Increased Capacity per Unit-Cost by Oversampling

Tobias Koch
University of Cambridge
Cambridge, UK
Email: tobi.koch@eng.cam.ac.uk

Amos Lapidoth
ETH Zurich
Zurich, Switzerland
Email: lapidoth@isi.ee.ethz.ch

***Abstract*—It is demonstrated that doubling the sampling rate recovers some of the loss in capacity incurred on the bandlimited Gaussian channel with a one-bit output quantizer.**

I. Introduction

We study the capacity of the continuous-time, bandlimited, additive white Gaussian noise (AWGN) channel with one-bit output quantization. Our focus is on the capacity at low transmit powers, i.e., on the capacity per unit-cost, which is defined as the slope of the capacity-vs-input-power curve at zero. We show that increasing the sampling rate reduces the loss in capacity per unit-cost caused by the quantization.

The capacity of the continuous-time AWGN channel without output quantization was studied by Shannon [1]. He showed that if the channel input is bandlimited to W Hz and satisfies the average-power constraint P, and if the additive Gaussian noise is of double-sided power spectral density $\mathsf{N}_0/2$, then the capacity (in nats per second) is given by (see also [2])

$$C(\mathsf{P}) = \mathsf{W}\log\left(1+\frac{\mathsf{P}}{\mathsf{W}\mathsf{N}_0}\right) \tag{1}$$

where $\log(\cdot)$ denotes the natural logarithm function. This capacity can be achieved by transmitting

$$X(t) = \sum_{\ell=-\infty}^{\infty} X_\ell \operatorname{sinc}(2\mathsf{W}t-\ell), \quad t\in\mathbb{R} \tag{2}$$

(where $\mathbb{R}$ denotes the set of real numbers), and by sampling the output $Y(\cdot)$ at Nyquist rate $2\mathsf{W}$. Here $\{X_\ell,\ \ell\in\mathbb{Z}\}$ (where $\mathbb{Z}$ denotes the set of integers) is a sequence of independent and identically distributed (IID) Gaussian random variables of zero mean and variance P, and $t\mapsto\operatorname{sinc}(t)$ denotes the sinc-function, i.e., $\operatorname{sinc}(0)=1$ and $\operatorname{sinc}(t)=\sin(\pi t)/(\pi t)$ for $t\neq 0$.

The above (capacity-achieving) transmission scheme reduces the continuous-time channel to a discrete-time AWGN channel with inputs $\{X_\ell,\ \ell\in\mathbb{Z}\}$ and outputs $\{Y(\ell/(2\mathsf{W})),\ \ell\in\mathbb{Z}\}$. Yet, it is often required that the channel inputs and outputs be not only discrete in time, but also take on a discrete value, i.e., take value in a finite set rather than in $\mathbb{R}$. This is, for example, the case if the transmitter and receiver use digital signal processing techniques. To ensure that the channel inputs are discrete-valued, we can simply restrict ourselves to finite input alphabets. This restriction is not critical for small input powers P. Indeed, it is well-known that binary inputs achieve the capacity per unit-cost of the AWGN channel [1]. To ensure that the channel outputs are discrete-valued, we have to employ a quantizer (analog-to-digital converter), which approximates the continuous-valued output by a finite number of bits.

The capacity (in nats per channel use) of the discrete-time AWGN channel with binary symmetric output quantization—where the quantizer produces 1 for a nonnegative output and -1 for a negative output—is given by

$$\log 2 - H_b\left(Q\left(\sqrt{\mathsf{P}}/\sigma\right)\right) \tag{3}$$

where σ^2 denotes the variance of the additive noise, $H_b(\cdot)$ the binary entropy function, and $Q(\cdot)$ the Q-function; see [3, (3.4.18)], [4, p. 107], [5, Thm. 2]. To the best of our knowledge, there exists no closed-form expression for the capacity of the discrete-time AWGN channel with nonbinary output quantization. However, numerical results are, for example, given in [5]. Furthermore, there exist analytical results concerning the capacity per-unit cost. For example, it was demonstrated that if a binary symmetric quantizer is employed, then the capacity per unit-cost equals $\frac{1}{\pi}\frac{1}{\sigma^2}$ [3, (3.4.20)]. It was further demonstrated that for an octal quantizer with uniform quantization the capacity per unit-cost is not less than $0.475\frac{1}{\sigma^2}$ [3, (3.4.21)]. Thus, at low transmit power, employing a binary quantizer causes a loss of a factor of $2/\pi$ relative to the capacity per unit-cost $\frac{1}{2}\frac{1}{\sigma^2}$ for unquantized decoding [1]. In contrast, by quantizing the output with 3 bits, a capacity per unit-cost can be achieved that is close to the capacity per unit-cost for unquantized decoding. (Note that the capacity of discrete-time channels is measured in nats per channel use, whereas the capacity of continuous-time channels is measured in nats per second. Since with a continuous-time signal of bandwidth W Hz we can approximately transmit $2\mathsf{W}$ samples per second, we have that one nat per channel use corresponds to $2\mathsf{W}$ nats per second. By further noting that lowpass filtering and sampling Gaussian noise of double-sided power spectral density $\mathsf{N}_0/2$ yields Gaussian noise-samples of variance $\mathsf{W}\mathsf{N}_0$, it follows that the capacity per unit-cost of the continuous-time channel corresponds to $2\mathsf{W}$ times the capacity

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 252663.

978-1-4244-8682-3/10/$26.00 ©2010 IEEE

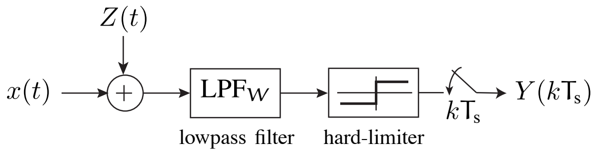

Fig. 1. System model.

per unit-cost of the discrete-time channel with σ^2 replaced by $W\mathsf{N}_0$.)

The above results suggest that, in order to reduce the loss in capacity per unit-cost caused by the quantization, one needs to increase the quantizer's resolution. However, while this clearly holds for the discrete-time channel, this does not necessarily hold for the underlying continuous-time channel. Indeed, in contrast to the unquantized channel output, the quantized output is not bandlimited, and it is therefore *prima facie* not clear, whether sampling the quantized output at Nyquist rate is optimal. One might thus increase the capacity of the continuous-time channel by oversampling the quantized output, i.e., by sampling the quantized output at rates higher than the Nyquist rate.

When there is no additive noise, it was shown by Gilbert [6] and by Shamai [7] that oversampling indeed increases the capacity. In this paper, we demonstrate that oversampling also increases the capacity when the noise power is large relative to the transmit power. In particular, we show that, for binary symmetric output quantization, sampling the quantized output at twice the Nyquist rate yields a capacity per unit-cost that is not less than $0.75\frac{1}{\mathsf{N}_0}$, which is strictly larger than the capacity per unit-cost $\frac{2}{\pi}\frac{1}{\mathsf{N}_0} \approx 0.64\frac{1}{\mathsf{N}_0}$ that can be achieved by sampling the quantized output at Nyquist rate.

The rest of this paper is organized as follows. Section II describes the channel model. Section III defines channel capacity and the capacity per unit-cost and presents the main result. Section IV outlines the proof of the main result. Section V concludes the paper with a discussion of our results.

II. Channel Model

We consider the communication channel depicted in Figure 1 whose input $x(\cdot)$ is bandlimited to W Hz and satisfies the average-power constraint P. The channel output $Y(k\mathsf{T}_\mathsf{s})$ at integer multiples of the sampling interval $\mathsf{T}_\mathsf{s} > 0$ is

$$Y(k\mathsf{T}_\mathsf{s}) = \operatorname{sgn}\left\{\left((\mathbf{x}+\mathbf{Z})\star \mathrm{LPF}_W\right)(k\mathsf{T}_\mathsf{s})\right\}, \quad k \in \mathbb{Z} \tag{4}$$

where $\operatorname{sgn}\{\cdot\}$ denotes the sign function; $(\mathbf{a}\star\mathbf{b})(t)$ the convolution between $a(\cdot)$ and $b(\cdot)$ at time t; and $\mathrm{LPF}_W(\cdot)$ is the impulse response of the ideal unit-gain lowpass filter of cutoff frequency W. The hard-limiter is a binary symmetric quantizer that produces 1 for a nonnegative input and -1 for a negative input. We assume that $\{Z(t),\, t\in\mathbb{R}\}$ is zero-mean white Gaussian noise of double-sided power spectral density $\mathsf{N}_0/2$.

Without loss of optimality, we restrict ourselves to signals $x(\cdot)$ of the form

$$x(t) = \frac{1}{\sqrt{2W}}\sum_{\ell=-\infty}^{\infty} x_\ell\, g\left(t - \frac{\ell}{2W}\right), \quad t \in \mathbb{R} \tag{5}$$

where $g(\cdot)$ is some unit-energy waveform that is bandlimited to W Hz. Indeed, by the Sampling Theorem [8, Thm. 8.4.5], any signal $x(\cdot)$ that is bandlimited to W Hz can be written as (5) with

$$x_\ell = x\left(\frac{\ell}{2W}\right), \quad \ell \in \mathbb{Z}$$

and

$$g\,(t) = \sqrt{2W}\operatorname{sinc}(2Wt), \quad t \in \mathbb{R}.$$

III. Channel Capacity and Capacity per Unit-Cost

We define the capacity (in nats per second) as

$$C_{\mathsf{T}_\mathsf{s}}(\mathsf{P}) \triangleq \varliminf_{n\to\infty} \sup \frac{2W}{n} I\left(X_1^n;\mathbf{Y}_1^n\right) \tag{6}$$

where the supremum is over all unit-energy waveforms $g(\cdot)$ that are bandlimited to W Hz and over all joint distributions on $(X_1, X_2, \ldots, X_n)$ satisfying $\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right] \le \mathsf{P}$. Here we use A_m^n to denote the sequence $A_m, A_{m+1}, \ldots, A_n$ and

$$\mathbf{Y}_k \triangleq \left(Y\left(\left\lceil\frac{2k-1}{4W\mathsf{T}_\mathsf{s}}\right\rceil \mathsf{T}_\mathsf{s}\right), \ldots, Y\left(\left\lfloor\frac{2k+1}{4W\mathsf{T}_\mathsf{s}}\right\rfloor \mathsf{T}_\mathsf{s}\right)\right)$$

(with $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ denoting the ceiling and the floor function). A more general definition of channel capacity for continuous-time channels can be found in [2, Sec. 8.1]. For the above channel (4), the capacity $C_{\mathsf{T}_\mathsf{s}}(\mathsf{P})$ defined by (6) is a lower bound on the capacity defined in [2, Sec. 8.1]. The two capacities coincide, for example, for the continuous-time AWGN channel (without output quantization).

That oversampling can increase the capacity of the above channel has been demonstrated in the noiseless case, i.e., when $\mathsf{N}_0 = 0$. In particular, Gilbert [6] showed that, for a Gaussian input $X(\cdot)$, sampling the output at twice the Nyquist rate yields an information rate of $2.14W$ bits per second, which is strictly larger than the $2W$ bits per second that can be achieved by sampling the output at Nyquist rate. Shamai [7] further showed, *inter alia*, that by sampling the output at η-times the Nyquist rate, rates of $2W\log(1+\eta)$ nats per second are achievable by transmitting a bandlimited process that possesses a single real zero within each Nyquist interval. In the absence of noise it is thus possible to trade amplitude resolution versus time resolution.

In this paper, we focus on the case where the variance of the additive noise is large relative to the transmit power. In particular, we study the capacity per unit-cost, defined as

$$\dot{C}_{\mathsf{T}_\mathsf{s}}(0) \triangleq \varlimsup_{\mathsf{P}\downarrow 0} \frac{C_{\mathsf{T}_\mathsf{s}}(\mathsf{P})}{\mathsf{P}}. \tag{7}$$

By the Data Processing Inequality [9, Thm. 2.8.1] it follows that quantizing the output does not increase the capacity. This implies that the capacity per unit-cost is upper bounded by the capacity per unit-cost of the continuous-time AWGN channel (without output quantization)

$$\dot{C}_{\mathsf{T_s}}(0) \leq \lim_{\mathsf{P}\to\infty} \frac{W \log\left(1+\frac{\mathsf{P}}{W\mathsf{N}_0}\right)}{\mathsf{P}} = \frac{1}{\mathsf{N}_0}. \tag{8}$$

For the case where the output is sampled at Nyquist rate $1/\mathsf{T_s} = 2W$, it was shown that the capacity per unit-cost is given by [3, (3.4.20)]

$$\dot{C}_{\frac{1}{2W}}(0) = \frac{2}{\pi}\frac{1}{\mathsf{N}_0} \approx 0.637\frac{1}{\mathsf{N}_0}. \tag{9}$$

Thus, when we sample the output at Nyquist rate, hard-limiting causes a loss of a factor of $2/\pi$. This loss can be reduced by sampling the output at twice the Nyquist rate:

Theorem 1 (Main Result): Sampling the output at rate $4W$ yields

$$\begin{aligned}\dot{C}_{\frac{1}{4W}}(0) \geq \frac{2}{\pi}\frac{1}{\mathsf{N}_0}\Bigg[&\frac{\left(\frac{g_1}{2}+\frac{g_0}{4}+\frac{g_1}{\pi}\vartheta_1-\frac{g_0}{2\pi}\vartheta_2\right)^2}{\frac{1}{4}+\frac{1}{\pi}\arcsin(\rho)} \\ &+8\left(\frac{g_0}{4}-\frac{g_1}{\pi}\vartheta_1+\frac{g_0}{2\pi}\vartheta_2\right)^2 \\ &+\frac{\left(\frac{g_1}{2}-\frac{g_0}{4}-\frac{g_1}{\pi}\vartheta_1+\frac{g_0}{2\pi}\vartheta_2\right)^2}{\frac{1}{4}-\frac{1}{\pi}\arcsin(\rho)}\Bigg] \\ \approx 0.747\frac{1}{\mathsf{N}_0}\end{aligned} \tag{10}$$

where $\rho = \frac{2}{\pi}$, $\vartheta_1 = \arcsin\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)$, $\vartheta_2 = \arcsin\left(\frac{\rho^2}{1-\rho^2}\right)$, and

$$g_0 = \left.\frac{1+\frac{2}{\pi}\lambda}{\sqrt{\frac{1}{2}\lambda^2+\frac{4}{\pi}\lambda+1}}\right|_{\lambda=1.4}$$

$$g_1 = \left.\frac{\frac{2}{\pi}+\frac{1}{2}\lambda}{\sqrt{\frac{1}{2}\lambda^2+\frac{4}{\pi}\lambda+1}}\right|_{\lambda=1.4}.$$

Proof: The proof is outlined in Section IV. A full proof can be found in [10, Sec. 4.2]. ■

The main ingredients in the proof of Theorem 1 are expansions of the complementary cumulative distribution function (CCDF) of bivariate and trivariate Gaussian vectors around the orthant probability.[1] We present these expansions in the following two propositions.

Proposition 2: Let $(x,y) \mapsto \phi_{\mathbf{0},\mathsf{K}}(x,y)$ denote the probability density function (PDF) of the bivariate, zero-mean, Gaussian vector of covariance matrix

$$\mathsf{K} = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}$$

for $|\varrho| < 1$. Then, for every $\mathsf{A} \geq 0$, $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$,

$$\begin{aligned}&\int_{-\alpha\mathsf{A}}^{\infty}\int_{-\beta\mathsf{A}}^{\infty}\phi_{\mathbf{0},\mathsf{K}}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &\quad= \frac{1}{4}+\frac{1}{2\pi}\arcsin(\varrho)+\frac{\alpha+\beta}{2}\frac{\mathsf{A}}{\sqrt{2\pi}}+\Delta(\mathsf{A},\alpha,\beta)\end{aligned} \tag{11}$$

where

$$|\Delta(\mathsf{A},\alpha,\beta)| \leq \mathsf{A}^2\eta(\mathsf{A},\alpha,\beta)$$

and where $\eta(\mathsf{A},\alpha,\beta) = \eta(\mathsf{A},|\alpha|,|\beta|)$ is monotonically increasing in $(\mathsf{A},|\alpha|,|\beta|)$ and is bounded for every finite A, α, and β.

Proof: See [10, Sec. 4.1.1]. ■

Proposition 3: Let $(x,y,z) \mapsto \phi_{\mathbf{0},\mathsf{K}}(x,y,z)$ denote the PDF of the trivariate, zero-mean, Gaussian vector of covariance matrix

$$\mathsf{K} = \begin{pmatrix} 1 & \varrho_{12} & \varrho_{13} \\ \varrho_{12} & 1 & \varrho_{23} \\ \varrho_{13} & \varrho_{23} & 1 \end{pmatrix}$$

for $|\varrho_{12}| < 1$, $|\varrho_{13}| < 1$, $|\varrho_{23}| < 1$ satisfying $\det(\mathsf{K}) > 0$ (where $\det(\mathsf{K})$ denotes the determinant of the matrix K). Then, for every $\mathsf{A} \geq 0$, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, and $\gamma \in \mathbb{R}$,

$$\begin{aligned}&\int_{-\alpha\mathsf{A}}^{\infty}\int_{-\beta\mathsf{A}}^{\infty}\int_{-\gamma\mathsf{A}}^{\infty}\phi_{\mathbf{0},\mathsf{K}}(x,y,z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x \\ &\quad= \frac{1}{8}+\frac{1}{4\pi}\big(\arcsin(\varrho_{12})+\arcsin(\varrho_{13})+\arcsin(\varrho_{23})\big) \\ &\qquad+\frac{\mathsf{A}}{\sqrt{2\pi}}\Bigg[\frac{\alpha+\beta+\gamma}{4} \\ &\qquad\quad+\frac{\alpha}{2\pi}\arcsin\left(\frac{\varrho_{23}-\varrho_{12}\varrho_{13}}{\sqrt{(1-\varrho_{12}^2)(1-\varrho_{13}^2)}}\right) \\ &\qquad\quad+\frac{\beta}{2\pi}\arcsin\left(\frac{\varrho_{13}-\varrho_{12}\varrho_{23}}{\sqrt{(1-\varrho_{12}^2)(1-\varrho_{23}^2)}}\right) \\ &\qquad\quad+\frac{\gamma}{2\pi}\arcsin\left(\frac{\varrho_{12}-\varrho_{13}\varrho_{23}}{\sqrt{(1-\varrho_{13}^2)(1-\varrho_{23}^2)}}\right)\Bigg] \\ &\qquad+\Delta(\mathsf{A},\alpha,\beta,\gamma)\end{aligned} \tag{12}$$

where

$$|\Delta(\mathsf{A},\alpha,\beta,\gamma)| \leq \mathsf{A}^2\eta(\mathsf{A},\alpha,\beta,\gamma)$$

and where $\eta(\mathsf{A},\alpha,\beta,\gamma) = \eta(\mathsf{A},|\alpha|,|\beta|,|\gamma|)$ is monotonically increasing in $(\mathsf{A},|\alpha|,|\beta|,|\gamma|)$ and is bounded for every finite A, α, β, and γ.

Proof: See [10, Sec. 4.1.2]. ■

Due to space limitations, we omit the proofs of Propositions 2 and 3, and we provide only an outline of the proof of Theorem 1. A full proof of Theorem 1 and of Propositions 2 and 3 can be found in [10].

IV. Proof Outline

To prove Theorem 1, we derive a lower bound on $C_{\frac{1}{4W}}(\mathsf{P})$ and compute its ratio to P in the limit as P tends to zero. To

[1] The orthant probability is the probability that all components of a random vector have the same sign.

this end, we evaluate $(2W)/n\, I(X_1^n; \mathbf{Y}_1^n)$ for $\{X_k,\, k \in \mathbb{Z}\}$ being a sequence of IID, binary random variables with

$$X_k = \begin{cases} \sqrt{P}, & \text{with probability } \frac{1}{2} \\ -\sqrt{P}, & \text{with probability } \frac{1}{2}. \end{cases}$$

We shall restrict ourselves to waveforms $g(\cdot)$ that satisfy

$$\sum_{\ell \neq 0} \left| g\left(\frac{\ell - 1/2}{2W}\right) \right| < \infty \tag{13}$$

$$\sum_{\ell \neq 0} \left| g\left(\frac{\ell}{2W}\right) \right| < \infty \tag{14}$$

$$\sum_{\ell \neq 0} \left| g\left(\frac{\ell + 1/2}{2W}\right) \right| < \infty. \tag{15}$$

By the chain rule for mutual information [9, Thm. 2.5.2]

$$\begin{aligned} \frac{2W}{n} I(X_1^n; \mathbf{Y}_1^n) &= \frac{2W}{n} \sum_{k=1}^{n} I(X_k; \mathbf{Y}_1^n \mid X_1^{k-1}) \\ &\geq 2W\, I(X_1; Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}}) \end{aligned} \tag{16}$$

where we define $Y_\tau \triangleq Y\left(\frac{\tau}{2W}\right)$.

A. The Joint Law of $(X_1, Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}})$

In order to evaluate the right-hand side (RHS) of (16), we first compute the conditional probability of $(Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}})$, conditioned on $X_{-\infty}^{\infty}$. Let $N_\tau \triangleq \frac{1}{\sqrt{W N_0}} (Z \star \mathrm{LPF}_W) \left(\frac{\tau}{2W}\right)$, and let

$$\alpha \triangleq \frac{1}{\sqrt{\mathsf{P}(2W)(W\mathsf{N}_0)}} \sum_{\ell=-\infty}^{\infty} x_\ell\, g\left(\frac{1/2 - \ell}{2W}\right)$$

$$\beta \triangleq \frac{1}{\sqrt{\mathsf{P}(2W)(W\mathsf{N}_0)}} \sum_{\ell=-\infty}^{\infty} x_\ell\, g\left(\frac{1 - \ell}{2W}\right)$$

and

$$\gamma \triangleq \frac{1}{\sqrt{\mathsf{P}(2W)(W\mathsf{N}_0)}} \sum_{\ell=-\infty}^{\infty} x_\ell\, g\left(\frac{3/2 - \ell}{2W}\right).$$

It follows from (4) and Proposition 3 that

$$\begin{aligned} &\Pr\left(Y_{\frac{1}{2}} = 1, Y_1 = 1, Y_{\frac{3}{2}} = 1 \mid X_{-\infty}^{\infty} = x_{-\infty}^{\infty}\right) \\ &\quad = \Pr\left(N_{\frac{1}{2}} \geq -\alpha\sqrt{\mathsf{P}}, N_1 \geq -\beta\sqrt{\mathsf{P}}, N_{\frac{3}{2}} \geq -\gamma\sqrt{\mathsf{P}}\right) \\ &\quad = \int_{-\alpha\sqrt{\mathsf{P}}}^{\infty} \int_{-\beta\sqrt{\mathsf{P}}}^{\infty} \int_{-\gamma\sqrt{\mathsf{P}}}^{\infty} \phi_{0,\mathsf{K}}(x,y,z)\, \mathrm{d}z\, \mathrm{d}y\, \mathrm{d}x \\ &\quad = \frac{1}{8} + \frac{1}{2\pi} \arcsin(\rho) \\ &\qquad + \sqrt{\frac{\mathsf{P}}{2\pi}} \left[\frac{\alpha + \beta + \gamma}{4} + \frac{\alpha + \gamma}{2\pi} \vartheta_1 - \frac{\beta}{2\pi} \vartheta_2 \right] \\ &\qquad + \Delta\left(\sqrt{\mathsf{P}}, \alpha, \beta, \gamma\right) \end{aligned} \tag{17}$$

where ρ, ϑ_1, and ϑ_2 are as in Theorem 1, and where

$$\mathsf{K} = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}.$$

Averaging the RHS of (17) over $X_{-\infty}^{\infty}$ and over $(X_{-\infty}^{0}, X_2^{\infty})$ yields the probability of the event $(Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}}) = (1,1,1)$ and the conditional probability of $(Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}}) = (1,1,1)$, conditioned on X_1. It thus follows from Bayes' law that

$$\begin{aligned} &\Pr\left(X_1 = \sqrt{\mathsf{P}} \mid Y_{\frac{1}{2}} = 1, Y_1 = 1, Y_{\frac{3}{2}} = 1\right) \\ &\quad = \frac{1}{2} + \sqrt{\frac{\mathsf{P}}{2\pi}} \frac{\frac{\alpha_0 + \beta_0 + \gamma_0}{4} + \frac{\alpha_0 + \gamma_0}{2\pi}\vartheta_1 - \frac{\beta_0}{2\pi}\vartheta_2}{\frac{1}{4} + \frac{1}{\pi}\arcsin(\rho)} + o\left(\sqrt{\mathsf{P}}\right) \end{aligned} \tag{18}$$

where $o(x)$ satisfies $\lim_{x \downarrow 0} o(x)/x = 0$, and where

$$\alpha_0 \triangleq \mathsf{E}\left[\alpha \,\middle|\, X_1 = \sqrt{\mathsf{P}}\right] = \frac{1}{\sqrt{(2W)(W\mathsf{N}_0)}} g\left(-\frac{1}{4W}\right) \tag{19}$$

$$\beta_0 \triangleq \mathsf{E}\left[\beta \,\middle|\, X_1 = \sqrt{\mathsf{P}}\right] = \frac{1}{\sqrt{(2W)(W\mathsf{N}_0)}} g(0) \tag{20}$$

$$\gamma_0 \triangleq \mathsf{E}\left[\gamma \,\middle|\, X_1 = \sqrt{\mathsf{P}}\right] = \frac{1}{\sqrt{(2W)(W\mathsf{N}_0)}} g\left(\frac{1}{4W}\right). \tag{21}$$

The remaining conditional probabilities of $X_1 = \sqrt{\mathsf{P}}$ can be computed in a similar way.

B. Evaluating $I(X_1; Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}})$

Let

$$\begin{aligned} &\wp(y_{\frac{1}{2}}, y_1, y_{\frac{3}{2}}) \\ &\quad \triangleq \Pr\left(X_1 = \sqrt{\mathsf{P}} \mid Y_{\frac{1}{2}} = y_{\frac{1}{2}}, Y_1 = y_1, Y_{\frac{3}{2}} = y_{\frac{3}{2}}\right). \end{aligned}$$

By noting that

$$H_b\left(\frac{1}{2} + \xi\right) = \log 2 - 2\xi^2 + o(\xi^2), \quad |\xi| \leq \frac{1}{2} \tag{22}$$

where $H_b(p) \triangleq -p\log p - (1-p)\log(1-p)$, $0 \leq p \leq 1$ (with $0 \log 0 \triangleq 0$) denotes the binary entropy function, we obtain

$$\begin{aligned} &I(X_1; Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}}) \\ &\quad = \log 2 - \mathsf{E}\left[H_b\left(\wp(Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}})\right)\right] \\ &\quad = 2\,\mathsf{E}\left[\left(\wp(Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}}) - \frac{1}{2}\right)^2\right] + o(\mathsf{P}). \end{aligned} \tag{23}$$

Applying the expressions for the conditional probabilities $\wp(y_{\frac{1}{2}}, y_1, y_{\frac{3}{2}})$ to (23) yields

$$\begin{aligned} &I(X_1; Y_{\frac{1}{2}}, Y_1, Y_{\frac{3}{2}}) \\ &= \frac{\mathsf{P}}{\pi} \left[\frac{\left(\frac{\alpha_0 + \beta_0 + \gamma_0}{4} + \frac{\alpha_0 + \gamma_0}{2\pi}\vartheta_1 - \frac{\beta_0}{2\pi}\vartheta_2\right)^2}{\frac{1}{4} + \frac{1}{\pi}\arcsin(\rho)} \right. \\ &\quad + 4\left(\frac{\alpha_0 + \beta_0 - \gamma_0}{4} - \frac{\alpha_0 + \gamma_0}{2\pi}\vartheta_1 + \frac{\beta_0}{2\pi}\vartheta_2\right)^2 \\ &\quad + \frac{\left(\frac{\alpha_0 - \beta_0 + \gamma_0}{4} - \frac{\alpha_0 + \gamma_0}{2\pi}\vartheta_1 + \frac{\beta_0}{2\pi}\vartheta_2\right)^2}{\frac{1}{4} - \frac{1}{\pi}\arcsin(\rho)} \end{aligned}$$

$$+\,4\left(\frac{-\alpha_0+\beta_0+\gamma_0}{4}-\frac{\alpha_0+\gamma_0}{2\pi}\vartheta_1+\frac{\beta_0}{2\pi}\vartheta_2\right)^2\Bigg]+o(\mathsf{P}) \tag{24}$$

where α_0, β_0, and γ_0 are defined in (19)–(21).

Combining (24) with (16) and (6), and computing the ratio to P in the limit as P tends to zero, yields the lower bound on the capacity per unit-cost

$$\begin{aligned}
&\dot{C}_{\frac{1}{4W}}(0)\\
&\geq \frac{2W}{\pi}\left[\frac{\left(\frac{\alpha_0+\beta_0+\gamma_0}{4}+\frac{\alpha_0+\gamma_0}{2\pi}\vartheta_1-\frac{\beta_0}{2\pi}\vartheta_2\right)^2}{\frac{1}{4}+\frac{1}{\pi}\arcsin(\rho)}\right.\\
&\quad+4\left(\frac{\alpha_0+\beta_0-\gamma_0}{4}-\frac{\alpha_0+\gamma_0}{2\pi}\vartheta_1+\frac{\beta_0}{2\pi}\vartheta_2\right)^2\\
&\quad+\frac{\left(\frac{\alpha_0-\beta_0+\gamma_0}{4}-\frac{\alpha_0+\gamma_0}{2\pi}\vartheta_1+\frac{\beta_0}{2\pi}\vartheta_2\right)^2}{\frac{1}{4}-\frac{1}{\pi}\arcsin(\rho)}\\
&\quad\left.+4\left(\frac{-\alpha_0+\beta_0+\gamma_0}{4}-\frac{\alpha_0+\gamma_0}{2\pi}\vartheta_1+\frac{\beta_0}{2\pi}\vartheta_2\right)^2\right].
\end{aligned} \tag{25}$$

Note that this lower bound holds for all unit-energy waveforms $g(\cdot)$ that are bandlimited to W Hz and that satisfy (13)–(15).

C. *Choosing a Waveform*

Any choice of $g(\cdot)$ satisfying the above conditions yields a lower bound on $\dot{C}_{\frac{1}{4W}}(0)$. We shall choose $g(\cdot)$ to be of Fourier Transform

$$\hat{g}(f)=\frac{1}{\sqrt{2W}}\frac{1+\lambda\cos\left(\pi\frac{f}{2W}\right)}{\sqrt{\frac{1}{2}\lambda^2+\frac{4}{\pi}\lambda+1}}\,\mathrm{I}\left\{|f|\leq W\right\},\quad f\in\mathbb{R}$$

where $\mathrm{I}\{\cdot\}$ denotes the indicator function, i.e., $\mathrm{I}\{\text{statement}\}$ is 1 if the statement is true and 0 otherwise. This yields

$$\alpha_0=\gamma_0=\frac{1}{\sqrt{W\mathsf{N}_0}}\frac{\frac{2}{\pi}+\frac{1}{2}\lambda}{\sqrt{\frac{1}{2}\lambda^2+\frac{4}{\pi}\lambda+1}},\quad \lambda\in\mathbb{R} \tag{26}$$

and

$$\beta_0=\frac{1}{\sqrt{W\mathsf{N}_0}}\frac{1+\frac{2}{\pi}\lambda}{\sqrt{\frac{1}{2}\lambda^2+\frac{4}{\pi}\lambda+1}},\quad \lambda\in\mathbb{R}. \tag{27}$$

Note that, for the above choice, $g(\cdot)$ does not satisfy (13)–(15). However, there exist waveforms that satisfy (13)–(15) and that yield $(\alpha_0,\beta_0,\gamma_0)$ that are arbitrarily close to (26) and (27); see [10, App. B].

Applying (26) and (27) to (25) with $\lambda=1.4$ yields the lower bound (10) and proves thus Theorem 1.

It was demonstrated in [10, Sec. 4.2.3] that among all tuples $(\alpha_0,\beta_0,\gamma_0)$ satisfying $\alpha_0=\gamma_0$, the above choice (26) and (27) with $\lambda=1.4$ maximizes the RHS of (25) and yields thus the largest lower bound on $\dot{C}_{\frac{1}{4W}}(0)$.

V. Summary and Conclusion

We demonstrated that doubling the sampling rate recovers some of the loss in capacity per unit-cost incurred on the bandlimited Gaussian channel with a one-bit output quantizer. Indeed, when the channel output is sampled at Nyquist rate $2W$, it is well-known that the capacity per unit-cost is given by $\frac{2}{\pi}\frac{1}{\mathsf{N}_0}\approx 0.64\frac{1}{\mathsf{N}_0}$ [3], which is a factor of $\frac{2}{\pi}$ smaller than the capacity per unit-cost of the same channel but without output quantizer. We showed that, by sampling the output at twice the Nyquist rate, a capacity per unit-cost not less than $0.75\frac{1}{\mathsf{N}_0}$ can be achieved. This can be viewed as a very-noisy counterpart of the work by Gilbert [6] and by Shamai [7], which demonstrated that oversampling increases the capacity of the above channel when there is no additive noise.

The conclusions that can be drawn from this result are twofold. Firstly, we demonstrated that in order to reduce the loss in capacity per unit-cost caused by the quantization, one can either increase the quantization resolution or the sampling rate. Thus, it is possible to trade amplitude resolution (quantization) versus time resolution (sampling rate). Secondly, we observe that while sampling the output at Nyquist rate is optimal for the AWGN channel (without output quantization), this does not hold when the output is quantized. Thus, a communication scheme that is optimal in the sense that it achieves the capacity need not be optimal anymore if the channel output is processed by a noninvertible operation (such as quantization).

Acknowledgment

Fruitful discussions with Michèle Wigger are gratefully acknowledged.

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[3] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[4] R. J. McEliece, *The Theory of Information and Coding*, 2nd ed. Cambridge University Press, 2002.

[5] J. Singh, O. Dabeer, and U. Madhow, “Transceiver design with low-precision analog-to-digital conversion: An information theoretic perspective,” 2009, to appear in *IEEE Transactions on Communications*. [Online]. Available: http://arxiv.org/abs/0804.1172v1

[6] E. N. Gilbert, “Increased information rate by oversampling,” *IEEE Transactions on Information Theory*, vol. 39, pp. 1973–1976, November 1993.

[7] S. Shamai (Shitz), “Information rates by oversampling the sign of a bandlimited process,” *IEEE Transactions on Information Theory*, vol. 40, pp. 1230–1236, July 1994.

[8] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2009.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley & Sons, 1991.

[10] T. Koch and A. Lapidoth, “Increased capacity per unit-cost by oversampling,” September 2010. [Online]. Available: http://arxiv.org/abs/1008.5393