Capacity of Gaussian Channels with Duty Cycle and Power Constraints

Lei Zhang and Dongning Guo
Department of Electrical Engineering and Computer Science
Northwestern University, Evanston, IL 60208, USA
Email: {lzhang, dguo}@northwestern.edu

***Abstract*—In many wireless communication systems, radios are subject to *duty cycle* constraint, that is, a radio only actively transmits signals over a fraction of the time. For example, it is desirable to have a small duty cycle in some low power systems; a half-duplex radio cannot keep transmitting if it wishes to receive useful signals; and a cognitive radio needs to listen and detect primary users frequently. This work studies the capacity of scalar discrete-time Gaussian channels subject to duty cycle constraint as well as average transmit power constraint. The duty cycle constraint can be regarded as a requirement on the minimum fraction of nontransmission or zero symbols in each codeword. A unique *discrete* input distribution is shown to achieve the channel capacity. In many situations, numerical results demonstrate that using the optimal input can improve the capacity by a large margin compared to using Gaussian input, which is capacity-achieving in the absence of the duty cycle constraint. This is because the positions of the nontransmission symbol in a codeword can convey information. The results suggest that, under duty cycle constraint, departing from the usual paradigm of intermittent packet transmissions may yield substantial gain.**

I. Introduction

In many wireless communication systems, a radio is designed to transmit actively only for a fraction of the time, which is known as its *duty cycle*. For example, the ultra-wideband system proposed in [1] employs impulse radio to transmit short bursts of signals to trade bandwidth for power savings. The physical half-duplex constraint also requires a radio to stop transmission from time to time if it wishes to receive useful signals. Hence wireless relays are subject to duty cycle constraint, so do cognitive radios which have to listen to the channel frequently to avoid causing interference to primary users. The *de facto* standard solution under duty cycle constraint is to transmit packets intermittently.

This work studies the question of what is the optimal signaling for a Gaussian channel with duty cycle constraint as well as average transmission power constraint. For simplicity, we consider a discrete-time scalar additive white Gaussian noise (AWGN) channel. We assume the analog waveform corresponding to each symbol is of the length of exactly one symbol interval to keep the discussion concise. An important observation is that the signaling in nontransmission periods can be regarded as transmission of the special symbol *zero*, so that the duty cycle constraint is equivalent to a requirement on the minimum fraction of zero symbols in each transmitted codeword. The mathematical model of the AWGN channel and input constraints is described in Section II.

Determining the capacity of a channel subject to various input constraints is a classical problem. It is well-known that Gaussian signaling achieves the capacity of a Gaussian channel with input power constraint only. Smith [2] investigated the capacity of a scalar AWGN channel under both peak power constraint and average power constraint. The input distribution that achieves the capacity is shown to be discrete with a finite number of probability mass points. The discreteness of capacity-achieving distributions for various channels, including quadrature Gaussian channels, and Rayleigh-fading channels is also established in [3]–[7]. Chan [8] studied the capacity-achieving input distribution for conditional Gaussian channels which form a general channel model for many practical communication systems.

The impact of duty cycle constraint on capacity-achieving signaling is underexplored in the literature. In Section III of this paper, we use a similar approach as in [2] and [8] to show that the capacity-achieving input distribution for an AWGN channel with duty cycle and power constraints is discrete. The optimal distribution has an infinite number of probability mass points, whereas only a finite number of the points are found in every bounded interval. This enables efficient numerical optimization of the input distribution.

Numerical results in Section IV demonstrate that significant gain is possible using discrete signaling with finite probability mass points compared to using Gaussian signaling. For example, in case the radio is allowed to transmit no more than half the time, i.e., the duty cycle is no greater than 50%, a near-optimal discrete input achieves 50% higher rate than Gaussian signaling at 10 dB signal-to-noise ratio (SNR). This suggests that, compared to intermittently transmitting packets using Gaussian or Gaussian-like signaling, it is more efficient to disperse nontransmission symbols within each packet to form codewords, which results in a form of *on-off* signaling.

A key reason for the superiority of on-off signaling is that the positions of nontransmission symbols can be used to convey a substantial amount of information, especially in case of low SNR and low duty cycle. This has been observed in the past. For example, as shown in [9] (see also [10], [11]), time sharing or time-division duplex (TDD) can fall considerably short of the theoretical limits in a relay network: The capacity of a cascade of two noiseless binary bit pipes through a half-duplex relay is 1.14 bits per channel use, which far exceeds the 0.5 bit achieved by TDD and even the 1 bit upper bound on the rate of binary signaling.

Besides that duty cycle constraint is frequently seen in practice, another motivation of this study is a recent work reported in [12], in which on-off signaling is proposed for a clean-slate design of wireless ad hoc networks formed by half-duplex radios. Using this signaling scheme, called rapid on-off-division duplex (RODD), a node listens to the channel and receives useful signals during its own off symbols within each frame. Each node can transmit and receive messages at the same time over one frame interval, thereby achieving (virtual) full-duplex communication. Understanding the impact of duty cycle constraint is crucial to characterizing the fundamental limits of such wireless networks.

II. System Model

Consider digital communication systems where coded data are mapped to waveforms for transmission. Usually there is a collection of waveforms, all of one symbol duration, where each waveform represents a symbol (or letter) from a discrete alphabet. Without loss of capacity, we assume some linear modulation scheme is used, such as pulse amplitude modulation (PAM) or quadrature amplitude modulation (QAM). We view nontransmission over a symbol interval as transmitting the all zero waveform. In other words, a symbol interval of nontransmission can be regarded as transmitting the symbol 0, which carries no energy.

For simplicity, we consider the baseband discrete-time model for the AWGN channel. The received signal over a block of n symbols can be described by $Y_i = X_i + N_i$, $i = 1, \dots, n$, where X_i denotes the transmitted symbol at time i and $N_1, \dots, N_n$ are independent standard Gaussian random variables. The constraint that the duty cycle is no greater than $1-q$, where $q \in (0,1)$, can be considered as a constraint that the fraction of symbol 0 in the input codeword is no less than q. That is, every codeword $(x_1, x_2, \dots, x_n)$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq 0\}} \leq 1-q \tag{1}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function. In addition, we consider a frequently-used constraint that the average input power or the SNR of the channel is no greater than P.

III. The Capacity-Achieving Input

Let μ denote the distribution (a.k.a. probability measure) of the channel input X. The corresponding output probability density function of $Y = X + N$ with standard Gaussian N exists and is

$$p_Y(y;\mu) = \int p_{Y|X}(y|x) d\mu = \mathsf{E}_\mu \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-X)^2}{2}} \right\} .$$

The differential entropy of Y, expressed as

$$h_Y(\mu) = -\int_{-\infty}^{\infty} p_Y(y;\mu) \log p_Y(y;\mu) dy \tag{2}$$

relates to the mutual information $I(\mu) = I(X;Y)$ by the following:

$$I(\mu) = h_Y(\mu) - \frac{1}{2}\log(2\pi e). \tag{3}$$

We define a set of probability measures $\Lambda = \{\mu : \mu(\{0\}) \geq q, \mathsf{E}_\mu\{X^2\} < \infty\}$ and its subset $\Lambda_0 = \{\mu : \mu(\{0\}) \geq q, \mathsf{E}_\mu\{X^2\} \leq P\}$.

In the following, we apply a similar approach as in [2] and [8] to study the channel capacity problem under two constraints. A key difference here is that the input is not necessarily bounded due to the lack of a peak power constraint. The existence and uniqueness of the capacity-achieving input distribution are guaranteed by the following result.

Theorem 1: The capacity of an AWGN channel with duty cycle no greater than $1-q$ and SNR no greater than P is

$$C = \max_{\mu \in \Lambda_0} I(\mu), \tag{4}$$

which is achieved by a unique probability measure $\mu_0 \in \Lambda_0$, which is symmetric about 0.

There are three parts to Theorem 1: (i) the existence and uniqueness of the maximizer of (4), (ii) the achievability of C, and (iii) the converse. Suppose for now (i) is true and denote the unique maximizer of (4) by μ_0. Since the mirror reflection of μ_0 about the origin is evidently also a maximizer of (4), the uniqueness requires that μ_0 be symmetric. The achievability of $I(\mu_0)$ can be shown using Shannon's random codebook generated according to distribution μ_0. To establish the converse, we view the duty cycle constraint (1) as a per-letter cost constraint on the input and find that higher rates than C are unachievable due to Fano's inequality and concavity of $I(\mu)$. The techniques for showing (ii) and (iii) are standard in information theory, so the details are omitted due to space limitations. In the following, we provide a detailed proof of (i).

Proof (The existence and uniqueness of the maximizer): We shall establish the existence using the fact that a continuous function achieves its maximum in a compact set of a metric space. The uniqueness follows by strict concavity of $I(\mu)$.

Let $(\mathbb{R}, d)$ be one-dimensional Euclidean metric space with Borel sigma algebra $\mathcal{B}$. Let $\mathcal{P}$ denote the collection of all probability measures on the measurable space $(\mathbb{R}, \mathcal{B})$. Given $\mu, \nu \in \mathcal{P}$, define the Lévy-Prohorov metric [13] as

$$\begin{aligned} L(\mu,\nu) = \inf \big\{ \delta : & \mu(F) \leq \nu(F^{(\delta)}) + \delta \text{ and} \\ & \nu(F) \leq \mu(F^{(\delta)}) + \delta \text{ for all } F \subseteq \mathcal{B} \big\} \end{aligned} \tag{5}$$

where $F^{(\delta)}$ denotes the set of all $x \in \mathbb{R}$ which lie a d-distance less than δ from F. Then L is a complete metric for $\mathcal{P}$. Since $\mathbb{R}$ is a Polish space, convergence of probability measures in the Lévy-Prohorov metric is equivalent to weak convergence of measures. The following two lemmas are useful:

Lemma 1: Λ_0 is compact in the metric space $(\mathcal{P}, L)$.

Lemma 2: $I(\mu)$ is continuous on Λ_0.

Due to space limitations, the proof of all lemmas in this paper is omitted. Interested readers are referred to a longer version of this work [14] for more details. By Lemmas 1 and 2, the mutual information $I(\mu)$ achieves its maximum in Λ_0. The existence of the capacity-achieving input distribution then follows.

To show the uniqueness, let $\mu_0, \mu_1 \in \Lambda_0$ be two capacity-achieving input distributions. For any $\theta \in (0,1)$, define $\mu_\theta = \theta\mu_0 + (1-\theta)\mu_1$, which is also in Λ_0 due to its convexity, then

$$p_Y(y;\mu_\theta) = \theta p_Y(y;\mu_0) + (1-\theta)p_Y(y;\mu_1). \tag{6}$$

By the strict concavity of the function $-x\log x$, we have

$$h_Y(\mu_\theta) \geq \theta h_Y(\mu_0) + (1-\theta)h_Y(\mu_1) \tag{7}$$

with equality if and only if $p_Y(y;\mu_0) = p_Y(y;\mu_1)$. Since μ_0 and μ_1 are both capacity-achieving measures, (7) holds with equality, i.e., $p_Y(y;\mu_0) = p_Y(y;\mu_1)$. Since the Fourier transform of the probability density of the Gaussian noise N is nonzero everywhere, $p_Y(y;\mu)$ and μ are in one-to-one correspondence (see [8]). Hence the uniqueness of the capacity-achieving input distribution is proved. ∎

A point $x \in \mathbb{R}$ is said to be a point of increase of a probability measure μ if $\mu(O) > 0$ for every open subset O of $\mathbb{R}$ containing x. Let S_μ be the set of points of increase of μ.

Theorem 2: The capacity-achieving input probability measure μ_0 is discrete with an infinite number of probability mass points, i.e., S_{μ_0} is countably infinite. Moreover, the number of probability mass points inside any bounded interval is finite.

To prove Theorem 2, we need the following result.

Theorem 3: Let $\mu_0 \in \Lambda_0$ and $h_N = \frac{1}{2}\log(2\pi e)$. Then μ_0 is capacity achieving if and only if there exists $\lambda \geq 0$ such that for all $x \in \mathbb{R}$,

$$qd(0) + (1-q)d(x) \leq 0 \tag{8}$$

where

$$d(x) = h(x;\mu_0) - h_N - I(\mu_0) - \lambda(x^2 - P) \tag{9}$$

and $h(x;\mu_0) = -\int_{-\infty}^{\infty} p_{Y|X}(y|x)\log p_Y(y;\mu_0)dy$. Furthermore, the equality of (8) holds for all $x \in S_{\mu_0}\backslash\{0\}$.

Proof of Theorem 3: We need the following result:

Lemma 3: If two input probability measures μ_1, μ_2 satisfy that $\mathsf{E}_{\mu_1}\left\{X^2\right\} < \infty$ and $\mathsf{E}_{\mu_2}\left\{X^2\right\} < \infty$, then $-\int_{-\infty}^{\infty} p_Y(y;\mu_2)\log p_Y(y;\mu_1)dy < \infty$.

It is easy to see that Λ is a convex set, and $I(\mu)$ is finite for any $\mu \in \Lambda$ by Lemma 3. Define the Lagrangian $J(\mu) = I(\mu) - \lambda\mathsf{E}_\mu\left\{X^2 - P\right\}$, where λ is the Lagrange multiplier. Then μ_0 is capacity achieving if and only if there exists $\lambda \geq 0$ such that the following two conditions hold [15]:

(a) $\lambda\mathsf{E}_{\mu_0}\left\{X^2 - P\right\} = 0$;
(b) for all $\mu \in \Lambda$, $J(\mu_0) \geq J(\mu)$.

Due to the concavity of $h_Y(\mu)$ by (7), $J(\mu)$ is also concave. Condition (b) is then equivalent to that $J'_{\mu_0}(\mu) \leq 0$ for all $\mu \in \Lambda$, where $J'_{\mu_0}(\mu)$ is the weak derivative of $J(\mu)$ at μ_0 defined as

$$J'_{\mu_0}(\mu) = \lim_{\theta\to 0^+} \frac{J\left((1-\theta)\mu_0 + \theta\mu\right) - J(\mu_0)}{\theta}. \tag{10}$$

The following result is useful and a similar result with proof can be found in [5], [8].

Lemma 4: Let $\mu_0 \in \Lambda$, the weak derivative of the mutual information function $I(\mu)$, $\mu \in \Lambda$ at μ_0 is

$$I'_{\mu_0}(\mu) = \int h(x;\mu_0)d\mu - h_N - I(\mu_0). \tag{11}$$

Now by Lemma 4, the linearity of $\mathsf{E}_\mu\left\{X^2 - P\right\}$ w.r.t. μ and Condition (a), $J'_{\mu_0}(\mu)$ can be calculated as

$$\begin{aligned} J'_{\mu_0}(\mu) &= \int h(x;\mu_0)d\mu - h_N - I(\mu_0) - \lambda\mathsf{E}_\mu\left\{X^2 - P\right\} \\ &= \mathsf{E}_\mu\left\{d(X)\right\} \end{aligned} \tag{12}$$

where $h(x;\mu_0)$ is well defined due to the following result:

Lemma 5: For any $\mu \in \Lambda$ and $w \in \mathbb{C}$,

$$h(w;\mu) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-w)^2}{2}} \log p_Y(y;\mu)dy \tag{13}$$

is well defined and it is a holomorphic function of w on $\mathbb{C}$. Consequently, $h(x;\mu) = -\int_{-\infty}^{\infty} p_{Y|X}(y|x)\log p_Y(y;\mu)dy$ is a continuous function of x on $\mathbb{R}$.

Therefore, according to Conditions (a) and (b), μ_0 is capacity achieving if and only if $\lambda\mathsf{E}_{\mu_0}\left\{X^2 - P\right\} = 0$ and $\mathsf{E}_\mu\left\{d(x)\right\} \leq 0$ for all $\mu \in \Lambda$.

The necessity part of Theorem 3 is shown as follows. Suppose μ_0 is capacity achieving and λ is chosen to satisfy $\lambda\mathsf{E}_{\mu_0}\left\{X^2 - P\right\} = 0$ and $\mathsf{E}_\mu\left\{d(X)\right\} \leq 0$ for all $\mu \in \Lambda$. For any $x \in \mathbb{R}\backslash\{0\}$, choose μ such that $\mu(\{0\}) = q$ and $\mu(\{x\}) = 1-q$, so by the fact that $\mu \in \Lambda$, we have

$$0 \geq \mathsf{E}_\mu\left\{d(X)\right\} = qd(0) + (1-q)d(x). \tag{14}$$

Due to the continuity of $h(x;\mu_0)$ by Lemma 5, $d(x)$ is also continuous so that (14) holds for all $x \in \mathbb{R}$.

Define

$$a(x) = qd(0) + (1-q)d(x), \tag{15}$$

then $a(x) \leq 0$ for all $x \in \mathbb{R}$ and evidently $a(0) = d(0)$. Next we show that $a(x) = 0$ for all $x \in S_{\mu_0}\backslash\{0\}$. Let $x_0 \in S_{\mu_0}\backslash\{0\}$, and suppose, to the contrary, that $a(x_0) = -\epsilon < 0$. By the continuity of $a(x)$, there exists an open set $\mathcal{B}$ containing x_0 such that $0 \notin \mathcal{B}$ and $a(x) \leq -\frac{\epsilon}{2}$ for all $x \in \mathcal{B}$. Then,

$$\int a(x)d\mu_0 \leq qa(0) + \left(-\frac{\epsilon}{2}\right)\mu_0(\mathcal{B}) < qd(0). \tag{16}$$

On the other hand, however, by (9) and (15),

$$\begin{aligned} \int a(x)d\mu_0 &= qd(0) + (1-q)\Big(\int h(x;\mu_0)d\mu_0 \\ &\qquad - h_N - I(\mu_0) - \lambda\mathsf{E}_{\mu_0}\left\{X^2 - P\right\}\Big) \\ &= qd(0) + (1-q)\left(h_Y(\mu_0) - h_N - I(\mu_0)\right) \quad (17) \\ &= qd(0) \quad (18) \end{aligned}$$

where we use the fact that $\lambda\mathsf{E}_{\mu_0}\left\{X^2 - P\right\} = 0$ in (17). A contradiction occurs, so it is proved that $a(x) = 0$ for all $x \in S_{\mu_0}\backslash\{0\}$, which implies the necessity part.

The sufficiency part of Theorem 3 is established next.

Suppose $a(x) = qd(0) + (1-q)d(x) \leq 0$ for all $x \in \mathbb{R}$. By integrating $a(x)$ w.r.t. μ_0, we have, by (15),

$$\begin{aligned} qa(0) &\geq \int a(x) d\mu_0 \\ &= qa(0) - (1-q)\lambda \mathsf{E}_{\mu_0}\left\{X^2 - P\right\} \quad (19) \\ &\geq qa(0) \quad (20) \end{aligned}$$

where (19) is due to $\mathsf{E}_{\mu_0}\{h(x;\mu_0) - h_N - I(\mu_0)\} = 0$ and $a(0) = d(0)$, and (20) follows from $\mathsf{E}_{\mu_0}\left\{X^2\right\} \leq P$ since $\mu_0 \in \Lambda_0$. Hence, $\lambda \mathsf{E}_{\mu_0}\left\{X^2 - P\right\} = 0$ due to the fact that $q < 1$. Furthermore, for any $\mu \in \Lambda$, by integrating $a(x)$ w.r.t. μ, we have, again by (15),

$$qa(0) \geq \int a(x) d\mu = qd(0) + (1-q)\mathsf{E}_{\mu}\{d(X)\}.$$

Because $a(0) = d(0)$, we have $\mathsf{E}_{\mu}\{d(X)\} \leq 0$. Together with $\lambda \mathsf{E}_{\mu_0}\left\{X^2 - P\right\} = 0$, μ_0 must be capacity achieving. ■

With the necessary and sufficient condition for the capacity-achieving input distribution established as in Theorem 3, we next prove Theorem 2.

Proof of Theorem 2: First we extend the function $d(x)$ in Theorem 3 to be defined on the whole complex plane $\mathbb{C}$ as

$$d(w) = h(w;\mu_0) - h_N - I(\mu_0) - \lambda(w^2 - P) \quad (21)$$

where $h(w;\mu_0)$ is defined in (13) and $\lambda \geq 0$ satisfies the condition (8). By Lemma 5, $h(w;\mu_0)$ is a holomorphic function of w on $\mathbb{C}$. Therefore, $a(w) = qd(0) + (1-q)d(w)$ is also holomorphic on $\mathbb{C}$. According to Theorem 3, each element in the set $S_{\mu_0}\backslash\{0\}$ is a zero of the function $a(w)$.

Next we show that for any bounded interval L of $\mathbb{R}$, $S_{\mu_0} \bigcap L$ is a finite set. Suppose, to the contrary, $S_{\mu_0} \bigcap L$ is infinite, then it has a limit point in $\mathbb{R}$ by the Bolzano-Weierstrass Theorem [16] and hence, $a(w) = 0$ on the whole complex plane $\mathbb{C}$ by the Identity Theorem [17]. Then, by (9) and (15),

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{(x-y)^2}{2}}}{\sqrt{2\pi}} \left(\log p_Y(y;\mu_0) + D + \lambda(y^2 - P - 1)\right) dy = 0 \quad (22)$$

for every $x \in \mathbb{R}$, where $D = h_N + I(\mu_0) - \frac{q}{1-q}d(0)$ is a constant. This is to say that the convolution of a standard Gaussian density and $g(y) = \log p_Y(y;\mu_0) + D + \lambda(y^2 - P - 1)$ is equal to the zero function.

By Jensen's inequality, we have

$$\begin{aligned} p_Y(y;\mu) &= \mathsf{E}_{\mu}\left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{(y-X)^2}{2}}\right\} \\ &\geq \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mathsf{E}_{\mu}\left\{(y-X)^2\right\}} = e^{-\frac{1}{2}y^2 - ay - b} \quad (23) \end{aligned}$$

where $a = -\mathsf{E}_{\mu}\{X\}$ and $b = \frac{1}{2}\left(\mathsf{E}_{\mu}\left\{X^2\right\} + \log(2\pi)\right)$. By the assumption that $\mathsf{E}_{\mu}\left\{X^2\right\} < \infty$, it is easy to see that $a, b \in \mathbb{R}$, so $|\log P_Y(y;\mu)| \leq \frac{1}{2}y^2 + ay + b$. As a result, there exist some $\alpha, \beta > 0$ such that $|g(y)| \leq \alpha y^2 + \beta$. Therefore, according to [8, Corollary 9], $g(y)$ is the zero function, which suggests that capacity-achieving output distribution $p_Y(y;\mu_0)$ is Gaussian. This requires X to be Gaussian, which has no probability mass at 0 as desired. Therefore, $S_{\mu_0} \bigcap L$ must be a finite set for any bounded interval L, which further implies that S_{μ_0} is at most countable because $S_{\mu_0} = \bigcup_{n=1}^{\infty} (S_{\mu_0} \bigcap (-n, n))$.

Finally, we show that S_{μ_0} is countably infinite. Suppose, to the contrary, $S_{\mu_0} = \{x_i\}_{i=1}^N$ is a finite set with $\mu_0(\{x_i\}) = p_i$ and $|x_i| \leq B_1$ for all $i = 1, 2, \ldots, N$. For any $y > B_1$, we have

$$p_Y(y;\mu_0) = \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x_i)^2}{2}} p_i \leq e^{-\frac{(y-B_1)^2}{2}}. \quad (24)$$

For any $\epsilon > 0$, choose $B_2 > 0$ such that $\int_{-B_2}^{B_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx > 1 - \epsilon$. Now according to (8) and (9), for any $x > B_1 + B_2$,

$$\begin{aligned} 0 &\geq -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}} \log p_Y(y;\mu_0) dy - D - \lambda(x^2 - P) \\ &\geq \int_{x-B_2}^{x+B_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}} \cdot \frac{1}{2}(y - B_1)^2 dy - D - \lambda(x^2 - P) \\ &\geq \frac{1}{2}(x - B_1)^2 (1 - \epsilon) - D - \lambda(x^2 - P). \quad (25) \end{aligned}$$

In order to make (25) hold for very large x, λ must satisfy $\lambda \geq \frac{1}{2}$. On the other hand, however, it can be proved that $\lambda < \frac{1}{2}$ for any $P > 0$. A contradiction occurs, thus it follows that $\boldsymbol{S}_{\mu_0}$ is countably infinite.

To finish the proof, it suffices to show that $\lambda < \frac{1}{2}$ for any $P > 0$. For fixed $q \in (0, 1)$, denote the channel capacity defined in (4) by $C(P)$ and the lagrange multiplier as $\lambda(P)$. Denote $C_G(P) = \frac{1}{2}\log(1 + P)$, which is the channel capacity of a Gaussian channel with the average power constraint only. By the envelope theorem [15], $\lambda(P)$ is the derivative of $C(P)$ w.r.t. P. Since $C(0) = C_G(0) = 0$ and the derivative of $C_G(P)$ at $P = 0$ is $\frac{1}{2}$, we have $\lambda(0) \leq \frac{1}{2}$, otherwise we could find a small enough P such that $C(P)$ would exceed $C_G(P)$ which is obviously impossible. Next we show that $C(P)$ is strictly concave for $P \geq 0$. Suppose μ_1 and μ_2 are the capacity-achieving input distributions of (4) for different P_1 and P_2, respectively. For any $\theta \in (0, 1)$, define $\mu_\theta = \theta\mu_1 + (1-\theta)\mu_2$. It is easy to see that μ_θ satisfies that the duty cycle is no greater than $1 - q$ and the average input power is no greater than $\theta P_1 + (1-\theta)P_2$. As argued in the proof of Theorem 1, we have

$$C(\theta P_1 + (1-\theta)P_2) \geq I(\mu_\theta) \geq \theta I(\mu_1) + (1-\theta)I(\mu_2), \quad (26)$$

where the second inequality becomes equality if and only if $\mu_1 = \mu_2$. This is, however, impossible by the following arguments. Without loss of generality, let $P_2 > P_1$ and $a = \sqrt{P_2/P_1} > 1$. Let μ_2' be the input distribution of a random variable aX, where the distribution of X is μ_1. It is obvious that

$$\begin{aligned} C(P_2) &\geq I(\mu_2') = h_Y(\mu_1) - \frac{1}{2}\log(2\pi e/a^2) \\ &> I(\mu_1) = C(P_1). \quad (27) \end{aligned}$$

Therefore, the strict concavity of $C(P)$ for $P \geq 0$ follows by (26). Hence, $\lambda(P)$ i.e., the derivative of $C(P)$ is strictly

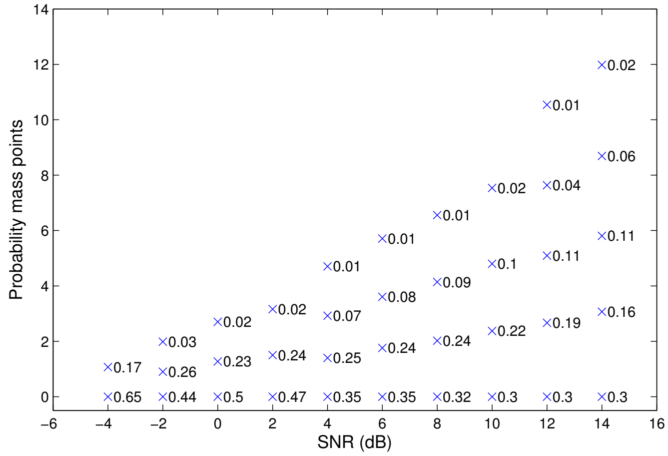

Fig. 1. Suboptimal input distribution for $P(X=0) \geq q = 0.3$

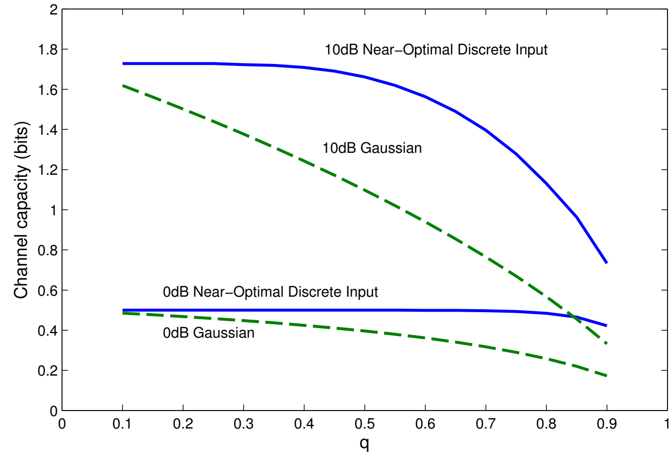

Fig. 2. Capacity under duty cycle constraint for 0dB and 10dB SNRs

decreasing for $P \geq 0$, which implies that $\lambda(P) < \lambda(0) = \frac{1}{2}$ for all $P > 0$. ■

IV. Numerical Results

One implication of Theorem 2 is that directly computing the capacity-achieving input distribution requires solving an optimization problem with infinite variables which is prohibitive. Assuming any upper bound on the number of probability mass points, however, a numerical optimization over the mutual information can yield a suboptimal input distribution and a lower bound on the channel capacity. As we increase the number of mass points allowed, the lower bound can be further refined. We take this approach to numerically compute a good approximation of the channel capacity by optimizing over a sufficient number of probability mass points. Given the duty cycle and power constraints, we first numerically optimize the mutual information by a 3-point input distribution (including a mass at 0), then increase the number of probability mass points by 2 at a time to improve the mutual information, until the improvement is less than 10^{-3}.

First consider the case that the duty cycle is no greater than 70%, i.e., $P(X=0) \geq q = 0.3$. For different SNRs, the mass points of the near-optimal input distribution with finite support along with the corresponding probability masses are shown in Fig. 1. Due to symmetry, only one half of the input distribution is plotted. We can see that as the SNR increases, more masses are put on higher-amplitude points, whereas the probability mass at zero achieves its lower bound 0.3 eventually.

In Fig 2, we compare rate achieved by the near-optimal input distribution and rate achieved by the conventional TDD scheme, which is $(1-q)$ times the Gaussian channel capacity without duty cycle constraint. It is shown in the figure that there is substantial gain for both 0 dB and 10 dB SNRs by using discrete input over Gaussian signaling. For example, at 10 dB SNR, given the duty cycle is no more than 50%, the discrete input distribution achieves 50% higher rate than TDD. Hence departing from the usual paradigm of intermittent packet transmissions may yield substantial gain.

Acknowledgement

The authors would like to thank Terence Chan for sharing the codes for the numerical results in [8].

References

[1] D. Julian and S. Majumdar, "Low power personal area communication," in *Proc. Inform. Theory Appl. Workshop*, La Jolla, CA, USA, 2011.

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

[3] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *Proc. IEE Communications, Speech and Vision*, vol. 137, pp. 424–430, Dec. 1990.

[4] S. Shamai 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.

[5] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.

[6] M. Katz and S. Shamai, "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] M. Gursoy, H. Poor, and S. Verdu, "The noncoherent Rician fading channel-part I: structure of the capacity-achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2193–2206, 2005.

[8] 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.

[9] T. Lutz, C. Hausl, and R. Kötter, "Coding strategies for noise-free relay cascades with half-duplex constraint," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 2385–2389, Toronto, ON, Canada, July 2008.

[10] G. Kramer, "Communication strategies and coding for relaying," *Wireless Networks*, vol. 143 of The IMA Volumes in Mathematics and its Applications, pp. 163–175, 2007.

[11] T. Lutz, G. Kramer, and C. Hausl, "Capacity for half-duplex line networks with two sources," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 2393–2397, Austin, TX, USA, June 2010.

[12] D. Guo and L. Zhang, "Rapid on-off-division duplex for mobile ad hoc neworks," in *Proc. Allerton Conf. Commun., Control, & Computing*, Monticello, IL, USA, 2010.

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

[14] L. Zhang and D. Guo, "Capacity of Gaussian channels with duty cycle and power constraints," *available on arxiv*.

[15] D. G. Luenberger, *Optimization by Vector Sapce Methods*. New York: Wiley, 1969.

[16] S. Lang, *Complex Analysis*. New York: Springer-Verlag, 1999.

[17] W. Rudin, *Real and Complex Analysis*. McGraw-Hill Science Engineering, 1986.