Molecular Communication in Fluid Media: The Additive Inverse Gaussian Noise Channel

K. V. Srinivas, Andrew W. Eckford, *Member, IEEE*, and Raviraj S. Adve, *Senior Member, IEEE*

***Abstract*—In this paper, we consider molecular communication, with information conveyed in the time of release of molecules. These molecules propagate to the transmitter through a fluid medium, propelled by a positive drift velocity and Brownian motion. The main contribution of this paper is the development of a theoretical foundation for such a communication system; specifically, the additive inverse Gaussian noise (AIGN) channel model. In such a channel, the information is corrupted by noise that follows an IG distribution. We show that such a channel model is appropriate for molecular communication in fluid media. Taking advantage of the available literature on the IG distribution, upper and lower bounds on channel capacity are developed, and a maximum likelihood receiver is derived. Results are presented which suggest that this channel does not have a single quality measure analogous to signal-to-noise ratio in the additive white Gaussian noise channel. It is also shown that the use of multiple molecules leads to reduced error rate in a manner akin to diversity order in wireless communications. Finally, some open problems are discussed that arise from the IG channel model.**

***Index Terms*—Molecular communication, mutual information, nanobiotechnology.**

I. Introduction

MODERN communication systems are almost exclusively based on the propagation of electromagnetic (or acoustic) waves. Of growing recent interest, nanoscale networks, or nanonetworks, are systems of communicating devices, where both the devices themselves and the gaps between them are measured in nanometers [1]. Due to the limitations on the available size, energy, and processing power, it is difficult for these devices to communicate through conventional means. Thus, communication between nanoscale devices will substantially differ from the well-known wired/wireless communication scenarios.

In this paper, we address communication in a nanonetwork operating in an aqueous environment; more precisely, we consider communication between two nanomachines connected through a fluid medium, where molecules are used as the information carriers. In such a communication system, the transmitter encodes the information onto the molecules and releases them into the fluid medium; the molecules propagate through the fluid medium and the receiver, upon receiving the molecules, decodes the information by processing or reacting with the molecules. This approach, known as *molecular communication* [2], is inspired by biological micro-organisms which exchange information through molecules. Information can be encoded onto the molecules in different ways, such as using timing, concentration, or the identities of the molecules themselves.

Molecular communication has recently become a rapidly growing discipline within communications and information theory. The existing literature can be divided into two broad categories: in the first category, components and designs to implement molecular communication systems are described; for example, communications based on calcium ion exchange [3] and liposomes [4] have been proposed. These are commonly used by living cells to communicate. Other work (e.g., [5] and [6]) has explored the use of molecular motors to actively transport information-bearing molecules. A considerable amount of work has been done in related directions, much of which is beyond the scope of this paper; a good review is found in [7].

In the second category, channel models are proposed and information-theoretic limits have been studied, largely via simulations. Our own prior work falls in this category: in [8], idealized models and mutual information bounds were presented for a Wiener process model of Brownian motion without drift; while in [9] and [10], a net positive drift was added to the Brownian motion and mutual information between transmitter and receiver calculated using simulations. Mutual information has been calculated for simplified transmission models (e.g., on-off keying) in [11] and [12], while communication channel models for molecular *concentration* have been presented in [13], and mutual information is calculated in [14]. Less closely related to the current paper, information-theoretic analysis has been done to evaluate multiuser molecular communication channels [15], and to evaluate the capacity of calcium-signaling relay channels [16]. Related work also includes information-theoretic literature on the trapdoor channel [17], [18], and the queue-timing channel [19], [20].

As in [10], in this paper we consider molecular communication with *timing modulation*; i.e., information is encoded in the release times of molecules into the fluid medium. We also assume that molecules propagate via Brownian motion with a net drift from transmitter to receiver, generally called *positive drift*. Brownian motion is physically realistic for nanodevices, since these devices have dimensions broadly on the same scale as individual molecules. Further, we choose positive drift since it arises in our applications of interest (e.g., communications

Manuscript received November 30, 2010; revised September 22, 2011; accepted February 07, 2012. Date of publication April 27, 2012; date of current version June 12, 2012.

K. V. Srinivas is with Samsung India Electronics Pvt. Ltd., Noida 201301, UP, India (e-mail: kvsrinivas@gmail.com).

A. W. Eckford is with the Department of Computer Science and Engineering, York University, Toronto, ON M3J 1P3, Canada (e-mail: aeckford@yorku.ca).

R. S. Adve is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: rsadve@comm.utoronto.ca).

Communicated by M. Gastpar, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2012.2193554

that take advantage of the bloodstream). Our focus here is on the channel; we assume that the transmitter and receiver work perfectly. We further assume that the receiver has infinite time to guarantee that all transmitted molecules will arrive and that there are no "stray" particles in the environment. Therefore, in our system, communication is corrupted only by the inherent randomness due to Brownian motion.

The key contributions of this paper are as follows.

1) We show that molecular communication with timing modulation can be abstracted as an additive noise channel with the noise having *inverse Gaussian* (IG) distribution (see Section II); thus, the molecular communication is modeled as communication over an additive inverse Gaussian noise (AIGN) channel. This forms the basis of the theoretical developments that follow.
2) Using the AIGN framework, we obtain upper and lower bounds on the information theoretic capacity of a molecular communication system with timing modulation (see Theorem 1).
3) We investigate receiver design for molecular communication and present a maximum likelihood estimator (MLE) (see Theorem 2), and an upper bound on the symbol error probability (SEP) (see Theorem 3). We also show an effect similar to diversity order in wireless communications when multiple molecules are released simultaneously (see Theorem 4).

While the work in [10] is based largely on simulations, the AIGN framework developed here allows us to place molecular communications on a theoretical footing. However, we emphasize that this paper remains an initial investigation into the theory of molecular communications with timing modulation in fluid media.

This paper is organized as follows. Section II presents the system and channel model under consideration. Section III then uses this channel model to develop capacity bounds for this system. Section IV then develops a maximum likelihood (ML) receiver. Section V presents extensive discussion and few open problems and, finally, Section VI concludes this paper.

A. Notation

$h(X)$ denotes the differential entropy of the random variable X. $X \sim \exp(\gamma)$ implies that X is an exponentially distributed random variable with mean $1/\gamma$; i.e., $f_X(x) = \gamma e^{-\gamma x}, x > 0$. $\mathcal{L}(X)$ denotes the Laplace transform of the the probability density function (pdf) of the random variable X. $\Re(z)$ denotes the real part of z. Throughout this paper, $\log$ refers to the natural logarithm; hence, information is measured in nats. $\mathbb{R}_+$ represents the set of nonnegative real numbers.

II. System and Channel Model

The main components of a molecular communication system are the transmitter, receiver, molecules that convey the information and the environment or the medium in which the molecules propagate from the transmitter to the receiver. Fig. 1 illustrates the system under consideration.

The transmitter is a point source of identical molecules. It conveys a message X, where $X \in \mathcal{X}$ is a random variable with alphabet $\mathcal{X}$ having a finite cardinality $|\mathcal{X}|$, by releasing molecules into the fluid medium. Information can be encoded onto the molecules in different ways: in the release time of the molecules (*timing modulation*) [8], in the concentration of the molecules (number of molecules per unit area) [13], or in the number of molecules (*amplitude modulation*). We consider encoding the message in the release time of the molecule(s). The transmitter does not affect the propagation of the molecules once they are released into the medium.

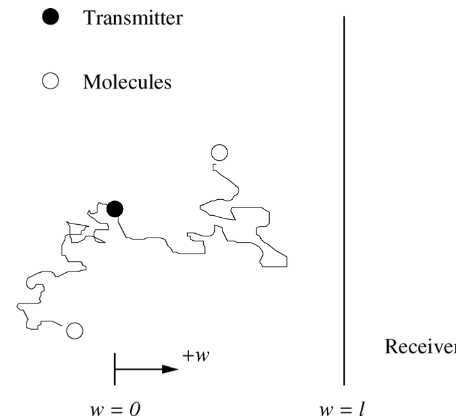

Fig. 1. System model with transmitter at $w = 0$ and receiver at $w = l$.

The medium is an aqueous solution and the molecules (released by the transmitter) propagate by Brownian motion from the transmitter to the receiver. Brownian motion can be characterized by two physical parameters of the fluid medium: the drift velocity and the diffusion coefficient [21].

Throughout this paper, concerning the emission and reception process of molecules, we use the ideal channel model assumptions.

1) The transmitter perfectly controls the release time and number of molecules released to represent a particular message.
2) The receiver perfectly measures the arrival times of the molecules. There is perfect synchronization between the transmitter and the receiver.
3) Once a molecule arrives at the receiver, it is absorbed and does not return to the medium.
4) The propagation environment is unlimited and that, other than the receiving boundary, nothing interferes with the free propagation of the molecule.
5) The trajectories of individual information carrying molecules are independent.

Using these assumptions, we are able to abstract away the physical details of molecular transmission and molecular reception. Further, in [8], it was shown that these assumptions are ideal in an information-theoretic sense, in that eliminating any of them results in a channel with lower capacity. Thus, capacity results obtained using these assumptions are as good as any molecular communication system can do. We consider only 1-D propagation of the molecules.

The *Wiener process* is an appropriate model for physical Brownian motion if friction is negligible [22]. Let $W(x)$ be a continuous-time random process which represents the position at time x of a molecule propagating via Brownian motion. Let $0 \leq x_1 < x_2 < \ldots < x_k$ represent a sequence of time instants, and let $R_i = W(x_i) - W(x_{i-1})$ represent the increments of the random process for $i \in \{1, 2, \ldots, k\}$. Then $W(x)$ is a Wiener process if the increments R_i are independent Gaussian random variables with variance $\sigma^2(x_i - x_{i-1})$. The Wiener process has *drift* if $E[R_i] = v(x_i - x_{i-1})$, where v is the drift velocity.

Consider a fluid medium with positive drift velocity v and diffusion coefficient[1] d, where the Wiener process variance is given by $\sigma^2 = d/2$. A molecule is released into this fluid at time $x = 0$ at position $w = 0$. Under the Wiener process, the probability density of the particle's position w at time $x > 0$ is given by [21]

$$f_W(w; x) = \frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left(-\frac{(w - vx)^2}{2\sigma^2 x}\right). \quad (1)$$

That is, treating the time x as a parameter, the pdf of the position w is Gaussian with mean vx and variance $\sigma^2 x$.

Since the receiver acts as a perfectly absorbing boundary, we are only concerned with the first arrival time N at the boundary. We assume that the transmitter is located at the origin, and in the axis of interest, the receiver is located at position $l > 0$. In this case, the first arrival time is given by

$$N = \min\{x : W(x) = l\}. \quad (2)$$

The key observation here is that if $v > 0$, the pdf of N, denoted by $f_N(n)$, is given by the IG distribution [24]

$$f_N(n) = \begin{cases} \sqrt{\frac{\lambda}{2\pi n^3}} \exp\left(-\frac{\lambda(n-\mu)^2}{2\mu^2 n}\right), & n > 0 \\ 0, & n \leq 0 \end{cases} \quad (3)$$

where

$$\mu = \frac{l}{v}, >> \text{and} \quad (4)$$

$$\lambda = \frac{l^2}{\sigma^2}. \quad (5)$$

The mean and the variance of N are given by $m_N = \mu$ and $\mathrm{Var}(N) = \frac{\mu^3}{\lambda}$, respectively. We will use $\mathrm{IG}(\mu, \lambda)$ as shorthand for this distribution; i.e., $N \sim \mathrm{IG}(\mu, \lambda)$ implies (3). It is important to note that if $v = 0$, the distribution of N is not IG. Furthermore, if $v < 0$, there is a nonzero probability that the particle never arrives at the receiving boundary. Throughout this paper, we will assume that $v > 0$.

Since the trajectories of the molecules are assumed to be mutually independent, the processes $W(x)$ for different molecules are also independent. As the information to be transmitted is encoded in the transmit time of each molecule, the message (or, the symbol) alphabet is $\mathcal{X} \subset \mathbb{R}_+$; further, the symbol $X = x$ represents a release of a single molecule at time x. This molecule has initial condition $W(x) = 0$; the molecule propagates via a Wiener process with drift velocity $v > 0$, and Wiener process variance σ^2. This process continues until arrival at the receiver, which occurs at time $Y \in \mathbb{R}_+$. Under our assumptions, for a single molecule released at time x

$$Y = x + N \quad (6)$$

[1]In [23], diffusion coefficient values between 1 and 10 $\mu\mathrm{m}^2/\mathrm{s}$ were considered realistic for signaling molecules.

where N is the first arrival time of the Wiener process. The probability density of observing channel output $Y = y$ given channel input $X = x$ is given by

$$f_{Y|X}(y|x) = \begin{cases} \sqrt{\frac{\lambda}{2\pi(y-x)^3}} \exp\left(-\frac{\lambda(y-x-\mu)^2}{2\mu^2(y-x)}\right), & y > x \\ 0, & y \leq x. \end{cases} \quad (7)$$

It is apparent that the channel is affected by additive noise, in the form of the random propagation time N; furthermore, by assumption, this is the only source of uncertainty or distortion in the system. As the additive noise N has the IG distribution, we refer to the channel defined by (6)–(7) as an AIGN channel. Note that we assume that the receiver can wait for infinite time to ensure that the molecule does arrive.

As an example of a molecular communication system with timing modulation, assume that the transmitter needs to convey one of the t messages; the message alphabet would be $\mathcal{X} = \{x_1, \ldots, x_t\}$ with $\Pr\{x_i\} = p_i$. With $p_i = 1/t$, the transmitter could convey up to $\log(t)$ nats per channel use. Corresponding to the ith message, $1 \leq i \leq t$, the transmitter releases a molecule into the fluid medium at time x_i. For the receiver, the transmitted message is a discrete random variable X with alphabet $\mathcal{X}$ and it observes $Y = x_i + N$ where N is the time taken by the molecule to arrive at the receiver. The receiver computes an estimate $\hat{X}$ of the transmitted message from Y, making use of other information such as the pdf of N and the *a priori* probabilities of the messages p_i, $i = 1, \ldots, t$. The transmission is successful when $\hat{X} = x_i$, and we declare an error when $\hat{X} \neq x_i$.

III. Capacity Bounds

A. Main Result

Equation (6) is reminiscent of the popular additive white Gaussian noise (AWGN) channel, a crucial parameter of which is the channel capacity. As in the AWGN case, the mutual information between the input and the output of the channel is given by

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X), \\ &= h(Y) - h(X + N|X) = h(Y) - h(N|X), \\ &= h(Y) - h(N), \end{aligned} \quad (8)$$

since X and N are independent. The capacity of the channel is the maximum mutual information, optimized over all possible input distributions $f_X(x)$. The set of all possible input distributions is determined by the constraints on the input signal X. With the information being encoded in the release time of the molecule, there is no immediate analog to input power for the AWGN channel; the constraints are application dependent; e.g., both peak-constrained and mean-constrained inputs appear reasonable. So far, peak constraints have not been analytically tractable; in this paper, we constrain the mean of the input signal such that

$$E[X] \leq m. \quad (9)$$

That is, on average we are only willing to wait m seconds to transmit our signal. (Throughout this paper, we also assume that the random variable X is nonnegative.)

We define capacity as follows.

Definition 1: The capacity of the AIGN channel with input X and mean constraint $E[X] \leq m$ is defined as

$$C = \max_{f_X(x):E[X]\leq m} I(X;Y). \tag{10}$$

From the receiver's perspective, $E[N]$ is finite as long as $v > 0$, so (9) ensures that the expected time of arrival of the molecule at the receiver is constrained; i.e., $E[Y] = E[X] + E[N] \leq m + E[N]$. Note that peak constraints are not possible at the receiver, since the support of the pdf of N is $[0, \infty)$.

Our main result in this section is an upper and lower bound on the capacity of the AIGN channel. Before stating this result, we need the following two properties of the IG distribution.

Property 1 (Differential Entropy of the IG Distribution): Let $h_{\mathrm{IG}(\mu,\lambda)}$ represent the differential entropy of the IG distribution with the parameters μ and λ. Then

$$h_{\mathrm{IG}(\mu,\lambda)} = \log\left(2K_{-1/2}(\lambda/\mu)\mu\right) + \frac{3}{2}\frac{\frac{\partial}{\partial\gamma}K_\gamma(\lambda/\mu)\left|_{\gamma=-1/2}\right.}{K_{-1/2}(\lambda/\mu)} + \frac{\lambda}{2\mu}\frac{K_{1/2}(\lambda/\mu)+K_{-3/2}(\lambda/\mu)}{K_{-1/2}(\lambda/\mu)} \tag{11}$$

where $K_\gamma(\cdot)$ is the order-γ modified Bessel function of the second kind. ■

This property is easily derived from the differential entropy of a generalized IG distribution; see Appendix A. The derivative of the Bessel function with respect to its order, needed in the second term of (11), is given by [25]

$$\frac{\partial}{\partial\gamma}K_\gamma(z)\left|_{\gamma=-1/2}\right. = \sqrt{\frac{\pi}{2z}}e^z\mathrm{Ei}(-2z) \tag{12}$$

where $\mathrm{Ei}(z) = \int_{-\infty}^{z}\frac{e^t}{t}dt$ is the exponential integral [25].

Property 2 (Additivity Property of the IG Distribution [24]): Let $N_i \sim \mathrm{IG}(\mu_i, \lambda_i)$, $i = 1, \ldots, k$, be k not necessarily independent IG random variables, and let $N = \sum_i c_i N_i$ for constants $c_i > 0$. If there exists a constant ν such that $\frac{\lambda_i}{c_i\mu_i^2} = \nu$ for all i, then $N \sim \mathrm{IG}(\sum_i c_i\mu_i, \nu(\sum_i c_i\mu_i)^2)$. ■

The bounds on the capacity C are then given by the following theorem.

Theorem 1: The capacity of the AIGN channel, defined in (10), is bounded as

$$h_{\mathrm{IG}(m+\mu,(\lambda/\mu^2)(m+\mu)^2)} - h_{\mathrm{IG}(\mu,\lambda)} \leq C \leq \log((\mu+m)e) - h_{\mathrm{IG}(\mu,\lambda)} \tag{13}$$

where $h_{\mathrm{IG}(\mu,\lambda)}$ is given by Property 1.

Proof: From (8)

$$I(X;Y) = h(Y) - h_{\mathrm{IG}(\mu,\lambda)} \tag{14}$$

with $h_{\mathrm{IG}(\mu,\lambda)}$ given by Property 1. From (10)

$$C = \max_{f_X(x):E[X]\leq m} h(Y) - h_{\mathrm{IG}(\mu,\lambda)} \tag{15}$$

$$= -h_{\mathrm{IG}(\mu,\lambda)} + \max_{f_X(x):E[X]\leq m} h(Y) \tag{16}$$

which follows since $h_{\mathrm{IG}(\mu,\lambda)}$ is independent of $f_X(x)$.

Upper Bound: Since X is nonnegative by assumption, and the IG-distributed first passage time is nonnegative by definition, Y is also nonnegative. Moreover, if $E[X] \leq m$, then $E[Y] \leq m + \mu$. Then

$$\max_{f_X(x):E[X]\leq m} h(Y) \leq \log((\mu+m)e) \tag{17}$$

which follows since $\log((\mu+m)e)$ is the entropy of an exponentially distributed random variable with mean $m+\mu$: the maximum-entropy distribution of a nonnegative random variable with a mean constraint [26]. Thus, substituting into (16)

$$C \leq \log((m+\mu)e) - h_{\mathrm{IG}(\mu,\lambda)}. \tag{18}$$

Lower Bound: Suppose the input signal X is IG distributed with mean equal to m, satisfying (9). Choose the second parameter of the IG distribution for the input signal X as $(\lambda/\mu^2)m^2$; i.e., $X \sim \mathrm{IG}(m, (\lambda/\mu^2)m^2)$. Then from Property 2, $Y \sim \mathrm{IG}(m+\mu, (\lambda/\mu^2)(m+\mu)^2)$ and $h(Y) = h_{\mathrm{IG}(m+\mu,(\lambda/\mu^2)(m+\mu)^2)}$. Then

$$\max_{f_X(x):E[X]\leq m} h(Y) \geq h_{\mathrm{IG}(m+\mu,(\lambda/\mu^2)(m+\mu)^2)} \tag{19}$$

since $\mathrm{IG}(m, (\lambda/\mu^2)m^2)$ is in the feasible set of the maximization. Thus, again substituting into (16)

$$C \geq h_{\mathrm{IG}(m+\mu,(\lambda/\mu^2)(m+\mu)^2)} - h_{\mathrm{IG}(\mu,\lambda)}. \tag{20}$$

The theorem follows from (18) and (20). ■

Note that if one could find a valid pdf for X (with $E[X] \leq m$) that resulted in an exponential distribution for Y (via convolution with the IG distribution of N), then the expression in (18) would be the true capacity for mean constrained inputs. For example, at asymptotically high velocities, i.e., as $v \to \infty$, $\mu = d/v \to 0$, and the variance $\mathrm{Var}(N) = \frac{\mu^3}{\lambda} \to 0$, i.e., the noise distribution tends to the Dirac delta function. The fact that $\frac{N}{\mu} \to 1$ as $v \to \infty$ is proven in [27]. The fact that Y is distributed exponentially then leads to the conclusion that, at high drift velocities, the optimal input X is also exponential; i.e., $X \sim \exp(1/m)$.

At low velocities, the situation is considerably more complicated. As shown in Appendix B, the deconvolution of the output (Y) and noise (N) pdfs leads to an invalid pdf; i.e., at asymptotically low velocities, this upper bound does not appear achievable.

B. Numerical Results

We now present numerical results by evaluating the mutual information of the AIGN channel and, in order to illustrate the upper and lower bounds stated by Theorem 1, we consider $Y \sim \exp(1/(m+\mu))$ and $Y \sim \mathrm{IG}(m+\mu, (\lambda/\mu^2)(m+\mu)^2)$. We present two more lower bounds on the capacity of the AIGN channel by considering the following.

1) Uniformly distributed input X in the range $[0, 2m]$,
2) Exponentially distributed input X with mean m; i.e., $X \sim \exp(1/m)$

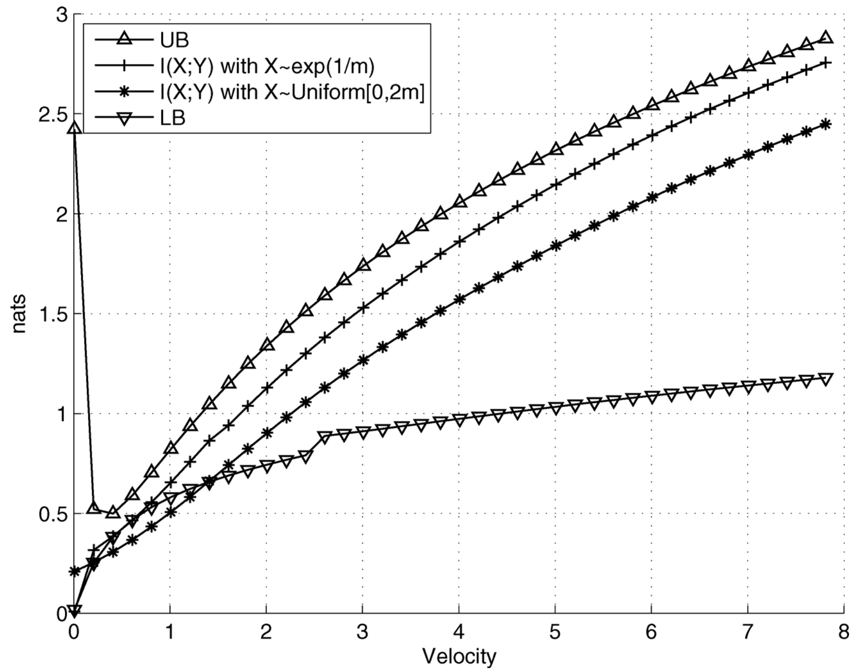

Fig. 2. Mutual information as a function of velocity; $\sigma^2 = 1$.

In these two cases, we only have closed-form expressions for $f_Y(y)$ but not for $h(Y)$ and we compute $h(Y)$ through numerical integration to plot $I(X;Y)$. Note that, for all the cases, $m = 1$.

In the case where X has the uniform distribution on $[0, 2m]$, convolving the input and noise distributions leads to

$$f_Y(y) = \begin{cases} \frac{1}{2m} F_N(y), & y \le 2m; \\ \frac{1}{2m}\left(F_N(y) - F_N(y-2m)\right), & y > 2m, \end{cases} \tag{21}$$

where $F_N(n)$ is the cumulative distribution function (cdf) of N and is given by [24]

$$F_N(n) = \Phi\left(\sqrt{\frac{\lambda}{n}}\left(\frac{n}{\mu} - 1\right)\right) + e^{2\lambda/\mu}\Phi\left(-\sqrt{\frac{\lambda}{n}}\left(\frac{n}{\mu} + 1\right)\right), \tag{22}$$

where $\Phi(z) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$ is the cdf of a standard Gaussian random variable Z.

In the case where $X \sim \exp(1/m)$ with $m \ge 2\sigma^2/v^2$, the convolution leads to [28]

$$f_Y(y) = \frac{1}{m} e^{\left(-\frac{y}{m} + d\frac{v}{\sigma^2}\right)} \left(e^{-kd/\sigma^2}\Phi \cdot \left(\frac{ky - d}{\sigma\sqrt{y}}\right) + e^{kd/\sigma^2}\Phi\left(-\frac{ky + d}{\sigma\sqrt{y}}\right)\right), \tag{23}$$

where $k = \sqrt{v^2 - \frac{2\sigma^2}{m}}$. When $m < 2\sigma^2/v^2$, the convolution of X and N results in the pdf of Y given by [28]

$$f_Y(y) = \frac{1}{m} e^{-\frac{(d - vy)^2}{2\sigma^2 y}} \Re\left(w\left(\frac{k'\sqrt{y}}{\sigma\sqrt{2}} + \jmath\frac{d}{\sigma\sqrt{2y}}\right)\right), \tag{24}$$

where $k' = \sqrt{\frac{2\sigma^2}{m} - v^2}$ and $w(z) = e^{-z^2}(1 - \operatorname{erf}(-\jmath z))$.

Fig. 2 plots the mutual information as a function of velocity for all four cases mentioned previously. It is only over a narrow range of velocities that the upper and lower bounds, given by (13), are close to each other. Further, the cases with exponential and uniform inputs track the upper bound, with the exponential input approaching the bound at high velocities. This is consistent with the discussion in the previous section. However, given its finite support, a uniform input may be closer to a practical signaling scheme.

Fig. 3 presents a closer look at the behavior of the bounds at low velocities ($\sigma^2 = 1$) by plotting $h(Y)$ and $h(N)$, the differential entropies of Y and N, respectively, along with the upper and lower bounds. The upper bound has a nonmonotonic behavior: it decreases with decreasing velocity but after a certain value of v, it increases as the velocity decreases. The lower bound decreases monotonically as the velocity goes to zero. This can be explained as follows. Consider the upper bound on capacity, i.e., $I(X;Y)$ with $Y \sim \exp(1/(m+\mu))$. Recall that $\mu = \frac{l}{v}$ and $h(Y) = \log((m+\mu)e)$. As v decreases, μ increases and hence $h(Y)$ also increases (as can be seen from the figure). $h(N)$ also increases as the velocity decreases,[2] but it does so at a lower rate than $h(Y)$. This results in a net improvement in $I(X;Y)$ with $Y \sim \exp(1/(m+\mu))$. On the other hand, when $Y \sim \mathrm{IG}(m+\mu, (\lambda/\mu^2)(m+\mu)^2)$, both $h(Y)$ and $h(N)$ have almost the same value at lower values of v, resulting in $I(X;Y)$ that goes to zero.

At low velocities, as shown in Appendix B, the result of deconvolving the output (Y) and noise (N) pdfs does not satisfy the necessary properties of a pdf. Hence, the upper bound does not appear achievable, and the upper bound may be rather loose at low velocities. As the differential entropy of the noise ($h(N)$)

[2]The monotonic decrease in $h(N)$ with increasing velocity satisfies the intuition that the propagation time has less variance as the velocity of the fluid medium increases.

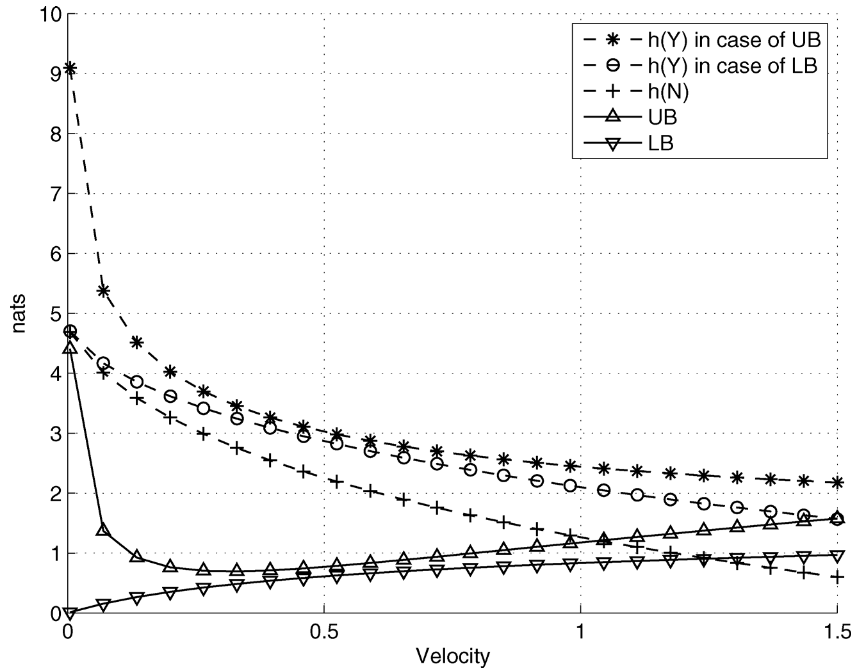

Fig. 3. Upper and lower bounds at low velocities for $\sigma^2 = 1$.

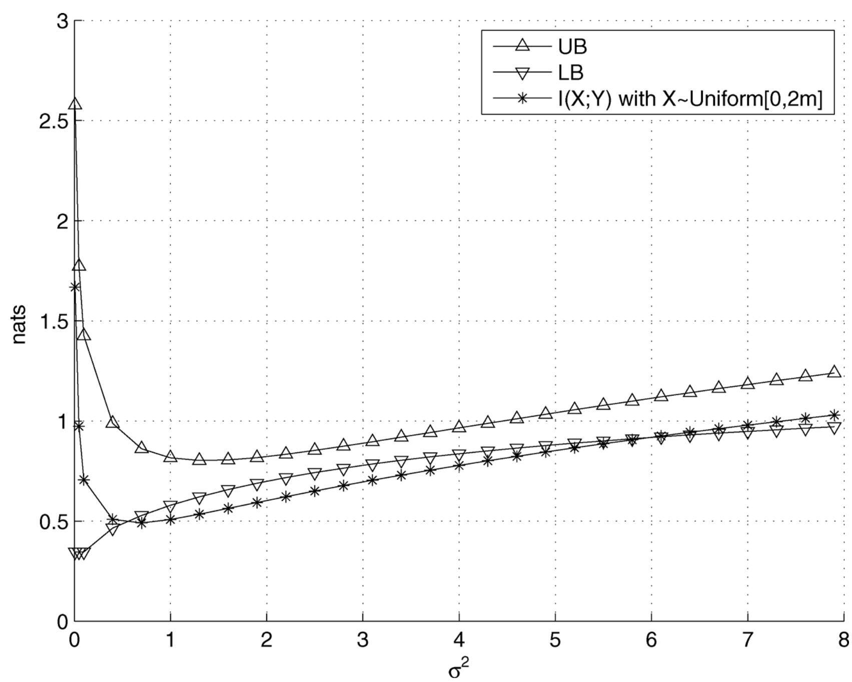

Fig. 4. Mutual information as a function of σ^2; $v = 1$.

decreases monotonically with increasing velocity, velocity is one indicator of the channel quality in an AIGN channel. However, it is instructive to compare these features to the AWGN channel, in which the channel capacity is monotonic with respect to a single parameter, namely the signal power to noise power ratio (SNR). In the AIGN channel, the mutual information cannot be similarly reduced to a function of a single parameter. The pdf in (3) is a function of both velocity (via μ) and diffusion coefficient $d = 2\sigma^2$ (via λ).

An example of this complex relationship is shown in Fig. 4, where $v = 1$. Both the upper bound and the mutual information with uniform inputs fall with increasing diffusion (randomness), but then further increasing diffusion increases mutual information. The increase in mutual information as a function of diffusion is, initially, counterintuitive since diffusion is assumed to be the source of randomness, and hence uncertainty. To understand this result, it is instructive to consider the zero-velocity (no drift) case. Without diffusion, molecules would forever remain stationary at the transmitter, and never arrive at the receiver, resulting in zero mutual information. In this case, increasing diffusion *helps* communication. Thus, while it is true that diffusion increases randomness, its impact is not monotonic. To illustrate

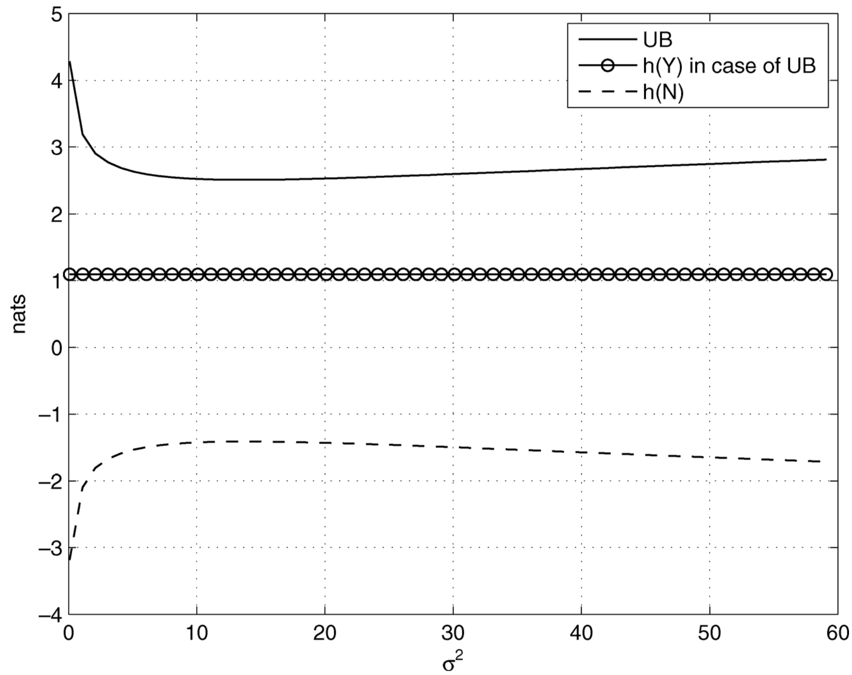

Fig. 5. Mutual information as a function of σ^2; $v = 10$.

this effect, consider Fig. 5. Here, the velocity is set relatively high ($v = 10$). The plots are the entropies and mutual information (upper bound) as a function of σ^2. The upper bound falls steeply until $\sigma^2 \approx 4$, very slowly until $\sigma^2 \approx 10$, and then *rises* slowly for increasing σ^2. This is because for relatively large values of σ^2, this velocity appears "low" and increasing diffusion increases mutual information. This is confirmed by the falling entropy of the noise term ($h(N)$).

To summarize, in this section we developed capacity bounds for the AIGN channel based on the IG distribution of the molecule propagation time. While increasing velocity increases mutual information, increasing diffusion beyond a point also increases mutual information. Unlike the AWGN channel, no single parameter captures the performance of the AIGN channel.

IV. Receiver Design

We now discuss receivers for the AIGN channel by recovering the transmitted message (transmission time) from the times the molecules are received. We develop both the ML estimator and the ML detector, and provide an error probability analysis for the ML detection.

A. Maximum Likelihood Estimator

The receiver observes $Y = X + N$ and needs to compute $\hat{X}$, an estimate of X. Given an observation $Y = y$, the ML estimator of X, denoted $\hat{X}_{\mathrm{ML}}$, is given by

$$\hat{X}_{\mathrm{ML}} = \arg\max_{x} f_{Y|X}(y|X = x), \tag{25}$$

where

$$f_{Y|X}(y|X = x) = \begin{cases} \sqrt{\frac{\lambda}{2\pi(y-x)^3}} \exp\left(-\frac{\lambda}{2\mu^2}\frac{((y-x)-\mu)^2}{(y-x)}\right), & y > x \\ 0, & y \le x. \end{cases} \tag{26}$$

The pdf given previously is commonly known as the shifted IG distribution, or the three-parameter IG distribution, and is denoted as $\mathrm{IG}(t_0, \mu, \lambda)$ where t_0 is the location parameter [24], or the threshold parameter [27].

Theorem 2: For a given observation $Y = y$, the ML estimator $\hat{X}_{\mathrm{ML}}$ of the transmitted symbol X in an AIGN channel is given by

$$\hat{X}_{\mathrm{ML}} = y + \frac{\mu^2}{\lambda}\left(\frac{3}{2} - \sqrt{\frac{9}{4} + \frac{\lambda^2}{\mu^2}}\right). \tag{27}$$

Proof: Let $\Lambda(x_i) = \log f_{Y|X}(y|X = x_i)$ represent the log-likelihood function. Since $\log$ is monotonic,

$$\hat{X}_{\mathrm{ML}} = \arg\max_{x_i} f_{Y|X}(y|X = x_i) = \arg\max_{x_i} \Lambda(x_i).$$

In our case,

$$\Lambda(x_i) = \begin{cases} -\frac{3}{2}\log(y - x_i) - \frac{\lambda}{2\mu^2}\frac{((y-x_i)-\mu)^2}{(y-x_i)}, & y > x_i \\ -\infty, & y \le x_i. \end{cases} \tag{28}$$

By setting $\frac{\partial \Lambda(x_i)}{\partial x} = 0$, and searching over values of $x_i < y$, we obtain the MLE given by (27). ■

This result is consistent with the expected high-velocity case ($v \to \infty$), wherein $\hat{X}_{\mathrm{ML}} = y$.

B. ML Detection: SEP Analysis

Analogous to the use of a signal constellation in AWGN channels, we now restrict the input to the channel, i.e, the transmission time, to take discrete values: for t-ary modulation, we have $X \in \{x_1, \ldots, x_t\}$, $0 \le x_1 < x_2, \ldots, < x_t$. We begin with the error probability for binary modulation with ML detection at the receiver. Let $X \in \{x_1, x_2\}$, $0 \le x_1 < x_2$, with

$\Pr(X = x_1) = p_1$ and $\Pr(X = x_2) = p_2 = 1 - p_1$. For a given observation $Y = y$, the *log-likelihood ratio* $L(y)$ is given by

$$L(y) = \log \frac{f_{Y|X}(y|X = x_2)}{f_{Y|X}(y|X = x_1)} = \Lambda(x_2) - \Lambda(x_1) = \begin{cases} -\infty, & y \le x_2 \\ \frac{3}{2}\log\frac{y-x_1}{y-x_2} + \frac{\lambda}{2\mu^2}\left(\mu^2\left(\frac{1}{y-x_2} - \frac{1}{y-x_1}\right) + x_1 - x_2\right), & y > x_2. \end{cases} \tag{29}$$

If $L(y)$ is positive (negative), then x_2 has higher (lower) likelihood than x_1. If $L(y) = 0$, then there is no preference between x_1 and x_2; we ignore this case, which occurs with vanishing probability. Thus, for ML detection, the decision rule is as follows.

- Pick $X = x_2$ if $L(y) > 0$; otherwise, pick $X = x_1$.

For MAP detection, we use the same decision rule, replacing $L(y) > 0$ with $L(y) > \log(p_1/p_2)$.

The SEP is given by

$$P_e = p_1 \Pr\{x_1 \to x_2\} + p_2 \Pr\{x_2 \to x_1\}, \tag{30}$$

where $\Pr\{x_i \to x_j\}$ is the probability of $\hat{X}_{\mathrm{ML}} = x_j$ when $X = x_i$.

$$\Pr\{x_1 \to x_2\} = \int_{y_{\mathrm{th}}}^{\infty} f_{Y|X}(y|X = x_1)dy, \tag{31}$$

where y_{th} is the decision threshold value of y, satisfying $L(y_{\mathrm{th}}) = 0$. Similarly,

$$\Pr\{x_2 \to x_1\} = \int_{x_2}^{y_{\mathrm{th}}} f_{Y|X}(y|X = x_2)dy. \tag{32}$$

We now give an upper bound on the error probability for the case when $p_1 \ge p_2$, which is simple to calculate and yet closely approximates the exact error probability.

Theorem 3: Let $X \in \{x_1, x_2\}$, $0 \le x_1 < x_2$, with $\Pr(X = x_1) = p_1$, $\Pr(X = x_2) = p_2$, and $p_1 \ge p_2$. The upper bound on the SEP of the ML detector in an AIGN channel with input X is given by

$$P_e < p_1(1 - F_N(x_2 - x_1)). \tag{33}$$

Proof: Let

$$\delta = \int_{x_2}^{\infty} f_{Y|X}(y|X = x_1)dy - \int_{y_{\mathrm{th}}}^{\infty} f_{Y|X}(y|x_1)dy = \int_{x_2}^{y_{\mathrm{th}}} f_{Y|X}(y|X = x_1)dy. \tag{34}$$

Then

$$\Pr\{x_1 \to x_2\} = \int_{y_{\mathrm{th}}}^{\infty} f_{Y|X}(y|X = x_1)dy = \int_{x_2}^{\infty} f_{Y|X}(y|X = x_1)dy - \delta. \tag{35}$$

Note that $\delta > 0$ since $y_{\mathrm{th}} > x_2$. Furthermore

$$\begin{aligned}\Pr\{x_2 \to x_1\} &= \int_{x_2}^{y_{\mathrm{th}}} f_{Y|X}(y|X = x_2)dy \\ &\le \int_{x_2}^{y_{\mathrm{th}}} f_{Y|X}(y|X = x_1)dy \qquad (36) \\ &= \delta \qquad (37)\end{aligned}$$

where (36) follows since, under ML detection, $f_{Y|X}(y|X = x_1) \le f_{Y|X}(y|X = x_2)$ when $y \le y_{\mathrm{th}}$. Now, (30) becomes

$$\begin{aligned}P_{\mathrm{e}} &= p_1 \Pr\{x_1 \to x_2\} + p_2 \Pr\{x_2 \to x_1\} \\ &\le p_1\left(\int_{x_2}^{\infty} f_{Y|X}(y|X = x_1)dy - \delta\right) + p_2\delta \qquad (38) \\ &= p_1 \int_{x_2}^{\infty} f_{Y|X}(y|X = x_1)dy - (p_1 - p_2)\delta \\ &\le p_1 \int_{x_2}^{\infty} f_{Y|X}(y|X = x_1)dy \qquad (39)\end{aligned}$$

where the last inequality follows since $p_1 \ge p_2$ (by assumption), and so $(p_1 - p_2)\delta$ is nonnegative. Finally, note that $\int_{x_2}^{\infty} f_Y(y|X = x_1)dy = 1 - F_N(x_2 - x_1)$, and (33) follows. ■

Corollary 1: The upper bound on the SEP given by (33) is asymptotically tight as $v \to \infty$; i.e.,

$$\lim_{v\to\infty} (P_{\mathrm{e}} - p_1(1 - F_N(x_2 - x_1))) = 0. \tag{40}$$

Proof: The error in bound (38) is at most $p_2\delta$, and the error in bound (39) is equal to $(p_1 - p_2)\delta$; thus, the total error is at most $p_1\delta$. Noting that $\mu \to 0$ as $v \to \infty$, we show that $\delta \to 0$ as $\mu \to 0$. For $y \ge x_2$, we have

$$\begin{aligned}f_{Y|X}(y|x = x_1) &= \sqrt{\frac{\lambda}{w\pi(y - x_1)^3}} \exp\left(-\frac{\lambda(y - x_1 - \mu)^2}{2\mu^2(y - x_1)}\right) \\ &= \sqrt{\frac{\lambda}{w\pi(y-x_1)^3}} \exp\left(-\frac{\lambda(y-x_1-2\mu)}{2\mu^2}\right)\exp\left(-\frac{\lambda}{2(y-x_1)}\right) \\ &\le \sqrt{\frac{\lambda}{w\pi(x_2 - x_1)^3}} \exp\left(-\frac{\lambda(x_2 - x_1 - 2\mu)}{2\mu^2}\right). \qquad (41)\end{aligned}$$

Finally, $\delta \to 0$ follows from substituting (41) into (34): since $x_2 - x_1 > 0$ (by assumption), $f_{Y|X}(y|X = x_1) \to 0$ for all $y \ge x_2$ as $\mu \to 0$, and (40) follows. ■

To illustrate this result, consider Fig. 6: δ is the area under the curve $f_{Y|X}(y|X = x_1)$ as y varies from x_2 to y_{th} and is always larger than $\int_{x_2}^{y_{\mathrm{th}}} f_{Y|X}(y|X = x_2)dy$, the area under the curve $f_{Y|X}(y|X = x_2)$ from x_2 to y_{th}.

This bound can easily be generalized to t-ary modulation. When $X \in \{x_1, \ldots, x_t\}$, $0 \le x_1 < x_2, \ldots, < x_t$, and $p_1 \ge p_2 \ge \ldots \ge p_t$, the upper bound on SEP is given by

$$P_{\mathrm{e}} < \sum_{i=1}^{t-1} p_i \left(1 - F_N\left(x_{i+1} - x_i\right)\right). \tag{42}$$

To compute the ML estimate, the receiver needs to know μ and λ, the parameters of the noise. One way to enable the receiver

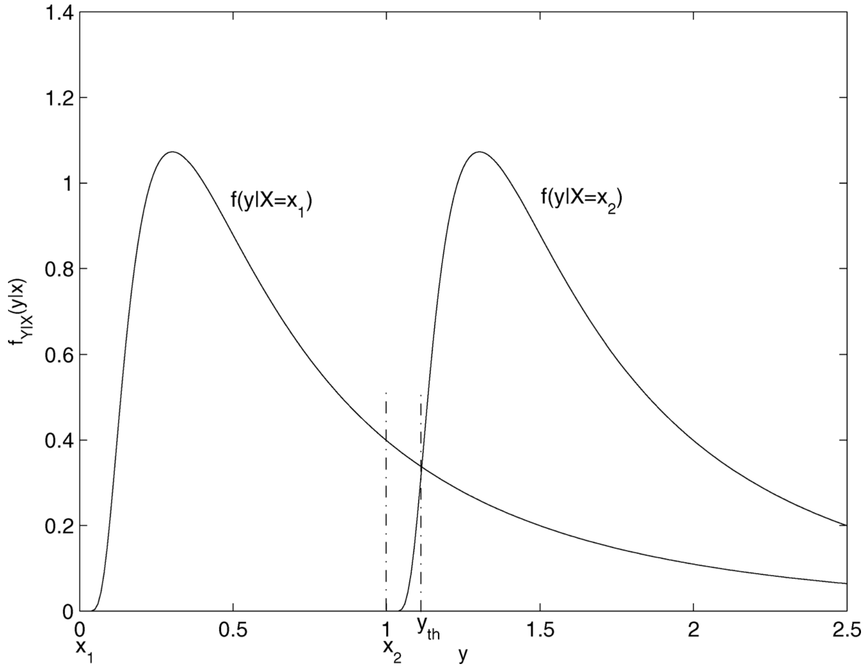

Fig. 6. Deriving the upper bound on SEP; $x_1 = 0, x_2 = 1, v = 1, \sigma^2 = 1$, and $l = 1$.

to acquire the knowledge of these parameters is by *training* as in a conventional communication system. Appendix C provides the ML estimates of these parameters based on the IG pdf.

C. Improving Reliability: Transmitting Multiple Molecules

The performance of a molecular communication system (the mutual information and the error rate performance) can be improved by transmitting multiple molecules to convey a message symbol. We assume that the trajectories of the molecules are independent, and that they do not interact with each other during their propagation from the transmitter to the receiver.

The transmitter releases $\kappa > 1$ molecules *simultaneously* to convey one of t messages, $X \in \{x_1, \ldots, x_t\}$. In [9], it was shown using simulations that if multiple molecules are available, releasing them simultaneously is the best strategy. Essentially, releasing them at different times leads to confusion at the receiver with molecules potentially arriving out of order. In the case of simultaneous transmissions, the receiver observes κ mutually independent arrival times

$$Y_j = X + N_j,\ j = 1, \ldots, \kappa \tag{43}$$

where N_j are i.i.d. with $N_j \sim \mathrm{IG}(\mu, \lambda), j = 1, \ldots, \kappa$.

1) Maximum Likelihood Estimation: We first consider ML detection of the symbol when multiple molecules are used. Assuming that the receiver knows the values of μ and λ through an earlier training phase, it can use the multiple observations $Y_j, j = 1, \ldots, \kappa$, to obtain $\hat{X}_{\mathrm{ML}}$.

The pdfs $f_{Y_j|X}(y_j|X = x_i), j = 1, \ldots, \kappa$, are i.i.d. with $f_{Y_j|X}(y_j|X = x_i)$ given by (26). The ML estimate, in this case, is given by (44), shown at the bottom of the page. Simplifying the aforementioned equation, the ML estimate can be expressed as

$$\hat{X}_{\mathrm{ML}} = \arg\max_{x_i} \Lambda_\kappa(x_i) \tag{45}$$

where

$$\Lambda_\kappa(x_i) = -\frac{3}{2}\sum_{j=1}^{\kappa} \log(y_j - x_i) - \frac{\lambda}{2\mu^2}\sum_{j=1}^{\kappa} \frac{((y_j - x_i) - \mu)^2}{(y_j - x_i)} y_j > x_i. \tag{46}$$

2) Linear Filter: The aforementioned approach estimates the transmitted message using a complicated ML detection filter that processes the received signal. Given the potential applications of this research, a simpler filter would be useful. One such filter is the linear average, which is optimal in an AWGN

$$\begin{aligned}\hat{X}_{\mathrm{ML}} &= \arg\max_{x_i} \prod_{j=1}^{\kappa} f_{Y_j|X}(y_j|X = x_i)\\ &= \arg\max_{x_i} \prod_{j=1}^{\kappa} (y_j - x_i)^{-3/2} \exp\left(-\frac{\lambda}{2\mu^2}\sum_{j=1}^{\kappa} \frac{((y_j - x_i) - \mu)^2}{(y_j - x_i)}\right), \quad y_j > x_i\end{aligned} \tag{44}$$

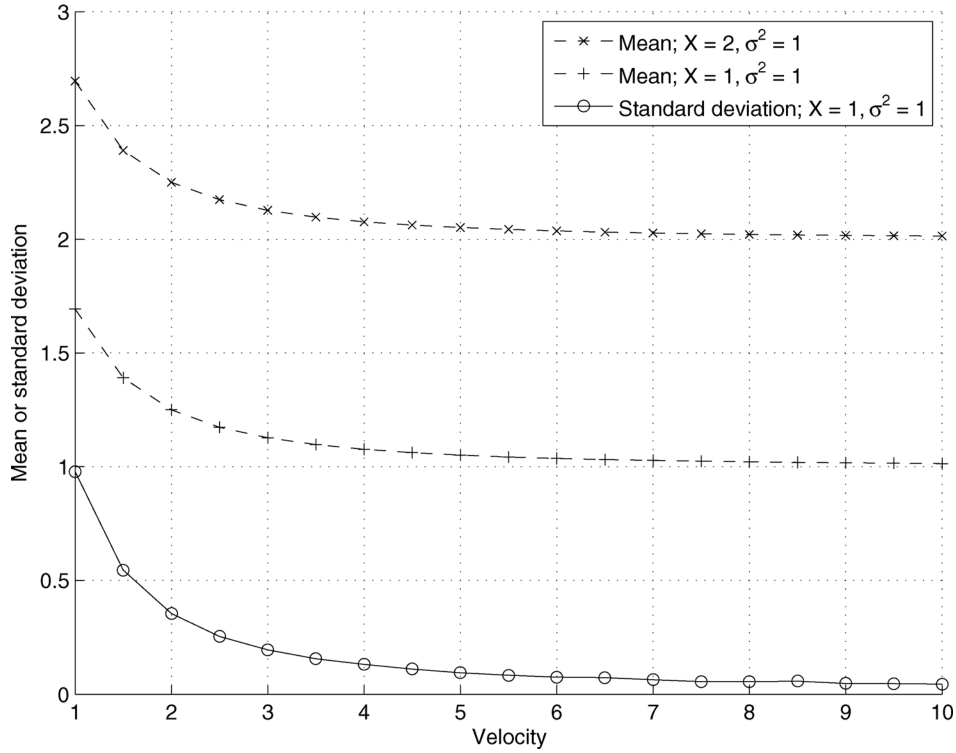

Fig. 7. Mean and standard deviation of $\hat{X}_{\mathrm{ML}}$.

channel [29]. In this case, the receiver averages the κ observations and performs a ML estimate with the sample mean as the test statistic. The receiver generates

$$Z = \frac{1}{\kappa}\sum_{j=1}^{\kappa} Y_j. \tag{47}$$

The linear filter has the following nice property: by the additivity property of IG distribution in Property 2, $Z \sim \mathrm{IG}(E[X]+\mu, \kappa\lambda)$. Now,

$$\hat{X}_{\mathrm{ML}} = \arg\max_{x_i} f_Z(z|X = x_i),$$

where

$$f_{Z|X}(z|X = x_i) = \sqrt{\frac{\kappa\lambda}{2\pi(z - x_i)^3}}\exp\left(-\frac{\kappa\lambda}{2\mu^2}\frac{((z - x_i) - \mu)^2}{(z - x_i)}\right) z > x_i. \tag{48}$$

The linear receiver, therefore, acts as if σ^2 is reduced by a factor of κ to σ^2/κ. At reasonably high velocities, this leads to better performance; however, we have seen in Section III that, at low velocities, diffusion can actually help communications.

At high drift velocities, the reduction in the effective diffusion results in an effect akin to the diversity order in wireless communication systems. This is shown in the following result.

Theorem 4: For 2-ary modulation with $X \in \{x_1, x_2\}, 0 \leq x_1 < x_2$, as drift velocity $v \to \infty$, the upper bound on SEP can be approximated as

$$\log(P_{\mathrm{e}}) < -c_1\frac{(cv^2 - vl)}{\sigma^2} + c_2 \log\frac{d^2}{cv^2} + c_3 \tag{49}$$

where c_1, c_2, and c_3 are constants.

Proof: The proof is found in Appendix D. ■

Furthermore, for κ molecules and detection using the linear filter,

$$\log(P_{\mathrm{e}}) < -c_1\frac{\kappa(cv^2 - vl)}{\sigma^2} + c_2 \log\frac{l^2}{cv^2} + c_3, \tag{50}$$

which is essentially (49) with σ^2 replaced by σ^2/κ.

Since, in both (49) and (50), the first term dominates at high velocities, a semilog plot of P_{e} versus velocity is asymptotically linear, with slope proportional to $-\kappa$.

D. Simulation Results

Fig. 7 shows how the variance and the mean of the ML estimate vary with velocity for a given σ^2. With increasing velocity, the estimator becomes unbiased and the variance approaches zero. Fig. 8 plots the SEP with t-ary modulation for different values t. The input alphabet employed for simulations is $X \in \{1 + \frac{i-1}{t-1}, i = 1, \ldots, t\}$. The figure also compares the upper bound on error probability, presented in Section IV-B, with the error probability obtained through Monte Carlo simulations.

The poor performance of t-ary modulation as shown in Fig. 8 motivates the multiple molecule system described in Section IV-C. Fig. 9 plots the error rate performance when $X \in \{1, 2\}$ and each symbol is conveyed by releasing multiple molecules. As expected, there is a effect akin to receive diversity in a wireless communication system. Here, the performance gain in the error probability increases with the number of molecules transmitted per message symbol.

Fig. 9 also compares the performance of the averaging filter with the ML estimation given by (45). The linear averaging filter is clearly suboptimal with performance worsening with increasing number of molecules transmitted per symbol (κ). This result again underlines the significant differences between the AIGN and AWGN channel models.

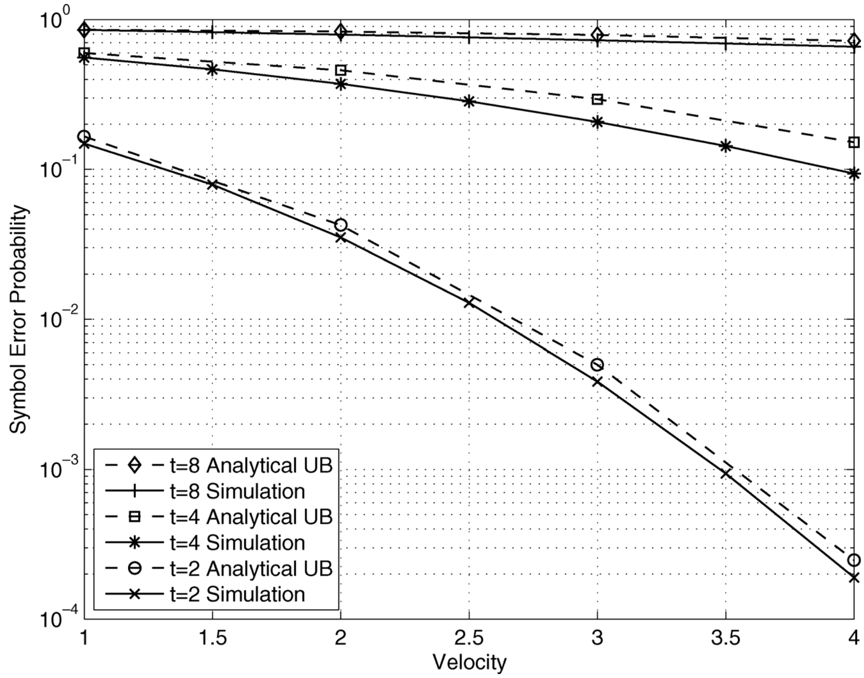

Fig. 8. Comparing the analytical upper bound and simulated error probability; single molecule case with t-ary modulation. Equiprobable symbols and $\sigma^2 = 1$.

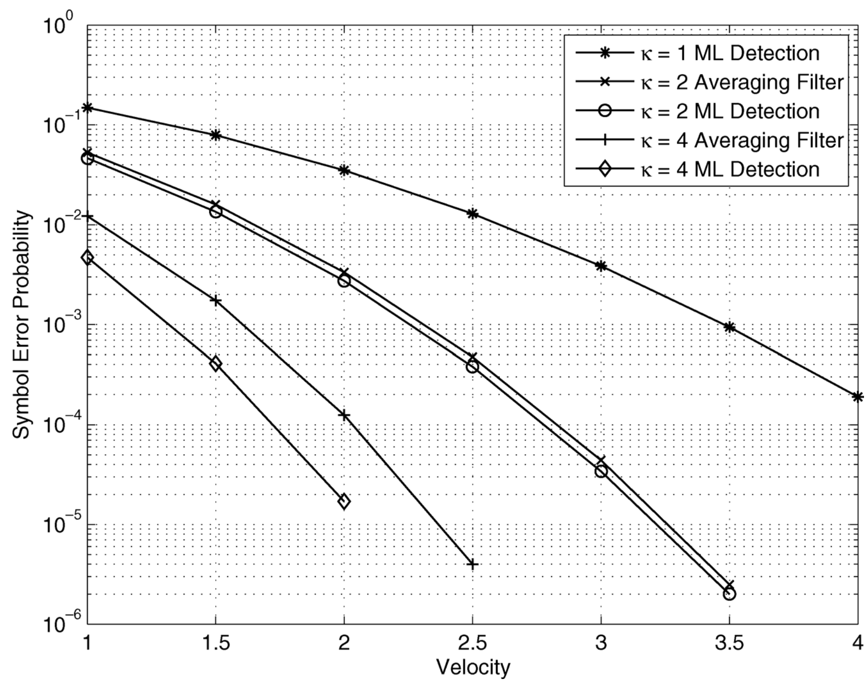

Fig. 9. Comparing the error probability of MLE with the averaging filter. Equal *a priori* probabilities and $\sigma^2 = 1$.

V. Discussion

In proposing a new channel model based on IG noise, we have analyzed the simplest possible interesting cases. In this regard, there are several issues left unresolved.

A. Single Versus Multiple Channel Uses

Throughout this paper, we have focused on the case of a single channel use, in which we use the channel to transmit a single symbol of information; our capacity results are measured in units of nats per channel use. Translating these results to nats per molecule is straightforward: each channel use consists of a deterministic number of molecules κ, where $\kappa \geq 1$; thus, we merely divide by κ. However, measuring capacity in nats per unit time is a more complicated issue, since the duration of the channel use is a random variable, dependent on both the input and the output. Following [19], where the capacity per unit time of a queue timing channel was calculated with respect to the average service time, here we can normalize our capacity results either with the average propagation time $E[N]$ or the average length of the communication session $E[Y]$. Since $E[Y] = E[X] + E[N]$, our decision to constrain the mean of the input distribution $f_X(x)$ would then have a natural interpretation in terms of the capacity per unit time.

Further, our system model excludes the possibility of other molecules propagating in the environment, except those transmitted as a result of the channel use; equivalently, we assume each channel use is orthogonal. This raises the question of how to use the channel repeatedly: if the signaling molecules are indistinguishable, then (under our formulation) the transmitter must wait until all M molecules have arrived before a new channel use can begin. On the other hand, if the signaling molecules are distinguishable, then channel uses can take place at any time, or even at the same time. This is because if there is no ambiguity in matching received molecules to channel uses, those channel uses are orthogonal.

B. Intersymbol Interference

Repeated channel uses also lead to a situation akin to intersymbol interference (ISI) in conventional communications. Since propagation time is not bounded, the transmitter may release the molecule corresponding to the "next" symbol while the "previous" molecule is still in transit. Molecules may, therefore, arrive out of order, introducing memory to the channel: to calculate the likelihood of the transmitted message, the optimal detector must consider all possible permutations of molecules' arrival. This problem is exacerbated if multiple molecules are released simultaneously to achieve diversity.

C. Synchronization and Differential Encoding

The system model and the analysis presented here assume perfect synchronization between the transmitter and the receiver. It is unclear how difficult it would be to achieve this with nanoscale devices. An information theoretic analysis of the effect of asynchronism in AWGN channels has been presented in [30]. Given the importance of timing in our model, extensions of such work to the AIGN channel would be useful. An interesting alternative would be to use differential modulation schemes such as *interval modulation* presented in [31].

D. Amplitude and Timing Modulation

The work presented here focuses on timing modulation, which leads naturally to the AIGN channel model. We can as well consider "amplitude" modulation wherein the number of molecules released by the transmitter is varied according to the message to be transmitted. In this context, it might be possible to leverage work on positive-only signaling channels such as in optics [32]. Amplitude modulation could be coupled with the timing modulation considered here. However, it is important to note that any amplitude information would be reproduced at the receiver faithfully since, in the model we have considered so far, the receiver is allowed to wait for all molecules to arrive before decoding. Therefore, to be useful, a reasonable model of amplitude modulation must include delay constraints and account for the issue of ISI as described previously.

E. Two-Way Communication and Negative Drifts

The AIGN channel model is valid only in the case of a positive drift velocity. In this regard, it does not support two-way communication between nanodevices. With negative drift velocities, it is not guaranteed that the molecule arrives at the receiver [24]. Molecular communications with negative drift velocities remain a completely open problem and one that is outside the scope of this paper. With negative drift, the pdf of the noise is distributed IG (albeit with a nonzero probability of particles never arriving), so the IG framework provided here might be useful to analyze such a channel.

VI. Conclusion

We have considered molecular communication between nanoscale devices connected through fluid medium, with information encoded in the release times of the molecules. We have proposed a mathematical framework to study such a communication system by modeling it as communication over an additive noise channel with noise following the IG distribution. We then obtained lower and upper bounds on the capacity of an AIGN channel and discussed receiver design. Our results illustrate the feasibility of molecular communication and show that it can be given a mathematical framework. Our key contribution here has been to provide this mathematical framework, making it possible to tackle some of the open problems in molecular communications.

Appendix A
Differential Entropy of the IG Distribution

Here, we prove Property 1. For a given μ and λ, the differential entropy of the noise $h(N)$ is fixed and can be computed from the generalized IG distribution (GIG). The GIG distribution is characterized by three parameters and the pdf of a random variable X distributed as GIG is given by [24]

$$f_X(x;\gamma,\mu,\lambda)=\frac{1}{2\mu^\gamma K_\gamma\left(\frac{\lambda}{\mu}\right)}\,x^{\gamma-1}\exp\left(-\frac{\lambda x^{-1}+(\lambda/\mu^2)x}{2}\right)$$
$$-\infty<\gamma<\infty,\mu>0,\lambda\geq 0,x>0 \quad (51)$$

where $K_\gamma(\cdot)$ is the modified Bessel function of the third kind of order γ. It is commonly denoted as $\mathrm{GIG}(\gamma,\mu,\lambda)$ and $\mathrm{IG}(\mu,\lambda)$ is a special case, obtained by substituting $\gamma=-1/2$ [24].

When $X\sim\mathrm{GIG}(\gamma,\mu,\lambda)$, its differential entropy, in nats, is given by [33]

$$h(X)=\log\left(2K_\gamma(\lambda/\mu)\mu\right)-(\gamma-1)\frac{\frac{\partial}{\partial\gamma}K_\gamma(\lambda/\mu)}{K_\gamma(\lambda/\mu)} +\frac{\lambda}{2\mu}\frac{K_{\gamma+1}(\lambda/\mu)+K_{\gamma-1}(\lambda/\mu)}{K_\gamma(\lambda/\mu)}. \quad (52)$$

Setting $\gamma=-1/2$, the differential entropy of $N\sim\mathrm{IG}(\mu,\lambda)$ is given by

$$h(N)=h_{\mathrm{IG}(\mu,\lambda)}=\log\left(2K_{-1/2}(\lambda/\mu)\mu\right) +\frac{3}{2}\frac{\frac{\partial}{\partial\gamma}K_\gamma(\lambda/\mu)\big|_{\gamma=-1/2}}{K_{-1/2}(\lambda/\mu)} +\frac{\lambda}{2\mu}\frac{K_{1/2}(\lambda/\mu)+K_{-3/2}(\lambda/\mu)}{K_{-1/2}(\lambda/\mu)} \quad (53)$$

and the property follows.

APPENDIX B
EVALUATING OPTIMAL INPUT DISTRIBUTION AT LOW VELOCITIES

If a pdf exists that leads to an exponentially distributed measured signal Y, it would be the capacity achieving input distribution. Furthermore, the pdf $f_Y(y)$ is the convolution of $f_X(x)$ and $f_N(n)$. We, therefore, attempt to evaluate $f_X(x)$ at asymptotically low velocities by deconvolving the exponential distribution (of the output signal Y) and the IG distribution (of the noise). The Laplace transform of the IG distribution is given by

$$\mathcal{L}(N) = E[e^{-sX}] = e^{\frac{\lambda}{\mu}\left(1-\sqrt{1+\frac{2\mu^2}{\lambda}s}\right)}. \qquad (54)$$

For given values of σ^2 and d, as $v \to 0$ we have $\mu \to \infty$ and λ is fixed. In such a case, $\mathcal{L}(N)$ can be approximated as

$$\mathcal{L}(N) \approx e^{-\sqrt{2\lambda s}}. \qquad (55)$$

As $Y = X + N$, $\mathcal{L}(X) = \mathcal{L}(Y)/\mathcal{L}(N)$. To achieve the upper bound on capacity, $f_Y(y) = \frac{1}{m_Y} e^{\frac{-y}{m_Y}}$, where $m_Y = E[Y] = E[X] + \mu$ and hence

$$\mathcal{L}(Y) = \frac{1/m_Y}{s + (1/m_Y)} \Rightarrow \mathcal{L}(X) = \frac{1/m_Y}{(1/m_Y) + s} e^{\sqrt{2\lambda s}} \qquad (56)$$

and the pdf of X can be obtained by computing the inverse Laplace transform $\mathcal{L}^{-1}(X)$. The inverse Laplace transform can be computed by making use of the following Laplace transform pair [34]:

$$\mathcal{L}^{-1}\left\{\frac{e^{-c\sqrt{s+b}}}{s-a}\right\} = \frac{e^{at}}{2}\left(e^{-c\sqrt{a+b}}\operatorname{erfc}\left(\frac{c}{2\sqrt{t}} - \sqrt{(a+b)t}\right) + e^{c\sqrt{a+b}}\operatorname{erfc}\left(\frac{c}{2\sqrt{t}} + \sqrt{(a+b)t}\right)\right) \qquad (57)$$

where a, b, and c are constants. Using(57), we obtain

$$\mathcal{L}^{-1}\left\{\frac{1/m_Y}{s + (1/m_Y)} e^{\sqrt{2\lambda s}}\right\} = \frac{(1/m_Y)e^{\frac{-1}{m_Y t}}}{2}\left(e^{\jmath\sqrt{2\lambda/m_Y}}\operatorname{erfc}\left(-\sqrt{\lambda/2t} - \jmath\sqrt{t/m_Y}\right) + e^{-\jmath\sqrt{2\lambda/m_Y}}\operatorname{erfc}\left(-\sqrt{\lambda/2t} + \jmath\sqrt{t/m_Y}\right)\right) \qquad (58)$$

where

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-z^2} dz.$$

Note that $\operatorname{erfc}(z)$ can be evaluated for complex values of its argument z and $\operatorname{erfc}(z^*) = (\operatorname{erfc}(z))^*$, where z^* is the complex conjugate of z. Hence

$$f_X(x) = \frac{e^{\frac{-1}{m_Y x}}}{m_Y}\Re\left\{e^{\jmath\sqrt{2\lambda/m_Y}}\operatorname{erfc}\left(-\sqrt{\lambda/2x} - \jmath\sqrt{x/m_Y}\right)\right\} \qquad (59)$$

where $\Re\{z\}$ denotes the real part of z. The pdf obtained previously, unfortunately, is not a valid pdf.

1) When There Is No Drift: To confirm the result in (59), we test the case of zero velocity. Note that in this case, the noise is not IG distributed; however, the zero velocity case converges in limit to the case without drift. Without drift, the pdf of the arrival time is given by [24]

$$f_N(n) = \sqrt{\frac{\lambda}{2\pi n^3}} e^{\frac{-\lambda}{2n}}, \qquad n > 0. \qquad (60)$$

Note that $N \sim \text{Inverse Gamma}(1/2, \lambda/2)$ and

$$\mathcal{L}(N) = \mathcal{L}[\text{Inv Gamma}(1/2, \lambda/2)] = \frac{2(s\lambda/2)^{1/4}}{\sqrt{\pi}} K_{1/2}(\sqrt{2\lambda s}). \qquad (61)$$

Substituting $K_{1/2}(z) = \sqrt{\frac{\pi}{2z}} e^{-z}$, we get

$$\mathcal{L}(N) = e^{-\sqrt{2\lambda s}}. \qquad (62)$$

This results in

$$\mathcal{L}(X) = \frac{1/m_Y}{s + (1/m_Y)} e^{\sqrt{2\lambda s}}. \qquad (63)$$

Note that (62) is the same as (55), and (63) is the same as (56). Hence, we get (59) when we try to obtain $f_X(x)$ by evaluating $\mathcal{L}^{-1}(X)$.

APPENDIX C
ESTIMATING NOISE PARAMETERS

To estimate the noise parameters, the transmitter releases k "training" molecules at known time x_0. Let the receiver observe $Y_j = x_0 + N_j, j = 1, 2, \ldots, k$, where $N_j \sim \text{IG}(\mu, \lambda)$ are i.i.d. and the receiver knows x_0 *a priori*. The pdfs of $(Y_j - x_0), j = 1, \ldots, k$, are i.i.d. and IG distributed as given by

$$f_{Y_j - x_0}(y_j - x_0) = \sqrt{\frac{\lambda}{2\pi(y_j - x_0)^3}} \exp\left(-\frac{\lambda}{2\mu^2}\frac{((y_j - x_0) - \mu)^2}{(y_j - x_0)}\right) y_j > x_0. \qquad (64)$$

In general, $-\infty < x_0 < \infty$; however, in our case, $0 < x_0 < \infty$. When $Y \sim \text{IG}(x_0, \mu, \lambda)$, $m_Y = E[Y] = \mu + x_0$. When the receiver knows the value of t_0, the ML estimates of the remaining two parameters μ and λ can be obtained as

$$\hat{\mu}(x_0) = \bar{Y} - x_0, \qquad (65)$$

where $\bar{Y} = \frac{1}{k}\sum_{j=1}^k Y_j$ is the sample mean and

$$\hat{\lambda}(x_0) = \left[\frac{1}{k}\sum_{j=1}^k\left(\frac{1}{Y_j - x_0} - \frac{1}{\bar{Y} - x_0}\right)\right]^{-1}. \qquad (66)$$

Assuming μ and λ does not change significantly from the time the receiver estimates the parameters and the time of actual communication, the receiver can obtain the ML estimate of the release times of the molecules.

APPENDIX D
UPPER BOUND ON ASYMPTOTIC ERROR RATE

Here, we prove Theorem 4. Recall that, for 2-ary modulation with $X \in \{x_1, x_2\}, 0 \leq x_1 < x_2$, the upper bound on SEP is given by

$$P_e < p_1(1 - F_N(x_2 - x_1)) \tag{67}$$

where

$$F_N(n) = \Phi\left(\sqrt{\frac{\lambda}{n}}\left(\frac{n}{\mu} - 1\right)\right) + e^{2\lambda/\mu}\Phi\left(-\sqrt{\frac{\lambda}{n}}\left(\frac{n}{\mu} + 1\right)\right) \tag{68}$$

where $\Phi(z) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$ is the cdf of a standard Gaussian random variable Z and

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-u^2} du. \tag{69}$$

For $z \gg 1$, $\text{erf}(z)$ can be approximated as

$$\text{erf}(z) \approx 1 - \frac{e^{-z^2}}{\sqrt{\pi}z}. \tag{70}$$

Now, we compute $F_N(c), c = x_2 - x_1$, and examine its behavior as $v \to \infty$. Recall that $\mu = \frac{l}{v}$ and $\lambda = \frac{l^2}{\sigma^2}$

$$F_N(c) = \Phi\left(\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right) + e^{2vl/\sigma^2}\Phi\left(-\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right). \tag{71}$$

Consider the first term in $F_N(c)$

$$\Phi\left(\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}}{\sqrt{2}}\right)\right). \tag{72}$$

When $v \to \infty$, we have $\sqrt{\frac{cv^2}{\sigma^2}} \to \infty$ and thus $\left(\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right) \gg 1$. Hence, we use (70) to obtain

$$\Phi\left(\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right) \approx 1 - \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}}e^{\left(-\frac{cv^2}{2\sigma^2} + \frac{vl}{\sigma^2} - \frac{l^2}{2c\sigma^2}\right)}. \tag{73}$$

Now, consider the second term in $F_N(c)$

$$\Phi\left(-\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right) = \frac{1}{2}\left(1 - \text{erf}\left(\frac{\sqrt{\frac{cv^2}{\sigma^2}} + \sqrt{\frac{l^2}{c\sigma^2}}}{\sqrt{2}}\right)\right). \tag{74}$$

When $v \to \infty$ we have $\left(\sqrt{\frac{cv^2}{\sigma^2}} + \sqrt{\frac{l^2}{c\sigma^2}}\right) \gg 1$ and, using (70), we obtain

$$e^{2vl/\sigma^2}\Phi\left(-\sqrt{\frac{cv^2}{\sigma^2}} - \sqrt{\frac{l^2}{c\sigma^2}}\right) \approx \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\frac{cv^2}{\sigma^2}} + \sqrt{\frac{l^2}{c\sigma^2}}}e^{\left(-\frac{cv^2}{2\sigma^2} + \frac{vl}{\sigma^2} - \frac{l^2}{2c\sigma^2}\right)}. \tag{75}$$

Thus

$$F_N(c) \approx 1 - \frac{1}{\sqrt{2\pi}}e^{\left(-\frac{cv^2}{\sigma^2} + \frac{vl}{\sigma^2} - \frac{l^2}{c\sigma^2}\right)}\left(\frac{2l^2}{c^2v^2 - l^2}\right). \tag{76}$$

At high velocities, $F_N(c)$ can be approximated as

$$F_N(c) \approx 1 - \sqrt{\frac{2}{\pi}}\frac{l^2}{c^2v^2}e^{\left(-\frac{cv^2}{\sigma^2} + \frac{vl}{\sigma^2}\right)}, \tag{77}$$

and hence, the upper bound on SEP can be approximated by

$$p_1\left(\sqrt{\frac{2}{\pi}}\frac{l^2}{c^2v^2}e^{\left(-\frac{cv^2}{\sigma^2} + \frac{vl}{\sigma^2}\right)}\right). \tag{78}$$

The theorem follows by taking the logarithm of this expression.

REFERENCES

[1] S. F. Bush, *Nanoscale Communication Networks*. Boston: Artech House, 2010.

[2] S. Hiyama *et al.*, "Molecular communication," in *Proc. 2005 Nano Sci. and Technol. Institute Conf.*, 2005, pp. 391–394.

[3] T. Nakano, T. Suda, M. Moore, R. Egashira, A. Enomoto, and K. Arima, "Molecular communication for nanomachines using intercellular calcium signaling," in *Proc. 5th IEEE Conf. Nanotechnol.*, Jul. 2005, pp. 478–481.

[4] Y. Moritani, S. M. Nomura, S. Hiyama, K. Akiyoshi, and T. Suda, "A molecular communication interface using liposomes with gap junction proteins," in *Proc. 1st Int. Conf. Bio-Inspired Models Netw., Inf. Comput. Syst.*,, Dec. 2006.

[5] M. Moore, A. Enomoto, T. Nakano, R. Egashira, T. Suda, A. Kayasuga, H. Kojima, H. Sakakibara, and K. Oiwa, "A design of a molecular communication system for nanomachines using molecular motors," in *Proc. 4th Annu. IEEE Conf. Pervasive Comput. Commun. Workshops*, Mar. 2006, p. 554.

[6] S. Hiyama, Y. Moritani, and T. Suda, "A biochemically engineered molecular communication system," presented at the presented at the 3rd Int. Conf. Nano-Networks, Boston, MA, 2008.

[7] S. Hiyama and Y. Moritani, "Molecular communication: Harnessing biochemical materials to engineer biomimetic communication systems," *Nano Commun. Netw.*, vol. 1, pp. 20–30, Mar. 2010.

[8] A. W. Eckford, "Molecular communication: Physically realistic models and achievable information rates," arXiv:0812.1554v1 [cs.IT] 8 December 2008.

[9] S. Kadloor and R. S. Adve, "Development of a framework to study a molecular communication system," in *Proc. 18th Int. Conf. Comput. Commun. Netw.*, 2009.

[10] S. Kadloor, R. S. Adve, and A. W. Eckford, "Molecular communication using Brownian motion with drift," *IEEE Trans. Nanobiosci.*, to be published.

[11] M. J. Moore, T. Suda, and K. Oiwa, "Molecular communication: Modeling noise effects on information rate," *IEEE Trans. Nanobiosci.*, vol. 8, no. 2, pp. 169–179, Jun. 2009.

[12] B. Atakan and O. Akan, "An information theoretical approach for molecular communication," in *Proc. 2nd Int. Conf. Bio-Inspired Models Netw., Inf., Comput. Syst.*, Budapest, Hungary, 2007, pp. 33–40.

[13] M. Pierobon and I. F. Akyildiz, "A physical end-to-end model for molecular communication in nanonetworks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 4, pp. 602–611, May 2010.

[14] P. Thomas, D. Spencer, S. Hampton, P. Park, and J. Zurkus, "The diffusion mediated biochemical signal relay channel," *Adv. Neural Inf. Process. Syst.*, vol. 16, pp. 1263–1270, 2004.

[15] B. Atakan and O. B. Akan, "Single and multiple-access channel capacity in molecular nanonetworks," presented at the presented at the 4th Int. Conf. Nano-Networks, Luzern, Switzerland, 2009.

[16] T. Nakano and J.-Q. Liu, "Design and analysis of molecular relay channels: An information theoretic approach," *IEEE Trans. Nanobiosci.*, vol. 9, no. 3, pp. 213–221, Sep. 2010.

[17] D. Blackwell, *Information Theory,* 1961, pp. 183–193, Modern Mathematics for the Engineer (Second Series).

[18] H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2008.

[19] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.

[20] R. Sundaresan and S. Verdú, "Capacity of queues via point-process channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2697–2709, Jun. 2006.

[21] I. Karatzas and S. E. Shreve*, Brownian Motion and Stochastic Calculus*, 2nd ed. New York: Springer-Verlag, 1991.

[22] S. Goldstein, "Mechanical models of Brownian motion," *Lecture Notes Phys.*, vol. 153, pp. 21–24, 1982.

[23] J. Berthier*, Microfluidics for Biotechnology*. Boston: Artech House, 2006.

[24] R. S. Chhikara and J. L. Folks*, The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. New York: Marcel Dekker, 1989.

[25] Y. A. Brychkov*, Handbook of Special Functions: Derivatives, Integrals, Series and Other Formulas*. Boca Raton, FL: Chapman Hall/CRC Press, 2008.

[26] T. M. Cover and J. A. Thomas*, Elements of Information Theory*. , NJ: Wiley, 2006.

[27] V. Seshadri*, The Inverse Gaussian Distribution: Statistical Theory and Applications*. New York: Springer-Verlag, 1999.

[28] W. Schwarz, "On the convolution of inverse Gaussian and exponential random variables," *Commun. Statistics: Theory Methods*, vol. 31, pp. 2113–2121, Dec. 2002.

[29] J. G. Proakis and M. Salehi*, Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[30] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4508–4528, Oct. 2009.

[31] S. Mukhtar and J. Bruck, "Interval modulation coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2002, p. 327.

[32] S. Hranilovic and F. Kschischang, "Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 784–795, May 2004.

[33] T. Kawamura and K. Iwase, "Characterizations of the distributions of power inverse Gaussian and others based on the entropy maximization principle," *J. Jpn. Statist. Soc.*, vol. 33, pp. 95–104, 2003.

[34] R. B. Hetnarski, "An algorithm for generating inverse Laplace transforms of exponential form," *J. Appl. Math. Phys.*, vol. 26, pp. 249–253, Mar. 1975.

K. V. Srinivas was born in Vijayawada, India. He received the B.E. degree in electronics and communications engineering from the Andhra University College of Engineering, Vishakhapatnam, India, in June 1996, the M.Tech. degree from the Indian Institute of Technology, Kanpur, India, in 1998, and the Ph.D degree from the Indian Institute of Technology Madras, Chennai, India, in 2009, both in electrical engineering.

Currently, he is a Chief Engineer in the Mobile R&D Division at Samsung Electronics, Noida, India. He was a Postdoctoral Fellow at the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, from March 2009 to October 2011. He was with Space Applications Centre, Indian Space Research Organization, Ahmedabad, India, as a Scientist, from October 1999 to September 2002. His research interests include wireless communications, with emphasis on physical layer algorithms, and the theoretical aspects of nanocommunication networks.

Andrew W. Eckford (M'96–S'97–M'04) is originally from Edmonton, AB, Canada. He received the B.Eng. degree in electrical engineering from the Royal Military College of Canada, Kingston, ON, Canada, in 1996, and the M.A.Sc. as well as Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1999 and 2004, respectively. In 2006, Dr. Eckford joined the faculty of the Department of Computer Science and Engineering, York University, Toronto, ON, Canada, and was promoted to his current rank of Associate Professor in 2010.

Raviraj S. Adve (S'88–M'97–SM'06) was born in Bombay, India. He received the B.Tech. degree in electrical engineering from IIT, Bombay, in 1990, and the Ph.D. degree from Syracuse University in 1996. Between 1997 and August 2000, he worked for Research Associates for Defense Conversion Inc. on contract with the Air Force Research Laboratory at Rome, NY. He joined the faculty at the University of Toronto in August 2000, where he is currently an Associate Professor. Dr. Adve's research interests include practical signal processing algorithms for multiple-input multiple-output wireless communications and distributed radar systems. In the area of wireless communications, he is currently focusing on precoding in various scenarios and cooperation in distributed wireless networks. In radar systems, he is particularly interested in waveform diversity and low-complexity space-time adaptive processing algorithms. He received the 2009 Fred Nathanson Young Radar Engineer of the Year award.