

Information Rates of ASK-Based Molecular Communication in Fluid Media

Siavash Ghavami, *Student Member, IEEE*, Raviraj Adve, *Senior Member, IEEE*
and Farshad Lahouti, *Senior Member, IEEE*

Abstract

This paper studies the capacity of molecular communications in fluid media, where the information is encoded in the number of transmitted molecules in a time-slot (amplitude shift keying). The propagation of molecules is governed by random Brownian motion and the communication is in general subject to inter-symbol interference (ISI). We first consider the case where ISI is negligible and analyze the capacity and the capacity per unit cost of the resulting discrete memoryless molecular channel and the effect of possible practical constraints, such as limitations on peak and/or average number of transmitted molecules per transmission. In the case with a constrained peak molecular emission, we show that as the time-slot duration increases, the input distribution achieving the capacity per channel use transitions from binary inputs to a discrete uniform distribution. In this paper, we also analyze the impact of ISI. Crucially, we account for the correlation that ISI induces between channel output symbols. We derive an upper bound and two lower bounds on the capacity in this setting. Using the input distribution obtained by an extended Blahut-Arimoto algorithm, we maximize the lower bounds. Our results show that, over a wide range of parameter values, the bounds are close.

Index Terms

Capacity, Inter-symbol interference, Molecular Communication, AIG channels, Amplitude-shift keying, Capacity bounds.

S. Ghavami and F. Lahouti are with the School of E&CE, College of Engineering, University of Tehran, e-mail: (s.ghavami,lahouti)@ut.ac.ir. F. Lahouti is a visiting professor of electrical engineering at California Institute of Technology, email: lahouti@caltech.edu.

R. Adve is with the Department of E&CE, University of Toronto, email: rsadve@ece.utoronto.ca.

I. INTRODUCTION

As nano-technology has received increasing attention, researchers have investigated communications based on the release of molecules in a fluid propagation medium and their detection at a receiver. Diffusion-based channels are of particular interest because diffusion is the basis of data transmission in most of cell signaling, e.g., calcium signaling, hormones, etc. [1]. Diffusion-based molecular communications systems encode information in the concentration [2], time of release [3], the number of molecules released in a time-slot (amplitude-shift keying (ASK)), the type of molecules and the ratio of different types of molecules [4]. The released molecules propagate in a fluid medium before arriving at the receiver. Of these choices, this paper focuses on amplitude shift keying and characterizes the capacity of a diffusion molecular communication channel in different settings.

The movement of molecules in a fluid medium is governed by Brownian motion [5], including when the medium itself moves with a drift velocity [6]. The propagation time between transmitter and receiver is, therefore, random. Several works have investigated the capacity of molecular communications using a fluid medium. In [7], the authors study the capacity of a discrete molecular diffusion-based channel when the information is encoded in the concentration of molecules. The work described in [8], analyzes the capacity of molecular communications with on-off keying. While most of the initial works assumed perfect synchronization, the work in [9] arrives at a more realistic model considering sequences of molecules. Recent results have also considered noise, memory and the impact of the associated physical properties in diffusion-based communications [10]. In addition to capacity, modulation, error probability, and symbol interval optimization have been investigated in, e.g., [11] [12]. The capacity of molecular communication when information is encoded in the concentration of molecules released is studied for binary communications in [13], [14] and for binary and 4-ary communications in [15]. In [13]–[15], the aggregate distribution of the number of arrived molecules in a time-slot is approximated by a Gaussian distribution.

In [3], the random propagation time between transmitter and receiver was shown to follow an additive inverse Gaussian (AIG) distribution for 1-dimensional propagation. Based on the AIG distribution, expressions and bounds for channel capacity are presented in [3], [16], [17], for the case when information is encoded in the release time of molecules.

In the model considered here, in which data transfer is based on ASK, the transmitter releases a

chosen number of molecules into a 1-D fluid medium with drift. The molecules then propagate to the receiver where they are detected and removed from the system. One of the main challenges in such a diffusion-based system is inter-symbol interference (ISI): due to the random propagation time, molecules may arrive over many time-slots. Indeed, if the system did not suffer from ISI, in our model wherein molecule release and detection are perfect, communications would also be perfect. The work in [5], disregarding ISI, studies a binary ASK scheme by considering the life expectancy of molecule with an AIG model for propagation time. Of interest here is a capacity analysis of an ASK molecular system suffering from the effects of ISI. In [18], an achievable rate and the probability of error of binary and 4-ary ASK molecular schemes are investigated in presence of ISI. Specifically, the analysis there assumes that output symbols are mutually independent, resulting in a lower bound on capacity [19].

Capacity analyses of conventional communication channels with ISI has a strong presence in the literature. Several methods have been proposed, starting from the seminal work by Hirt and Massey [20], which analyzes the capacity of the Gaussian ISI channel. The problem is solved for the case of an unrestricted input distribution, but remains open for a discrete input alphabet [21]. However, several bounds, both lower and upper, to the capacity with discrete inputs and Gaussian channels have been obtained [19]. A simulation-based approach to calculate capacity in the case of i.i.d. (independent and identically distributed) inputs with binary modulation is given in [22]. This approach is based on a trellis structure where the number of states relates to the memory of the ISI channel.

In this paper, with information encoded in the number of molecules released, the receiver counts the number of molecules received within each time-slot. We obtain the probability of molecules arriving within a specific time-slot using the AIG distribution; this leads to a binomial distribution on the number of molecules received in each time-slot. We begin by analyzing ASK over such a channel in the basic discrete memoryless setting. In Section III we model the system as a memoryless channel. This molecular discrete memoryless channel (DMC) may be motivated by a molecular communication system with life expectancy of molecules being equal to the time-slot duration. This implies that the molecules disappear or decompose after the duration of a time slot. Specifically, in Section III, the molecules do not cause ISI.

We also study the capacity per channel use, the capacity per unit time, and the capacity per unit cost and examine the effects of possible constraints on peak and average molecular cost. Capacity per

channel use and time are useful measures in applications such as drug delivery systems (DDS) [23] and communication over fluid channels. The capacity per unit cost is the objective function in cost efficient communications, where the information rate per molecular emission is more important. To the best of our knowledge, this is the first study of capacity per unit cost in molecular communications. Constraints on transmission cost address certain practical constraints in molecular communications. Specifically, the peak constraint acts as a toxicity constraint in DDS [23]. Similarly, the average constraint reflects a limitation on the average input emission or injection rate in molecular communications or DDS [24]. For DMC, we obtain the optimal input distribution and the resulting quantity for different capacity measures with average or peak molecular cost constraints. We also use this analysis to obtain the optimum transmission symbol interval as a function of the channel model parameters in each case.

We then move on to analyzing the impact of ISI in ASK-based molecular communications. As in [19], with some abuse of notation, we call $I_{i.i.d}$ the capacity with i.i.d. inputs, the channel capacity. We propose two lower bounds and an upper bound for the $I_{i.i.d}$ of such a channel. To provide a tractable exposition, we restrict the effect of ISI to one time-slot and analyze the average mutual information with correlated channel outputs. The lower bounds are maximized by optimizing the input distribution using a modified Blahut-Arimoto algorithm. As our results show, over a large range of parameter values, the lower and upper bounds are close, thereby well characterizing the capacity of such a molecular communication channel.

A crucial question left unanswered by the presented capacity analysis is what the impact of the simplifying assumption on one time-slot memory is. We are unable to directly evaluate the quality of this assumption since the true channel capacity is unknown. As a proxy, we develop the maximum likelihood (ML) detector as a possible implementation of the molecular communication receiver. We evaluate the error rates assuming ISI restricted to one time-slot and compare the performance to simulations which do not impose this assumption. As we will see, this allows us to characterize the range of parameters over which the assumption holds true.

The remainder of this paper is organized as follows: the system model and the problem statement are presented in Section II. Section III focuses on capacity analyses of a discrete memoryless ASK molecular communications, while Section IV presents our analysis in presence of ISI. Numerical results illustrating the analyses are presented in each section. Finally, Section V wraps up the paper with concluding remarks.

II. SYSTEM MODEL AND PROBLEM STATEMENT

A. System Model

The transmitter is a point source of identical molecules. Using ASK, at the beginning of every time-slot of length T , it transmits a message by releasing X , $0 \leq X \leq X_{\max}$, molecules into the fluid medium. Here, X_{\max} is the maximum number of molecules releasable in any time-slot. The transmitter does not affect the propagation of the molecules. The channel is one-dimensional. Molecules propagate between the transmitter and receiver by Brownian motion characterized by a diffusion constant d and (positive) drift velocity v .

At the receiver, all received molecules are absorbed and removed from the system. We consider two cases: first, we ignore ISI and assume that a molecule released in time-slot m arrives within the same time-slot or disappears; this leads to a DMC. In the second half of the paper, we consider one slot of ISI, i.e., we assume molecules that do not arrive within two time-slots have disappeared¹. Everything else within the system operates perfectly, i.e., the only randomness in our model is the propagation time (equivalently, the number of molecules that are received in a time-slot). If l denotes the distance between transmitter and receiver ², using the AIG analysis in [3], the cumulative distribution function (CDF) of the propagation time is given by

$$F_W(w) = \Phi\left(\sqrt{\frac{\lambda}{w}}\left(\frac{w}{\mu} - 1\right)\right) + e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{w}}\left(\frac{w}{\mu} + 1\right)\right), w > 0. \quad (1)$$

Here, $\Phi(\cdot)$ is the CDF of a standard Gaussian random variable, $\mu = l/v$, $\lambda = l^2/\sigma^2$ and $\sigma^2 = d/2$ is the variance of the associated Weiner process [3]. It is noteworthy that in a 3-dimensional environment the first passage process is transient, i.e., there is a non-zero probability of the molecule never arriving at the receiver [25]. On the other hand, this is not true for the 1-D propagation under consideration; the first passage process for the 1-D case is said to be “recurrent”.

When the transmitter releases $X = x$ molecules in a time-slot, the probability of receiving $Y = y$ of

¹Note that one could use our analysis to avoid this assumption. However, the exposition becomes unwieldy and exponentially complicated to deal with.

²The units of l and v are normally in μm and $\mu m/sec$ but but any scaled version of these units can be used

these molecules in the same time-slot is given by

$$\Pr(Y = y | X = x) = \begin{cases} \binom{x}{y} q_1^y (1 - q_1)^{x-y}, & 0 \leq y \leq x \\ 0, & y < 0, y > x, \end{cases} \quad (2)$$

where $q_1 = F_W(T)$ is the probability of a molecule arriving within the same time-slot of its release. In general, $q_k = F_W(kT) - F_W((k-1)T)$ denotes the probability of a molecule arriving in the k -th time-slot after transmission.

At the start of time-slot k , $X_k \in \{0, 1, \dots, X_{\max}\}$ molecules are released. Let $a_x = \Pr(X = x)$, $0 \leq x \leq X_{\max}$, denote the probability of releasing x molecules at the transmitter, and Z_k denote the number of molecules in time-slot k , that were released but *not received* in slot k . If Y_k denotes the number of molecules received in time-slot k , we have

$$Y_k = X_k - Z_k + N_{k-1}, \quad Y_k \in \{0, 1, \dots, kX_{\max}\}, \quad (3)$$

in which, $N_{k-1} \in \{0, 1, \dots, (k-1)X_{\max}\}$ denotes the total number of “interfering” molecules from the previous $k-1$ time-slots arriving in time-slot k . Considering $P_k(n) = \Pr(N_{k-1} = n)$, the channel transition probability when $X_k = x$ molecules are transmitted is given by

$$\begin{aligned} p_{Y_k|X_k}(y | X_k = x) &= (1 - q_1)^x P_k(y) + \binom{x}{1} q_1 (1 - q_1)^{x-1} P_k(y-1) + \dots \\ &\binom{x}{x-1} q_1^{x-1} (1 - q_1) P_k(y - (x-1)) + q_1^x P_k(y - x), \quad y = 0, \dots, X_{\max}k - (X_{\max} - x). \end{aligned} \quad (4)$$

In (4), $P_k(y)$ can be easily calculated by induction as

$$\begin{aligned}
P_k(y) &= \left(a_0 + (1 - q_k) a_1 + \dots + (1 - q_k)^{X_{\max}} a_{X_{\max}} \right) P_{k-1}(y) + \\
&\left(q_k a_1 + \binom{2}{1} q_k (1 - q_k) a_2 + \dots + \binom{X_{\max}}{1} q_k (1 - q_k)^{X_{\max}-1} a_{X_{\max}} \right) P_{k-1}(y-1) + \\
&\vdots \\
&\left(q_k^{X_{\max}-1} a_{X_{\max}-1} + \binom{X_{\max}}{X_{\max}-1} q_k^{X_{\max}-1} (1 - q_k) a_{X_{\max}} \right) P_{k-1}(y - (X_{\max} - 1)) + \\
&q_k^{X_{\max}} a_{X_{\max}} P_{k-1}(y - X_{\max})
\end{aligned} \tag{5}$$

where, as defined, $a_i = \Pr(X = i)$ for $i \in \{0, \dots, X_{\max}\}$. If, as later in Section IV, we were to assume that the ISI only affects the next time-slot, (5) is simplified to

$$P_2(0) = \left(a_0 + (1 - q_2) a_1 + \dots + (1 - q_2)^{X_{\max}} a_{X_{\max}} \right) P_1(0), \tag{6a}$$

$$P_2(1) = \left(q_2 a_1 + \binom{2}{1} q_2 (1 - q_2) a_2 + \dots + \binom{X_{\max}}{1} q_2 (1 - q_2)^{X_{\max}-1} a_{X_{\max}} \right) P_1(0), \tag{6b}$$

\vdots

$$P_2(X_{\max}) = q_2^{X_{\max}} a_{X_{\max}} P_1(0), \tag{6c}$$

where $P_1(0) = 1$ as 0 molecules are received from the previous 0 time-slots.

B. Problem Statement

For the ISI channel, the average mutual information per channel use is given by

$$I = \lim_{L \rightarrow \infty} \frac{1}{L} I(X^L; Y^{L+k}), \tag{7}$$

in which, $X^L = [X_1, \dots, X_L]$ and $Y^{L+k} = [Y_1, \dots, Y_{L+k}]$, denote the length- L input and $L + k$ output sequences, k is the length of ISI, and the mutual information is evaluated for a given joint input distribution $P_{X^L}(x^L)$. This determines the achievable rate of reliable communication through this channel with this specific input distribution; the channel capacity is the supremum of this mutual information over all allowed joint input distributions.

Obtaining the capacity of the ISI channel requires optimization over the joint input distribution, a

seemingly intractable problem. As suggested in [19], we focus on the case of i.i.d. inputs, i.e., $P_{X^L} = \mathbf{a} \times \dots \times \mathbf{a} = \mathbf{a}^L$, where $\mathbf{a} = [a_0, \dots, a_{X_{\max}}]$. We denote the resulting average mutual information as $I_{i.i.d}$ [19]. Note that this mutual information is a lower bound on the expression in (7), but, as in [19], with a slight abuse of notation we will call the resulting mutual information the ‘‘channel capacity’’. This capacity is given by

$$C = \lim_{L \rightarrow \infty} \frac{1}{L} \sup_{\mathbf{a}^L} I(X^L; Y^{L+k}), \quad (8)$$

Despite this significant simplification, calculating the channel capacity in (8) appears intractable and, in the next sections, we investigate two related scenarios. We first ignore ISI and obtain the capacity of the resulting DMC under various possible constraints. In Section IV, we consider an ISI channel with a memory of one time-slot.

III. CAPACITY ANALYSIS OF MOLECULAR DMC WITH ASK

We begin by analyzing several capacity measures with different cost constraints for the basic discrete memoryless molecular communication channel with ASK. As described in Section II, the DMC model may also be motivated by a molecular communication system where the life expectancy of molecules is equal to the time-slot duration. Specifically, we derive the capacity per channel use, per unit time and per unit cost with possible constraints on peak and average cost of transmission. In each case, the optimum input distribution and the resulting maximized capacity measure are also quantified.

A. Capacity Per Channel Use

The general capacity problem with only a constraint on the maximum number of transmitted molecules per channel use (akin to a peak constraint on the input) is as follows

$$\begin{aligned} & \sup_{P(X_m)} I(X_m; Y_m), \\ \text{subject to} & \quad 0 \leq X_m \leq X_{\max} \end{aligned} \quad (9)$$

where X_{\max} is the allowed maximum number of transmitted molecules per time-slot. Note that for this case, we allow the transmitter to not release any molecules in a time-slot, i.e., $X = 0$ is allowed. The objective function in (9) is concave with respect to the input probability vector \mathbf{a} and the constraint is

linear, hence the optimization problem is concave. Hence, the solution of problem in (9) can be obtained using the Blahut-Arimoto algorithm [26].

To describe the algorithm as applied here, using (2), we define an $(X_{\max} + 1) \times (X_{\max} + 1)$ channel transition matrix \mathbf{P} as $\mathbf{P}_{x,y} = [p(Y_m = y | X_m = x)]$, where, $y \in [0, \dots, X_{\max}]$ and $x \in [0, \dots, X_{\max}]$; and matrix \mathbf{Q} with $\mathbf{Q}_{y,x} = [p(X_m = x | Y_m = y)]$ with the same size as \mathbf{P} , where $\mathbf{P}_{x,y}$ ($\mathbf{Q}_{y,x}$) denotes the (x, y) ((y, x))-th element of \mathbf{P} (\mathbf{Q}). Let

$$J(\mathbf{a}, \mathbf{P}, \mathbf{Q}) = \sum_x \sum_y a_x \mathbf{P}_{x,y} \log \frac{\mathbf{Q}_{y,x}}{a_x}. \quad (10)$$

Then the following is true [26]:

- 1) $C = \max_{\mathbf{Q}} \max_{\mathbf{a}} J(\mathbf{a}, \mathbf{Q}, \mathbf{P})$.
- 2) For fixed \mathbf{a} , $J(\mathbf{a}, \mathbf{P}, \mathbf{Q})$ is maximized by

$$\mathbf{Q}_{y,x} = \frac{a_x \mathbf{P}_{x,y}}{\sum_j a_x \mathbf{P}_{x,y}}. \quad (11)$$

- 3) For fixed \mathbf{Q} , $J(\mathbf{a}, \mathbf{P}, \mathbf{Q})$ is maximized by

$$a_x = \frac{\exp\left(\sum_y \mathbf{P}_{x,y} \log \mathbf{Q}_{y,x}\right)}{\sum_x \exp\left(\sum_y \mathbf{P}_{x,y} \log \mathbf{Q}_{y,x}\right)}, \quad 0 \leq x \leq X_{\max}. \quad (12)$$

The Blahut-Arimoto algorithm begins with the transition probability matrix defined by (2) and an arbitrary, but valid, choice for \mathbf{a} . It then iterates between steps 2 and 3 above until convergence. Since, the mutual information in (9) is concave in terms of input probability, the output of the algorithm is the optimal, capacity-achieving, input probability distribution, $\hat{\mathbf{a}}$; as an aside, the matrix \mathbf{Q} is the corresponding output-input reverse transition matrix.

Figure 1 illustrates the results of using the Blahut-Arimoto algorithm in obtaining the optimal input distribution. Figure 1a plots the optimum input distribution, a_x , in terms of T with $l = 10^{-2}$, $v = 1$, $\sigma^2 = 1$ and $X_{\max} = 10$. Interestingly, though perhaps intuitively, the optimal distribution transitions from a bipolar distribution for small T to a uniform distribution for large T (system essentially perfect). In all cases in the figure, as T increases, the capacity converges to $\log_2(X_{\max} + 1)$.

Figure 1b plots the corresponding capacities as a function of the time-slot duration, T , for different values of σ^2 and the same, fixed, values of l and v . We observe that, for the given values of v and l , by increasing σ^2 and T , the capacity increases. While it is intuitive that increasing T should increase

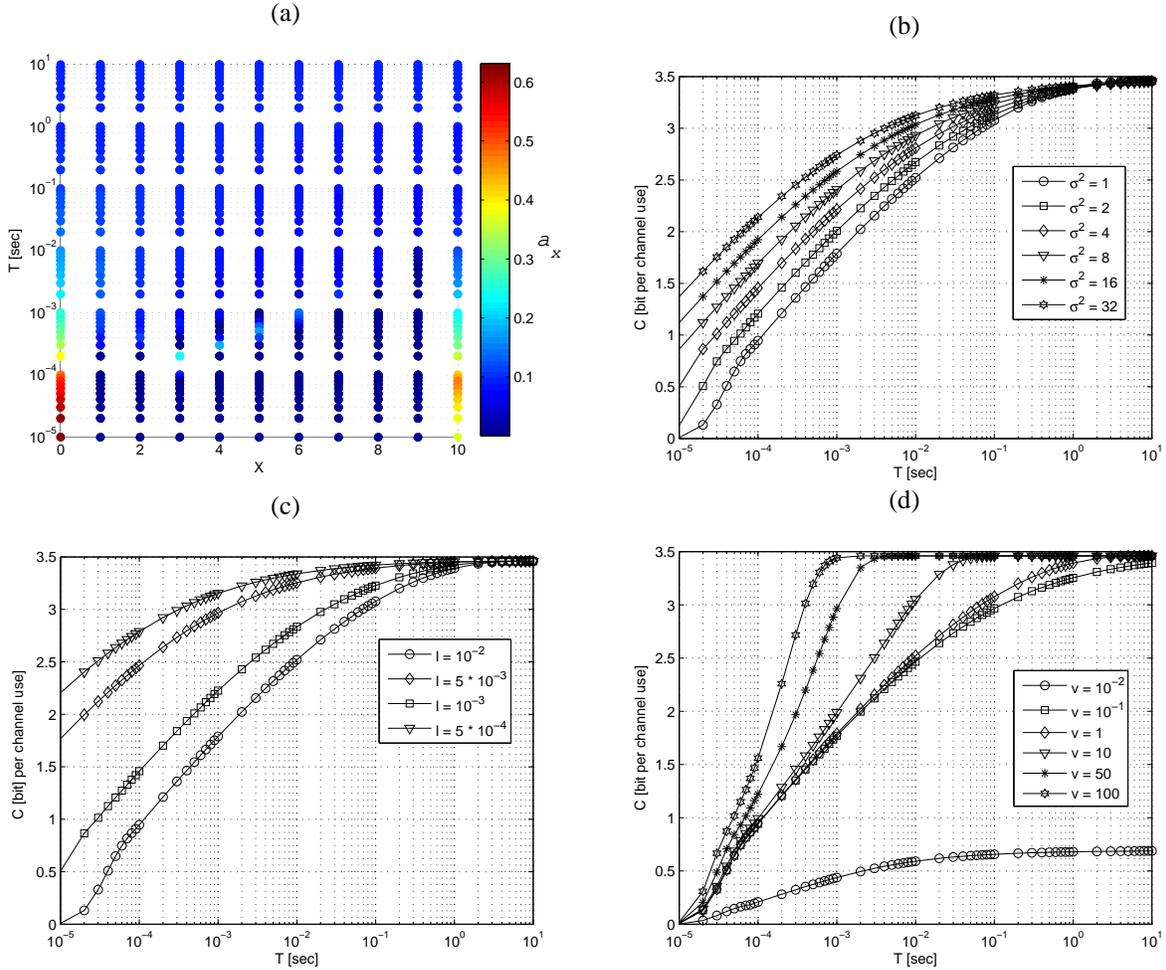


Fig. 1: (a) Optimum a_x in terms of T with $l = 10^{-2}$, $v = 1$, $\sigma^2 = 1$. (b-d) C in terms of T for different values of (b) σ^2 with $v = 1$ and $l = 10^{-2}$, (c) l with $\sigma^2 = 1$ and $v = 1$ and (d) v with $l = 10^{-2}$ and $\sigma^2 = 1$, all with $X_{\max} = 10$.

the capacity per channel use, by increasing the probability of receiving all molecules in one time-slot, the increase with increasing σ^2 is less intuitive. The explanation is that increasing the diffusion constant increases the randomness in the location of the molecules, and the randomness helps improve the chances of molecules arriving at the destination within one time-slot. This is consistent with the fact that $q_1 = F_W(T)$ is an increasing function of σ for low-to-medium values of v . For large v , $q_1 \simeq 1$ and capacity is near its maximum value. It is worth noting that when information is encoded in time-of-release as in [3], the mutual information is not monotonic in σ .

Figures 1c and 1d plot the capacity for different values of l and v , respectively. As expected, decreasing l (equivalently, increasing v) improves capacity significantly because, in each case, clearly, q_1 increases.

B. Capacity with Average Cost Constraint

A variation on the optimization in (9) is when the average cost of transmitted molecules per channel use is also constrained. The resulting optimization problem is

$$\sup_{p(X_m)} I(X_m; Y_m), \quad (13a)$$

$$\text{subject to } 0 \leq X_m \leq X_{\max} \quad (13b)$$

$$\mathbb{E}(e_{X_m}) \leq E \quad (13c)$$

where $e_{X_m=x} \geq 0$ is the cost of using x molecules as the ASK symbol. The vector $\mathbf{e} = [e_x]$, $x \in \{0, 1, \dots, X_{\max}\}$ specifies the cost vector. A concrete example is when the transmission cost is equal (proportional) to the number of transmitted molecules per channel use, i.e., $e_x = x$. The capacity with cost constraint E is then defined as

$$C(E) = \max_{\mathbf{a} \in A_E} \sum_x \sum_y a_x \mathbf{P}_{x,y} \log \frac{\mathbf{P}_{x,y}}{\sum_x a_x \mathbf{P}_{x,y}} = \max_{\mathbf{a} \in A_E} I(\mathbf{a}, \mathbf{P}) \quad (14)$$

where $A_E = \{\mathbf{a} \mid \sum_x a_x e_x \leq E\}$ denotes the set of all allowable input distributions that meet the cost constraint.

With a concave cost function and linear constraints, the optimization problem in (14) is concave. Hence, the solution can be obtained using the constrained Blahut-Arimoto algorithm [26]. Using Lagrange multipliers, the cost function can be parameterized as

$$C(E) = \max_{\mathbf{a}} \left[\sum_x \sum_y a_x \mathbf{P}_{x,y} \log \frac{\mathbf{P}_{x,y}}{\sum_x a_x \mathbf{P}_{x,y}} - s \left(\sum_x a_x e_x - E \right) \right], \quad (15)$$

where s denotes the Lagrange multiplier. The maximization is now over all input probability vectors \mathbf{a} . For any matrix \mathbf{Q} with size $(X_{\max} + 1) \times (X_{\max} + 1)$, let

$$J(s, \mathbf{a}, \mathbf{P}, \mathbf{Q}) = \sum_x \sum_y a_x \mathbf{P}_{x,y} \log \frac{\mathbf{Q}_{y,x}}{a_x} - s \sum_x a_x e_x \quad (16)$$

Hence, using (16), the following is true [26]

- 1) The constrained capacity $C = s \sum_x x a_x + \max_{\mathbf{Q}} \max_{\mathbf{a}} J(s, \mathbf{a}, \mathbf{P}, \mathbf{Q})$

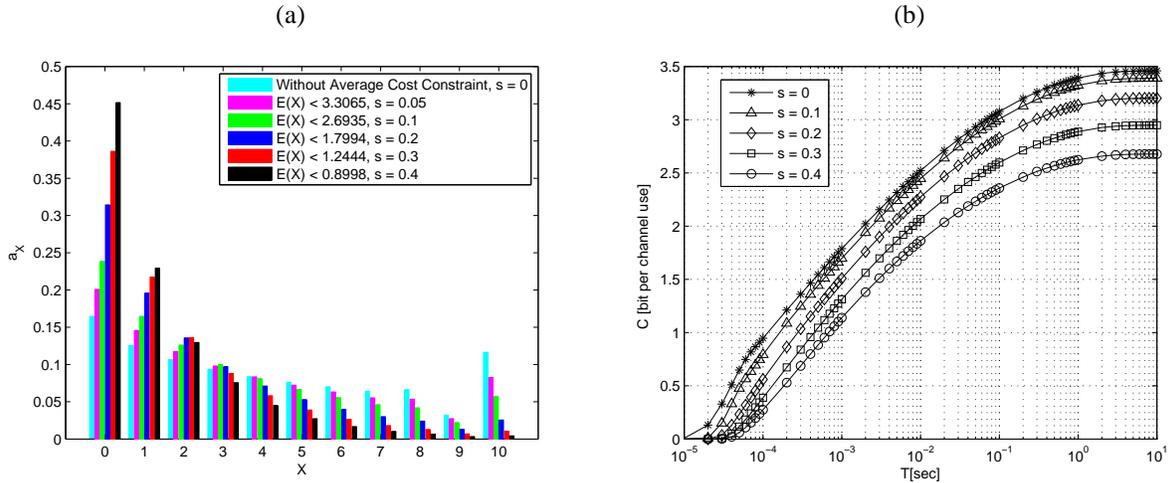


Fig. 2: (a) Optimum a_x for different values of average cost constraint with $T = 10 \text{ msec}$, (b) C in terms of average cost constraint for different values of T , all with $\sigma^2 = 1$, $v = 1$, $l = 10^{-2}$, $X_{\max} = 10$.

2) For fixed \mathbf{a} , $J(s, \mathbf{a}, \mathbf{P}, \mathbf{Q})$ is maximized by

$$\mathbf{Q}_{y,x} = \frac{a_x \mathbf{P}_{x,y}}{\sum_x a_x \mathbf{P}_{x,y}} \quad (17)$$

3) For fixed s , \mathbf{P} , $J(s, \mathbf{a}, \mathbf{P}, \mathbf{Q})$ is maximized by

$$a_x = \frac{\exp\left(\sum_y \mathbf{P}_{x,y} \log \mathbf{Q}_{y,x} - s e_x\right)}{\sum_x \exp\left(\sum_y \mathbf{P}_{x,y} \log \mathbf{Q}_{y,x} - s e_x\right)}. \quad (18)$$

4) The optimum s is the minimum value (≥ 0) that satisfies the constraint $\sum_x e_x a_x \leq E$.

The procedure to optimize for the input distribution is similar to that without the average cost constraint. The one additional step is to obtain the parameter s after updating the distribution vector \mathbf{a} . As per 4) above, the minimum value of s can be obtained by a bisection search. To obtain the optimal distribution, the algorithm is executed setting $E = \sum_x e_x a_x$. In the rest of the paper, unless otherwise mentioned, we consider $e_x = x$.

Figure 2 plots the result of including an average cost constraint within the optimization problem of (13). For ease of illustration, in Fig. 2a, with $T = 10 \text{ msec}$, we vary the Lagrange multiplier s , to vary the average transmission cost. Increasing the parameter s in (18) reduces $E = \mathbb{E}(e_x)$ as expected; further, reducing E , i.e., making transmissions expensive, shifts the optimal probability distribution to the lower values of x . Figure. 2b plots the constrained capacity in bits per channel use in terms of average cost constraint, i.e. E , in molecules, for different values of T . It is evident that by increasing E and T the

capacity increases, i.e., as expected, a stricter constraint reduces the achievable rate.

C. Capacity Per Unit Time

It is worth noting that the solutions to the optimization problems in Sections III-A and III-B can be used to solve other problems of interest such as to maximize the mutual information per unit time with constraints on the maximum and/or average number of transmitted molecules per unit time. The general optimization problem is as follows

$$\sup_{p(X_m)} \frac{I(X_m; Y_m)}{T} \quad (19a)$$

$$\text{subject to } 0 \leq \frac{X_m}{T} \leq P_{\max}, \quad (19b)$$

$$0 \leq E\left(\frac{e^{X_m}}{T}\right) \leq \bar{P} \quad (19c)$$

where, P_{\max} and \bar{P} are the maximum and the average number of transmitted molecules allowed *per unit time*. However, since T is a constant and independent of a_x , we can use the same approach as described above by setting X_{\max} to be the integer closest to $(P_{\max}T)$ and $E = \bar{P}T$.

We note that a related problem of interest is obtaining the optimal time-slot duration, T_{opt} , that maximizes the capacity per unit time for a given set of system parameters (v , l and σ^2). Unfortunately, this optimization problem is not concave and so we have resorted here to numerical solutions.

Figures 3a and 3b plot C/T in bits/sec and q_1 , respectively, as functions of time-slot duration T , for different values of v . Here, $\sigma^2 = 0.5$, $l = 10^{-2}$ and $X/T \leq 20 \times 10^3$ molecules/sec. As expected, there is an optimum value for time-slot duration, T_{opt} , which maximizes the $I(X_m; Y_m)/T$, though, interestingly, this duration does not vary significantly with drift velocity. As drift velocity increases, the optimal value slowly decreases. Locating the corresponding values of q_1 in Fig 3b we observe that the maximum value of C/T corresponds to a value of q_1 higher than 0.9. For such a large probability that the molecules arrive within the same time-slot, the assumed DMC model turns out to be essentially accurate. Hence, we plot C/T in Figure 3a for the range of T which corresponds to this value of q_1 .

Figure 3c plots T_{opt} , derived numerically, as a function of l for different values of σ^2 . We observe that increasing σ^2 , T_{opt} reduces while it increases with increasing l (to ensure that an adequate percentage of molecules arrive within the time-slot).

Figures 4a and 4b plot $E(X/T)$ and C/T in terms of T for different values of s . In the results of

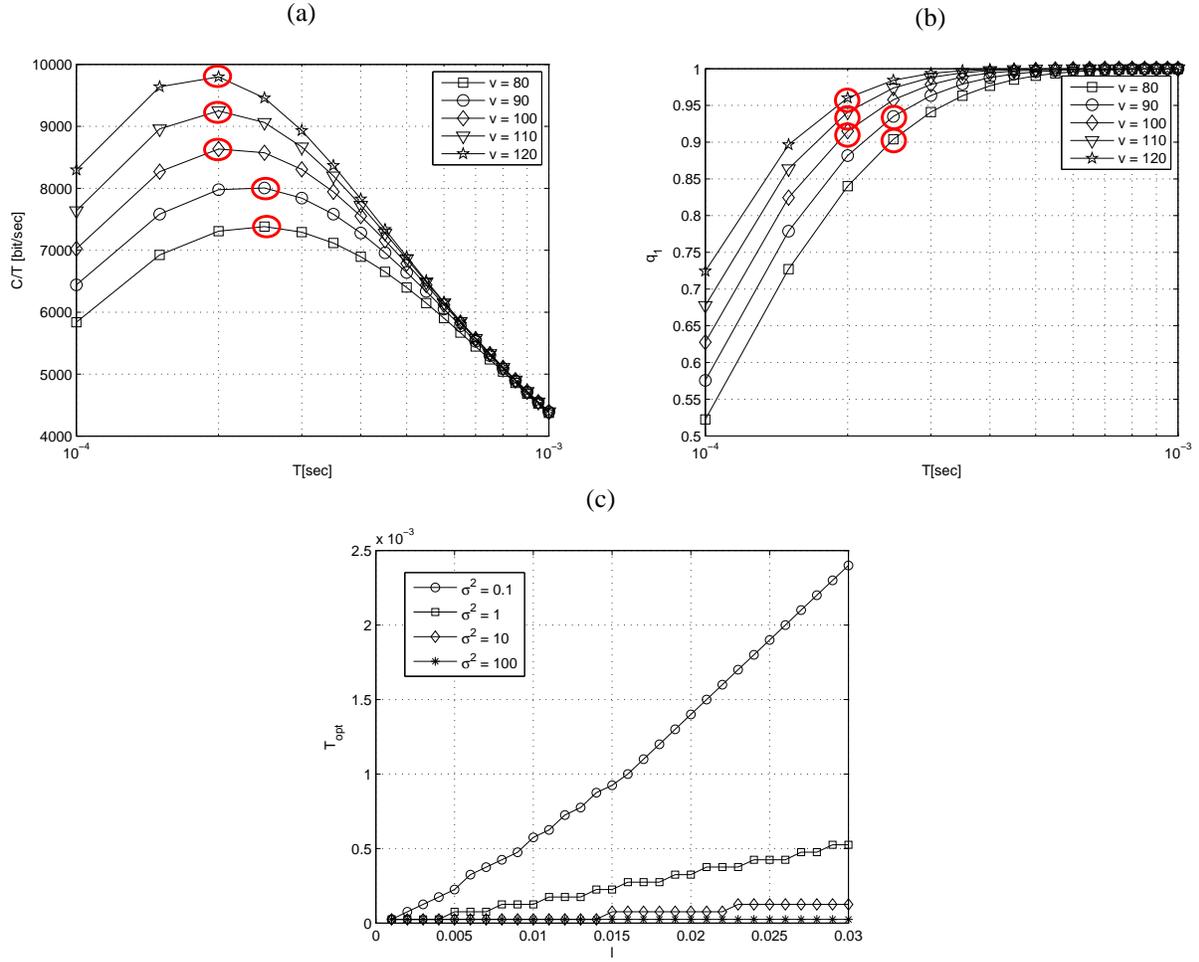


Fig. 3: (a) C/T and (b) q_1 in terms of T for different values of v with $\sigma^2 = 0.5$, $l = 10^{-2}$, (c) T_{opt} in terms of l for different values of σ^2 with $v = 1$

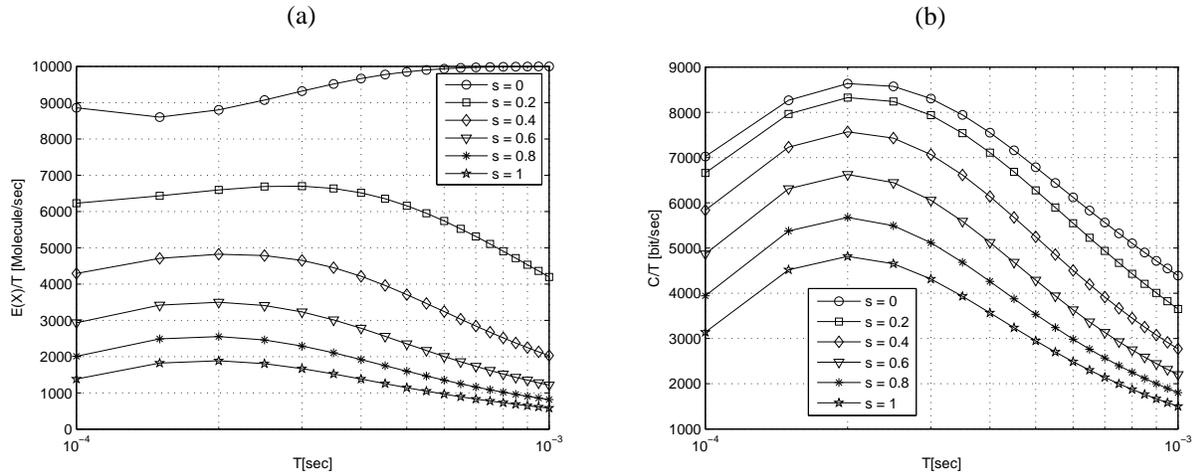


Fig. 4: (a) $E(X/T)$ (b) C/T in terms of T , for different values of s and $e(X) = X$, all with constraint on average transmuted molecules per second and $l = 10^{-2}$, $v = 100$, $\sigma^2 = 0.5$ and $X/T \leq 20 \times 10^3$ molecules/sec.

these figures, the input distribution is derived from optimization problem in (19). Here $E(X/T)$ and C/T decrease with increasing s . Since, s is a Lagrange multiplier, in (13), increasing its value tightens the constraint in (13c), i.e., $E(X/T)$ and C/T are also reduced.

D. Capacity Per Unit Cost

A final optimization problem within the DMC framework of interest is the relative capacity, i.e., the capacity per unit cost. This is akin to optimizing the “energy efficiency” in molecular communications. Here, we investigate maximizing the mutual information per average transmission cost when the maximum number of transmitted molecules per channel use is limited. The associated optimization problem is given by

$$\begin{aligned} & \sup_{p(X_m)} \frac{I(X_m; Y_m)}{E(e_{X_m})} & (20) \\ \text{subject to} & \quad 0 \leq X_m \leq X_{\max} \end{aligned}$$

The ratio of mutual information to transmission cost, as in (20), is defined in [27] as relative capacity, usually denoted by C_R .

In (20) the choice of cost $e_{X_m=x} = x$ is the average number of molecules transmitted, $E(X)$; allowing $X = 0$ implies that we allow for a symbol with zero cost. Clearly, any optimization would maximize use of the symbol ($a_0 = 1$). To avoid form zero information rate, we can add a constant, i.e., use $e_x = c_0 + x$, where c_0 is a small constant value, ensuring that there is no symbol with zero cost.

Since the relative capacity, $I(X_m; Y_m)/E(e_{X_m})$, is a continuous and quasi concave function of \mathbf{a} , it may be maximized using the Jimbo-Kunisawa algorithm [28]. The corresponding iterative procedure is as follows:

- 1) Initially, choose an arbitrary probability vector $\mathbf{a}^{(0)}$,
- 2) After the r th iteration, having obtained probability vector $\mathbf{a}^{(r)}$, construct the $(r + 1)$ th probability vector $\mathbf{a}^{(r+1)}$ as follows;

$$\tilde{a}_x^{(r+1)} = a_x^{(r)} \exp \left[\frac{e_0 D_x(\mathbf{a}^{(r)})}{e_x} \right], x = 0, \dots, X_{\max}, \quad (21)$$

where

$$D_x(\mathbf{a}^{(r)}) = \sum_y \mathbf{P}_{x,y} \log \frac{\mathbf{P}_{x,y}}{\sum_x a_x^{(r)} \mathbf{P}_{x,y}}. \quad (22)$$

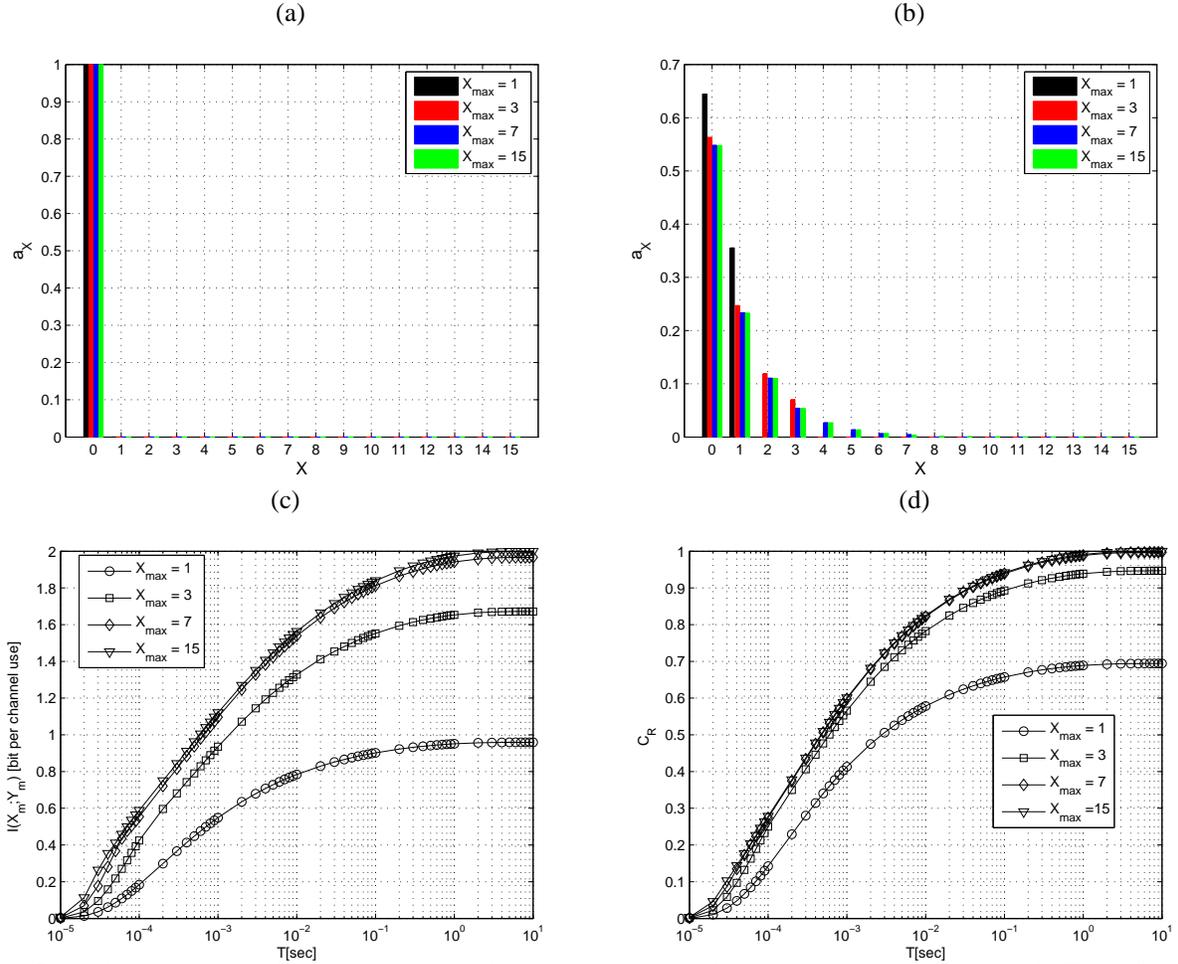


Fig. 5: (a) Optimum a_x which maximizes relative capacity for different values of X_{\max} with $T = 10\text{msec}$, $c_0 = 0$. (b) Same as (a) with $c_0 = 1$. (c) $I(X_m; Y_m)$ and (d) C_R in terms of T for different values of X_{\max} with $c_0 = 1$ and all with $\sigma^2 = 1$, $v = 1$, $l = 10^{-2}$.

Finally, we need to normalize the distribution using

$$a_x^{(r+1)} = \frac{\tilde{a}_x^{(r+1)}}{\sum_x \tilde{a}_x^{r+1}}. \quad (23)$$

We iterate until $|C_R^{(r+1)} - C_R^{(r)}| < \varepsilon$, where $C_R^{(r)}$ is the value of objective function in (20) at r th iteration and $\varepsilon > 0$ is arbitrary small number.

Figure 5 plots the results of the optimization problem in (20). Figure 5a plots the optimum input distribution which maximizes (20) without shifting the cost function, i.e., letting $c_0 = 0$. As is clear, for all values of X_{\max} , the solution is to never transmit any molecules!

The remaining figures in this set use $c_0 = 1$, i.e., $e_x = x + 1$. Figure 5b plots the optimum input distribution which maximizes (20) for $T = 10\text{ms}$. Comparing Figs. 5a and 5b, we observe that by

TABLE I: Summary of results for Capacity in DMC.

| Objective Function | Transmission Cost function | Solution | Parameters | | | | | |
|----------------------|--|-------------------|------------|----------|-----|-----|-----|--------|
| | | | X_{max} | σ | v | l | T | $E(X)$ |
| $I(X_m; Y_m)$ | $0 \leq X_m \leq X_{max}$ | (12) | ↑ | ↑ | ↑ | ↓ | ↑ | NA |
| | $X_m \leq X_{max}, E(X_m) \leq E$ | (18) | | | | | | ↑ |
| $I(X_m; Y_m)/T$ | $0 \leq X_m \leq X_{max}$ | (12) | ↑ | ↑ | ↑ | ↓ | ↗↘ | NA |
| | $0 \leq X_m \leq X_{max}, E(X_m) \leq E$ | (18) | | | | | | ↑ |
| $I(X_m; Y_m)/E(X_m)$ | $0 \leq X_m \leq X_{max}$ | (23) | ↓ | ↑ | ↑ | ↓ | ↑ | NA |
| P_e | $0 \leq X_m \leq X_{max}$ | (44) | ↑ | * | ↓ | ↑ | ↓ | NA |
| T_{opt} | $0 \leq X_m \leq X_{max}$ | Numerical results | – | ↓ | ↓ | ↑ | ↑ | NA |

↑ : Ascending function, ↓ : Descending function, ↗↘ : Function have a maximum,

* : Depends on T , for non large values of T is ↓, for large values of T is ↑

NA: Not applicable.

changing the cost function to $e(X) = X+1$, the optimum distribution is radically different, and probability of transmitting non-zero symbols increases (even though $X = 0$ remains the symbol with the highest probability). Figure 5c plots $I(X_m; Y_m)$ when the input distribution is obtained from (23) for different values of X_{max} with $e(X) = X + 1$. Observe that, by increasing X_{max} , $I(X_m; Y_m)$ increases. Figure 5d plots the relative capacity, C_R , for different values of X_{max} . We observe that the relative capacity increases as X_{max} increases, but that quickly saturates to a value of 1 bit/cost.

Table 1 summarizes the effect of parameters in the molecular medium, such as v , l and σ^2 , and the transmitter parameters, such as X_{max} and $E(X)$. Note that the table refers to error rates, an issue considered in Section IV-F.

IV. CAPACITY ANALYSIS OF MOLECULAR ISI CHANNEL

While the previous section analyzed several capacity measures for ASK communications over molecular DMC, in this section, we initiate an analysis in presence of ISI. As stated earlier, we focus on one time-slot of allowed memory. As with many problems dealing with ISI, we are unable to obtain exact results and resort to bounds. Specifically, we develop two lower bounds and an upper bound for the capacity of the molecular communication system under consideration. We emphasize that the inputs are i.i.d., i.e., the term "capacity" here refers to the simplified case where the inputs within each time-slot are chosen independently and from the same probability distribution. Finally, in this section we develop the maximum a posteriori (MAP) detector (as a possible implementation of a molecular communication

receiver) to evaluate the assumption of restricting memory to only one time-slot. We do this by comparing the numerical and simulation results of the performance measure, the probability of error. It is worth noting that one could consider ISI over multiple time-slots using (5). However, this adds exponential complexity to the analysis.

A. Lower bound 1

The first lower bound considers the effect of ISI on the mutual information between input and output symbols within the same time-slot. This is a lower bound on the channel capacity because this measure ignores the memory; essentially, we consider a DMC but with an additional source of measurement error due to molecules from the previous time-slot. This lower bound relates to a lower bound of $I_{i.i.d}$ in a discrete-input Gaussian channel with ISI [19]. Hence we have

$$\begin{aligned} I_{LB_1} &= I(X_m; Y_m) = H(Y_m) - H(Y_m | X_m) \\ &= - \sum_{y_m=0}^{2X_{\max}} p(y_m) \log(p(y_m)) + \sum_{x_m=0}^{X_{\max}} a_{x_m} \sum_{y_m=0}^{x_m+X_{\max}} p(y_m | x_m) \log p(y_m | x_m) \end{aligned} \quad (24)$$

where $p(y_m | x_m)$ is given by

$$p(y_m | x_m) = \begin{cases} \sum_{i=0}^{x_m} \binom{x_m}{i} (1-q_1)^{x_m-i} q_1^i \sum_{j=y_m-i}^{X_{\max}} \binom{j}{y_m-i} a_j \times \\ \quad q_2^{y_m-i} (1-q_2)^{j-(y_m-i)}, & y_m \leq x_m + X_{\max} \\ 0, & y_m > x_m + X_{\max}. \end{cases} \quad (25)$$

By averaging over x_m on $p(y_m | x_m)$, $p(y_m)$ is given by

$$p(y_m) = \sum_{x_m=0}^{X_{\max}} a_{x_m} \sum_{i=0}^{x_m} \binom{x_m}{i} (1-q)^{x_m-i} q^i \sum_{j=y-i}^{X_{\max}} \binom{j}{y_m-i} a_j p_2^{y_m-i} (1-p_2)^{j-(y_m-i)}. \quad (26)$$

Given an input probability vector \mathbf{a} , this lower bound can be easily evaluated.

B. Lower Bound 2

With the one time-slot memory model, the transmitted symbol in the current time-slot only affects the received molecules in the current and the next time-slot. We consider the mutual information between

transmitted symbol in current time-slot and received symbols in the current and next time-slot

$$\begin{aligned}
I_{LB_2} &= I(X_{m-1}; Y_{m-1}, Y_m) \stackrel{(a)}{=} H(Y_m, Y_{m-1}) - H(Y_m, Y_{m-1} | X_{m-1}) \\
&\stackrel{(b)}{=} H(Y_{m-1}) + H(Y_m | Y_{m-1}) - H(Y_{m-1} | X_{m-1}) - H(Y_m | X_{m-1}, Y_{m-1}) \quad (27)
\end{aligned}$$

where (a) and (b) are obtained based on the definitions of mutual information and joint entropy, respectively. We consider the channel in the steady state regime, hence $P(y_{m-1} | x_{m-1}) = p(y_m | x_m)$, which is given in (25). Also, $p(y_m | y_{m-1}, x_{m-1})$ is given by

$$\begin{aligned}
p(y_m | y_{m-1}, x_{m-1}) &\stackrel{(a)}{=} \sum_{x_m=0}^{X_{\max}} p(x_m | y_{m-1}, x_{m-1}) p(y_m | y_{m-1}, x_m, x_{m-1}) \\
&\stackrel{(b)}{=} \sum_{x_m=0}^{X_{\max}} p(x_m) p(y_m | y_{m-1}, x_m, x_{m-1}) \quad (28)
\end{aligned}$$

where (a) is obtained based on the law of total probability, (b) is obtained based on the independence of x_m from y_{m-1} and x_{m-1} which is due to the i.i.d assumption of the input distribution and causality. By averaging over x_{m-1} in (28), $p(y_m | y_{m-1})$ is given by

$$\begin{aligned}
p(y_m | y_{m-1}) &= \sum_{x_{m-1}=0}^{X_{\max}} p(x_{m-1} | y_{m-1}) p(y_m | y_{m-1}, x_{m-1}) \\
&\stackrel{(a)}{=} \sum_{x_{m-1}=0}^{X_{\max}} p(x_{m-1} | y_{m-1}) \sum_{x_m=0}^{X_{\max}} p(x_m) p(y_m | y_{m-1}, x_m, x_{m-1}) \quad (29) \\
&\stackrel{(b)}{=} \sum_{x_{m-1}=0}^{X_{\max}} \frac{p(y_{m-1} | x_{m-1}) p(x_{m-1})}{p(y_{m-1})} \sum_{x_m=0}^{X_{\max}} p(x_m) p(y_m | y_{m-1}, x_m, x_{m-1}).
\end{aligned}$$

where (a) and (b) are obtained based on the law of total probability, and Bayes rule, respectively. Also, $p(y_m | y_{m-1}, x_m, x_{m-1})$ is given by

$$\begin{aligned}
p(y_m | y_{m-1}, x_m, x_{m-1}) &\stackrel{(a)}{=} \sum_{y'_{m-2}=0}^{X_{\max}} p(y'_{m-2} | y_{m-1}, x_m, x_{m-1}) p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) \\
&\stackrel{(b)}{=} \sum_{y'_{m-2}=0}^{X_{\max}} p(y'_{m-2} | y_{m-1}, x_{m-1}) p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) \\
&\stackrel{(c)}{=} \sum_{y'_{m-2}=0}^{X_{\max}} \frac{p(y_{m-1}, x_{m-1} | y'_{m-2}) p(y'_{m-2})}{p(y_{m-1}, x_{m-1})} p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) \\
&\stackrel{(d)}{=} \sum_{y'_{m-2}=0}^{X_{\max}} \frac{p(y_{m-1} | y'_{m-2}, x_{m-1}) p(x_{m-1} | y'_{m-2}) p(y'_{m-2})}{p(y_{m-1} | x_{m-1}) p(x_{m-1})} p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) \\
&\stackrel{(e)}{=} \sum_{y'_{m-2}=0}^{X_{\max}} \frac{p(y_{m-1} | y'_{m-2}, x_{m-1}) p(x_{m-1}) p(y'_{m-2})}{p(y_{m-1} | x_{m-1}) p(x_{m-1})} p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) \\
&= \sum_{y'_{m-2}=0}^{X_{\max}} \frac{p(y_{m-1} | y'_{m-2}, x_{m-1}) p(y'_{m-2})}{p(y_{m-1} | x_{m-1})} p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) \\
&\stackrel{(f)}{=} \sum_{y'_{m-2}=0}^{y_{m-1}} \frac{p(x'_{m-1} | x_{m-1}) p(y'_{m-2})}{p(y_{m-1} | x_{m-1})} p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2})
\end{aligned} \tag{30}$$

where (a) is obtained based on the law of total probability and y'_{m-2} is the number of received molecules at the end of time-slot $m-1$ from transmitted molecules in time-slot $m-2$; (b) is obtained due to the independence of y'_{m-2} from x_m ; (c) is obtained based on Bayes' rule; (d) is obtained based on the joint probability formula; (e) is obtained based on independence of y'_{m-2} and x_{m-1} and the joint probability formula; (f) is obtained due to the fact that $y_{m-1} = x'_{m-1} + y'_{m-2}$, where x'_{m-1} denotes the number of absorbed molecules at time-slot $m-1$ at end of time-slot $m-1$. Based on definition of y'_{m-2} we have

$$p(y'_{m-2}) = \sum_{x_{m-2}=0}^{X_{\max}} a_{x_{m-2}} \binom{x_{m-2}}{y'_{m-2}} q_2^{y'_{m-2}} (1 - q_2)^{x_{m-2} - y'_{m-2}} \tag{31}$$

$$\begin{aligned}
p(y_m | y_{m-1}, x_m, x_{m-1}, y'_{m-2}) &\stackrel{(a)}{=} p(y_m | x_m, x''_{m-1} = x_m - (y_{m-1} - y'_{m-2})) = \\
&\begin{cases} \sum_{i=y_m - x_{m-1}}^{x_m} \binom{x_m}{i} (1 - q_1)^{x_m - i} q_1^i \binom{x''_{m-1}}{y_m - i} q_2^{y_m - i} (1 - q_2)^{x''_{m-1} - (y_m - i)}, & y_m < x_m + x''_{m-1} \\ 0 & , y_m > x_m + x''_{m-1} \end{cases}
\end{aligned} \tag{32}$$

(a) is obtained because $x_{m-1} = x'_{m-1} + x''_{m-1}$, where x''_{m-1} is the number of remaining molecules at end of time-slot $m-1$ from transmitted molecules in the same time-slot. Moreover, based on definition

of x'_{m-1} , $P(x'_{m-1}|x_{m-1})$ is given by

$$P(x'_{m-1}|x_{m-1}) = \binom{x_{m-1}}{x'_{m-1}} q_1^{x'_{m-1}} (1 - q_1)^{x_{m-1} - x'_{m-1}}. \quad (33)$$

While far more involved, as with the first lower bound in (24), given the input probability distribution, obtaining this bound is a simple matter.

C. Upper Bound

Having developed two lower bounds on capacity, we now develop an upper bound motivated by the matched filter upper bound in Gaussian ISI channels [19]. With one time-slot memory for the channel, if we transmit symbols and wait two time-slots before transmitting the next symbol and remove any subsequently arriving molecules, we have an interference-free channel and arrive at an upper bound to the capacity per channel use. This is equivalent to the DMC case with the binomial transition probabilities of (2) where q_1 is replaced by $q_U = F_W(2T)$. The mutual information of this channel is concave which can be maximized using the Blahut-Arimoto algorithm [26]. We therefore have the upper bound

$$I_{UB} = I(X_m; Y_m) = H(Y_m) - H(Y_m | X_m) \quad (34)$$

where $P(y_m|x_m)$ is given by

$$P(y_m|x_m) = \begin{cases} \binom{x_m}{y_m} q_U^{y_m} (1 - q_U)^{x_m - y_m}, & y_m \leq x_m \\ 0 & , y_m > x_m. \end{cases} \quad (35)$$

and $q_U = F_W(2T)$. Clearly,

$$\begin{aligned} I_{LB_2} &= I(X_{m-1}; Y_{m-1}, Y_m) \stackrel{(a)}{=} I(X_{m-1}; Y_{m-1}) + I(X_{m-1}; Y_m | Y_{m-1}) \\ &\stackrel{(b)}{\geq} I(X_{m-1}; Y_{m-1}) = I_{LB_1} \end{aligned} \quad (36)$$

where (a) is obtained from chain rule in mutual information and (b) is obtained based on the non-negativity of mutual information. Hence, we have the following result

$$I_{LB_1} \leq I_{LB_2} \leq I_{i.i.d} \leq I_{UB},$$

as long as the input distribution for the upper bound is optimum. We discuss optimizing the bounds next.

D. Optimizing Lower and Upper Bounds

We are unable to show the concavity of the lower bounds I_{LB_1} and I_{LB_2} with respect to the input distribution vector \mathbf{a} ; however, we can modify the Blahut-Arimoto algorithm [26] to find a local maximum, and as a result, we can optimize, to within a local maximum, these two lower bounds. The channel transition matrices $\mathbf{P}^{(LB_h)}$, $h \in \{1, 2\}$ for lower bounds 1 and 2 are

$$\mathbf{P}_{x_m, y_m}^{(LB_1)} = [p(Y_m = y_m | X_m = x_m)] \quad (37)$$

$$\mathbf{P}_{x_{m-1}, y_{m-1} y_m}^{(LB_2)} = [p(Y_m = y_m, Y_{m-1} = y_{m-1} | X_{m-1} = x_{m-1})] \quad (38)$$

where, importantly, the size of $\mathbf{P}^{(LB_1)}$ and $\mathbf{P}^{(LB_2)}$ are $(X_{\max} + 1) \times (2X_{\max} + 1)$ and $(X_{\max} + 1) \times (2X_{\max} + 1)^2$ respectively. Also, $y_m, y_{m-1} \in [0, \dots, 2X_{\max}]$ and $x_m \in [0, \dots, X_{\max}]$. For matrix \mathbf{Q} with size of $\mathbf{P}^{(LB_h)}$, $h \in \{1, 2\}$, let

$$J(\mathbf{a}, \mathbf{P}^{(LB_h)}, \mathbf{Q}) = \sum_j \sum_i a_j \mathbf{P}_{j,i}^{(LB_h)} \log \frac{\mathbf{Q}_{i,j}}{a_j}. \quad (39)$$

Then the following is true.

- 1) $I_{LB_h} = \max_{\mathbf{a}} \max_{\mathbf{Q}} J(\mathbf{a}, \mathbf{P}^{(LB_h)}, \mathbf{Q})$.
- 2) For fixed \mathbf{a} , $J(\mathbf{a}, \mathbf{P}^{(LB_h)}, \mathbf{Q})$ is locally maximized by

$$\mathbf{Q}_{k,j} = \frac{a_j \mathbf{P}_{j,k}^{LB^{(h)}}}{\sum_j a_j \mathbf{P}_{j,k}^{LB^{(h)}}}. \quad (40)$$

- 3) For fixed \mathbf{Q} , $J(\mathbf{a}, \mathbf{P}^{(LB_h)}, \mathbf{Q})$ is locally maximized by

$$a_j = \frac{\exp\left(\sum_i \mathbf{P}_{j,i}^{(LB_h)} \log \mathbf{Q}_{i,j}\right)}{\sum_j \exp\left(\sum_i \mathbf{P}_{j,i}^{(LB_h)} \log \mathbf{Q}_{i,j}\right)} \quad (41)$$

The algorithm iterates between a_j derived in (41) and the transition probability matrices $\mathbf{P}^{(LB_h)}$ in (37) or (38). This procedure is repeated until the convergence of the probabilities a_j .

In contrast to the lower bounds, I_{UB} is a concave function in terms of \mathbf{a} . Hence, using the standard Blahut-Arimoto algorithm [26] a unique distribution globally maximizing the upper bound can be obtained (note that this is critical for I_{UB} to be a valid upper bound).

E. MAP and ML Detectors

In this subsection, we derive the detection performance of a molecular ASK receiver. Since the true channel capacity is unknown, we are unable to directly evaluate the quality of the assumption of one time-slot memory in molecular ISI channel on the capacity analysis. As a proxy, we examine the effect of this assumption on the detection performance using simulations and analysis. In general, one may resort to a sequence detector over an ISI channel for improved performance [29]–[31], however, in molecular communications the implementation complexity for nano-machines is likely to be prohibitive. Hence, here we consider a MAP symbol-by-symbol detector at the receiver to estimate the number of molecules transmitted. Hence, here we consider a MAP symbol-by-symbol detector at the receiver to estimate the number of molecules transmitted. The decision rule is given by

$$\hat{x} = \arg \max_x p_{Y_k|X_k}(x|y) = \arg \max_x p_{Y_k|X_k}(y|x) p(x) \quad (42)$$

By replacing (4) in (42), we have

$$\hat{x} = \arg \max_x \sum_{i=0}^{y-x} a_x \binom{x}{i} q_1^i (1-q_1)^{x-i} P_k(y-i) \quad (43)$$

If x molecules are transmitted, the decision region for detecting x at the receiver is denoted by $\{Y_x\}$.

Hence, probability of error is denoted by

$$P_e = \sum_{x=0}^{X_{\max}} P(y \notin \{Y_x\} | x) a_x \quad (44)$$

As always, using a uniform input distribution in a_x , makes the MAP detector a ML detector.

F. Numerical Results

We now evaluate and compare the derived bounds for various medium parameters such as l , v and σ . Figures 6a and 6b plot the two lower bounds. We quantify the lower bounds for three input distributions: (i) optimized input distribution from maximizing $I(X_m; Y_m)$ ($I(X_m; Y_m, Y_{m+1})$), denoted by "Optimized $I(X_m, Y_m)$ " ("Optimized $I(X_m; Y_m, Y_{m+1})$ ") in the legend for I_{LB_1} (I_{LB_2}), (ii) optimized input distribution from maximizing the mutual information of DMC in (9), denoted by "Optimized DMC" in the legend, and (iii) uniform input distribution. This study shows the impact of selecting input distribution on the numerical value of capacity bounds. One sees that using the optimized input

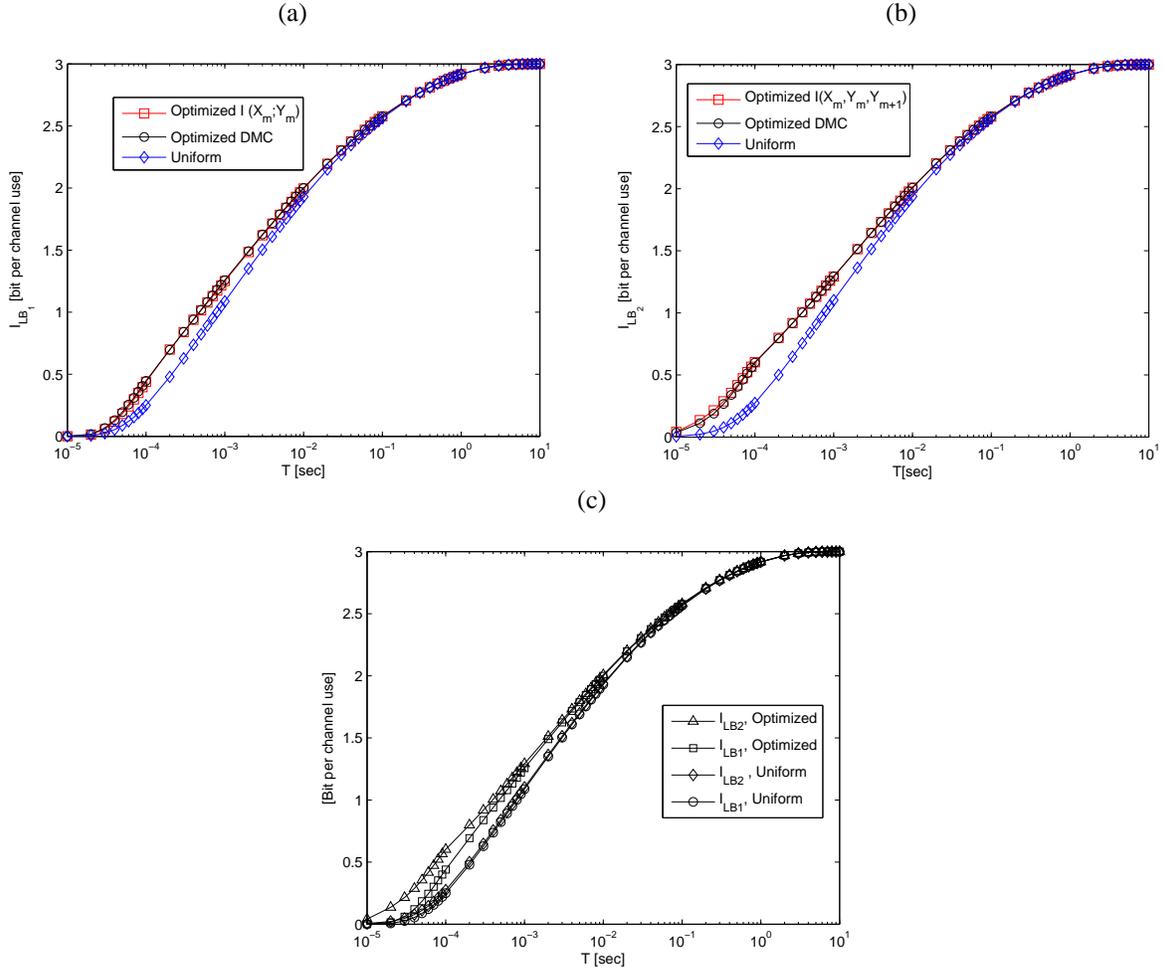


Fig. 6: (a-b) Two lower bounds versus T (c) Comparing the lower bounds for uniform and optimized input distributions. $v = 1$, $l = 10^{-2}$, $\sigma^2 = 1$, $X_{max} = 7$.

distributions, the bounds I_{LB_1} and I_{LB_2} improve noticeably in comparison to the cases with uniform input distribution. This is while a uniform distribution maximizes the source entropy and the capacity of a molecular channel without ISI (here an error free channel). Another important observation here is that each of the two proposed bounds for the molecular ISI channel amounts to almost an equal rate with either its corresponding optimized distribution or that from the molecular DMC case in Section III. To directly compare the two bounds, Figure 6c compares I_{LB_1} and I_{LB_2} as a function of T for the optimized input distribution from (41) (labeled by "Optimized"), and a uniform input distribution.

Figures 7a-7c plot I_{LB_2} , I_{LB_1} , C_{DMC} and I_{UB} versus T for different values of the transmitter-receiver distance, l , the diffusion constant σ and drift velocity v , respectively. As evident, by decreasing l all bounds increase and converge to $\log_2(X_{max} + 1) = 3$, which is the entropy of the source. Similarly, all capacity bounds are increasing functions of σ . As also observed in [3] (for the case when information is

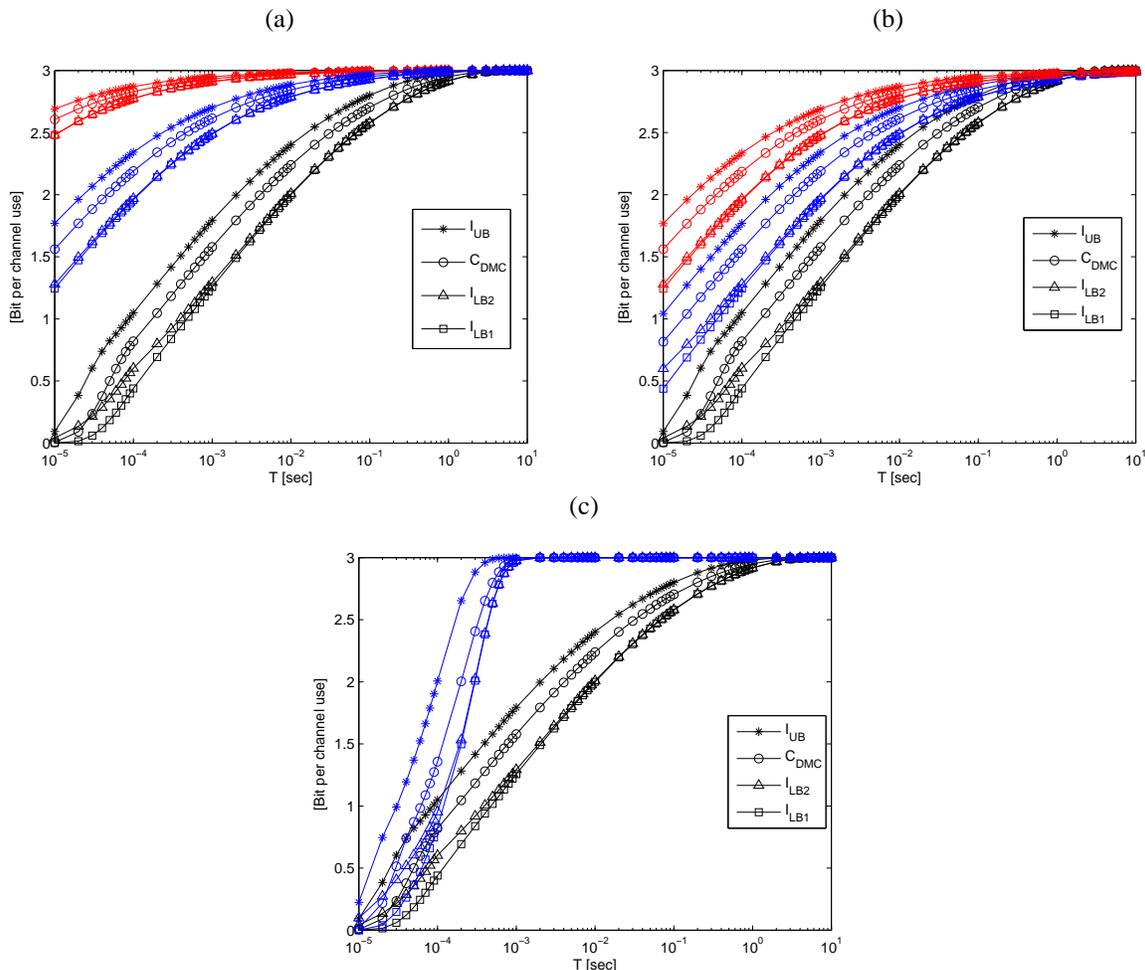


Fig. 7: I_{LB_2} , I_{LB_1} , C_{DMC} and I_{UB} in terms of T for different values of (a) l (Black curves, $l = 10^{-2}$, blue curves, $l = 10^{-3}$, red curves, $l = 10^{-4}$) with $v = 1$, $\sigma^2 = 1$, (b) σ^2 (Black curves, $\sigma^2 = 1$, blue curves, $\sigma^2 = 10$, red curves, $\sigma^2 = 100$) with $v = 1$, $l = 10^{-2}$, (c) v (Black curves, $v = 1$, blue curves, $v = 100$) with $l = 10^{-2}$, $\sigma^2 = 1$ all with $X_{max} = 7$.

encoded in the time of release) increasing drift velocity increases mutual information due to reduced ISI. Comparing the results of the three recent sub-figures, we note that the bounds are most sensitive to the transmitter-receiver distance l and drift velocity v while not being as sensitive to the diffusion constant σ . Crucially, over wide ranges of T , the upper and lower bounds are close. Also, due to the reduced ISI, reducing l reduces the gap between all the derived bounds.

The capacity of an ASK-based molecular communication channel is in general unknown. To test the validity of the one time-slot memory assumption, we compare the probability of error of the ML detector in three cases: no ISI (the DMC of Section III), including ISI up to one time-slot (the case analyzed in Section IV) and simulation results that track the arrivals of all molecules. We reason that, if our

assumption is invalid, there should be a noticeable difference between the error rates for the three cases. In all cases, we use a uniform input distribution (making the MAP and ML detectors equivalent).

Figure 8 plots the probability of error, P_e , of the ML detector versus T . In this figure, (i) the solid curves denote the error probability for the molecular DMC case, (ii) the dashed curves denote the error assuming single time-slot memory (STM) ISI model, and (iii) the markers denote simulation results based on tracking all molecules (labeled as the multiple time-slot ISI model (MTM)). By comparing whether the markers match the dashed or solid curves in these figures, one can examine the range of parameters in which each of the DMC, STM or MTM models is valid. In Fig. 8a, P_e is depicted versus T for different values of v with $l = 10^{-2}$ and $\sigma^2 = 1$. One sees that with $v = 10$, for $T > 2 \times 10^{-2}$, the numerical results of DMC match the simulation results and for $3 \times 10^{-3} < T < 2 \times 10^{-2}$, numerical results of STM match the simulation results; hence for $v = 10$, the DMC and STM models are valid in the said ranges, respectively. Also, Figure 8b shows P_e in terms of T for different values of σ^2 with $v = 1$ and $l = 10^{-2}$. We observe that for small to medium values of T , increasing σ^2 reduces P_e , while for large values of T , increasing σ^2 , increases P_e instead. This is consistent with the behavior of q_1 as a function of σ^2 . For small and medium values of T , in which ISI is important, q_1 is an increasing function in terms of σ^2 , i.e., the increased variance in position increases the probability of the molecules arriving within two time-slots after release. Moreover, one sees in this figure that for $\sigma^2 = 0.1$, the DMC and STM models are valid for $T > 3 \times 10^{-1}$ and $10^{-3} < T < 3 \times 10^{-1}$, respectively.

It is evident in Figure 8, that the numerical results of the DMC case (the solid curves) match well to the simulation results for higher values of T , while those of the one time-slot memory model (the dashed curves) match the simulation results for lower values of T . The results show that accounting for one time-slot of memory is adequate for design and capacity analysis purposes in an interesting range of molecular channel parameters and error probabilities. Specifically, the corresponding results are very close to the more accurate case of memory spread over multiple time-slots. For more accurate analysis in these cases, one may opt to extend the presented analysis to the case where molecules arriving with more than one time-slot delay are also taken into account in modeling ISI. Naturally, this is only achieved at the cost of substantial (exponentially growing) computational complexity in the analysis.

For small values of T the duration of ISI is more than one time-slot, and the probability of error performance of the ML detector deteriorates. This is naturally assuming that the delayed molecules do

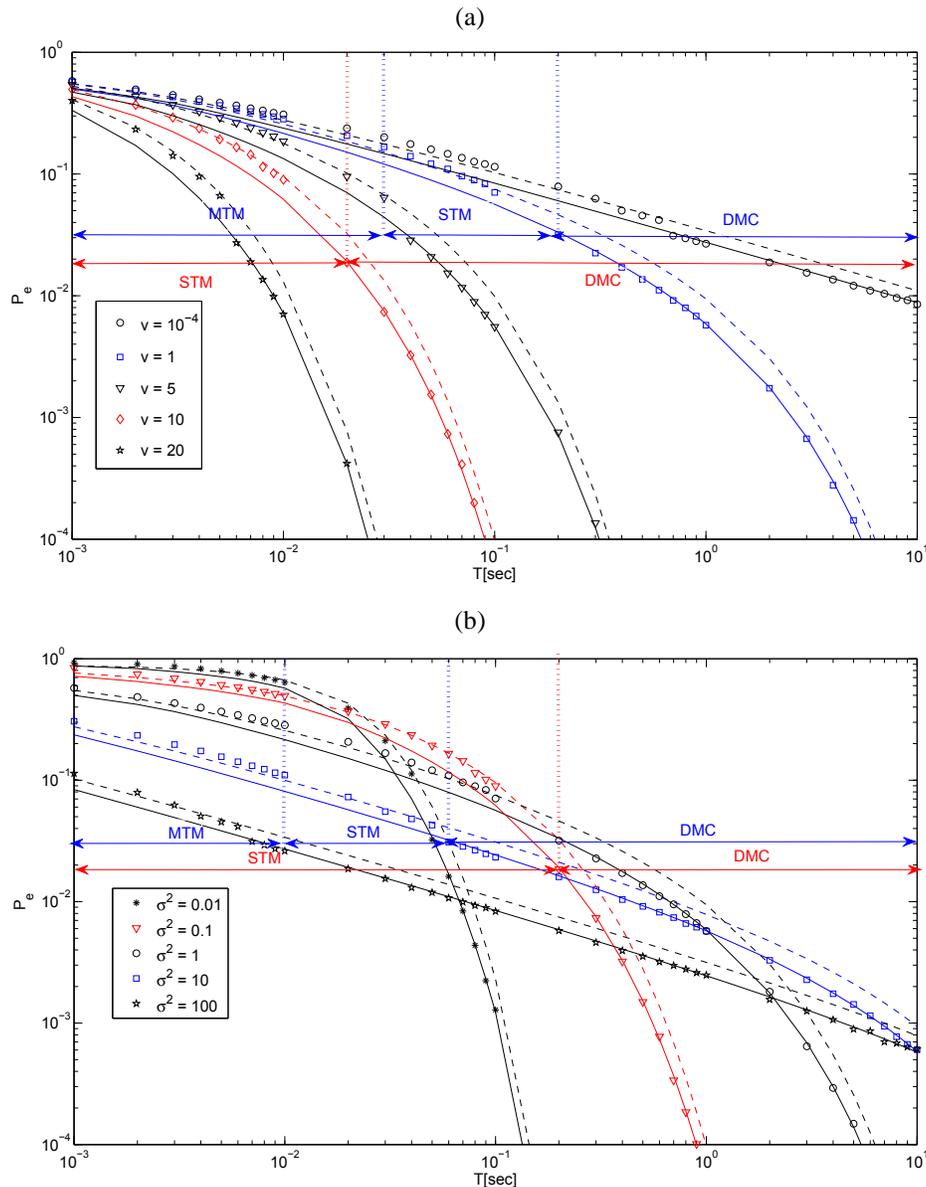


Fig. 8: P_e in terms of T using ML detector for different values of (a) v with $l = 10^{-2}$, $\sigma^2 = 1$ (b) σ^2 with $l = 10^{-2}$, $v = 1$, all with $X_{\max} = 7$. (Solid lines: numerical results of calculated P_e over DMC, Dashed lines: numerical results of calculated P_e with single time-slot memory (STM), markers: Simulation results (multiple time-slot memory (MTM)).)

not expire over time. As stated, one may resort to more advanced and complex sequence detectors over such harsh ISI channels to increase the rate and yet maintain a desired error performance. The analysis and design of such schemes may be investigated in future studies.

It is noteworthy that depending on the characteristics of the fluid media, the quantity of molecules and their life expectancy, the slot duration should be chosen such that a statistical majority of molecules emitted in a slot arrive within a time-slot. Of course, the presented model, detectors and analyses will

serve most effectively in such a practical setting.

V. CONCLUSIONS

In this paper, we analyze the capacity of a molecular communications channel when information is encoded in the number of transmitted molecules (molecular ASK). The molecules propagation is governed by Brownian motion and the probability of arrival within a specific time-slot is derived using the additive inverse Gaussian model for the transmission time. We analyzed the capacity in both DMC and ISI cases.

For the DMC, the optimized input distribution was derived to achieve the capacity per channel use, per unit time and per unit cost. Our results show that for small (large) values of transmission time-slots, the distribution achieving the capacity per channel use converges to bipolar (uniform) distribution. Also, imposing a limitation on average number of transmitted molecules per channel use reduces the capacity per channel use and makes symbols with lower transmission cost more probable. As a result in this case, the optimum input distribution, maximizing the capacity per channel use, deviates from a uniform distribution even for large values of time-slot duration. By considering the capacity per unit time as the objective function, the optimum time-slot duration, which maximizes the capacity with peak and/or average constraints on the number of transmitted molecules per unit time, was obtained numerically. We studied the dependence of the optimum time-slot duration with parameters of the molecular medium. This paper also offered the first study on the capacity per unit cost in molecular communications, which is a capacity measure conscious of its relative efficiency with respect to the molecular injection rate. For the case where molecular alphabet only consists of symbols with non-zero cost, a non-zero capacity per unit cost is feasible and we obtained the corresponding optimized input distribution.

The second half of this paper analyzes the case with ISI; specifically, with ISI restricted to one time-slot. Two lower bounds and an upper bound for ASK-based molecular communication with ISI were derived. Our results quantified how the lower bounds improve when the corresponding input symbol distributions are optimized. Importantly, our results showed that over a wide range of parameter values the gap between the lower and upper bounds are small and so can provide a good measure of capacity. The results also showed that the optimum distribution obtained assuming a DMC provides close-to-optimal results for the ISI case as well.

To test the validity of the one time-slot memory model for the molecular ISI channel, we compared the performance of the ML detector in this case, with that assuming a DMC model and the results from the simulations. As our results show, the one time-slot memory model is valid over an interesting range

of molecular channel parameters. One may extend the presented analysis to the case with multiple time-slot memory models at the cost of increased computational complexity. Also, the design and analysis of efficient detectors in these settings is another possible future research direction.

REFERENCES

- [1] F. Sako and M. Ueda, *Cell Signaling Reactions, Single-Molecular Kinetic Analysis*. Springer, 2011.
- [2] B. Atakan and O. Akan, "An information theoretical approach for molecular communication," in *Bio-Inspired Models of Network, Information and Computing Systems*, Dec 2007, pp. 33–40.
- [3] K. V. Srinivas, A. Eckford, and R. Adve, "Molecular communication in fluid media: The additive inverse gaussian noise channel," *IEEE Trans. on Inf. Theory*, vol. 58, no. 7, pp. 4678–4692, July 2012.
- [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 Commun.*, vol. 31, no. 12, pp. 847–856, December 2013.
- [5] T. Nakano, Y. Okaie, and J.-Q. Liu, "Channel model and capacity analysis of molecular communication with Brownian motion," *IEEE Commun. Letters*, vol. 16, no. 6, pp. 797–800, June 2012.
- [6] T. Nakano, A. W. Eckford, and H. T., *Molecular communication*. Cambridge, UK: Cambridge University Press, Oct. 2013.
- [7] A. Einolghozati, M. Sardari, A. Beirami, and F. Fekri, "Capacity of discrete molecular diffusion channels," in *IEEE Int. Symp. on Inf. Theory*, July 2011, pp. 723–727.
- [8] Q. Liu, K. Yang, and P. He, "Channel capacity analysis for molecular communication with continuous molecule emission," in *Int. Conf. on Wireless Commun. Signal Proc.*, Oct 2013, pp. 1–6.
- [9] B. Atakan, S. Galmes, and O. Akan, "Nanoscale communication with molecular arrays in nanonetworks," *IEEE Trans. on NanoBioscience*, vol. 11, no. 2, pp. 149–160, June 2012.
- [10] M. Pierobon and I. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise," *IEEE Trans. on Inf. Theory*, vol. 59, no. 2, pp. 942–954, Feb 2013.
- [11] H. Arjmandi, A. Gohari, M. Kenari, and F. Bateni, "Diffusion-based nanonetworking: A new modulation technique and performance analysis," *IEEE Commun. Letters*, vol. 17, no. 4, pp. 645–648, April 2013.
- [12] N.-R. Kim, A. Eckford, and C.-B. Chae, "Symbol interval optimization for molecular communication with drift," *IEEE Trans. on NanoBioscience*, vol. 13, no. 3, pp. 223–229, Sept 2014.
- [13] L.-S. Meng, P.-C. Yeh, K.-C. Chen, and I. Akyildiz, "A diffusion-based binary digital communication system," in *IEEE Int. Conf. on Comm.*, June 2012, pp. 4985–4989.
- [14] M. Kabir and K. Kwak, "Molecular nanonetwork channel model," *Electronics Letters*, vol. 49, no. 20, pp. 1285–1287, September 2013.
- [15] M. Kuran, H. Yilmaz, T. Tugcu, and I. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks," in *IEEE Int. Conf. on Commun.*, June 2011, pp. 1–5.
- [16] H. Li, S. M. Moser, and D. Guo, "Capacity of the memoryless additive inverse gaussian noise channel," *IEEE Journal on Selected Areas in Commun.*, vol. 32, no. 12, pp. 2315–2329, Dec. 2012.
- [17] A. Eckford, K. Srinivas, and R. Adve, "The peak constrained additive inverse gaussian noise channel," in *IEEE Int. Symp. on Inf. Theory*, July 2012, pp. 2973–2977.
- [18] A. Singhal, R. Mallik, and B. Lall, "Molecular communication with brownian motion and a positive drift: performance analysis of amplitude modulation schemes," *IET Commun.*, vol. 8, no. 14, pp. 2413–2422, Sept 2014.
- [19] S. Shamai, L. Ozarow, and A. Wyner, "Information rates for a discrete-time gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. on Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov 1991.
- [20] W. Hirt and J. Massey, "Capacity of the discrete-time gaussian channel with intersymbol interference," *IEEE Trans. on Inf. Theory*, vol. 34, no. 3, pp. 38–38, May 1988.
- [21] G. Taricco and J. Boutros, "An asymptotic approximation of the isi channel capacity," in *Inf. Theory and App. Workshop*, Feb 2014, pp. 1–5.
- [22] D.-M. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. on Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug 2006.
- [23] Y. Chahibi and I. Akyildiz, "Molecular communication noise and capacity analysis for particulate drug delivery systems," *IEEE Trans. on Commun.*, vol. PP, no. 99, pp. 1–13, Dec 2014.

- [24] Y. N. Dou, J. Zheng, W. D. Foltz, R. Weersink, N. Chaudary, D. A. Jaffray, and C. Allen, "Heat-activated thermosensitive liposomal cisplatin (htlc) results in effective growth delay of cervical carcinoma in mice," *Journal of Controlled Release*, vol. 178, pp. 69 – 78, 2014.
- [25] R. M. Ziff, S. N. Majumdar, and A. Comtet, "Capture of particles undergoing discrete random walks," *The Journal of Chemical Physics*, vol. 130, no. 20, 2009.
- [26] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. on Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul 1972.
- [27] F. Reza, *An introduction to information theory*. McGraw-Hill, 1961.
- [28] M. Jimbo and K. Kunisawa, "Iteration method for calculating the relative capacity," *IEEE Trans. on Inf. Theory*, vol. 43, no. 2, pp. 216–223, Nov 1979.
- [29] D. Kilinc and O. Akan, "Receiver design for molecular communication," *IEEE Journal on Selected Areas in Commun.*, vol. 31, no. 12, pp. 705–714, December 2013.
- [30] A. Noel, K. Cheung, and R. Schober, "Optimal receiver design for diffusive molecular communication with flow and additive noise," *IEEE Trans. on NanoBioscience*, vol. 13, no. 3, pp. 350–362, Sept 2014.
- [31] H. ShahMohammadian, G. G. Messier, and S. Magierowski, "Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels," *Nano Communication Networks*, vol. 3, no. 3, pp. 183 – 195, 2012.