Capacity of the Memoryless Additive Inverse Gaussian Noise Channel

Hui Li, Stefan M. Moser, *Senior Member, IEEE*, and Dongning Guo, *Senior Member, IEEE*

Abstract—The memoryless additive inverse Gaussian noise channel model describing communication based on the exchange of chemical molecules in a drifting liquid medium is investigated for the situation of simultaneously an average-delay and a peak-delay constraint. Analytical upper and lower bounds on its capacity in bits per molecule use are presented. These bounds are shown to be asymptotically tight, i.e., for the delay constraints tending to infinity with their ratio held constant (or for the drift velocity of the fluid tending to infinity), the asymptotic capacity is derived precisely. Moreover, characteristics of the capacity-achieving input distribution are derived that allow the exact numerical computation of capacity. It is argued that the optimal input is a mixed continuous and discrete distribution.

Index Terms—Additive inverse Gaussian noise, average- and peak-delay constraints, Brownian motion, channel capacity, molecular communication.

I. INTRODUCTION

SINCE nanoscale devices have very limited size, energy, and processing power, molecular communication becomes a promising approach for nanonetworks. In this new paradigm of communication, the transmitter releases information carrying molecules, which then propagate through a fluid medium, until the receiver captures them and decodes the information [1]. As an example consider the communication between two nanoscale devices in a blood vessel. Such a system, of course, will behave fundamentally differently from the usual information transmission systems. It is therefore a very interesting task to model this communication scenario and analyze it.

Consider a molecular communication system where the transmitter encodes the message in the release timing of every molecule. For example, if a binary message is encoded as the release time T_0 for symbol "0", and T_1 for symbol "1", the receiver will capture the molecule at time $T_0 + N$ or $T_1 + N$ respectively, where N is a random propagation time,

known as the *first arrival time*: the first passage time for a molecule traveling from the transmitter to the receiver. In a one-dimensional fluid medium and under the assumption of Brownian motion, the first arrival time is inverse Gaussian (IG) distributed [2]. Therefore, such a system can be modeled as the so-called *additive inverse Gaussian noise (AIGN) channel*. This model was first introduced in [3], where also some upper and lower bounds on the capacity under an average-delay constraint were presented and a maximum likelihood (ML) receiver design was investigated. Additional bounds on the capacity have been developed under either a peak-delay constraint [4] or an average-delay constraint [5]. The capacity-achieving input distribution of the AIGN channel with both average- and peak-delay constraints was studied in [6].

In this paper, we study the capacity of the AIGN channel under either both an average- and a peak-delay constraint (which includes the special case of only a peak-delay constraint), or under only an average-delay constraint. We provide tight analytical upper and lower bounds that coincide asymptotically when both average and peak delay tend to infinity with their ratio held constant (which is equivalent to the situation when the drift velocity of the liquid tends to infinity). Thus, for the asymptotic capacity, we derive an exact analytical expression. We also include improvements on some of the bounds in [4] [5]. Moreover, we establish some properties of the capacity-achieving input distribution, including sufficient and necessary conditions for an optimal distribution. We show that an optimal input distribution must be a mixture of a continuous distribution and two point masses at "0" and the peak, respectively. This is very different from the results seen in common channel models under both average- and peak-power constraints, like, e.g., the additive white Gaussian noise (AWGN) channel [7], quadrature Gaussian channels, or Rayleigh-fading channels [8]–[15], where the capacity-achievable input is discrete. The presented insights into the capacity-achieving input also allow us to numerically compute the capacity. These numerical results correspond very well with the analytical bounds presented in this paper and in [3] [5].

The remainder of this paper is structured as follows. After a few comments about notation, in Section II, we will introduce the system model and its underlying assumptions more in detail. In Section III, we present new analytical upper and lower bounds on capacity and the exact asymptotic capacity formulas. In Section IV, we then investigate the capacity-achieving input distribution. We prove existence and uniqueness, derive some properties, and we demonstrate that the capacity-achieving input can be found through the method of

Lagrangian multipliers for nonlinear optimization. Section V then focuses on numerical methods for the calculation of the capacity and the capacity-achieving input distribution. Finally, we conclude in Section VI.

We try to use uppercase Romans (e.g., N) for random variables and lowercase Romans (e.g., n) for their realizations. For nonrandom quantities we use a sans-serif font in uppercase, e.g., M or T. All logarithms $\log(\cdot)$ refer to the natural logarithm, i.e., all results are stated in nats. Following engineering custom, though, we have scaled the plots given in all figures to be in bits.

II. SYSTEM MODEL

We assume that a transmitter emits a molecule at a certain time into a one-dimensional fluid that itself is drifting with constant velocity v toward the receiver, which is at distance d from the transmitter. The molecule is transported by the fluid and its inherent Brownian motion. Once it reaches the receiver, the molecule is removed from the fluid. In this diffusion-mediated molecular communication system, the message is encoded in the release timing of the molecules. The molecule is released at time $X \geq 0$, travels for a random time $N \geq 0$ and arrives at the receiver at time

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

The Brownian motion is modeled by a Wiener process with infinitesimal variance σ^2. Over any time interval τ, the displacement of the molecule is Gaussian distributed with mean $v\tau$ and variance $\sigma^2\tau$, independently of any other nonoverlapping time intervals.

The random time N has an *inverse Gaussian distribution* with the corresponding probability density function (PDF)[1]

$$p_N(n) = \sqrt{\frac{\lambda}{2\pi}}\, n^{-\frac{3}{2}} \exp\left(-\frac{\lambda(n-c)^2}{2c^2n}\right) \mathsf{I}\{n > 0\}. \tag{2}$$

Here c denotes the average travel time and λ relates to the Brownian motion parameter σ^2:

$$c = \frac{d}{v}, \qquad \lambda = \frac{d^2}{\sigma^2}. \tag{3}$$

Throughout this paper we will assume that $v > 0$. For the case $v = 0$—which corresponds to the case where the drift velocity of the fluid is zero and where the molecules are transported through the fluid only because of the inherent Brownian motion—the random travel time N is not inverse Gaussian distributed, but Lévy distributed:

$$p_N(n) = \sqrt{\frac{\lambda}{2\pi}}\, n^{-\frac{3}{2}} \exp\left(-\frac{\lambda}{2n}\right) \mathsf{I}\{n > 0\}. \tag{4}$$

Furthermore, if $v < 0$, the fluid drifts away from the receiver, which means that there is a positive probability that a molecule will never arrive at the receiver. Both these cases need a separate investigation and are not considered in this paper.

For notational convenience, we shall redefine the units of time and distance as follows without loss of generality. Let

---

[1] The function $\mathsf{I}\{\cdot\}$ denotes the indicator function that takes on the values 1 or 0 depending on whether the statement holds true or not.

one unit of distance be the distance between the transmitter and receiver. Let one unit of time be the average travel time from the transmitter to the receiver due to the drift of the fluid. Under this new system, the distance $d = 1$ and the average travel time $c = 1$. The only free parameter is λ, which is the ratio of the distance squared and the infinitesimal variance of the Wiener process, a dimensionless quantity. We have

$$\mathsf{E}[\log N] = e^{2\lambda}\,\mathrm{Ei}(-2\lambda) \tag{5}$$
$$h(N) = \frac{1}{2}\log\frac{2\pi e}{\lambda} + \frac{3}{2}\, e^{2\lambda}\,\mathrm{Ei}(-2\lambda) \tag{6}$$

with $\mathrm{Ei}(\cdot)$ denoting the exponential integral function

$$\mathrm{Ei}(-t) = -\int_t^\infty \frac{e^{-\tau}}{\tau}\, d\tau. \tag{7}$$

It is interesting to note that the inverse Gaussian distribution is differential-entropy maximizing under the constraints $\mathsf{E}[N] = \alpha_1$, $\mathsf{E}[N^{-1}] = \alpha_2$, and $\mathsf{E}[\log N] = \alpha_3$.

The molecule channel with ℓ molecule uses consists of ℓ independent uses of the random transformation (1), where the ℓ release times form a codeword. It is referred to as *the additive inverse Gaussian noise (AIGN) channel*. The receiver decodes the message from the ℓ arrival times. The following idealized assumptions are made to simplify our channel model: 1) The transmitter and receiver are perfectly synchronized; 2) the transmitter perfectly controls the release time of the molecules; 3) the receiver perfectly measures the arrival time of the molecules; 4) on first arrival at the receiver, the 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. The last assumption guarantees that the channel is memoryless. (If the molecules were indistinguishable, the order of arrivals would be in general unknown, which introduces memory into the channel. The capacity of such a channel with memory is an open problem.)

The conditional PDF of the output Y given the input X is $p_{Y|X}(y|x) = p_N(y - x)$. Let μ denote the distribution of the channel input X. In general, μ is a probability measure defined on the Borel algebra on the nonnegative real numbers, denoted by $\mathcal{B}([0,\infty])$. The PDF of the channel output exists and is given as

$$p_Y(y;\mu) = \int p_{Y|X}(y|x)\, \mathrm{d}\mu(x) = \mathsf{E}_\mu[p_N(y - X)]. \tag{8}$$

The corresponding mutual information between channel input and output per channel use (molecule use) then is

$$I(\mu) \triangleq I(X;Y) \tag{9}$$
$$= \iint_0^\infty p_{Y|X}(y|x) \log\frac{p_{Y|X}(y|x)}{p_Y(y;\mu)}\, \mathrm{d}y\, \mathrm{d}\mu(x). \tag{10}$$

We define $Q(x;\mu)$ as

$$Q(x;\mu) \triangleq -\int_0^\infty p_{Y|X}(y|x) \log p_Y(y;\mu)\, \mathrm{d}y \tag{11}$$
$$= -\int_x^\infty p_N(y - x) \log p_Y(y;\mu)\, \mathrm{d}y \tag{12}$$

such that the differential entropy of Y for the given input distribution μ is

$$h(Y;\mu) = -\int_0^\infty p_Y(y;\mu) \log p_Y(y;\mu)\,\mathrm{d}y \qquad (13)$$

$$= -\int_0^\infty \int p_N(y-x) \log p_Y(y;\mu)\,\mathrm{d}\mu(x)\,\mathrm{d}y \quad (14)$$

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

Together with the conditional entropy $h(Y|X) = h(N)$, $I(\mu)$ can now be expressed as

$$I(\mu) = h(Y;\mu) - h(N) = \int Q(x;\mu)\,\mathrm{d}\mu(x) - h(N). \quad (16)$$

We consider two kinds of constraints on the release time of the molecules. First, the transmitter is subject to an average-delay constraint M:

$$\mathsf{E}[X] \le \mathsf{M}. \qquad (17)$$

Second, in addition to the average-delay constraint M, we assume that the molecules can only be released during the limited time interval $[0, \mathsf{T}]$, i.e., we impose the peak-delay time constraint T where, without loss of generality, $\mathsf{T} \ge \mathsf{M}$. Both M and T are in the units of the average arrival time. We define the average-to-peak delay ratio α as

$$\alpha \triangleq \frac{\mathsf{M}}{\mathsf{T}}. \qquad (18)$$

Note that $0 < \alpha \le 1$, where $\alpha = 1$ corresponds to the situation with only a peak-delay constraint.

We denote the set of distributions with the peak- and average-delay constraint T and M by

$$\Lambda(\mathsf{T}, \mathsf{M}) \triangleq \{\mu \colon \mu([0, \mathsf{T}]) = 1, \ \mathsf{E}_\mu[X] \le \mathsf{M}\}. \quad (19)$$

Similarly, we define the set of distributions with only an average-delay constraint M as

$$\Lambda_\mathrm{m}(\mathsf{M}) \triangleq \Lambda(\infty, \mathsf{M}). \qquad (20)$$

(Here the subscript "m" stands for "mean time delay.")

The capacity of the AIGN channel under both a peak- and an average-delay constraint is given as

$$\mathsf{C}(\mathsf{T}, \mathsf{M}) = \sup_{\mu \in \Lambda(\mathsf{T}, \mathsf{M})} I(\mu) \quad \text{nats per molecule use.} \quad (21)$$

For simplicity, we omit the unit of the capacity in the sequel. Similarly, the capacity of the AIGN channel with only an average-delay constraint is

$$\mathsf{C}_\mathrm{m}(\mathsf{M}) = \sup_{\mu \in \Lambda_\mathrm{m}(\mathsf{M})} I(\mu). \qquad (22)$$

III. BOUNDS ON CAPACITY

A. Known Bounds on Capacity

In [3], two analytical bounds on capacity are presented. First, an upper bound is derived based on the fact that differential entropy $h(Y)$ under an expectation constraint $\mathsf{E}[Y] \le \mathsf{M}+1$ is maximized by an exponential distribution: $h(Y) \le \log\big(e(\mathsf{M}+1)\big)$. This leads to the bound

$$\mathsf{C}_\mathrm{m}(\mathsf{M}) = \max_{\mu \in \Lambda_\mathrm{m}(\mathsf{M})} h(Y;\mu) - h(N) \qquad (23)$$

$$\le \frac{1}{2}\log\frac{\lambda e(\mathsf{M}+1)^2}{2\pi} - \frac{3}{2}e^{2\lambda}\operatorname{Ei}(-2\lambda). \quad (24)$$

Second, a lower bound is given that is based on the suboptimal input $X \sim \mathrm{IG}\big(\mathsf{M}, \lambda\mathsf{M}^2\big)$ chosen in such a way that the channel output is again inverse Gaussian distributed: $Y \sim \mathrm{IG}\big(\mathsf{M}+1, \lambda(\mathsf{M}+1)^2\big)$. This yields

$$\mathsf{C}_\mathrm{m}(\mathsf{M}) \ge \frac{1}{2}\log(\mathsf{M}+1) + \frac{3}{2}e^{2\lambda(\mathsf{M}+1)}\operatorname{Ei}(-2\lambda(\mathsf{M}+1))$$
$$- \frac{3}{2}e^{2\lambda}\operatorname{Ei}(-2\lambda). \qquad (25)$$

B. Useful Mappings

The following mappings and their properties will turn out to be useful for the derivations of the bounds given in Section III-C.

Definition 1. For $a > 0$ and $\zeta \ge 0$, we define

$$g(a, \zeta) \triangleq 1 - \mathcal{Q}\Big(\sqrt{\zeta}\big(\sqrt{a} - \sqrt{a^{-1}}\big)\Big)$$
$$+ e^{2\zeta}\mathcal{Q}\Big(\sqrt{\zeta}\big(\sqrt{a} + \sqrt{a^{-1}}\big)\Big) \qquad (26)$$

$$\bar{g}(a, \zeta) \triangleq 1 - \mathcal{Q}\Big(\sqrt{\zeta}\big(\sqrt{a} - \sqrt{a^{-1}}\big)\Big)$$
$$- e^{2\zeta}\mathcal{Q}\Big(\sqrt{\zeta}\big(\sqrt{a} + \sqrt{a^{-1}}\big)\Big) \qquad (27)$$

where the $\mathcal{Q}$-function is defined as

$$\mathcal{Q}(\xi) \triangleq \frac{1}{\sqrt{2\pi}}\int_\xi^\infty e^{-\frac{t^2}{2}}\,\mathrm{d}t. \qquad (28)$$

Lemma 2. *The mappings $a \mapsto g(a, \zeta)$ and $a \mapsto \bar{g}(a, \zeta)$ are both monotonically increasing and satisfy*

$$\lim_{a \downarrow 0} g(a, \zeta) = \lim_{a \downarrow 0} \bar{g}(a, \zeta) = 0 \qquad (29)$$

$$\lim_{a \uparrow \infty} g(a, \zeta) = \lim_{a \uparrow \infty} \bar{g}(a, \zeta) = 1 \qquad (30)$$

for all $\zeta \ge 0$. Moreover, for all $\zeta \ge 0$ and $a \ge 1$,

$$g(a, \zeta) \ge \frac{1}{2}. \qquad (31)$$

Proof: The derivatives of $g(a, \zeta)$ and $\bar{g}(a, \zeta)$ are positive:

$$\frac{\partial}{\partial a}g(a, \zeta) = \sqrt{\frac{\zeta}{2\pi a^3}}\,e^{-\frac{\zeta}{2}\left(\sqrt{a}-\sqrt{a^{-1}}\right)^2} > 0, \qquad (32)$$

$$\frac{\partial}{\partial a}\bar{g}(a, \zeta) = \sqrt{\frac{\zeta}{2\pi a}}\,e^{-\frac{\zeta}{2}\left(\sqrt{a}-\sqrt{a^{-1}}\right)^2} > 0. \qquad (33)$$

The limiting behavior follows directly from the properties of the $\mathcal{Q}$-function. Finally, using monotonicity,

$$g(a, \zeta) \ge g(1, \zeta) = \frac{1}{2} + e^{2\zeta}\mathcal{Q}(2\sqrt{\zeta}) \ge \frac{1}{2}. \qquad (34)$$

■

Lemma 3. *The mapping* $a \mapsto \log g(a, \zeta)$ *is monotonically increasing and concave.*

Proof: The monotonicity follows directly from Lemma 2 and the monotonicity of the logarithm. To confirm its concavity, we derive the second derivative:

$$\frac{\partial^2}{\partial a^2} \log g(a, \zeta)$$

$$= -\sqrt{\frac{\zeta}{8\pi a^7}} \frac{e^{-\frac{\zeta}{2}\left(\sqrt{a} - \sqrt{a^{-1}}\right)^2}}{g^2(a, \zeta)}$$

$$\cdot \left(g(a, \zeta)\left(\zeta a^2 + 3a - \zeta\right) + \sqrt{\frac{2\zeta a}{\pi}} \, e^{-\frac{\zeta}{2}\left(\sqrt{a} - \sqrt{a^{-1}}\right)^2} \right) \tag{35}$$

which is seen to be nonpositive once we show that

$$r(a) \triangleq g(a, \zeta)\left(\zeta a^2 + 3a - \zeta\right) + \sqrt{\frac{2\zeta a}{\pi}} \, e^{-\frac{\zeta}{2}\left(\sqrt{a} - \sqrt{a^{-1}}\right)^2} \tag{36}$$

$$\geq 0. \tag{37}$$

This is indeed the case because $\lim_{a \downarrow 0} r(a) = 0$ and because

$$\frac{\partial r(a)}{\partial a} = (3 + 2a\zeta)g(a, \zeta) + \sqrt{\frac{8\zeta}{\pi a}} \, e^{-\frac{\zeta}{2}\left(\sqrt{a} - \sqrt{a^{-1}}\right)^2} \tag{38}$$

$$\geq 0. \tag{39}$$

$\blacksquare$

C. New Bounds

In the following we will present our bounds on capacity. The corresponding derivations can be found in Appendix A.

1) Peak- and Average-Delay Constraint with $\frac{1}{2} \leq \alpha \leq 1$:

Theorem 4. *For an average-to-peak delay ratio $\frac{1}{2} \leq \alpha \leq 1$, the capacity $\mathsf{C}(\mathsf{T}, \alpha\mathsf{T})$ is lower-bounded by*

$$\mathsf{C}(\mathsf{T}, \alpha\mathsf{T}) \geq \frac{1}{2} \log \frac{\lambda}{2\pi e} - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda) + \log \mathsf{T}$$
$$- \log g(\alpha\mathsf{T} + 1, \lambda), \quad \forall \mathsf{T} \geq 0 \tag{40}$$

and upper-bounded by

$$\mathsf{C}(\mathsf{T}, \alpha\mathsf{T}) \leq \frac{1}{2} \log \frac{\lambda}{2\pi e} - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda) + \log(\mathsf{T} + \delta) + 1$$
$$- g(\delta, \lambda) - 2\left(1 - g(\delta, \lambda)\right) \log\left(1 - g(\delta, \lambda)\right)$$
$$- g(\delta, \lambda) \log g(\delta, \lambda), \quad \forall \mathsf{T} \geq 0. \tag{41}$$

Here, $g(\cdot, \cdot)$ is given in Definition 1 and $\delta \geq 1$ is a free parameter.

Proof: The derivations of the upper and lower bound can be found in Appendices A-A2 and A-B2, respectively. $\blacksquare$

The bound (41) can be optimized numerically over δ. A suboptimal but useful choice of δ is

$$\delta = 2\log(1 + \mathsf{T}) + \frac{1}{\alpha}. \tag{42}$$

The bounds of Theorem 4 are depicted in Fig. 1.

As T increases without bound, the capacity increases as $\log \mathsf{T}$. Intuitively, if the release time is chosen uniformly from the set $\{0, t, 2t, \ldots, \mathsf{T}\}$ where t is a large constant, then the release time can be estimated with low probability of
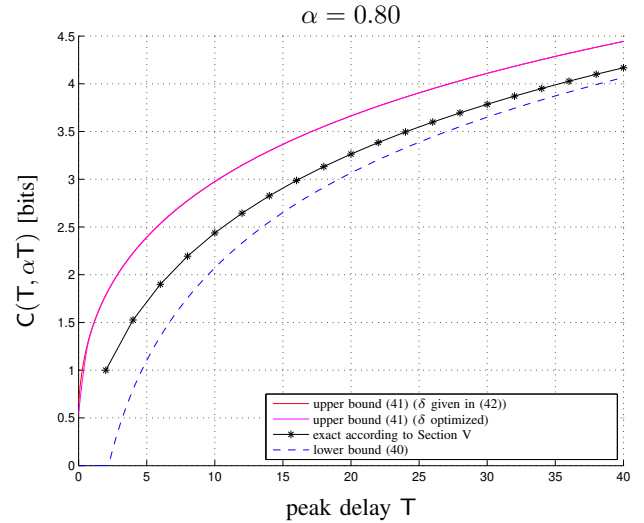

Fig. 1. Capacity and bounds on capacity as a function of T with $\alpha = 0.80$ and $\lambda = \frac{1}{2}$. The choice of the free parameter δ is either according to (42) or numerically optimized.

error, hence T/t messages can be sent reliably. (Note that the average delay is $\mathsf{T}/2$, i.e., the average-delay constraint is satisfied.) The following result gives the precise gap between the capacity and $\log \mathsf{T}$ as $\mathsf{T} \to \infty$.

Corollary 5 (Asymptotics). *For $\frac{1}{2} \leq \alpha \leq 1$,*

$$\lim_{\mathsf{T} \uparrow \infty} \left\{ \mathsf{C}(\mathsf{T}, \alpha\mathsf{T}) - \log \mathsf{T} \right\} = \frac{1}{2} \log \frac{\lambda}{2\pi e} - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda). \tag{43}$$

Proof: This follows from Theorem 4, the choice (42), and the asymptotic behavior of $g(\cdot, \cdot)$ given in Lemma 2. $\blacksquare$

Note that for $\alpha \geq \frac{1}{2}$, the asymptotic capacity does not depend on the average-delay constraint M and is identical to the capacity under a peak-delay constraint only.

2) Peak- and Average-Delay Constraint with $0 < \alpha < \frac{1}{2}$:

Theorem 6. *For an average-to-peak delay ratio $0 < \alpha < \frac{1}{2}$, the capacity $\mathsf{C}(\mathsf{T}, \alpha\mathsf{T})$ is lower-bounded by*

$$\mathsf{C}(\mathsf{T}, \alpha\mathsf{T}) \geq \frac{1}{2} \log \frac{\lambda}{2\pi e} - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda) + \log \mathsf{T}$$
$$+ \log \frac{1 - e^{-\beta^*}}{\beta^*} + \alpha\beta^* + \frac{\beta}{\mathsf{T}} - \lambda + \lambda\sqrt{1 - \frac{2\beta^*}{\lambda\mathsf{T}}}$$
$$- \log g\left((\alpha\mathsf{T} + 1)\sqrt{1 - \frac{2\beta^*}{\lambda\mathsf{T}}}, \lambda\sqrt{1 - \frac{2\beta^*}{\lambda\mathsf{T}}} \right),$$
$$\forall \mathsf{T} > \frac{2\beta^*}{\lambda} \tag{44}$$

and upper-bounded by

$$\mathsf{C}(\mathsf{T}, \alpha\mathsf{T}) \leq \frac{1}{2} \log \frac{\lambda}{2\pi e} - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda) + \log(\alpha\mathsf{T} + 1)$$
$$+ \alpha\beta^* - \log(\alpha\beta^*) + \log\left(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)}{\alpha\mathsf{T}+1}} \right) + 1$$
$$- g(\delta, \lambda) - 2\left(1 - g(\delta, \lambda)\right) \log\left(1 - g(\delta, \lambda)\right)$$
$$- 2g(\delta, \lambda) \log g(\delta, \lambda), \quad \forall \mathsf{T} \geq 0. \tag{45}$$

Here $g(\cdot, \cdot)$ is given in Definition 1, β^* is the nonnegative solution to[2]

$$\alpha = \frac{1}{\beta^*} - \frac{e^{-\beta^*}}{1 - e^{-\beta^*}} \qquad (46)$$

and $\delta \geq 1$ is a free parameter.

Proof: The derivations of the upper and lower bound can be found in Appendices A-A1 and A-B1, respectively. ∎

The bound (45) can be optimized numerically over δ, and again, a suboptimal but useful choice of δ is given in (42). The bounds of Theorem 6 are depicted in Fig. 2.

Similar asymptotics as Corollary 5 can be obtained, except that with $\alpha < 1/2$, the release time cannot be uniformly chosen from equally spaced points between 0 and T.

Corollary 7 (Asymptotics). *For* $0 < \alpha < \frac{1}{2}$,

$$\lim_{\mathsf{T}\uparrow\infty} \left\{ \mathsf{C}(\mathsf{T}, \alpha\mathsf{T}) - \log \mathsf{T} \right\}$$
$$= \frac{1}{2}\log\frac{\lambda}{2\pi e} - \frac{3}{2}e^{2\lambda}\operatorname{Ei}(-2\lambda) + \alpha\beta^* - \log\frac{\beta^*}{1 - e^{-\beta^*}}. \quad (47)$$

Proof: This follows from Theorem 6, the choice (42), and the asymptotic behavior of $g(\cdot, \cdot)$ given in Lemma 2. ∎

3) Average-Delay Constraint Only:

Theorem 8. *If the input is constrained by an average-delay constraint only, the capacity is lower-bounded by*

$$\mathsf{C}_{\mathrm{m}}(\mathsf{M}) \geq \frac{1}{2}\log\frac{\lambda}{2\pi e} - \frac{3}{2}e^{2\lambda}\operatorname{Ei}(-2\lambda) + \log\mathsf{M}$$
$$+ 1 + \frac{1}{\mathsf{M}} - \lambda + \lambda\sqrt{1 - \frac{2}{\lambda\mathsf{M}}}$$
$$- \log g\left((\mathsf{M}+1)\sqrt{1 - \frac{2}{\lambda\mathsf{M}}}, \lambda\sqrt{1 - \frac{2}{\lambda\mathsf{M}}} \right),$$
$$\forall\, \mathsf{M} > \frac{2}{\lambda} \quad (48)$$

and upper-bounded by either of the following three bounds:

$$\mathsf{C}_{\mathrm{m}}(\mathsf{M}) \leq \frac{1}{2}\log\left(1 + \frac{\lambda\mathsf{M}}{\mathsf{M}+1}\right) + \frac{3}{2}\log\left(1 + \mathsf{M}\left(1 + \frac{1}{\lambda}\right)\right),$$
$$\forall\, \mathsf{M} \geq 0 \quad (49)$$

$$\mathsf{C}_{\mathrm{m}}(\mathsf{M}) \leq \frac{1}{2}\log\lambda + \frac{\eta - 1}{2}e^{2\lambda}\operatorname{Ei}(-2\lambda)$$
$$+ \frac{\eta+2}{2}\log\left(1 + \mathsf{M}\left(1 + \frac{1}{\lambda}\right)\right) - \log\eta$$
$$+ \frac{1}{2}\log\left(\sqrt{\frac{2\lambda}{\pi}}\, e^\lambda\, \mathsf{K}_{\eta+\frac{1}{2}}(\lambda) - (\mathsf{M}+1)^{-\eta}\right),$$
$$\forall\, \mathsf{M} \geq 0 \quad (50)$$

$$\mathsf{C}_{\mathrm{m}}(\mathsf{M}) \leq \frac{1}{2}\log\lambda - e^{2\lambda}\operatorname{Ei}(-2\lambda)$$
$$+ \frac{1}{2}\log\left(1 + \mathsf{M}\left(1 + \frac{1}{\lambda}\right)\right)$$
$$+ \frac{1}{2}\log\left(1 + \mathsf{M} - \frac{\lambda}{1+\lambda}\right), \qquad \forall\, \mathsf{M} \geq 0. \quad (51)$$

---

[2]It is straightforward to show that for $0 < \alpha < \frac{1}{2}$, the solution to (46) always exists and is unique.

Here, $\operatorname{Ei}(\cdot)$ is defined in (7), $\mathsf{K}_\nu(\cdot)$ represents the order-ν modified Bessel function of the second kind, and $0 < \eta \leq 1$ is a free parameter.

Proof: The derivations of the upper bounds are given in Appendix A-A3, and the derivation of the lower bound can be found in Appendix A-B3. ∎

The second bound (50) can be optimized numerically over η. A suboptimal choice for η is

$$\eta \triangleq \min\left\{ \frac{2}{\log\left(4 + \frac{4\mathsf{M}}{\lambda}\right)}, 1 \right\}. \quad (52)$$

The bounds of Theorem 8 are depicted in Fig. 3.

Corollary 9 (Asymptotics). *For the AIGN channel under only an average-delay constraint,*

$$\lim_{\mathsf{M}\uparrow\infty} \left\{ \mathsf{C}_{\mathrm{m}}(\mathsf{M}) - \log\mathsf{M} \right\}$$
$$= \frac{1}{2}\log\frac{\lambda}{2\pi e} - \frac{3}{2}e^{2\lambda}\operatorname{Ei}(-2\lambda) + 1. \quad (53)$$

Proof: This follows from the lower bound in Theorem 8, the upper bound (24), and the asymptotic behavior of $g(\cdot, \cdot)$ given in Lemma 2. ∎

This asymptotic result agrees with the approximations given in [16].

IV. CAPACITY-ACHIEVING INPUT

In this section we study the properties of the capacity-achieving input under either an average-delay constraint or both a peak- and an average-delay constraint.

A. Existence and Uniqueness

Let $\mathcal{P}$ denote the collection of all Borel probability measures defined on $\left(\mathbb{R}^+ \cup \{0\}, \mathcal{B}(\mathbb{R}^+ \cup \{0\})\right)$, which is a topological space with the topology of weak convergence [17]. In order to prove the existence of the capacity-achieving input distribution under an average-delay constraint or under both a peak- and an average-delay constraint, we first show that the set of distributions subject to an average-delay constraint $\Lambda_{\mathrm{m}}(\mathsf{M})$ and the set of distributions subject to both a peak- and an average-delay constraint $\Lambda(\mathsf{T}, \mathsf{M})$ are compact in the topological space $\mathcal{P}$.

Lemma 10. $\Lambda_{\mathrm{m}}(\mathsf{M})$ *is compact in the topological space* $\mathcal{P}$.

Proof: According to [17], the topology of weak convergence on $\mathcal{P}$ is metrizable. Therefore, by Prokhorov's theorem [18], in order to prove that $\Lambda_{\mathrm{m}}(\mathsf{M})$ is compact in $\mathcal{P}$, it suffices to show that it is both tight and closed.

For any $\varepsilon > 0$, there exists an $a_\varepsilon > 0$ such that for all $\mu \in \Lambda_{\mathrm{m}}(\mathsf{M})$,

$$\mu\left(|X| \geq a_\varepsilon\right) \leq \frac{\mathsf{E}[X]}{a_\varepsilon} \leq \frac{\mathsf{M}}{a_\varepsilon} < \varepsilon \quad (54)$$

by Markov's inequality. Choose $\mathcal{K}_\varepsilon = [0, a_\varepsilon]$, then $\mathcal{K}_\varepsilon$ is compact in $\mathbb{R}$ and $\mu(\mathcal{K}_\varepsilon) \geq 1 - \varepsilon$ for all $\mu \in \Lambda_{\mathrm{m}}(\mathsf{M})$, thus $\Lambda_{\mathrm{m}}(\mathsf{M})$ is tight.
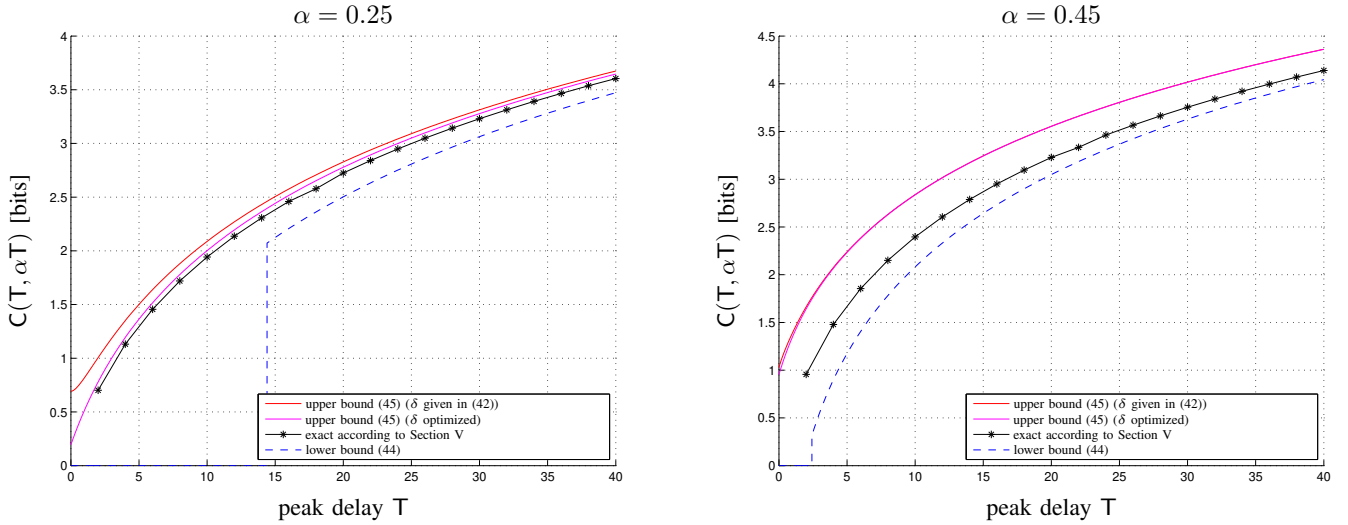
Fig. 2. Capacity and bounds on capacity as a function of T with $\alpha = 0.25$ (left) or $\alpha = 0.45$ (right), and $\lambda = \frac{1}{2}$. The choice of the free parameter δ is either according to (42) or numerically optimized.
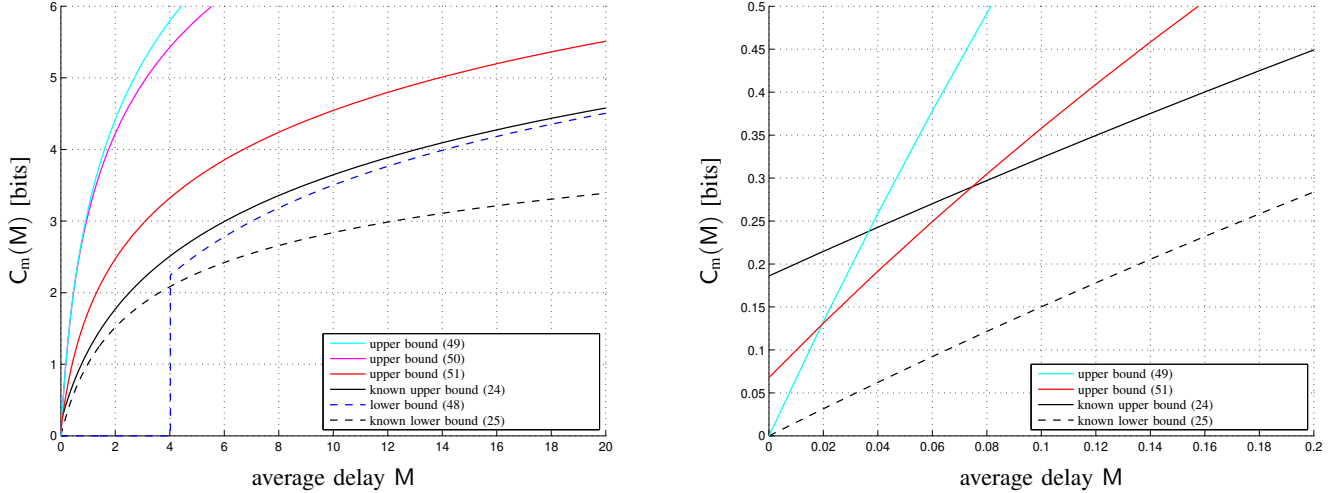

Fig. 3. Bounds on capacity with only an average-delay constraint as a function of M with $\lambda = \frac{1}{2}$. The right plot is zoomed in at low values of M. The choice of the free parameter η is according to (52).

Let $\{\mu_n\}_{n=1}^{\infty}$ be a convergent sequence in $\Lambda_m(M)$ with limit μ_0. Let $f(x) = x$, which is continuous and bounded below. By weak convergence [17, Section 3.1], we have

$$\mathsf{E}_{\mu_0}[X] = \int f \, \mathrm{d}\mu_0 \le \varliminf_{n \to \infty} \int f \, \mathrm{d}\mu_n \le M. \quad (55)$$

Therefore, $\mu_0 \in \Lambda_m(M)$, i.e., $\Lambda_m(M)$ is closed, and the compactness of $\Lambda_m(M)$ follows. ∎

Corollary 11. $\Lambda(T, M)$ *is compact in the topological space* $\mathcal{P}$.

Proof: Following a similar procedure as in the proof of Lemma 10, we can prove that $\Lambda(T, M)$ is closed. Since a closed subset of a compact space is also compact, $\Lambda(T, M) \subset \Lambda_m(M)$ is compact. ∎

Since the mutual information $I(\mu)$ is continuous on $\mathcal{P}$ [19, Theorem 7], it must achieve its maximum on the compact set

$\Lambda_m(M)$ or $\Lambda(T, M)$. Hence the capacity-achieving distribution μ_0 exists under an average-delay constraint or under both a peak- and an average-delay constraint.

In the remainder of Section IV, we will study the properties of the capacity-achieving distribution in a compact space $\bar{\Lambda}$, which can be either $\Lambda_m(M)$ or $\Lambda(T, M)$.

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

Theorem 12. *The input–output mutual information* $I(\mu)$ *of the AIGN channel is strictly convex in* $\bar{\Lambda}$.

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

$$p_Y(y; \mu_\theta) = (1 - \theta) \, p_Y(y; \mu_0) + \theta \, p_Y(y; \mu_1). \quad (56)$$

Thus,

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

$$= h(Y; \mu_\theta) - (1 - \theta)h(Y; \mu_1) - \theta h(Y; \mu_2) \quad (57)$$

$$= -\int p_Y(y; \mu_\theta) \log p_Y(y; \mu_\theta) \, dy$$

$$+ (1 - \theta) \int p_Y(y; \mu_1) \log p_Y(y; \mu_1) \, dy$$

$$+ \theta \int p_Y(y; \mu_2) \log p_Y(y; \mu_2) \, dy \quad (58)$$

$$= (1 - \theta)D\big(p_Y(\cdot; \mu_1) \big\| p_Y(\cdot; \mu_\theta)\big)$$

$$+ \theta D\big(p_Y(\cdot; \mu_2) \big\| p_Y(\cdot; \mu_\theta)\big) \quad (59)$$

where $D(\cdot \| \cdot)$ denotes the relative entropy [22]. Using the nonnegativity property of the relative entropy [22], we have

$$I(\mu_\theta) \geq (1 - \theta)I(\mu_0) + \theta I(\mu_1) \quad (60)$$

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 prove strict convexity, we show that if μ_0, μ_1 both achieve the maximum, then μ_0 and μ_1 are identical. This is implied if we show that $p_Y(y; \mu_0) = p_Y(y; \mu_1)$ for all y. Let $\varphi_N(t)$ be the characteristic function of $N \sim \mathrm{IG}(1, \lambda)$ and $\varphi_\mu(t)$ be the characteristic function of distribution μ. We know that $p_Y(y; \mu_1) = p_Y(y; \mu_2)$ is equivalent to $\varphi_N(t)\varphi_{\mu_1}(t) = \varphi_N(t)\varphi_{\mu_2}(t)$. We have [2]

$$\varphi_N(t) = \exp\left\{\lambda\left(1 - \left(1 - 2\mathrm{i}\lambda^{-1}t\right)^{\frac{1}{2}}\right)\right\} \quad (61)$$

which is nonzero for all t, i.e., $p_Y(y; \mu_1) = p_Y(y; \mu_2)$ if and only if $\varphi_{\mu_1}(t) = \varphi_{\mu_2}(t)$. This requires that $\mu_1 = \mu_2 = \mu_\theta$. ∎

B. Characterization of the Capacity-Achieving Distribution

We will next show that the capacity-achieving input distribution has the important property of assigning a positive probability to any ε-environment around the zero input.

Definition 13. 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 14. *The supremum in* (21) *or* (22) *is a maximum and is achieved by a unique* $\mu_0 \in \bar{\Lambda}$. *Moreover,* 0 *is a point of increase of the optimal input distribution, i.e., for any* $\varepsilon > 0$,

$$\mu_0\big([0, \varepsilon)\big) > 0. \quad (62)$$

Proof: Due to the strict convexity of $I(\mu)$ established in Theorem 12 and the compactness of $\bar{\Lambda}$ established in Lemma 10 and Corollary 11, the maximizer in (21) or (22) must exist and be unique. We next prove (62). By contradiction, suppose that the optimal input is lower-bounded by $\varepsilon > 0$, i.e., $\mu_0\big([0, \varepsilon)\big) = 0$. Then consider the 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 belong to $\bar{\Lambda}$, it thus also is an optimal input distribution. This contradicts the uniqueness of the optimal input. ∎

By Theorem 14, the capacity-achieving input distribution subject to either an average-delay constraint or subject to both a peak- and an average-delay constraint can be found respectively in the following distribution sets:

$$\Lambda_{\mathrm{m},0}(\mathsf{M}) = \big\{\mu \colon \mathsf{E}_\mu[X] \leq \mathsf{M}, \, \mu[0, \varepsilon) > 0, \, \forall \varepsilon > 0\big\} \quad (63)$$

$$\Lambda_0(\mathsf{T}, \mathsf{M}) = \big\{\mu \colon \mu\big([0, \mathsf{T}]\big) = 1, \mathsf{E}_\mu[X] \leq \mathsf{M},$$
$$\mu[0, \varepsilon) > 0, \, \forall \varepsilon > 0\big\}. \quad (64)$$

In the following we will use $\bar{\Lambda}_0$ to represent either $\Lambda_{\mathrm{m},0}(\mathsf{M})$ or $\Lambda_0(\mathsf{T}, \mathsf{M})$.

Lemma 15. *Consider an AIGN channel with noise* $N_1 \sim \mathrm{IG}(1, \lambda)$ *and input* $X_1 \sim \mu_1 \in \Lambda_{\mathrm{m}}(\mathsf{M})$, *and another AIGN channel with noise* $N_2 \sim \mathrm{IG}(a, a\lambda)$ *and input* $X_2 \sim \mu_2 \in \Lambda_{\mathrm{m}}(a\mathsf{M})$, *where* $a > 0$. *Then the capacity between input and output of both channels are identical.*

Proof: Since inverse Gaussians are closed under scaling, $aN_1 \sim \mathrm{IG}(a, a\lambda)$. Hence, $aX_1 + aN_1$ is equivalent to $X_2 + N_2$. ∎

A similar result can be derived for the case of both a peak- and an average-delay constraint.

We now define the set of all distributions with zero being a point of increase and with an arbitrary finite average delay as

$$\Lambda_{\mathrm{m},0}^\cup \triangleq \bigcup_{0 \leq \mathsf{M} < \infty} \Lambda_{\mathrm{m},0}(\mathsf{M}) \quad (65)$$

for the case of an average-delay constraint, and

$$\Lambda_0^\cup(\mathsf{T}) \triangleq \bigcup_{0 \leq \mathsf{M} \leq \mathsf{T}} \Lambda_0(\mathsf{T}, \mathsf{M}) \quad (66)$$

for the case of both a peak- and an average-delay constraint. We will use $\bar{\Lambda}_0^\cup$ to refer to either of them.

Lemma 16. *The mutual information function* $I(\mu)$ *is weakly differentiable in* $\bar{\Lambda}_0^\cup$. *The weak derivative* $I'_{\mu_0}(\mu)$, *defined as*

$$I'_{\mu_0}(\mu) \triangleq \lim_{\theta \downarrow 0} \frac{I\big((1 - \theta)\mu_0 + \theta\mu\big) - I(\mu_0)}{\theta} \quad (67)$$

can be expressed as

$$I'_{\mu_0}(\mu) = \int Q(x; \mu_0) \, d\mu(x) - I(\mu_0) - h(N). \quad (68)$$

Proof: See Appendix B-A. ∎

The following Theorem gives necessary and sufficient conditions on the channel input in order to be capacity achieving. This is particularly useful for the numerical results presented in Section V.

Theorem 17. *Let* μ_0 *be an admissible input probability measure, i.e.,* $\mu_0 \in \bar{\Lambda}$. *Then* μ_0 *is capacity achieving if and only if there exists a* $\nu \geq 0$ *such that for every* $x \geq 0$

$$Q(x; \mu_0) - h(N) - I(\mu_0) - \nu(x - \mathsf{M}) \leq 0. \quad (69)$$

Furthermore, (69) *holds with equality for all points of increase of* μ_0.

Proof: See Appendix B-B. ∎

V. NUMERICAL CALCULATION OF THE CAPACITY-ACHIEVING INPUT

Based on the insights obtained in Section IV, the capacity under a peak- and an average-delay constraint can be achieved in the set $\Lambda_0(\mathsf{T}, \mathsf{M})$:

$$\mathsf{C}(\mathsf{T}, \mathsf{M}) = \max_{\mu \in \Lambda_0(\mathsf{T}, \mathsf{M})} I(\mu). \quad (70)$$

In [7], Smith proves that only a discrete distribution can achieve the capacity of an additive white Gaussian noise channel with amplitude and variance constraints. The discreteness of the capacity-achieving input hinges on the fact that the function $Q(x; \mu_0)$ defined in (12) analytically extends to the entire complex plane in the setting of [7]. However, for the AIGN channel, it can be proved that $Q(z; \mu_0)$ is not analytic at $z = 0$. As demonstrated below, the resulting optimal input for the AIGN channel appears to be a mixed discrete and continuous distribution.

From Theorem 17, we can calculate the capacity-achievable input μ_0 through the method of Lagrangian multipliers on the nonlinear program problem (70). Since a direct search of the optimal input on an infinite dimensional functional space is difficult, the following discrete approximation of (70) is solved to estimate the optimal input:

$$\underset{P_X}{\text{maximize}} \ I(X; Y) \tag{71}$$

$$\text{subject to } \mathsf{E}[X] \leq \mathsf{M} \tag{72}$$

$$X \in \mathcal{S} = \{x_1, \ldots, x_n\}$$

$$\text{where } 0 = x_1 < \cdots < x_n = \mathsf{T}$$

where $\mathcal{S}$ is the alphabet set of the input distribution. Evidently, the set of distribution P_X on $\mathcal{S}$ satisfying the average constraint in (72) is a convex set. Since the mutual information $I(X; Y)$ is concave in P_X for fixed $P_{Y|X}$ and since the average constraint is linear, the maximization problem in (71) is a general convex programming problem. We solve this problem by the interior point method. For $p_i = P_X(x_i)$, the output distribution is

$$p_Y(y) = \sum_{i=1}^{n} p_i p_N(y - x_i) \tag{73}$$

and we can solve (71) by maximizing

$$h(Y) = \int_0^{\infty} p_Y(y) \log \frac{1}{p_Y(y)} \, \mathrm{d}y \tag{74}$$

under the average-delay constraint

$$\sum_{i=1}^{n} p_i x_i - \mathsf{M} \leq 0. \tag{75}$$

Plots of the rate achieved by this optimized input are shown in Figs. 1 and 2 in combination with the bounds derived in Section III. Evidently, the results are very close to capacity.

Fig. 4 shows the optimized input distribution (top) and the corresponding $a(x; \mu_0)$ (bottom) defined in (171), where $\lambda = 1$, $\mathsf{M} = 1$, $\mathsf{T} = 4$ and $x_i = 0.01i$, $i = 0, \ldots, 400$. It is shown that $a(x; \mu_0)$ is essentially nonpositive. This indicates that the optimized input distribution in Fig. 4 satisfies (69) in Theorem 17. Thus the optimized input distribution in (71) 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-delay time. This mixture distribution is in contrast
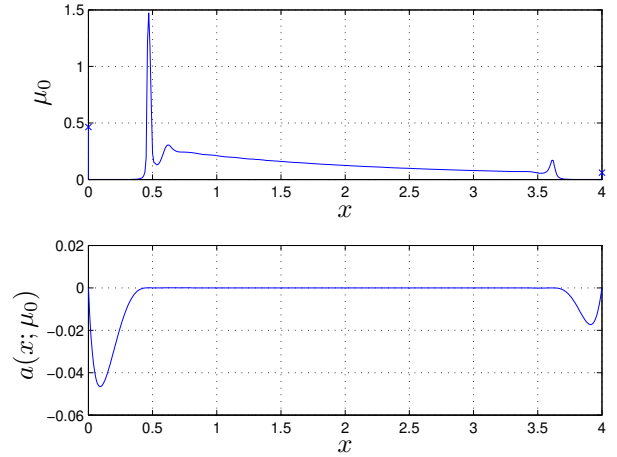

Fig. 4. Optimized distribution (top) and $a(x; \mu_0)$ (bottom) for $\lambda = 1$, $\mathsf{M} = 1$, and $\mathsf{T} = 4$.

with the optimal distribution for additive white Gaussian noise channels and Rayleigh-fading channels with an average-power and a peak-amplitude constraint [7] [12] [11].

VI. CONCLUSION

We have studied the capacity of the memoryless additive inverse Gaussian noise channel with an average-delay constraint or with both an average- and a peak-delay constraint. New upper and lower bounds on the capacity in nats (or bits) per molecule use are presented and properties of the capacity-achieving input distribution are derived. The latter allows the exact numerical computation of the capacity. Note that all results are analytical results and are not based on simulations.

The numerical results show that the optimal input is a mixed continuous and discrete distribution, which deserves further study. It might also be helpful for improving on the bounds in the low-delay region. Another promising approach for a lower bound on capacity is to use the entropy power inequality [22, Theorem 17.7.3] in combination with the chosen inputs given in Appendix A-B.

In this paper, we assume that the molecules released by the transmitter can be distinguished. It is an interesting and open question how the results change when this assumption does not hold.

APPENDIX A
DERIVATIONS OF SECTION III-C

A. Upper Bounds

The upper bounds on capacity are all based on the duality technique that has successfully been used in previous works, see, e.g., [20] or [21]. For an arbitrary choice of a distribution $f(\cdot)$ on the channel output alphabet, we have

$$\mathsf{C} \leq \mathsf{E}_{\mu^*} \left[D\big(p_{Y|X}(\cdot|X) \big\| f(\cdot)\big) \right] \tag{76}$$

where $D(\cdot\|\cdot)$ is the relative entropy [22] and μ^* denotes the (unknown!) capacity-achieving input distribution. Note that Theorem 14 guarantees the existence and uniqueness of μ^*.

The challenge then lies in finding an elaborate choice of $f(\cdot)$ that is simple enough to allow the evaluation of (76), but that at the same time is good enough to result in a close bound. Moreover, we need to find ways of dealing with the expectation over the unknown μ^*.

1) Case with $0 < \alpha < \frac{1}{2}$: We make the following choice for $f(\cdot)$:

$$f(y) \triangleq \begin{cases} \dfrac{1}{\mathsf{T}} \dfrac{\beta(1-p)}{1 - e^{-\beta\left(1+\frac{\delta}{\mathsf{T}}\right)}} e^{-\frac{\beta}{\mathsf{T}}y} & 0 \leq y \leq \mathsf{T} + \delta, \\ p\eta\, e^{-\eta(y - \mathsf{T} - \delta)} & y > \mathsf{T} + \delta \end{cases} \tag{77}$$

where $\beta > 0, \eta > 0, \delta \geq 0$, and $0 \leq p \leq 1$ are free parameters. Equation (76) now evaluates as follows:

$$\mathsf{C}(\mathsf{T}, \alpha\mathsf{T})$$
$$\leq \mathsf{E}_{\mu^*}\Bigg[\bigg(\frac{\beta X}{\mathsf{T}} + \log \mathsf{T} + \log\Big(1 - e^{-\beta\left(1+\frac{\delta}{\mathsf{T}}\right)}\Big) - \log\beta$$
$$- \log(1-p)\bigg)\underbrace{\int_0^{\mathsf{T}+\delta-X}\sqrt{\frac{\lambda}{2\pi t^3}}\, e^{-\frac{\lambda(t-1)^2}{2t}}\, dt}_{\triangleq\, r_1(X)}\Bigg]$$
$$+ \mathsf{E}_{\mu^*}\Bigg[\frac{\beta}{\mathsf{T}}\underbrace{\int_0^{\mathsf{T}+\delta-X} t\sqrt{\frac{\lambda}{2\pi t^3}}\, e^{-\frac{\lambda(t-1)^2}{2t}}\, dt}_{\triangleq\, r_2(X)}\Bigg]$$
$$+ \mathsf{E}_{\mu^*}\Bigg[\eta\int_{\mathsf{T}+\delta-X}^{\infty} t\sqrt{\frac{\lambda}{2\pi t^3}}\, e^{-\frac{\lambda(t-1)^2}{2t}}\, dt\Bigg]$$
$$+ \mathsf{E}_{\mu^*}\Bigg[\Big(\eta X - \log p - \log\eta - \eta(\mathsf{T}+\delta)\Big)$$
$$\cdot\int_{\mathsf{T}+\delta-X}^{\infty}\sqrt{\frac{\lambda}{2\pi t^3}}\, e^{-\frac{\lambda(t-1)^2}{2t}}\, dt\Bigg] - h(N) \tag{78}$$
$$= \mathsf{E}_{\mu^*}\Bigg[\bigg(\frac{\beta X}{\mathsf{T}} + \log \mathsf{T} + \log\Big(1 - e^{-\beta\left(1+\frac{\delta}{\mathsf{T}}\right)}\Big) - \log\beta$$
$$- \log(1-p)\bigg)r_1(X) + \frac{\beta}{\mathsf{T}}r_2(X)$$
$$+ \big(\eta X - \log p - \log\eta - \eta(\mathsf{T}+\delta)\big)\big(1 - r_1(X)\big)$$
$$+ \eta\big(1 - r_2(X)\big)\Bigg] - h(N). \tag{79}$$

The integrals r_1 and r_2 can be evaluated as follows:

$$r_1(x) = g(\mathsf{T} + \delta - x, \lambda) \tag{80}$$
$$r_2(x) = \bar{g}(\mathsf{T} + \delta - x, \lambda) \tag{81}$$

where $g(\cdot, \cdot)$ and $\bar{g}(\cdot, \cdot)$ are given in Definition 1. From Lemma 2 now immediately follow the following bounds on r_1 and r_2.

Lemma 18. *The expressions $r_1(\cdot)$ and $r_2(\cdot)$ given in (81) satisfy:*

$$0 \leq g(\delta, \lambda) \leq r_1(x) \leq 1, \qquad \forall\, x \in [0, \mathsf{T}] \tag{82}$$
$$0 \leq \bar{g}(\delta, \lambda) \leq r_2(x) \leq 1, \qquad \forall\, x \in [0, \mathsf{T}]. \tag{83}$$

Moreover, for $\delta \geq 1$ and irrespective of λ,

$$g(\delta, \lambda) \geq \frac{1}{2}. \tag{84}$$

Proof: The lower bounds follow from the monotonicity of $g(\cdot, \cdot)$ and $\bar{g}(\cdot, \cdot)$ and the fact that $\mathsf{T} + \delta - x \geq \delta$. The other statements are given directly in Lemma 2. ∎

We now choose

$$p \triangleq 1 - \mathsf{E}_{\mu^*}[r_1(X)] \tag{85}$$
$$\eta \triangleq \frac{1 - \mathsf{E}_{\mu^*}[r_1(X)]}{\alpha\mathsf{T} + 1 - \mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]} \tag{86}$$

where from Lemma 18 it follows that $0 \leq p \leq 1$ and $\eta \geq 0$ as required. Then (79) looks as follows:

$$\mathsf{C}(\mathsf{T}, \alpha\mathsf{T})$$
$$\leq -h(N) - (1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\cdot\left(\frac{(\mathsf{T}+\delta)(1 - \mathsf{E}_{\mu^*}[r_1(X)])}{\alpha\mathsf{T} + 1 - \mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]} - 1\right)$$
$$- (1 - \mathsf{E}_{\mu^*}[r_1(X)])\log\frac{1 - \mathsf{E}_{\mu^*}[r_1(X)]}{\alpha\mathsf{T} + 1 - \mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]}$$
$$- (1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$+ \frac{\beta}{\mathsf{T}}\mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]$$
$$+ \mathsf{E}_{\mu^*}[r_1(X)]\Big(\log\mathsf{T} + \log\big(1 - e^{-\beta(1+\delta/\mathsf{T})}\big) - \log\beta$$
$$- \log\mathsf{E}_{\mu^*}[r_1(X)]\Big) \tag{87}$$
$$= -h(N) + (1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\cdot\left(1 - \underbrace{\frac{(\mathsf{T}+\delta)(1 - \mathsf{E}_{\mu^*}[r_1(X)])}{\alpha\mathsf{T} + 1 - \mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]}}_{\geq 0}\right)$$
$$- 2(1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$+ (1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\cdot\log\big(\alpha\mathsf{T} + 1 - \underbrace{\mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]}_{\geq 0}\big)$$
$$- \mathsf{E}_{\mu^*}[r_1(X)]\log\mathsf{E}_{\mu^*}[r_1(X)]$$
$$+ \frac{\beta}{\mathsf{T}}\mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]$$
$$+ \mathsf{E}_{\mu^*}[r_1(X)]\Big(\log\mathsf{T} - \log\beta + \log\big(1 - e^{-\beta(1+\delta/\mathsf{T})}\big)\Big) \tag{88}$$
$$\leq -h(N) + 1 - \mathsf{E}_{\mu^*}[r_1(X)]$$
$$- 2(1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$+ (1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(\alpha\mathsf{T} + 1)$$
$$- \mathsf{E}_{\mu^*}[r_1(X)]\log\mathsf{E}_{\mu^*}[r_1(X)]$$
$$+ \frac{\beta}{\mathsf{T}}\mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]$$
$$+ \mathsf{E}_{\mu^*}[r_1(X)]\Big(\log\mathsf{T} - \log\beta + \log\big(1 - e^{-\beta(1+\delta/\mathsf{T})}\big)\Big). \tag{89}$$

Further we choose

$$\beta \triangleq \frac{\alpha\mathsf{T}\, \mathsf{E}_{\mu^*}[r_1(X)]\beta^*}{\alpha\mathsf{T} + 1} > 0 \tag{90}$$

where β^* is the nonnegative solution to (46). It is not difficult to see that for $\alpha < 1/2$, this solution always exists and is

unique. With this choice we get

$$\frac{\beta}{\mathsf{T}}\mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]$$

$$= \alpha\beta^*\mathsf{E}_{\mu^*}[r_1(X)]\frac{\mathsf{E}_{\mu^*}[Xr_1(X) + r_2(X)]}{\alpha\mathsf{T} + 1} \quad (91)$$

$$\le \alpha\beta^*\mathsf{E}_{\mu^*}[r_1(X)] \quad (92)$$

and

$$-\log\beta = -\log(\alpha\mathsf{T}\beta^*) - \log\mathsf{E}_{\mu^*}[r_1(X)] + \log(\alpha\mathsf{T} + 1) \quad (93)$$

and

$$\log\left(1 - e^{-\beta(1+\delta/\mathsf{T})}\right) = \log\left(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)\mathsf{E}_{\mu^*}[r_1(X)]}{\alpha\mathsf{T}+1}}\right) \quad (94)$$

$$\le \log\left(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)}{\alpha\mathsf{T}+1}}\right). \quad (95)$$

Hence we obtain from (89)

$$C(\mathsf{T}, \alpha\mathsf{T})$$
$$\le -h(N) + 1 - \mathsf{E}_{\mu^*}[r_1(X)]$$
$$\quad - 2(1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\quad + (1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(\alpha\mathsf{T} + 1)$$
$$\quad - 2\mathsf{E}_{\mu^*}[r_1(X)]\log\mathsf{E}_{\mu^*}[r_1(X)]$$
$$\quad + \mathsf{E}_{\mu^*}[r_1(X)]\Big(\alpha\beta^* - \log\alpha\beta^* + \log(\alpha\mathsf{T} + 1)$$
$$\quad\quad + \log\Big(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)}{\alpha\mathsf{T}+1}}\Big)\Big) \quad (96)$$
$$= -h(N) + 1 - \mathsf{E}_{\mu^*}[r_1(X)]$$
$$\quad - 2(1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\quad - 2\mathsf{E}_{\mu^*}[r_1(X)]\log\mathsf{E}_{\mu^*}[r_1(X)] + \log(\alpha\mathsf{T} + 1)$$
$$\quad + \mathsf{E}_{\mu^*}[r_1(X)]\Big(\alpha\beta^* - \log\alpha\beta^* + \log\Big(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)}{\alpha\mathsf{T}+1}}\Big)\Big). \quad (97)$$

Next we point out that the mapping $z \mapsto 1 - z - 2(1 - z)\log(1 - z) - 2z\log z$ is monotonically decreasing for all $z \ge \frac{1}{\sqrt{e}+1} \approx 0.378$. Since by Lemma 18

$$\mathsf{E}_{\mu^*}[r_1(X)] \ge g(\delta, \lambda) \ge \frac{1}{2} > \frac{1}{\sqrt{e}+1} \quad (98)$$

as long as $\delta \ge 1$, we can bound

$$1 - \mathsf{E}_{\mu^*}[r_1(X)] - 2(1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\quad - 2\mathsf{E}_{\mu^*}[r_1(X)]\log\mathsf{E}_{\mu^*}[r_1(X)]$$
$$\quad \le 1 - g(\delta, \lambda) - 2(1 - g(\delta, \lambda))\log(1 - g(\delta, \lambda))$$
$$\quad\quad - 2g(\delta, \lambda)\log g(\delta, \lambda). \quad (99)$$

Finally, in order to be able to bound $\mathsf{E}_{\mu^*}[r_1(X)] \le 1$, we derive conditions that guarantee that

$$\alpha\beta^* - \log\alpha\beta^* + \log\Big(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)}{\alpha\mathsf{T}+1}}\Big) \ge 0. \quad (100)$$

To that goal, we define $\xi \triangleq \frac{\mathsf{T}+\delta}{\alpha\mathsf{T}+1}$ and note from (46) that

$$\alpha\beta^* = 1 - \frac{\beta^* e^{-\beta^*}}{1 - e^{-\beta^*}} \in [0, 1]. \quad (101)$$

With $\phi \triangleq \frac{\beta^* e^{-\beta^*}}{1-e^{-\beta^*}}$, where $0 \le \phi \le 1$, we then get that

$$\alpha\beta^* - \log\alpha\beta^* + \log\Big(1 - e^{-\frac{\alpha\beta^*(\mathsf{T}+\delta)}{\alpha\mathsf{T}+1}}\Big)$$
$$= 1 - \phi - \log(1 - \phi) + \log\Big(1 - e^{-\xi(1-\phi)}\Big) \ge 0 \quad (102)$$

if

$$\xi \ge \frac{\log\big(1 - (1-\phi)e^{\phi-1}\big)}{\phi - 1} \quad (103)$$

$$= 1 - \frac{1}{1-\phi}\log\big(e^{1-\phi} - (1-\phi)\big). \quad (104)$$

Note that

$$e^{1-\phi} - (1-\phi)$$
$$= 1 + \frac{1}{2}(1-\phi)^2 + \frac{1}{6}(1-\phi)^3 + \frac{1}{24}(1-\phi)^4 + \cdots \quad (105)$$
$$\ge 1 \quad (106)$$

such that Condition (104) (and therefore also (102)) is satisfied if $\xi \ge 1$, which is satisfied if $\delta \ge 1$.

2) Case with $\frac{1}{2} \le \alpha \le 1$: The derivation of this case is very similar to the derivations given in Section A-A1. Roughly speaking, we take over the results with $\beta^* \downarrow 0$. More precisely, we choose

$$f(y) \triangleq \begin{cases} \frac{1-p}{\mathsf{T}+\delta} & 0 \le y \le \mathsf{T}+\delta, \\ p\eta e^{-\eta(y-\mathsf{T}-\delta)} & y > \mathsf{T}+\delta \end{cases} \quad (107)$$

with $\delta \ge 0$ a free parameter, and with p and η chosen as in (85) and (86). Then (89) changes to

$$C(\mathsf{T}, \alpha\mathsf{T}) \le -h(N) + 1 - \mathsf{E}_{\mu^*}[r_1(X)]$$
$$\quad - 2(1 - \mathsf{E}_{\mu^*}[r_1(X)])\log(1 - \mathsf{E}_{\mu^*}[r_1(X)])$$
$$\quad - \mathsf{E}_{\mu^*}[r_1(X)]\log\mathsf{E}_{\mu^*}[r_1(X)] + \log(\alpha\mathsf{T} + 1)$$
$$\quad + \mathsf{E}_{\mu^*}[r_1(X)]\log\frac{\mathsf{T}+\delta}{\alpha\mathsf{T}+1}. \quad (108)$$

Using that the mapping $z \mapsto 1 - z - 2(1-z)\log(1-z) - z\log z$ is monotonically decreasing for all $z \ge \frac{(\sqrt{5}-1)^2}{4} \approx 0.382$, the remaining derivation follows then analogously to Section A-A1.

3) Case with Average-Delay Constraint Only: As discussed in Section III-A, the basic idea of the upper bound (24) was to upper-bound the output entropy by its maximum possible value, which will be achieved if the output happens to be exponentially distributed. It therefore does not come as a surprise that if we choose $f(\cdot)$ to be an exponential distribution, we can fully recover the upper bound (24) from (76).

The upper bounds (49), (50), and (51) are based on a choice of $f(\cdot)$ being a *power inverse Gaussian distribution* [23]:

$$f(y) = \sqrt{\frac{\delta}{2\pi\beta^3}}\left(\frac{\beta}{y}\right)^{1+\frac{\eta}{2}}$$
$$\quad \cdot \exp\left(-\frac{\delta}{2\eta^2\beta}\left(\left(\frac{y}{\beta}\right)^{\frac{\eta}{2}} - \left(\frac{\beta}{y}\right)^{\frac{\eta}{2}}\right)^2\right) \mathsf{I}\{y > 0\} \quad (109)$$

where $\beta, \delta > 0$, and $\eta \in \mathbb{R} \setminus \{0\}$ are free parameters. The family of power inverse Gaussian distributions contains the IG distribution as a special case for the choice $\eta = 1$.

We use the choice (109) in (76) and get

$$
\begin{aligned}
\mathsf{C_m}(\mathsf{M}) \\
\leq -h(N) &- \left(1 + \frac{\eta}{2}\right) \log \beta + \left(1 + \frac{\eta}{2}\right) \mathsf{E}_{\mu^*}[\log(X + N)] \\
&+ \frac{1}{2} \log \frac{2\pi}{\delta} + \frac{3}{2} \log \beta + \frac{\delta}{2\eta^2 \beta^{1+\eta}} \mathsf{E}_{\mu^*}[(X + N)^\eta] \\
&+ \frac{\delta}{2\eta^2 \beta^{1-\eta}} \mathsf{E}_{\mu^*}[(X + N)^{-\eta}] - \frac{\delta}{\eta^2 \beta}.
\end{aligned}
\tag{110}
$$

To minimize this upper bound we choose

$$
\delta \triangleq \frac{\eta^2 \beta}{\beta^{-\eta} \mathsf{E}_{\mu^*}[(X + N)^\eta] + \beta^\eta \mathsf{E}_{\mu^*}[(X + N)^{-\eta}] - 2}
\tag{111}
$$

$$
\beta \triangleq \mathsf{E}_{\mu^*}[(X + N)^\eta]^{\frac{1}{\eta}}
\tag{112}
$$

which yields

$$
\begin{aligned}
\mathsf{C_m}(\mathsf{M}) \\
\leq -h(N) &+ \left(1 + \frac{\eta}{2}\right) \mathsf{E}_{\mu^*}[\log(X + N)] + \frac{1}{2} \log \frac{2\pi e}{\eta^2} \\
&+ \frac{1}{2} \log \left(\mathsf{E}_{\mu^*}[(X + N)^{-\eta}] - \frac{1}{\mathsf{E}_{\mu^*}[(X + N)^\eta]} \right).
\end{aligned}
\tag{113}
$$

The first upper bound (49) now follows by picking $\eta \triangleq 1$ and continue bounding

$$
\mathsf{E}_{\mu^*}\left[\frac{1}{X + N}\right] \leq \mathsf{E}\left[\frac{1}{N}\right] = \frac{1}{\lambda} + 1
\tag{114}
$$

and

$$
\begin{aligned}
\mathsf{E}_{\mu^*}[\log(N + X)] \\
&\leq \mathsf{E}[\log N] + \mathsf{E}_{\mu^*}\left[\log\left(\mathsf{E}\left[1 + \frac{X}{N} \Big| X = x\right]\right)\right] \tag{115} \\
&\leq e^{2\lambda} \operatorname{Ei}(-2\lambda) + \log\left(1 + \mathsf{M}\left(1 + \frac{1}{\lambda}\right)\right) \tag{116}
\end{aligned}
$$

where we used twice Jensen's inequality.

For the second upper bound (50), we restrict the parameter η to $0 < \eta \leq 1$ such that $\xi \mapsto \xi^\eta$ is a concave function such that by Jensen's inequality:

$$
\mathsf{E}_{\mu^*}[(X + N)^\eta] \leq (\mathsf{E}_{\mu^*}[X + N])^\eta \leq (\mathsf{M} + 1)^\eta.
\tag{117}
$$

Moreover, for $\eta > 0$ we have

$$
\mathsf{E}_{\mu^*}[(X + N)^{-\eta}] \leq \mathsf{E}[N^{-\eta}] = \sqrt{\frac{2\lambda}{\pi}} e^\lambda \mathsf{K}_{\frac{1}{2}+\eta}(\lambda).
\tag{118}
$$

Plugging (117) and (118) into (113) then yields the second upper bound (50).

The third upper bound (51) follows from (113) with the choice $\eta \triangleq -1$ and the bounds (114) and (116).

B. Lower Bounds

For the lower bounds, we simply omit the maximization in the definition of capacity and choose a particular input distribution μ:

$$
\mathsf{C} \geq I(X; Y)\big|_{X \sim \mu} = h(Y)\big|_{X \sim \mu} - h(N).
\tag{119}
$$

To make the lower bound as tight as possible, we choose μ to satisfy the delay constraints with equality. Moreover, motivated by the fact that under a constraint on $\mathsf{E}[X]$, the differential entropy $h(X)$ is maximized by X being exponentially distributed, we choose μ to be a (truncated) exponential distribution.

1) Case with $0 < \alpha < \frac{1}{2}$: We choose

$$
\mu(x) \triangleq \frac{\beta^*}{\mathsf{T}(1 - e^{-\beta^*})} e^{-\frac{\beta^*}{\mathsf{T}} x} \mathsf{I}\{0 \leq x \leq \mathsf{T}\}
\tag{120}
$$

where β^* is chosen such that $\mathsf{E}_\mu[X] = \alpha\mathsf{T}$, i.e., β^* is the nonnegative solution to (46). For this given input distribution, the distribution at the channel output can now be derived as follows:[3]

$$
\begin{aligned}
p_Y(y) \\
&= \int_{-\infty}^{\infty} \mu(x)\, p_{Y|X}(y|x)\, \mathrm{d}x \tag{121} \\
&= \int_{-\infty}^{\infty} \frac{\beta^*}{\mathsf{T}(1 - e^{-\beta^*})} e^{-\frac{\beta^*}{\mathsf{T}} x} \mathsf{I}\{0 \leq x\} \mathsf{I}\{x \leq \mathsf{T}\} \\
&\qquad \cdot \sqrt{\frac{\lambda}{2\pi(y - x)^3}} e^{-\frac{\lambda(y - x - 1)^2}{2(y - x)}} \mathsf{I}\{y > x\}\, \mathrm{d}x \tag{122} \\
&= \int_{-\infty}^{\infty} \frac{\beta^*}{\mathsf{T}(1 - e^{-\beta^*})} e^{-\frac{\beta^*}{\mathsf{T}}(y - t)} \mathsf{I}\{0 \leq y - t\} \\
&\qquad \cdot \mathsf{I}\{y - t \leq \mathsf{T}\} \sqrt{\frac{\lambda}{2\pi t^3}} e^{-\frac{\lambda(t - 1)^2}{2t}} \mathsf{I}\{t > 0\}\, \mathrm{d}t \tag{123} \\
&= \frac{\beta^*}{\mathsf{T}(1 - e^{-\beta^*})} e^{-\frac{\beta^*}{\mathsf{T}} y + \lambda} \\
&\qquad \cdot \underbrace{\int_0^{\infty} \sqrt{\frac{\lambda}{2\pi}} t^{-3/2} e^{-t\left(\frac{\lambda}{2} - \frac{\beta^*}{\mathsf{T}}\right) - \frac{\lambda}{2t}} \mathsf{I}\{y - \mathsf{T} \leq t \leq y\}\, \mathrm{d}t}_{\triangleq r_0(y)}
\end{aligned}
\tag{124}
$$

where in (123) we have made the substitution $t = y - x$.

For the evaluation of $r_0(y)$, we will assume that $\frac{\lambda}{2} - \frac{\beta^*}{\mathsf{T}} > 0$, i.e., $\mathsf{T} > \frac{2\beta^*}{\lambda}$. Let $a^2 \triangleq \frac{\lambda}{2}$ and $b^2 \triangleq \frac{\lambda}{2} - \frac{\beta^*}{\mathsf{T}}$ and assume firstly that $y \leq \mathsf{T}$. Then, using the substitution $u = t^{-1/2}$, we get

$$
\begin{aligned}
r_0(y) \\
&= \int_0^y \sqrt{\frac{\lambda}{2\pi}} t^{-3/2} e^{-\frac{a^2}{t} - b^2 t}\, \mathrm{d}t \tag{125} \\
&= \sqrt{\frac{\lambda}{2\pi}} 2 \int_{\frac{1}{\sqrt{y}}}^{\infty} e^{-a^2 u^2 - \frac{b^2}{u^2}}\, \mathrm{d}u \tag{126} \\
&= \sqrt{\frac{\lambda}{2}} \frac{e^{2ab}}{2a} \left(1 - \operatorname{erf}\left(\frac{a}{\sqrt{y}} + b\sqrt{y}\right)\right) \\
&\quad + \sqrt{\frac{\lambda}{2}} \frac{e^{-2ab}}{2a} \left(1 - \operatorname{erf}\left(\frac{a}{\sqrt{y}} - b\sqrt{y}\right)\right) \tag{127}
\end{aligned}
$$

---

[3] The following derivation is inspired by [24].

$$= e^{-\lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}}\left(1 - Q\left(\sqrt{\lambda}\left(\sqrt{y\left(1-\frac{2\beta^*}{\lambda T}\right)} - \sqrt{\frac{1}{y}}\right)\right)\right.$$

$$\left. + e^{2\lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}} Q\left(\sqrt{\lambda}\left(\sqrt{y\left(1-\frac{2\beta^*}{\lambda T}\right)} - \sqrt{\frac{1}{y}}\right)\right)\right) \quad (128)$$

$$= e^{-\lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}} g\left(y\sqrt{1-\frac{2\beta^*}{\lambda T}}, \lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}\right) \quad (129)$$

$$\triangleq r_1(y). \quad (130)$$

Here, (127) follows from [25, 7.4.33], and $g(\cdot, \cdot)$ is given in Definition 1. If $y > T$, we get

$$r_0(y) = \int_{y-T}^{y} \sqrt{\frac{\lambda}{2\pi}}\, t^{-3/2}\, e^{-\frac{a^2}{t}-b^2 t}\, \mathrm{d}t \quad (131)$$

$$= \sqrt{\frac{\lambda}{2\pi}}\, 2 \int_{\frac{1}{\sqrt{y}}}^{\frac{1}{\sqrt{y-T}}} e^{-a^2 u^2 - \frac{b^2}{u^2}}\, \mathrm{d}u \quad (132)$$

$$= \sqrt{\frac{\lambda}{2\pi}}\, 2 \int_{\frac{1}{\sqrt{y}}}^{\infty} e^{-a^2 u^2 - \frac{b^2}{u^2}}\, \mathrm{d}u$$

$$\quad - \sqrt{\frac{\lambda}{2\pi}}\, 2 \int_{\frac{1}{\sqrt{y-T}}}^{\infty} e^{-a^2 u^2 - \frac{b^2}{u^2}}\, \mathrm{d}u \quad (133)$$

$$= r_1(y) - r_2(y) \quad (134)$$

where $r_1(y)$ is given in (130) and where

$$r_2(y) \triangleq e^{-\lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}} g\left((y-T)\sqrt{1-\frac{2\beta^*}{\lambda T}}, \lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}\right). \quad (135)$$

Hence,

$$p_Y(y) = \begin{cases} \frac{\beta^*}{T(1-e^{-\beta^*})}\, e^{-\frac{\beta^*}{T} y + \lambda}\, r_1(y) & 0 \le y \le T, \\ \frac{\beta^*}{T(1-e^{-\beta^*})}\, e^{-\frac{\beta^*}{T} y + \lambda}\, (r_1(y) - r_2(y)) & y > T. \end{cases} \quad (136)$$

The lower bound (119) now reads as follows:

$$C(T, \alpha T)$$

$$\ge \mathsf{E}[-\log p_Y(Y)] - h(N) \quad (137)$$

$$= \Pr[Y \le T]\mathsf{E}[-\log p_Y(Y)|Y \le T]$$
$$\quad + \Pr[Y > T]\mathsf{E}[-\log p_Y(Y)|Y > T] - h(N) \quad (138)$$

$$= \log T - \log\beta^* + \log\left(1-e^{-\beta^*}\right) + \frac{\beta^*}{T}\mathsf{E}[Y] - \lambda$$
$$\quad - h(N) - \Pr[Y \le T]\mathsf{E}[\log r_1(Y)|Y \le T]$$
$$\quad - \Pr[Y > T]\mathsf{E}\big[\log\big(r_1(Y) - \underbrace{r_2(Y)}_{\ge 0}\big)\big|Y > T\big] \quad (139)$$

$$\ge \log T - \log\beta^* + \log\left(1-e^{-\beta^*}\right) + \frac{\beta^*}{T}\mathsf{E}[Y] - \lambda$$
$$\quad - h(N) - \mathsf{E}[\log r_1(Y)] + \lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}$$
$$\quad - \mathsf{E}\left[\log g\left(Y\sqrt{1-\frac{2\beta^*}{\lambda T}}, \lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}\right)\right] \quad (140)$$

$$\ge \log T - \log\beta^* + \log\left(1-e^{-\beta^*}\right) + \alpha\beta^* + \frac{\beta}{T} - \lambda$$

$$\quad + \lambda\sqrt{1-\frac{2\beta^*}{\lambda T}} - \frac{1}{2}\log\frac{2\pi e}{\lambda} - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda)$$
$$\quad - \log g\left((\alpha T + 1)\sqrt{1-\frac{2\beta^*}{\lambda T}}, \lambda\sqrt{1-\frac{2\beta^*}{\lambda T}}\right) \quad (141)$$

where in the last step we used concavity (Lemma 3) and Jensen's inequality.

2) Case with $\frac{1}{2} \le \alpha \le 1$: We choose

$$\mu(x) \triangleq \frac{1}{T}\, \mathsf{I}\{0 \le x \le T\}. \quad (142)$$

A derivation analogous to (121)–(135) then yields

$$p_Y(y) = \begin{cases} \frac{1}{T} e^\lambda\, \tilde{r}_1(y) & 0 \le y \le T, \\ \frac{1}{T} e^\lambda\, (\tilde{r}_1(y) - \tilde{r}_2(y)) & y > T \end{cases} \quad (143)$$

where

$$\tilde{r}_1(y) \triangleq e^{-\lambda}\, g(y, \lambda) \quad (144)$$
$$\tilde{r}_2(y) \triangleq e^{-\lambda}\, g(y - T, \lambda). \quad (145)$$

The remainder is analogous to Section A-B1.

3) Case with Average-Delay Constraint Only: Choosing

$$\mu(x) \triangleq \frac{1}{M}\, e^{-x/M}\, \mathsf{I}\{x \ge 0\} \quad (146)$$

we get analogously to (121)–(135) that

$$p_Y(y) = \frac{1}{M}\, e^{-\frac{y}{M}+\lambda}\, r_1(y)\, \mathsf{I}\{y > 0\} \quad (147)$$

and therefore

$$C_m(M)$$

$$\ge \log M + \frac{1}{M}\mathsf{E}[Y] - \lambda - \mathsf{E}[\log r_1(Y)] - h(N) \quad (148)$$

$$= \log M + \frac{M+1}{M} - \lambda + \lambda\sqrt{1-\frac{2}{\lambda M}} - \frac{1}{2}\log\frac{2\pi e}{\lambda}$$

$$\quad - \frac{3}{2} e^{2\lambda} \operatorname{Ei}(-2\lambda)$$

$$\quad - \mathsf{E}\left[\log g\left(Y\sqrt{1-\frac{2}{\lambda M}}, \lambda\sqrt{1-\frac{2}{\lambda M}}\right)\right]. \quad (149)$$

The lower bound (48) now follows from Jensen's inequality.

APPENDIX B
DERIVATIONS OF SECTION IV

A. Proof of Lemma 16

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

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

$$= h(Y; \mu_\theta) - h(Y; \mu_0) \quad (150)$$

$$= \int_0^\infty p_Y(y; \mu_\theta) \log\frac{1}{p_Y(y; \mu_\theta)}\, \mathrm{d}y$$
$$\quad - \int_0^\infty p_Y(y; \mu_0) \log\frac{1}{p_Y(y; \mu_0)}\, \mathrm{d}y \quad (151)$$

$$= -D\big(p_Y(\cdot; \mu_\theta)\big\|p_Y(\cdot; \mu_0)\big)$$
$$\quad + \theta\int_0^\infty \big(p_Y(y; \mu_0) - p_Y(y; \mu)\big) \log p_Y(y; \mu_0)\, \mathrm{d}y. \quad (152)$$

Based on [26, Lemma on p. 1023], for any pair of probability measures $P_1 \ll P_0$, we have

$$\lim_{\theta \downarrow 0} \frac{1}{\theta} D\big(\theta P_1 + (1-\theta)P_0 \big\| P_0\big) = 0. \tag{153}$$

Since for every pair of probability measures μ_0 and μ in $\bar{\Lambda}'_0$, $p_Y(y;\mu) \ll p_Y(y;\mu_0)$, we obtain

$$I'_{\mu_0}(\mu)$$
$$= \lim_{\theta \downarrow 0} -\frac{1}{\theta} D\big(p_Y(\cdot;\mu_\theta) \big\| p_Y(\cdot;\mu_0)\big)$$
$$\quad + \int_0^\infty \big(p_Y(y;\mu_0) - p_Y(y;\mu)\big) \log p_Y(y;\mu_0)\, \mathrm{d}y \tag{154}$$
$$= \int_0^\infty p_Y(y;\mu) \log \frac{1}{p_Y(y;\mu_0)}\, \mathrm{d}y - I(\mu_0) - h(N) \tag{155}$$
$$= \int Q(x;\mu_0)\, \mathrm{d}\mu(x) - I(\mu_0) - h(N). \tag{156}$$

B. Proof of Theorem 17

Define the Lagrangian

$$J(\mu) \triangleq I(\mu) - \nu \mathsf{E}_\mu[X - \mathsf{M}] \tag{157}$$

where ν is the Lagrangian multiplier. By the method of Lagrangian multipliers [27], μ_0 is capacity achieving if and only if μ_0 and ν are such that
(a) $\nu \mathsf{E}_{\mu_0}[X - \mathsf{M}] = 0$,
(b) for all $\mu \in \bar{\Lambda}'_0$, $J(\mu_0) \geq J(\mu)$.
Since $I(\mu)$ is a strictly concave function on the convex set $\bar{\Lambda}$, $J(\mu)$ is also concave and the maximizer of (21) or (22) must exist and be unique. Condition (b) is equivalent to $J'_{\mu_0}(\mu) \leq 0$, where $J'_{\mu_0}(\mu)$ is the weak derivative of $J(\mu)$ at μ_0 defined as

$$J'_{\mu_0}(\mu) \triangleq \lim_{\theta \downarrow 0} \frac{J\big((1-\theta)\mu_0 + \theta\mu\big) - J(\mu_0)}{\theta} \tag{158}$$
$$= I'_{\mu_0}(\mu) - \nu\big(\mathsf{E}_\mu[X] - \mathsf{E}_{\mu_0}[X]\big) \tag{159}$$

where $I'_{\mu_0}(\mu)$ is given in (68). Then

$$J'_{\mu_0}(\mu) = \int Q(x;\mu_0)\, \mathrm{d}\mu(x) - h(N) - I(\mu_0)$$
$$\quad - \nu\big(\mathsf{E}_\mu[X] - \mathsf{E}_{\mu_0}[X]\big). \tag{160}$$

We first prove the necessity of (69). Suppose μ_0 achieves the capacity. For any given $x^* \geq 0$ for the case of an average-delay constraint and $0 \leq x^* \leq \mathsf{T}$ for the case of a peak- and an average-delay constraint, let μ be the point measure at x^*, i.e., $\mu(\{x^*\}) = 1$, which clearly satisfies $\mu \in \bar{\Lambda}$. Substituting μ into (160), we have, by the optimality of μ_0 and the equivalence of Condition (b) and $J'_{\mu_0}(\mu) \leq 0$,

$$0 \geq J'_{\mu_0}(\mu) \tag{161}$$
$$= Q(x^*;\mu_0) - h(N) - I(\mu_0) - \nu(x^* - \mathsf{M})$$
$$\quad + \nu \mathsf{E}_{\mu_0}[X - \mathsf{M}] \tag{162}$$
$$= Q(x^*;\mu_0) - h(N) - I(\mu_0) - \nu(x^* - \mathsf{M}) \tag{163}$$

where the in the last equality we used Condition (a). This establishes (69).

We next prove the sufficiency of (69). Suppose inequality (69) is satisfied. We integrate both side of (69) with respect to μ_0 and obtain

$$0 \geq \int Q(x;\mu_0)\, \mathrm{d}\mu_0(x) - h(N) - I(\mu_0)$$
$$\quad - \nu \mathsf{E}_{\mu_0}[X - \mathsf{M}] \tag{164}$$
$$= -\nu \mathsf{E}_{\mu_0}[X - \mathsf{M}] \tag{165}$$
$$\geq 0, \tag{166}$$

where (165) follows from (16), and (166) holds because $\mu_0 \in \bar{\Lambda}_0$. Hence $\nu \mathsf{E}_{\mu_0}[X - \mathsf{M}] = 0$. Furthermore, if we integrate (69) with respect to an arbitrary measure $\mu \in \bar{\Lambda}_0^\cup$, we have

$$0 \geq \int Q(x;\mu_0)\, \mathrm{d}\mu(x) - h(N) - I(\mu_0)$$
$$\quad - \nu \int (x - \mathsf{M})\, \mathrm{d}\mu(x) \tag{167}$$
$$= I'_{\mu_0}(\mu) - \nu \mathsf{E}_\mu[X - \mathsf{M}] \tag{168}$$
$$= I'_{\mu_0}(\mu) - \nu\big(\mathsf{E}_\mu[X] - \mathsf{E}_{\mu_0}[X]\big) - \nu\big(\mathsf{E}_{\mu_0}[X] - \mathsf{M}\big) \tag{169}$$
$$= J'_{\mu_0}(\mu), \tag{170}$$

where (168) follows from (156), and (170) follows from (159) and (166). Since Condition (b) is equivalent to $J'_{\mu_0}(\mu) \leq 0$, Conditions (a) and (b) are thus both satisfied. Hence μ_0 achieves the capacity.

Finally, we prove that inequality (69) is satisfied with equality at all points of increase. Define

$$a(x;\mu_0) \triangleq Q(x;\mu_0) - h(N) - I(\mu_0) - \nu(x - \mathsf{M}) \leq 0 \tag{171}$$

and by contradiction suppose that the inequality holds strictly at a point of increase x^*, i.e., $a(x^*;\mu_0) = -\varepsilon$ for some $\varepsilon > 0$. Since $p_Y(y;\mu_0) > 0$ for $y > 0$ and $p_N(y-x)$ is continuous for $y \geq x$, $Q(x;\mu_0) = -\int_0^\infty p_N(y-x) \log p_Y(y;\mu_0)\, \mathrm{d}y$ is continuous on $x \geq 0$ and thus $a(x;\mu_0)$ is also continuous on $x \geq 0$. 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 (69) with respect to μ_0, we obtain

$$0 = \int a(x;\mu_0)\, \mathrm{d}\mu_0(x) \leq \int_\mathbb{O} a(x;\mu_0)\, \mathrm{d}\mu_0(x) \tag{172}$$
$$< -\frac{1}{2}\varepsilon\mu_0(\mathbb{O}). \tag{173}$$

Contradiction arises because $\varepsilon\mu_0(\mathbb{O}) > 0$, hence Theorem 17 is proved.

REFERENCES

[1] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Comp. Networks*, vol. 52, no. 12, pp. 2260–2279, 2008.

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

[3] K. V. Srinivas, A. W. Eckford, and R. S. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4678–4692, Jul. 2012.

[4] A. W. Eckford, K. V. Srinivas, and R. S. Adve, "The peak constrained additive inverse Gaussian noise channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6, 2012, pp. 2983–2987.

[5] H.-T. Chang and S. M. Moser, "Bounds on the capacity of additive inverse Gaussian noise channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6, 2012, pp. 304–308.

[6] H. Li and D. Guo, "On the Capacity-achieving Input for Additive Inverse Gaussian Channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 7–12, 2013, pp. 1829–1833.

[7] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, no. 3, pp. 203–219, Feb. 1971.

[8] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," *Proc. IEE Commun. Speech and Vision*, vol. 137, no. 6, pp. 424–430, Dec. 1990.

[9] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.

[10] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[11] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

[12] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.

[13] 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, no. 5, pp. 2193–2206, Sep. 2005.

[14] J. Huang and S. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2336–2351, Jul. 2005.

[15] L. Zhang, H. Li, and D. Guo, "Capacity of Gaussian channels with duty cycle and power constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1615–1629, Mar. 2014.

[16] M. N. Khormuji, "On the capacity of molecular communication over the AIGN channel," in *Proc. Annu. Conf. Inf. Sci. Syst.*, Baltimore, MD, USA, Mar. 23–25, 2011, pp. 1–4.

[17] D. W. Stroock, *Probability Theory, an Analytic View*. New York: Cambridge Univ. Press, 1993.

[18] Y. V. Prokhorov, "Convergence of random processes and limit theorems in probability theory," *Theory Probab. Appl.*, vol. 1, pp. 157–214, 1956.

[19] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.

[20] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[21] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley-Interscience, 2006.

[23] T. Kawamura and K. Iwase, "Characterizations of the distributions of power inverse Gaussian and others based on the entropy maximization principle," *J. Japan Statist. Soc.*, vol. 33, no. 1, pp. 95–104, Jan. 2003.

[24] W. Schwarz, "On the convolution of inverse Gaussian and exponential random variables," *Commun. in Statistics — Theory and Methods*, vol. 31, no. 12, pp. 2113–2121, 2002.

[25] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, 10th ed. National Bureau of Standards, 1964.

[26] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[27] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
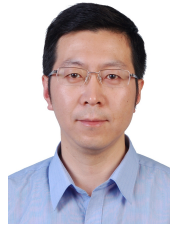
Hui Li received the B.Eng. degree in 1999 and Ph.D. degree in 2004, all from the University of Science and Technology of China, Hefei, China. He joined the faculty of University of Science and Technology of China, Hefei, China, in 2004, where he is currently an Associate Professor in the Department of Electronic Engineering and Information Science. He has held visiting positions at Northwestern University, Evanston, IL, in 2011–2012. His research interests include wireless communication and information theory.

Stefan M. Moser (S'01–M'05–SM'10) received the diploma (M.Sc.) in electrical engineering in 1999, the M.Sc. degree in industrial management (M.B.A.) in 2003, and the Ph.D. degree (Dr. sc. techn.) in the field of information theory in 2004, all from ETH Zürich, Switzerland.

From 1999 to 2003, he was a Research and Teaching Assistant, and from 2004 to 2005, he was a Senior Research Assistant with the Signal and Information Processing Laboratory at ETH Zürich. From 2005 to 2013, he has been a Professor with the Department of Electrical and Computer Engineering at National Chiao Tung University (NCTU), Hsinchu, Taiwan. Currently he is a Senior Researcher with the Signal and Information Processing Laboratory at ETH Zürich and an Adjunct Professor with the Department of Electrical and Computer Engineering at National Chiao Tung University. His research interests are in information theory and digital communications.

Dr. Moser is recipient of the Wu Ta-You Memorial Award by the National Science Council of Taiwan in 2012, and the Best Paper Award for Young Scholars by the IEEE Communications Society Taipei and Tainan Chapters and the IEEE Information Theory Society Taipei Chapter in 2009. Further he received various awards from the National Chiao Tung University, so two awards for outstanding teaching in 2007 and 2012, and he was presented with the Willi Studer Award of ETH and the ETH Medal both in 1999, and with the Sandoz (Novartis) Basler Maturandenpreis in 1993.
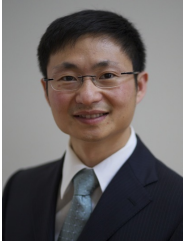
Dongning Guo (S'97–M'05–SM'11) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Eng. degree from the National University of Singapore, Singapore, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ.

In 2004, he joined the faculty of Northwestern University, Evanston, IL, where he is currently an Associate Professor in the Department of Electrical Engineering and Computer Science. From 1998 to 1999, he was an R&D Engineer in the Center for Wireless Communications (now the Institute for Infocom Research), Singapore. He has held visiting positions at the Norwegian University of Science and Technology in summer 2006 and the Chinese University of Hong Kong during 2010–2011. His research interests include information theory, communications, and networking.

Dr. Guo has been an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY and a Guest Editor of a Special Issue of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is an Editor of Foundations and Trends in Communications and Information Theory. He received the Huber and Suhner Best Student Paper Award in the International Zurich Seminar on Broadband Communications in 2000 and is the corecipient of the IEEE Marconi Prize Paper Award in Wireless Communications in 2010 (with Y. Zhu and M. L. Honig). He is also the recipient of the National Science Foundation Faculty Early Career Development (CAREER) Award in 2007.