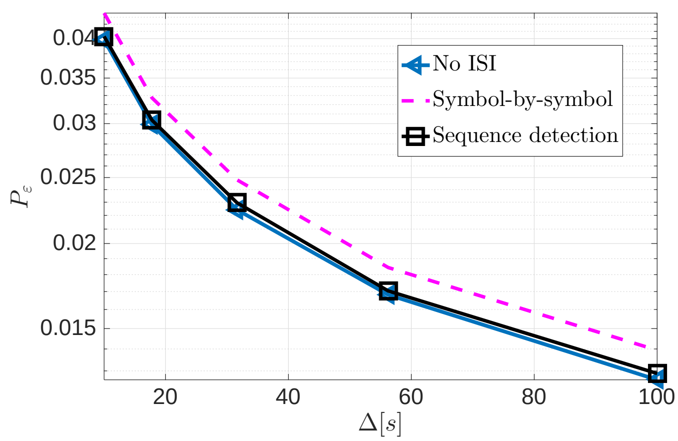

Figure 4: P_ε vs. Δ for $c = 0.1$ and $T_s = 15\Delta$. The sequence detection is implemented assuming $L = 2$. The block length is $K = 100$ and $\kappa = 2$.

ML detector is not presented due to its exponential computational complexity. The sequence detector was implemented with $L = 2$ and $\kappa = 2$. It can be observed that the performance of the sequence detector is very close to the optimal detector when no ISI is present. This is due to the relatively large spacing between transmissions. On the other hand, for $T_s = 15\Delta$, one can again observe a significant gain of the sequence detector compared to the symbol-by-symbol detector.

Figure 5 depicts the probability of error (averaged over all trials) for the bit indices $k = 1, 8$, and the average probability of error, for $c = 2$ and $T_s = 1.5\Delta$. K was set to 100. The sequence detector was implemented with $L = 2$ and $\kappa = 2$. It can first be observed that the probabilities of error are different for different bits. Furthermore, it can be observed that the average probability of error is higher than the average probability of error of the sequence detector in Figure 3 due to the longer block length. As the ISI experienced by each of the bits increases with the bit index, see Prop. 2, the probability of error also increases with the bit index. Hence, transmission over the time-slotted timing channel has the inherent property of unequal protection, namely, some of the bits are more protected than others. This property can be leveraged to transmit a bit-sequence in which some of the bits are more important than others. For instance, in some scenarios control bits are more important than data bits as the data cannot be decoded without the control bits. Another example is describing a real-valued measurement using a finite precision bit-sequence. In this case the most-significant-bits are more important than the least-significant-bits as an error in these bits leads to higher total error (distortion). Finally, unequal error control codes can equalize the probability of error in this case by protecting the bits with higher error probability

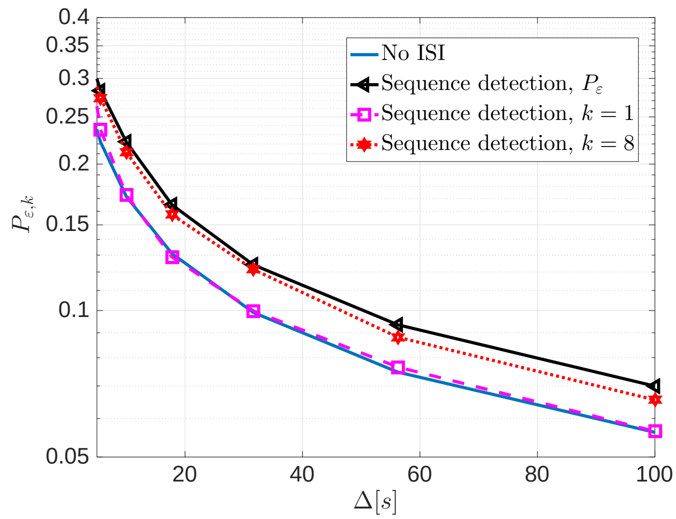

Figure 5: $P_{\varepsilon,k}$ vs. Δ for different bit indices. The simulated parameters are $c = 2$ and $T_s = 1.5\Delta$. All curves correspond to the sequence detector with $L = 2$.

through stronger coding.

Next, we numerically compare time-slotted communication with one-shot communication studied in [26].

6.2. A Comparison with One-shot Communication

We first briefly recall the problem formulation presented in Section 2. The transmitter is required to send K bits to the receiver, over K time slots of length T_s seconds each, and by using a *single* particle in each time-slot. The receiver observation window consists of $K + \kappa$ time-slots. Thus, K bits are sent, the communication duration is $(K + \kappa)T_s$ seconds, and K particles are used. A natural question that arises is: *Can a different approach for sending the K bits using K particles, with communication duration of $(K + \kappa)T_s$ seconds, perform better, i.e., achieve a lower bit error rate?*

In this section we compare the time-slotted communication approach with the one-shot transmission scheme studied in [26]. In the one-shot transmission scheme the transmitter divides the interval $[0, (K + \kappa)T_s]$ into 2^K sub-intervals of equal length. Then, it uses the k bits to select a sub-interval[9] and *simultaneously* releases all the K particles at the beginning of this sub-interval. Figure 6 illustrates this transmission scheme. The receiver decodes the transmitted symbol based on the first arrival (FA)

[9]We use a Gray coding mapping of the symbols to bits [46].

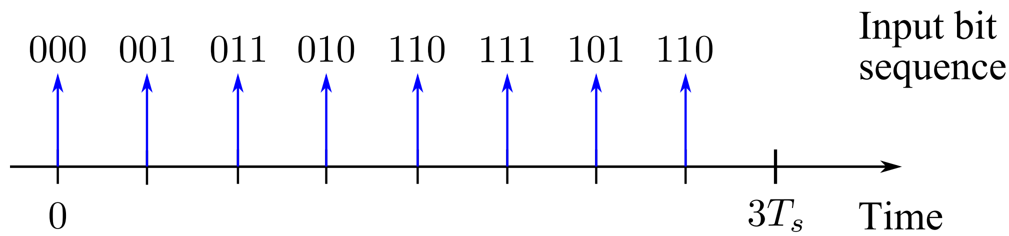

Figure 6: Illustration of the one-shot transmission scheme for $K = 3$ and $\kappa = 0$. For example, assuming the input sequence 110 the transmitter releases $K = 3$ particles at time $\frac{3T_s}{2}$.

among the K transmitted particles.[10] In [26, Sec. IV] we show that the PDF of the FA concentrates towards the release timing, thus, simultaneous release of multiple particles can significantly decrease the probability of error. Note that [26] assumes a memoryless channel; thus, in this scheme the transmitter sends *all* the bits via a single release and then remains quiet for a long period of time. In [26, Sec. V] it was also shown that the performance of the FA detector approaches the performance of the ML detector for the one-shot transmission scheme, and that FA decoding can be implemented in linear time. In the following we numerically compare the performance of the time-slotted and one-shot transmission schemes.

Figure 7 depicts the probability of bit decoding error for transmitting $K = 12$ bits, when $\kappa = 2$ and $T_s = 1.5\Delta$, for several values of Δ. In order to compare the two schemes the horizontal axis is the *normalized* time required to send a single bit. Three values of c are presented: $c = 0.1, 1$, and 2. The lines labeled with `Slotted SD` correspond to time-slotted transmission with the sequence detector presented in Section 5 implemented with $L = 2$, while the lines labeled with `One-Shot FA` correspond to one-shot transmission with the FA detector. It can first be observed that regardless of the transmission scheme, the probability of error improves when the value of c decreases. It can further be observed that for $K = 12$ the one-shot scheme significantly outperforms the time-slotted scheme. Thus, it efficiently takes advantage of the simultaneous release of multiple particles. The gap increases when the value of c decreases. In Figure 8 the same setting is presented, with the exception that $T_s = 15\Delta$. It can be observed that the performance improvement is much larger for the one-shot scheme. The reason is that for the time-slotted scheme, for each value of Δ, increasing T_s reduces the amount of ISI, yet, the performance is lower bounded by the performance of the no-ISI scenario with the respective Δ value. On the other hand, in the one-shot scheme, increasing T_s increases the total constellation interval and therefore also increases the spacing between

[10] In [26] it is shown that for small values of K, the FA and the ML detectors, see [26, Prop. 2], achieve very close performance.

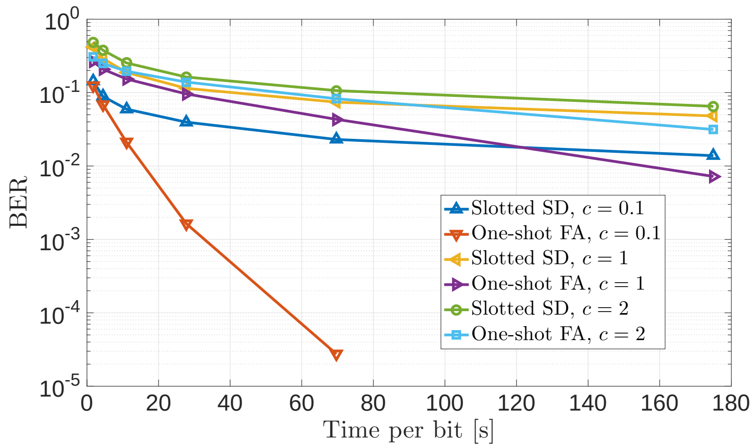

Figure 7: Probability of bit decoding error vs. the normalized time required to send a single bit. $K = 12, \kappa = 2$, and $T_s = 1.5\Delta$. All `Slotted SD` curves correspond to the sequence detector with $L = 2$.

the constellation points. Thus, increasing T_s leads to a major performance gain. Figures 7 and 8 indicate that for a small number of bits the one-shot scheme is preferable over the time-slotted scheme. A natural question is *does this relationship extend for larger values of K?*

Figure 9 depicts the probability of bit decoding error for transmitting $K = 50$ bits, when $\kappa = 2$ and $T_s = 15\Delta$. It can be observed that with the increase in the number of bits the time-slotted scheme significantly outperforms the one-shot scheme. This follows as the time-slotted scheme effectively eliminates (most of) the ISI, thus achieving performance close to the no-ISI case. On the other hand, in the one-shot scheme the number of constellation points increases exponentially while the number of particles increases only linearly. Thus, the increase in the number of particles cannot compensate for the vanishing spacing between the constellation points, and when the number of bits is moderate to high the time-slotted scheme is preferable over the one-shot scheme.

7. Conclusions and Future Work

In this work we have studied time-slotted communication over diffusion-based MT channels. Our focus was on the practical setting in which the receiver has a bounded observation window. This implies that not only do the transmitted particles arrive out-of-order due to the random propagation, but, in addition, some of them do not arrive at the receiver within the observation time. We derived three detectors for this communication system. The first is the ML detector which is based on calculating the permanent of a matrix involving the observed arrivals, and achieves *optimal* detection

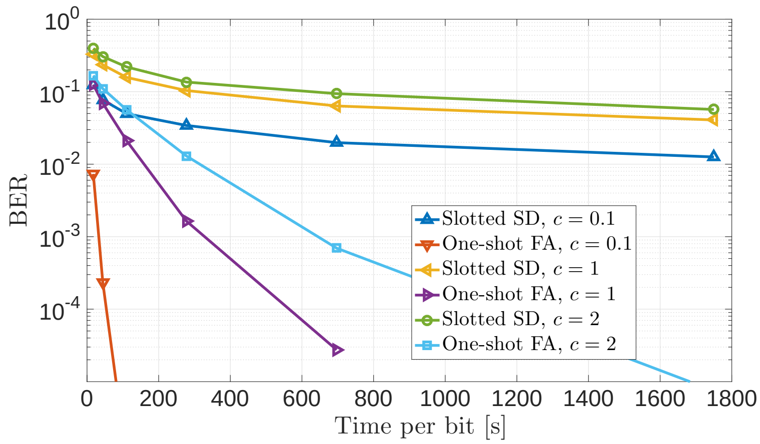

Figure 8: Bit error rate (BER) vs. the normalized time required to send a single bit. $K = 12, \kappa = 2$, and $T_s = 15\Delta$. All `Slotted SD` curves correspond to the sequence detector with $L = 2$.

performance at the cost of *exponential* complexity in the number of transmitted bits K. The second is a simple symbol-by-symbol detector which does not account for the ISI. The third detector, which accounts for the ISI while using reasonable computational complexity, is a sequence detection algorithm that is based on the VA. Yet, in contrast to the traditional VA algorithm in EM communications which assumes a linear channel model, the proposed VA is based on counting the number of arrivals in each time slot. Numerical simulations indicate that the proposed sequence detector achieves significantly better performance compared to the symbol-by-symbol detector. For a short number of transmitted symbols, the numerical simulations indicate that the performance of the proposed sequence detector is very close to the performance of the exponentially-complex ML detector. Finally, we numerically compared the time-slotted transmission scheme with a one-shot transmission scheme derived in [26] which uses the same amount of particles and requires the same total time. The comparison shows that for small number of transmitted bits, the one-shot scheme is preferable. Yet, when the number of bits increases, e.g., more than 20 bits, the proposed time-slotted transmission achieves lower bit error rate.

As a future research direction we recall the observation that the improved performance of the one-shot scheme is due to the fact that it *simultaneously* releases multiple particles, and the density of the FA concentrates towards the release timing. This motivates studying time-slotted communication in which multiple particles are released in each slot. On the other hand, releasing multiple particles in each slot leads to an increased ISI. Investigating this tradeoff is one of our future research directions.

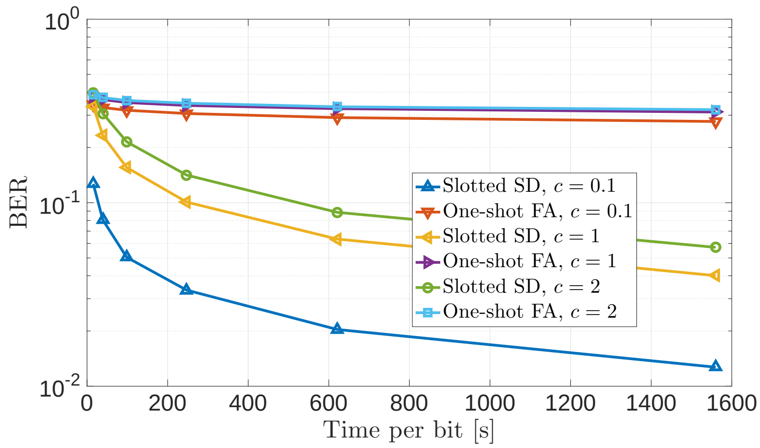

Figure 9: BER vs. the normalized time required to send a single bit. $K = 12, \kappa = 2$, and $T_s = 15\Delta$. All `Slotted SD` curves correspond to the sequence detector with $L = 2$.

Acknowledgment

This research was supported in part by the NSF Center for Science of Information (CSoI) under grant CCF-0939370, and the NSERC Postdoctoral Fellowship fund PDF-471342-2015.

Appendix A. Proof of Theorem 2

We first note that Eq. (2) implies that P_ε is the average of the probabilities of errors for different bit indices. Hence, we focus on the analysis of $P_{\varepsilon,k}$ which can be written as $P_{\varepsilon,k} = 0.5 \cdot (\Pr\{\hat{s}_k \neq s_k | s_k = 0\} + \Pr\{\hat{s}_k \neq s_k | s_k = 1\})$. Next, recall the decision rule (12), repeated here for ease of reference:

$$\hat{s}_k(\alpha_k, \beta_k) = \begin{cases} 0, & \alpha_k > 0 \text{ and } \beta_k = 0 \\ 1, & \alpha_k = 0 \\ q_k, & \alpha_k > 0 \text{ and } \beta_k > 0, \end{cases}$$

where α_k denotes the number of arrivals in the time interval $[kT_s, kT_s + \theta)$, β_k denotes the number of arrivals in the time interval $[kT_s + \theta, (k+1)T_s)$, and θ is the decision threshold from Prop 1. We begin our analysis with the case of $s_k = 0$, and note that an error occurs if one of the following two events takes place:

1. $\alpha_k = 0$ and $\beta_k \geq 0$.

2. $\alpha_k > 0$, $\beta_k > 0$, and $q_k = 1$.

We separately analyze the probabilities of each of these events.

Appendix A.1. Analysis of $\Pr\{\alpha_k = 0, \beta_k \geq 0 | s_k = 0\}$

Observe that up to time kT_s, k information particles were released. The first $k-1$ particles constitute the ISI, while the k^{th} particle is the (current) desired one. Assumption **A3**) implies that the propagation of each transmitted information particle is independent, and thus the arrival times are also independent. Therefore, the probability that no particle will arrive at the time interval $[kT_s, kT_s+\theta)$ is the product of the individual probabilities for each of the particles:

$$\Pr\{\alpha_k = 0, \beta_k \geq 0 | s_k = 0\} = \left(\prod_{n=1}^{k-1} \Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta)\}\right) \times \Pr\{y_k \neq [(k-1)T_s, (k-1)T_s+\theta) | s_k = 0\}. \quad \text{(A.1)}$$

To calculate $\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta)\}$ we first write:

$$\begin{aligned} &\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta)\} \\ &\quad = 0.5 \cdot \big(\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta) | s_n = 0\} \\ &\qquad + \Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta) | s_n = 1\}\big). \end{aligned} \quad \text{(A.2)}$$

Now, for $\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta) | s_n = 0\}$ we recall that the information particle was transmitted at time $(n-1)T_s$, and for $Z \sim \mathscr{L}(c, \mu)$, where μ is an offset parameter, we have:

$$\Pr\{Z \leq z\} = \mathsf{erfc}\left(\sqrt{\frac{c}{2(z-\mu)}}\right).$$

Hence, we write:

$$\begin{aligned} &\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta) | s_n = 0\} \\ &\quad = \Pr\{y_n < kT_s | s_n = 0\} + \Pr\{y_n \geq kT_s + \theta | s_n = 0\} \\ &\quad = \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s)}}\right) + 1 - \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s+\theta)}}\right). \end{aligned} \quad \text{(A.3)}$$

Similarly, for $\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta) | s_n = 1\}$ we recall that the information particle was transmitted at time $(n-1)T_s + \Delta$ and write:

$$\begin{aligned} &\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta) | s_n = 1\} \\ &\quad = \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s-\Delta)}}\right) + 1 - \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s-\Delta+\theta)}}\right). \end{aligned} \quad \text{(A.4)}$$

Thus, plugging (A.3) and (A.4) into (A.2) we obtain:

$$\begin{aligned} &\Pr\{y_n \neq [(k-1)T_s, (k-1)T_s+\theta)\} \\ &\quad = 0.5 \cdot \Bigg(2 + \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s)}}\right) + \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s-\Delta)}}\right) \\ &\qquad - \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s+\theta)}}\right) - \mathsf{erfc}\left(\sqrt{\frac{c}{2((k-n)T_s-\Delta+\theta)}}\right)\Bigg). \end{aligned}$$

We conclude that the first term on the RHS of (A.1) is given by $\psi(c, T_s, \Delta, k, 0, \theta)$, see (13). Following arguments similar to those leading to (A.3) we also obtain:

$$\Pr\{y_k \neq [(k-1)T_s, (k-1)T_s + \theta) | s_k = 0\} = 1 - \mathsf{erfc}\left(\sqrt{\frac{c}{2\theta}}\right),$$

which implies that (A.1) is given by:

$$\Pr\{\alpha_k = 0, \beta_k \geq 0 | s_k = 0\} = \psi(k, 0, \theta) \cdot \left(1 - \mathsf{erfc}\left(\sqrt{\frac{c}{2\theta}}\right)\right) = \rho_{k,3}^{(0)} + \rho_{k,1}^{(0)}. \tag{A.5}$$

Note that in (A.5) we split the event $\{\alpha_k = 0, \beta_k \geq 0 | s_k = 0\}$ into two distinct events: $\rho_{k,3}^{(0)}$ is the probability of the event $\{\alpha_k = 0, \beta_k > 0 | s_k = 0\}$, while $\rho_{k,1}^{(0)}$ is the probability of the event $\{\alpha_k = 0, \beta_k = 0 | s_k = 0\}$.[11] This approach is useful in the analysis of $\Pr\{\alpha_k > 0, \beta_k = 0 | s_k = 1\}$.

Finally we note that the probability of arrival of past symbols in any sub-interval of the interval $[(k-1)T_s, kT_s)$ can be calculated using the function $\psi(k, \eta_1, \eta_2)$ by properly setting the boundary parameters η_1 and η_2.

Appendix A.2. Analysis of $\Pr\{\alpha_k > 0, \beta_k > 0, q_k = 1 | s_k = 0\}$

Since q_k is independent of everything else we write:

$$\Pr\{\alpha_k > 0, \beta_k > 0, q_k = 1 | s_k = 0\} = 0.5 \cdot \Pr\{\alpha_k > 0, \beta_k > 0 | s_k = 0\}.$$

We also observe that:

$$\begin{aligned}\Pr\{\alpha_k > 0, \beta_k > 0 | s_k = 0\} = 1 - \big(&\Pr\{\alpha_k = 0, \beta_k = 0 | s_k = 0\} \\ &+ \Pr\{\alpha_k = 0, \beta_k > 0 | s_k = 0\} + \Pr\{\alpha_k > 0, \beta_k = 0 | s_k = 0\}\big).\end{aligned}$$

Hence, noting that $\Pr\{\alpha_k = 0, \beta_k = 0 | s_k = 0\} = \rho_{k,1}^{(0)}$ and that $\Pr\{\alpha_k > 0, \beta_k = 0 | s_k = 0\} = \rho_{k,2}^{(0)}$, we obtain:

$$\Pr\{\alpha_k > 0, \beta_k > 0, q_k = 1 | s_k = 0\} = 0.5 \cdot \left(1 - \sum_{j=1}^{3} \rho_{k,j}^{(0)}\right) = 0.5 \cdot \rho_{k,4}^{(0)}. \tag{A.6}$$

Combining (A.5) and (A.6) we have:

$$\Pr\{\hat{s}_k \neq s_k | s_k = 0\} = \rho_{k,1}^{(0)} + \rho_{k,3}^{(0)} + 0.5 \cdot \rho_{k,4}^{(0)}.$$

[11] This can be easily verified by following steps similar to (A.1)–(A.5)

By following arguments similar to the analysis of $\Pr\{\hat{s}_k \neq s_k | s_k = 0\}$ it can be shown that:

$$\Pr\{\hat{s}_k \neq s_k | s_k = 1\} = \rho^{(1)}_{k,2} + 0.5 \cdot \rho^{(1)}_{k,4}.$$

Finally, recalling that $P_{\varepsilon,k} = 0.5 \cdot (\Pr\{\hat{s}_k \neq s_k | s_k = 0\} + \Pr\{\hat{s}_k \neq s_k | s_k = 1\})$ and summing over k we obtain (15).

Appendix B. Proof of Proposition 2

The expression for $p^{\text{ISI}}_{k,n}$ can be obtained by following lines similar to those leading to (A.3). Based on the assumption **A3**), and noting that we consider the probability of a union of independent events, we conclude that $p^{\text{ISI}}_{k} = \sum_{n=0}^{k-1} p^{\text{ISI}}_{k,n}$. Finally, the sequence $p^{\text{ISI}}_{k}, k = 0, 1, \dots, K-1$ is an increasing sequence as when k is increased, only positive terms are added to the summation.

Appendix C. An Example for the Sequence Detection Algorithm

In this section we explicitly describe the sequence detector presented in Section 5 for the case of $L = 1$. In such case the trellis consists of two states, and the set $\mathcal{V}_{\text{tm}}$ is given by $\mathcal{V}_{\text{tm}} = \{(0,0),(0,1),(1,0),(1,1),(0,2),(2,0)\}$, while the set $\mathcal{V}_{\text{res}}$ contains the pairs $\{(\alpha,\beta) : (\alpha,\beta) \notin \mathcal{V}_{\text{tm}}, \alpha \in \mathfrak{N}, \beta \in \mathfrak{N}\}$, e.g., the pair $(2,1)$.

To calculate (21), we need to evaluate $\Pr\{V_k | \{x_i\}_{i=k-L}^{k}\}$, which for $L = 1$ is specialized to $\Pr\{V_k | x_k, x_{k-1}\}$. We implement this calculation using a pre-calculated table, which in the general setting contains $N_{\text{p}} = 2^{L+1} \cdot (|\mathcal{V}_{\text{tm}}| + 1) = 2^{L+1} \cdot \left(\frac{(L+2)(L+3)}{2} + 1\right)$ values (probabilities). In our case $N_{\text{p}} = 28$. Next, we explicitly calculate the above probabilities for the case of $\{x_k, x_{k-1}\} = \{0,0\}$; the analysis for the other cases follows similar lines.

As stated in Remark 4, we account only for the first $L+1$ time slots, while *the probability of the tail is mapped to the probability of the out of model event.* Therefore, each of the following calculations accounts for probability densities in the first two slots which follow the respective transmission.

1. $(\alpha_k, \beta_k) = (0,0)$: This case corresponds to the scenario in which no information particles arrive at the current slot. The probability that the information particle from the previous slot will not arrive in the current slot, given $x_{k-1} = 0$, is given by: $\mathsf{erfc}\left(\sqrt{\frac{c}{2T_s}}\right)$. Note that here *we ignore the tail* corresponding to $y > (k+1)T_s$. The probability that the information particle of the current slot will not arrive, given $x_k = 0$, is given by: $\mathsf{erfc}\left(\sqrt{\frac{c}{4T_s}}\right) - \mathsf{erfc}\left(\sqrt{\frac{c}{2T_s}}\right)$. Here, again, *we evaluate the*

probability density over only two time slots. Therefore, as we seek the probability of intersection of two independent events (recall assumption **A3**)), we obtain:[12]

$$\Pr\{V_k = (0,0)|0,0\} = \mathsf{erfcs}(T_s, 0) \cdot (\mathsf{erfcs}(2T_s, 0) - \mathsf{erfcs}(T_s, 0)) \,. \tag{C.1}$$

2. $(\alpha_k, \beta_k) = (0, 1)$: This event corresponds to one of the following scenarios: 1) The previous particle arrived at the second part of the slot while the current particle didn't. 2) The previous particle didn't arrive at the second part of the slot while the current particle did. Again, using assumption **A3**) we obtain:

$$\begin{aligned} \Pr\{V_k = (0,1)|0,0\} = {} & (\mathsf{erfcs}(2T_s, 0) - \mathsf{erfcs}(T_s, -\theta)) (\mathsf{erfcs}(2T_s, 0) - \mathsf{erfcs}(T_s, 0)) \\ & + \mathsf{erfcs}(T_s, 0) (\mathsf{erfcs}(T_s, 0) - \mathsf{erfcs}(\theta, 0)) \,. \end{aligned} \tag{C.2}$$

Following similar arguments for the other $(\alpha_k, \beta_k) \in \mathcal{V}_{\mathrm{tm}}$ pairs we obtain:

$$\begin{aligned} \Pr\{V_k = (1,0)|0,0\} = {} & (\mathsf{erfcs}(T_s, -\theta) - \mathsf{erfcs}(T_s, 0)) (\mathsf{erfcs}(2T_s, 0) - \mathsf{erfcs}(T_s, 0)) \\ & + \mathsf{erfcs}(T_s, 0)\mathsf{erfcs}(\theta, 0). \end{aligned} \tag{C.3}$$

$$\begin{aligned} \Pr\{V_k = (1,1)|0,0\} = {} & (\mathsf{erfcs}(T_s, -\theta) - \mathsf{erfcs}(T_s, 0)) (\mathsf{erfcs}(T_s, 0) - \mathsf{erfcs}(\theta, 0)) \\ & + (\mathsf{erfcs}(2T_s, 0) - \mathsf{erfcs}(T_s, -\theta)) \,\mathsf{erfcs}(\theta, 0). \end{aligned} \tag{C.4}$$

$$\Pr\{V_k = (0,2)|0,0\} = (\mathsf{erfcs}(2T_s, 0) - \mathsf{erfcs}(T_s, -\theta)) (\mathsf{erfcs}(T_s, 0) - \mathsf{erfcs}(\theta, 0)) \,. \tag{C.5}$$

$$\Pr\{V_k = (2,0)|0,0\} = (\mathsf{erfcs}(T_s, -\theta) - \mathsf{erfcs}(T_s, 0)) \,\mathsf{erfcs}(\theta, 0). \tag{C.6}$$

Finally, the probability assigned to $v \in \mathcal{V}_{\mathrm{res}}$ is calculated as in (21) using the expressions derived in (C.1)–(C.6).

References

[1] Y. Murin, N. Farsad, M. Chowdhury, and A. Goldsmith. On time-slotted communication over molecular timing channels. In *ACM International Conference on Nanoscale Computing and Communication*, Sep. 2016.

[2] T. Nakano, A. Eckford, and T. Haraguchi. *Molecular communication.* Cambridge University Press, 2013.

[3] N. Farsad, H. B. Yilmaz, A. Eckford, C. B. Chae, and W. Guo. A comprehensive survey of recent advancements in molecular communication. *IEEE Communications Surveys Tutorials*, 18(3):1887–1919, 2016. ISSN 1553-877X. doi: 10.1109/COMST.2016.2527741.

[12] Recall the definition $\mathsf{erfcs}(y, x) \triangleq \mathsf{erfc}\left(\sqrt{\frac{c}{2(y-x)}}\right)$.

[4] N.R Kim and C.B Chae. Novel modulation techniques using isomers as messenger molecules for nano communication networks via diffusion. *IEEE Journal on Selected Areas in Communications*, 31(12):847–856, Dec. 2013. ISSN 0733-8716. doi: 10.1109/JSAC.2013.SUP2.12130017.

[5] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz. Interference effects on modulation techniques in diffusion based nanonetworks. *Nano Communication Networks*, 3(1):65–73, March 2012.

[6] T. Nakano, Y. Okaie, and Jian-Qin Liu. Channel model and capacity analysis of molecular communication with brownian motion. *IEEE Communications Letters*, 16(6):797–800, June 2012. ISSN 1089-7798. doi: 10.1109/LCOMM.2012.042312.120359.

[7] N. Farsad, A. Eckford, and S. Hiyama. Design and optimizing of on-chip kinesin substrates for molecular communication. *IEEE Transactions on Nanotechnology*, 14(4):699–708, July 2015.

[8] A. Eckford. Nanoscale communication with brownian motion. In *Proc. of 41st Ann. Conf. on Inf. Sci. and Sys.*, pages 160–165, Baltimore, MD, 2007.

[9] N. Farsad, Y. Murin, A. Eckford, and A. Goldsmith. Capacity limits of diffusion-based molecular timing channels. In *IEEE International Symposium on Information Theory*, pages 1023–1027, Jul. 2016.

[10] Y. Murin, N. Farsad, M. Chowdhury, and A. Goldsmith. Communication over diffusion-based molecular timing channels. In *IEEE Global Commun. Conf.*, 2016.

[11] N. Farsad, W. Guo, and A. W. Eckford. Tabletop molecular communication: Text messages through chemical signals. *PLOS ONE*, 8(12):e82935, Dec 2013. doi: 10.1371/journal.pone.0082935.

[12] C. Lee, B. Koo, N.R Kim, H. B. Yilmaz, N. Farsad, A. Eckford, and C.B Chae. Molecular MIMO communication link. In *IEEE Conference on Computer Communications*, pages 1–2, Apr. 2015.

[13] M. Pierobon and I. F. Akyildiz. A physical end-to-end model for molecular communication in nanonetworks. *IEEE Jour. on Selec. Areas in Commun.*, 28(4):602–611, 2010.

[14] D. Kilinc and O.B. Akan. Receiver design for molecular communication. *IEEE Jour. on Selec. Areas in Commun.*, 31(12):705–714, Dec. 2013. ISSN 0733-8716. doi: 10.1109/JSAC.2013.SUP2.1213003.

[15] G. D. Forney. The viterbi algorithm. *Proc. IEEE*, 61(3):268–278, Mar. 1973.

[16] L-S. Meng, P-C Yeh, K-C Chen, and I. F. Akyildiz. On receiver design for diffusion-based molecular communication. *IEEE Trans. on Sig. Proc.*, 62(22):6032–6044, Nov. 2014.

[17] V. Jamali, N. Farsad, R. Schober, and A. Goldsmith. Non-coherent multiple-symbol detection for diffusive molecular communications. In *ACM International Conference on Nanoscale Computing and Communication*, Sep. 2016.

[18] A. Noel, K.C. Cheung, and R. Schober. Optimal receiver design for diffusive molecular communication with flow and additive noise. *IEEE Trans. on NanoBioscience*, 13(3):350–362, Sep. 2014.

[19] H. Arjmandi, A. Gohari, M. N. Kenari, and F. Bateni. Diffusion-based nanonetworking: A new modulation technique and performance analysis. *IEEE Communications Letters*, 17 (4):645–648, 2013.

[20] A. Borst and F. E. Theunissen. Information theory and neural coding. *Nature Neuroscience*, 2(11):947–957, Nov. 1999.

[21] B. Krishnaswamy, C. M. Austin, J. P. Bardill, D. Russakow, G. L. Holst, B. K. Hammer, C. R. Forest, and R. Sivakumar. Time-elapse communication: bacterial communication on a microfluidic chip. *IEEE Trans. on Commun.*, 61(12):5139–5151, Dec. 2013.

[22] V. Anantharam and S. Verdu. Bits through queues. *IEEE Transactions on Information Theory*, 42(1):4–18, Jan 1996. ISSN 0018-9448. doi: 10.1109/18.481773.

[23] B. Atakan. Optimal transmission probability in binary molecular communication. *IEEE Communications Letters*, 17(6):1152–1155, June 2013. ISSN 1089-7798. doi: 10.1109/LCOMM.2013.041813.130255.

[24] M. Pierobon and I.F. Akyildiz. Capacity of a diffusion-based molecular communication system with channel memory and molecular noise. *IEEE Transactions on Information Theory*, 59(2):942–954, 2013. ISSN 0018-9448. doi: 10.1109/TIT.2012.2219496.

[25] N. Farsad, W. Guo, C-B Chae, and A. W. Eckford. Stable distributions as noise models for molecular communication. In *IEEE Global Commun. Conf.*, 2015.

[26] Y. Murin, N. Farsad, M. Chowdhury, and A. Goldsmith. Optimal oetection for diffusion-based molecular timing channels. *IEEE Tran. on Sig. Proc.*, Feb. 2017. To be submitted to, available at http://web.stanford.edu/∼moriny.

[27] K. V. Srinivas, A.W. Eckford, and R.S. Adve. Molecular communication in fluid media: The additive inverse gaussian noise channel. *IEEE Trans. on Inf. Theory*, 58(7):4678–4692, Jul. 2012. ISSN 0018-9448. doi: 10.1109/TIT.2012.2193554.

[28] Hui Li, S.M. Moser, and Dongning Guo. Capacity of the memoryless additive inverse gaussian noise channel. *IEEE Journal on Selected Areas in Communications*, 32(12):2315–2329, Dec 2014. ISSN 0733-8716. doi: 10.1109/JSAC.2014.2367673.

[29] N. Farsad, Y. Murin, w. Guo, C. B. Chae, A. Eckford, and A. Goldsmith. On the impact of time-synchronization in molecular timing channels. In *IEEE Global Commun. Conf.*, 2016.

[30] C. Rose and I. S. Mian. Signaling with identical tokens: Upper bounds with energy constraints. In *IEEE International Symposium on Information Theory*, pages 1817–1821, 2014.

[31] C. Rose and I. S. Mian. A fundamental framework for molecular communication channels: timing & payload. In *IEEE Int. Conf. on Commun.*, pages 1043–1048, 2014.

[32] H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, 3rd edition, 2003.

[33] M. Marcus and H. Mine. Permanents. *The American Mathematical Monthly*, 72(6):577–591, Jun. 1965.

[34] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston, 2015. In progress, Ch. 1 at academic2.american.edu/∼jpnolan.

[35] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 1991.

[36] H. Birkan Yilmaz, A. Cem Heren, Tuna Tugcu, and Chan-Byoung Chae. Three-dimensional channel characteristics for molecular communications with an absorbing receiver. *IEEE Commun. Lett.*, 18(6):929–932, 2014.

[37] A. J. Goldsmith. *Wireless Communications*. Cambridge University Press, 1st edition, 2005.

[38] V. Jamali, A. Ahmadzade, C. Jardin, H. Sticht, and R. Schober. Channel estimation techniques for diffusion-based molecular communications. In *IEEE Int. Conf. on Commun.* IEEE, 2016.

[39] R. J. Vaughan and W. N. Venables. Permanent expressions for order statistic densities. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):308–310, 1972.

[40] A. Eckford. Molecular communication: Physically realistic models and achievable information rates. *arXiv:0812.1554*, Dec. 2008.

[41] H. Minc. Permanents. *Encyclopedia of Mathematics and Its Applications*, 6, 1982.

[42] H. J. Ryser. Combinatorial mathematics. *The Carus Mathematical Monographs*, 14, 1963.

[43] G. A. Rempala and J. Wesolowski. *Symmetric functionals on random matrices and random matching problems*. Springer-Verlag, 1st edition, 2008.

[44] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, 51(4):671–697, Jul. 2004.

[45] G. D. Forney. Maximum-likelihood sequence estimation of digital sequences in the presence of inter-symbol interference. *IEEE Transactions on Information Theory*, 18(3):363–378, May 1972.

[46] E. Agrell, J. Lassing, E. G. Storm, and T. Ottosson. On the optimality of the binary reflected gray code. *IEEE Transactions on Information Theory*, 50(12):3170–3182, Dec. 2004.