On the Capacity-achieving Input for Additive Inverse Gaussian Channels

Hui Li
Dept. of Electronic Engineering and Information Science
University of Science and Technology of China
Hefei, Anhui 230027, China

Dongning Guo
Dept. of Electrical Engineering and Computer Science
Northwestern University
Evanston, IL 60208, USA

***Abstract*—In a molecular communication system, molecules convey the information by traversing from the transmitter to the receiver through the medium, which is often liquid. The time for a molecule to travel a fixed distance according to Brownian motion with a constant drift has the inverse Gaussian distribution. Hence the molecular communication channel is modeled by an additive inverse Gaussian noise channel, the input of which is the release times of the molecules. This paper studies the capacity-achieving input distribution for such a channel, where the release time is subject to both peak and average constraints. Several properties of the capacity-achieving input are established. A numerical method for computing the optimal input distribution is developed. The result complements some existing bounds on the capacity of molecular channel.**

I. Introduction

In one form of molecular communication, information is modulated in the manner that the transmitter releases molecules, which then propagate through the fluid medium, until the receiver captures the molecules and decodes the information [1]. One scenario of molecular communication is between nanoscale devices in blood vessels.

A molecule released at time X is received at time $X+N$, where N is the random propagation time, known as the first arrival time. In a one-dimensional fluid medium, the first arrival time has the inverse Gaussian distribution. An additive inverse Gaussian noise (AIGN) channel was introduced in [1] to describe the molecular communication with release timing encoder. The capacity of the AIGN channel and what distribution achieves the capacity are open questions. Upper and lower bounds on the capacity of the AIGN channel have been developed in [1]. In a more recent paper [2], a close-form upper bound on the capacity of peak constrained AIGN channel is also provided. Another lower bound on the capacity of AIGN channel is given in [3].

In this paper, we investigate the capacity-achieving input of the AIGN channel with both average and peak release time constraints. Several properties of the capacity-achieving input distribution are established, including a sufficient and necessary condition for the optimal distribution obtained using the method of Lagrange multipliers. Numerical results strongly suggest a mixture of a continuous distribution and two point masses at “0” and the peak, respectively. This is in contrast to the optimal input for the additive white Gaussian noise channel with both average power and peak power constraints [4] and the optimal inputs for a number of other channels [5]–[11], where the capacity-achievable input is discrete. The optimized achievable rates are consistent with existing upper and lower bounds [1], [3]. To the best of our knowledge, this work is the first recorded attempt to compute the capacity-achieving input distribution for the AIGN channel.

[1]This material is based on work supported in part by the NSFC under Grant No. 41074055 and in part by the National Science Foundation under Grant No. 1231828.

II. System Model

Consider a diffusion-mediated molecular communication system which encodes a message in the release times of molecules. Suppose the molecular communication system is in a one-dimensional medium and the distance between the transmitter and the receiver is d. A molecule released by the transmitter at time $X \geq 0$ is captured by the receiver at time Y, which is expressed as

$$Y = X + N \tag{1}$$

where the positive random variable N is called is the *first arrival time*. The movement of a molecule in a fluid medium can be described as Brownian motion with drift velocity v and infinitesimal variance σ^2, i.e., the displacement of the molecule at time t is

$$S_t = vt + \sigma W_t \tag{2}$$

where W_t is a Wiener process, i.e., a Gaussian process with independent increments and $W_0 = 0$, $W_t - W_s \sim \mathcal{N}(0, t-s)$ for all $t \geq s$.

Then the first arrival time has an inverse Gaussian distribution, whose probability density function (pdf) is

$$f(x) = \sqrt{\frac{\lambda}{2\pi}} x^{-\frac{3}{2}} \exp\left(-\frac{\lambda(x-c)^2}{2c^2 x}\right), \quad x > 0 \tag{3}$$

where c denote the average travel time

$$c = \frac{d}{v}, \tag{4}$$

and λ relates to the Brownian motion parameter σ^2 as

$$\lambda = \frac{d^2}{\sigma^2}. \tag{5}$$

In this paper, we assume that the molecular communication system is ideal in the following sense:

1) The transmitter and the receiver have a common clock;
2) The transmitter perfectly controls the release times of the molecules;
3) The receiver perfectly measures the first arrival time of the molecules;
4) Upon first arrival at the receiver, molecules are captured and removed from the system;
5) The movement of every molecule is independent;
6) Every molecule can be distinguished from the others, so that the arrival order of the molecules is known at the receiver.

III. The Channel Capacity

The conditional pdf of output Y of the AIGN channel given input X is

$$p_{Y|X}(y|x) = f(y-x). \tag{6}$$

Let μ denote the distribution of channel input X. In general μ is a probability measure defined on the Borel algebra on the non-negative real number set, denoted by $\mathcal{B}[0,+\infty)$. The pdf of the channel output exists and is

$$p_Y(y;\mu) = \int p_{Y|X}(y|x)\mu(\mathrm{d}x) = \mathsf{E}_\mu\left\{f(y-X)\right\}. \tag{7}$$

The mutual information per molecule use is

$$I(\mu) = \int\int_0^\infty p_{Y|X}(y|x)\left(\log\frac{p_{Y|X}(y|x)}{p_Y(y;\mu)}\right)\mathrm{d}y\,\mu(\mathrm{d}x). \tag{8}$$

Let $h(Y|X=x)$ denote the conditional differential entropy of Y given $X=x$, then

$$h(Y|X=x) = h(N+X|X=x) = h(N) = h(f) \tag{9}$$

where $h(f)$ is the differential entropy of pdf $f(x)$, which can be calculated as [3]

$$h(f) = \frac{1}{2}\log\frac{2\pi ec^3}{\lambda} + \frac{3}{2}e^{\frac{2\lambda}{c}}\,\mathrm{Ei}\left(-\frac{2\lambda}{c}\right). \tag{10}$$

where $\mathrm{Ei}(\cdot)$ denotes the exponential integral function

$$\mathrm{Ei}(-t) = -\int_t^\infty \frac{e^{-\tau}}{\tau}d\tau. \tag{11}$$

Since $h(Y|X=x)$ does not depend on x

$$h(Y|X) = \mathsf{E}_\mu\left\{h(Y|X=x)\right\} = h(f). \tag{12}$$

Define $Q(x;\mu)$ as

$$Q(x;\mu) = \int_0^\infty p_{Y|X}(y|x)\log\frac{1}{p_Y(y;\mu)}\,\mathrm{d}y. \tag{13}$$

Then the differential entropy of Y given the input distribution μ is

$$h(Y;\mu) = \int p_Y(y;\mu)\log\frac{1}{p_Y(y;\mu)}\,\mathrm{d}y \tag{14}$$

$$= \int Q(x;\mu)\mu(\mathrm{d}x). \tag{15}$$

Then the mutual information can be expressed as

$$I(\mu) = h(Y;\mu) - h(f) \tag{16}$$

$$= \int Q(x;\mu)\mu(\mathrm{d}x) - h(f). \tag{17}$$

Suppose the molecules can only be released in a limit time interval $[0,W]$, where W is the peak release time constraint, and the released molecules must be observed at the receiver in time $(0,\infty)$. Then we denote the set of distribution with the peak and average release time constraint W and M by

$$\Lambda(W,M) = \{\mu : \mu[0,W] = 1, \mathsf{E}_\mu\left\{X\right\} \le M\}. \tag{18}$$

Without loss of generality, it is assumed that $W \ge M$, because if $W < M$, then the average release time constraint is nonbinding, and it is inconsequential to decrease the average time constraint M to W. Then the capacity of AIGN channel with average and peak release time constraints is

$$C(W,M) = \max_{\mu\in\Lambda(W,M)} I(\mu). \tag{19}$$

IV. The Optimal Release Time Distribution

A. The existence and uniqueness of the optimal distribution

Since [0,W] is a compact space, $\Lambda(W,M)$ defined on $[0,W]$ is tight [12]. According to Prokhorov's theorem the tight set $\Lambda(W,M)$ is compact [12]. Since the inverse Gaussian noise has a finite variance, mutual information $I(\mu)$ is continuous [13] and it must achieve its maximum on compact set $\Lambda(W,M)$. Hence the maximizer μ_0 of (19) exists.

In order to prove the uniqueness of the capacity-achieving distribution, we first prove the strict convexity of $I(\mu)$.

Theorem 1: The input–output mutual information $I(\mu)$ of the AIGN channel is strictly convex in μ.

Proof: For any μ_0 and μ_1 in $\Lambda(W,M)$, let $\mu_\theta = (1-\theta)\mu_0 + \theta\mu_1$, $\theta\in(0,1)$. Then

$$p_Y(y;\mu_\theta) = (1-\theta)p_Y(y;\mu_0) + \theta p_Y(y;\mu_1). \tag{20}$$

Then

$$I(\mu_\theta) - (1-\theta)I(\mu_0) - \theta I(\mu_1)$$

$$= h(Y;\mu_\theta) - (1-\theta)h(Y;\mu_1) - \theta h(Y;\mu_2) \tag{21}$$

$$= -\int p_Y(y;\mu_\theta)\log p_Y(y;\mu_\theta)\mathrm{d}y + (1-\theta)\int p_Y(y;\mu_1)\log p_Y(y;\mu_1)\mathrm{d}y + \theta\int p_Y(y;\mu_2)\log p_Y(y;\mu_2)\mathrm{d}y \tag{22}$$

$$= (1-\theta)D(p_Y(\cdot;\mu_1)\|p_Y(\cdot;\mu_\theta)) + \theta D(p_Y(\cdot;\mu_2)\|p_Y(\cdot;\mu_\theta)). \tag{23}$$

Using properties of the relative entropy, we have

$$I(\mu_\theta) \ge (1-\theta)I(\mu_0) + \theta I(\mu_1) \tag{24}$$

with equality if and only if $p_Y(y;\mu_1) = p_Y(y;\mu_0) = p_Y(y;\mu_\theta)$. Hence the convexity of $I(\mu)$ is established.

To see the strict convexity, we show that if μ_0, μ_1 both achieve the maximum, then μ_0 and μ_1 are identical. It is enough to show that $p_Y(y;\mu_0) = p_Y(y;\mu_1)$ for all y implies that $\mu_0 = \mu_1$. Let $C_f(t)$ be the characteristic function of $f(x)$ and $C_\mu(t)$ be the characteristic function of distribution μ. It can be shown [14] that

$$C_f(t) = \exp\left\{\frac{\lambda}{c}\left(1 - \left(1 - 2ic^2\lambda^{-1}t\right)^{\frac{1}{2}}\right)\right\}, \quad (25)$$

which is nonzero for all t. If $p_Y(y;\mu_1) = p_Y(y;\mu_2)$, then

$$C_f(t)C_{\mu_1}(t) = C_f(t)C_{\mu_2}(t), \quad (26)$$

so that $C_{\mu_1}(t) = C_{\mu_2}(t)$. This requires that $\mu_1 = \mu_2 = \mu_\theta$ by bijection between distribution functions and characteristic functions. ■

Corollary 1: The maximum of (19) is achieved by a unique $\mu_0 \in \Lambda(W, M)$. Then the optimal input distribution must have probability mass arbitrarily close to 0, i.e., for any $\varepsilon > 0$,

$$\mu_0[0, \varepsilon) > 0. \quad (27)$$

Proof: By the strict convexity of $I(\mu)$ established in Theorem 1, it has a unique maximizer on compact set $\Lambda(W, M)$.

We next show (27) must hold. Suppose, to the contrary, the optimal input is lower bounded by $\varepsilon > 0$, i.e., $\mu_0[0, \varepsilon) = 0$. Consider a different input, $X_1 = X - \varepsilon$. Evidently, $I(\mu_1) = I(X_1; X_1 + N) = I(X; X + N) = I(\mu_0)$. Since μ_1 also satisfies the peak and average release time constraints, this contradicts the uniqueness of μ_0. ■

Define $\Lambda_0(W, M)$ as

$$\begin{aligned}\Lambda_0(W, M) = \{\mu : \mu[0, W] = 1, \mu[0, \varepsilon) > 0, \forall \varepsilon > 0,\\ \mathsf{E}_\mu\{X\} \le M\},\end{aligned} \quad (28)$$

Since $\Lambda_0(W, M)$ is a convex set, the capacity of AIGN channel with average and peak release time constraints can be achieved in $\Lambda_0(W, M)$,

$$C(W, M) = \max_{\mu \in \Lambda_0(W,M)} I(\mu). \quad (29)$$

B. Calculation of μ_0

In order to find the capacity-achievable input μ_0, we first prove that the achievable rate $I(\mu)$ is weakly differentiable on $\Lambda_0(W)$ (Lemma 1), where $\Lambda_0(W)$ is defined as

$$\Lambda_0(W) = \{\mu : \mu[0, W] = 1, \mu[0, \varepsilon) > 0, \forall \varepsilon > 0\}. \quad (30)$$

Then we can apply the method of Lagrangian multiplier to solve the nonlinear optimization problem (29). We also develop a sufficient and necessary condition the optimal input must satisfy (Theorem 2), which is then used to verify the validity of the optimal input obtained by numerical methods.

Lemma 1: The mutual information function $I(\mu)$ is weakly differentiable in $\Lambda_0(W)$. The weak derivative $I'_{\mu_0}(\mu)$, defined as

$$I'_{\mu_0}(\mu) = \lim_{\theta \to 0^+} \frac{I((1-\theta)\mu_0 + \theta\mu) - I(\mu_0)}{\theta} \quad (31)$$

can be expressed as

$$I'_{\mu_0}(\mu) = \int Q(x;\mu_0)\mu(\mathrm{d}x) - h(f) - I(\mu_0). \quad (32)$$

Proof: Let $\mu_\theta = (1-\theta)\mu_0 + \theta\mu$. Using (16), we have

$$\begin{aligned}
&I(\mu_\theta) - I(\mu_0)\\
&= h(Y;\mu_\theta) - h(Y;\mu_0) && (33)\\
&= \int p_Y(y;\mu_\theta)\log\frac{1}{p_Y(y;\mu_\theta)}\,\mathrm{d}y\\
&\quad - \int p_Y(y;\mu_0)\log\frac{1}{p_Y(y;\mu_0)}\,\mathrm{d}y && (34)\\
&= -D(Y_{\mu_\theta}\|Y_{\mu_0})\\
&\quad + \theta\int (p_Y(y;\mu_0) - p_Y(y;\mu))\log p_Y(y;\mu_0)\,\mathrm{d}y. && (35)
\end{aligned}$$

Based on the Lemma [15, page 1023], for any pair of probability measures $P_1 \ll P_0$, we have

$$\lim_{\theta \downarrow 0}\frac{1}{\theta}D(\theta P_1 + (1-\theta)P_0\|P_0) = 0. \quad (36)$$

Since for every pair of probability measures μ_0 and μ in $\Lambda_0(W)$, $P_Y(y;\mu) \ll P_Y(y;\mu_0)$, then

$$\begin{aligned}
I'_{\mu_0}(\mu) &= \lim_{\theta\to 0^+} -\frac{1}{\theta}D(Y_{\mu_\theta}\|Y_{\mu_0})\\
&\quad + \int (p_Y(y;\mu_0) - p_Y(y;\mu))\log p_Y(y;\mu_0)\,\mathrm{d}y && (37)\\
&= \int p_Y(y;\mu)\log\frac{1}{p_Y(y;\mu_0)}\mathrm{d}y - I(\mu_0) - h(f) && (38)\\
&= \int Q(x;\mu_0)\mu(\mathrm{d}x) - I(\mu_0) - h(f). && (39)
\end{aligned}$$

Lemma 1 is thus established. ■

Definition 1: A point x is said to be a *point of increase* of μ if for any open subset $\mathbb{O}$ containing x, $\mu(\mathbb{O}) > 0$.

Theorem 2: Let μ_0 be the admissible input probability measure, i.e., $\mu_0 \in \Lambda_0(W, M)$. Then μ_0 is capacity achieving if and only if there exists $\nu \ge 0$ such that for every $x \in [0, W]$

$$Q(x;\mu_0) - h(f) - I(\mu_0) - \nu(x - M) \le 0. \quad (40)$$

Furthermore, the equality of (40) is satisfied by all points of increase of μ_0.

Proof: Define Lagrangian

$$J(\mu) = I(\mu) - \nu\mathsf{E}_\mu\{X - M\} \quad (41)$$

where ν is the Lagrangian multiplier. By the method of Lagrangian multiplier, μ_0 is capacity achieving if and only if μ_0 and ν are such that

a) $\nu\mathsf{E}_{\mu_0}\{X - M\} = 0$,
b) for all $\mu \in \Lambda_0(W)$, $J(\mu_0) \ge J(\mu)$.

Since $I(\mu)$ is a strictly concave function on a convex set $\Lambda_0(W)$, $J(\mu)$ is also concave function and the maximizer of (29) must exist and be unique. Condition (b) is equivalent to that $J'_{\mu_0}(\mu) \le 0$ where $J'_{\mu_0}(\mu)$ is the weak derivative of $J(\mu)$ at μ_0, defined as

$$\begin{aligned}
J'_{\mu_0}(\mu) &= \lim_{\theta\to 0^+}\frac{J((1-\theta)\mu_0 + \theta\mu) - J(\mu_0)}{\theta} && (42)\\
&= I'_{\mu_0}(\mu) - \nu(\mathsf{E}_\mu\{X\} - \mathsf{E}_{\mu_0}\{X\}), && (43)
\end{aligned}$$

where $I'_{\mu_0}(\mu)$ is defined in (32). Then

$$J'_{\mu_0}(\mu) = \int Q(x;\mu_0)\mu(\mathrm{d}x) - h(f) - I(\mu_0) \\ - \nu(\mathsf{E}_\mu\{X\} - \mathsf{E}_{\mu_0}\{X\}). \quad (44)$$

We first prove the necessity of (40). Suppose μ_0 achieves the capacity. For any given x^* in $[0, W]$, let μ be the point measure at x^*, i.e., $\mu(\{x^*\}) = 1$, which clearly satisfies $\mu \in \Lambda_0(W)$. Substituting μ into (44), we have, by the optimality of μ_0,

$$\begin{aligned} 0 &\geq J'_{\mu_0}(\mu) && (45)\\ &= Q(x^*;\mu_0) - h(f) - I(\mu_0) \\ &\quad - \nu(x^* - M) + \nu\mathsf{E}_{\mu_0}\{X - M\} && (46)\\ &= Q(x^*;\mu_0) - h(f) - I(\mu_0) - \nu(x^* - M). && (47)\end{aligned}$$

which establishes (40).

We next prove the sufficiency of (40). Suppose inequality (40) is satisfied, we can integrate both side of (40) with respect to μ_0 to write

$$\begin{aligned} 0 &\geq \int Q(x;\mu_0)\mu_0(\mathrm{d}x) - h(f) - I(\mu_0) - \nu\mathsf{E}_{\mu_0}\{X - M\} && (48)\\ &= -\nu\mathsf{E}_{\mu_0}\{X - M\} && (49)\\ &\geq 0 && (50)\end{aligned}$$

where (49) is by (17) and (50) is because $\mu_0 \in \Lambda_0(W, M)$. Hence $\nu\mathsf{E}_{\mu_0}\{X - M\} = 0$. On the other hand, if we integrate (40) respect to an arbitrary measure $\mu \in \Lambda_0(W)$, we have

$$\begin{aligned} 0 &\geq \int Q(x;\mu_0)\mu(\mathrm{d}x) - h(f) - I(\mu_0) \\ &\quad - \nu\int(x - M)\mu(\mathrm{d}x) && (51)\\ &= I'_{\mu_0}(\mu) - \nu\mathsf{E}_\mu\{X - M\} && (52)\\ &= I'_{\mu_0}(\mu) - \nu(\mathsf{E}_\mu\{X\} - \mathsf{E}_{\mu_0}\{X\}) && (53)\\ &= J'_{\mu_0}(\mu) && (54)\end{aligned}$$

where (52) is by (39). Condition (a) and (b) are thus both satisfied. Hence μ_0 achieves the capacity.

Finally, we prove that inequality (40) is satisfied with equality at all points of increase. Suppose, on the contrary, the strict inequality is satisfied at x^*, which is a point of increase. Define $a(x;\mu_0)$ as

$$a(x;\mu_0) = Q(x;\mu_0) - h(f) - I(\mu_0) - \nu(x - M). \quad (55)$$

Then $a(x^*;\mu_0) = -\varepsilon$ for some $\varepsilon > 0$. Since $p_Y(y;\mu_0) > 0$ for $y > 0$ and $f(y - x)$ is continuous function for $y \geq x$, $Q(x;\mu_0) = \int f(y-x)\log p_Y(y;\mu_0)\mathrm{d}y$ is continuous on $[0, W]$, $a(x;\mu_0)$ is also continuous on $[0, W]$. Hence there is an open subset $\mathbb{O}$ containing x^* such that $a(x,\mu_0) < -\varepsilon/2$ for all $x \in \mathbb{O}$. Consequently, if we integrate both side of (40) with respect to μ_0, we have

$$0 = \int a(x;\mu_0)\mu_0(\mathrm{d}x) \leq \int_{\mathbb{O}} a(x;\mu_0)\mu_0(\mathrm{d}x) < -\frac{1}{2}\varepsilon\mu_0(\mathbb{O}) \quad (56)$$

Contradiction arises since $\varepsilon\mu_0(\mathbb{O}) > 0$, hence Theorem 2 is proved. ∎

In [4], Smith proved that only a discrete distribution can achieve the capacity of additive white Gaussian noise channel with amplitude and variance constraints. The discreteness of the capacity-achieving input hinges on the fact that function $Q(x;\mu_0)$ defined in (13) analytically extends to the entire complex plane in the setting of [4]. However, for the AIGN channel in this paper, it can be proved that $Q(z;\mu_0)$ is not analytic at $z = 0$. As demonstrated in Section V, the resulting optimal input for the AIGN channel appears to be a mixed discrete and continuous distribution.

V. Numerical Method

From Theorem 2, we can calculate the capacity-achievable input μ_0 through the method of Lagrangian multiplier on the nonlinear program problem (29). Since directly search the optimal input on an infinite dimensional functional space is difficult, in this paper the following discrete approximation of (19) is solved to estimate the optimal input:

$$\underset{P_X}{\text{maximize}} \quad I(X;Y), \quad (57)$$

$$\begin{aligned} \text{subject to} \quad & \mathsf{E}\{X\} \leq M, \\ & X \in S = \{x_1, \cdots, x_n\} \quad (58)\\ & \text{where } 0 = x_1 < \cdots < x_n = W, \end{aligned}$$

where S is the alphabet set of the input distribution. Evidently the set of distribution P_X on S with average constraints (58) is a convex set. The mutual information $I(X;Y)$ is concave in P_X for fixed $P_{Y|X}$. The average constraints is linear. Therefore the maximization problem in (57) is a general convex programming problem, we solve this problem by interior point method. Let $p_i = P_X(x_i)$, then the output distribution is

$$p_Y(y) = \sum_{i=1}^{n} p_i f(y - x_i). \quad (59)$$

Then we can solve (57) by maximizing

$$h(Y) = \int p_Y(y)\log\frac{1}{p_Y(y)}\mathrm{d}y \quad (60)$$

under the average release time constraint

$$\sum_{i=1}^{n} p_i x_i - M \leq 0. \quad (61)$$

Fig. 1 is the optimized achievable rate as a function of average release time constraint M, where $c = 1$, $\lambda = 1$, $W = 20$ and $x_n = 0.1n$, $n = 0, \cdots, 200$. The optimized achievable rate is consistent with the upper bound and lower bounds provided in [1] and [3]. Evidently, the achievable rate given in this paper is close to the capacity.

Fig. 2 is the optimized achievable rate as a function of peak release time constraint W, where $c = 1$, $\lambda = 1$ and $M = 1$. With the increasing of peak release time W, the optimized achievable rate also increase.

Fig. 3 is the optimized input distribution (top) and the corresponding $a(x;\mu_0)$ (bottom) defined in (55), where $c = 1$,

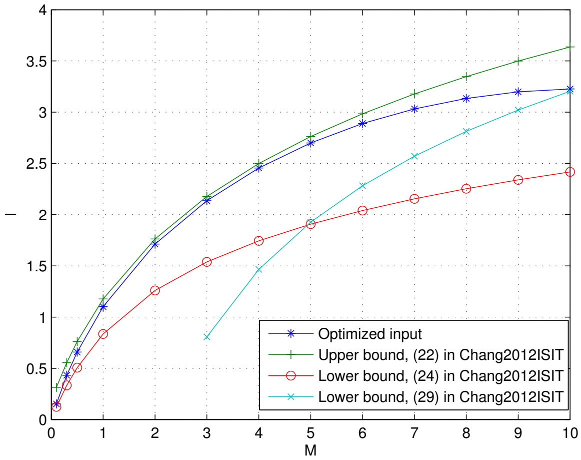

Fig. 1. Optimized achievable rate (bits per molecule use) versus average release time M with average travel time $c = 1$, $\lambda = 1$ and peak release time $W = 20$.

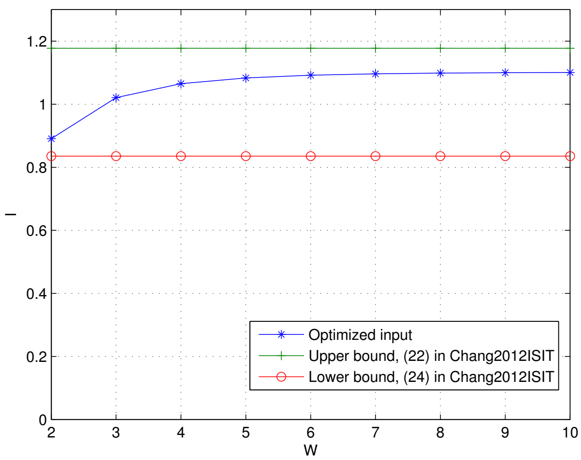

Fig. 2. Optimized achievable rate (bits per molecule use) versus peak release time W with average travel time $c = 1$, $\lambda = 1$ and average release time $M = 1$.

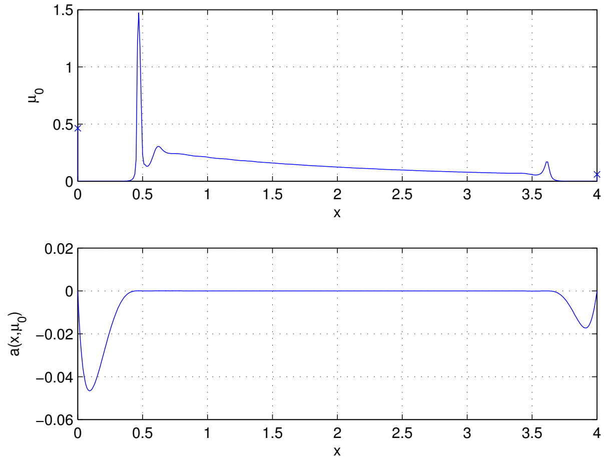

Fig. 3. Optimized distribution (top) and $a(x; \mu_0)$ (bottom) with the average travel time $c = 1$, $\lambda = 1$, average release time $M = 1$ and peak release time $W = 4$.

$\lambda = 1$, $M = 1$, $W = 4$ and $x_n = 0.01n$, $n = 0, \cdots, 400$. It is shown that $a(x; \mu_0)$ is essentially all non-positive. This indicates that the optimized input distribution in Fig. 3 satisfies (40) in Theorem 2. Thus the optimized input distribution in (57) is the capacity-achieving input distribution, to numerical precision. It is interesting to observe that $a(x; \mu_0)$ is zero over an interval where the input puts its probability masses (including $x = 0$). The numerical result also indicates that the optimized distribution is a mixture of a probability density and two discrete mass points, one at "0" and the other at the peak release time. This mixture distribution is in contrast with the optimal distribution for additive white Gaussian noise channels and Rayleigh-fading channels with average power constraint and peak amplitude constraint [4], [6], [7].

References

[1] K. Srinivas, A. Eckford, and R. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Trans. Inform. Theory*, vol. 58, pp. 4678–4692, july 2012.

[2] A. Eckford, S. Kothapalli, and R. Adve, "The peak constrained additive inverse Gaussian noise channel," in *Proc. IEEE Int. Symp. Information Theory*, (Cambridge, MA, USA), pp. 2983–2987, July 2012.

[3] H.-T. Chang and S. M. Moser, "Bounds on the capacity of additive inverse Gaussian noise channel," in *Proc. IEEE Int. Symp. Information Theory*, (Cambridge, MA, USA), pp. 304–308, July 2012.

[4] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, pp. 203–219, 1971.

[5] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1060–1071, July 1995.

[6] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2257–2270, Oct. 2004.

[7] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2073–2088, June 2005.

[8] M. C. Gursoy, H. V. Poor, and S. Verdú, "The noncoherent Rician fading channel-part I: Structure of the capacity-achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2193–2206, Sep. 2005.

[9] J. Huang and S. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2336–2351, July 2005.

[10] L. Zhang and D. Guo, "Capacity of Gaussian channels with duty cycle and power constraints," in *Proc. IEEE Int. Symp. Information Theory*, (St Petersburg, Russia), pp. 424–428, July 2011.

[11] L. Zhang, H. Li, and D. Guo, "Capacity of Gaussian channels with duty cycle and power constraints," *IEEE Trans. Inform. Theory*, under revision, 2013.

[12] P. Billingsley, *Convergence of probability measure*. John Wiley & Sons, 1968.

[13] Y. Wu and Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inform. Theory*, vol. 58, pp. 1289 –1301, march 2012.

[14] R. S. Chhikara and L. Folks, *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. New York: Marcel Dekker, INC, 1988.

[15] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, sep 1990.