Upper Bound on the Capacity of Discrete-Time Wiener Phase Noise Channels

Luca Barletta and Gerhard Kramer
Institute for Communications Engineering
Technische Universität München
D-80333 Munich, Germany
{luca.barletta, gerhard.kramer}@tum.de

Abstract—A discrete-time Wiener phase noise channel with an integrate-and-dump multi-sample receiver is studied. An upper bound to the capacity with an average input power constraint is derived, and a high signal-to-noise ratio (SNR) analysis is performed. If the oversampling factor grows as SNR^α for $0 \leq \alpha \leq 1$, then the capacity pre-log is at most $(1+\alpha)/2$ at high SNR.

I. INTRODUCTION

Instabilities of the oscillators used for up- and down-conversion of signals in communication systems give rise to the phenomenon known as *phase noise* [1]. The impairment on the system performance can be severe even for high-quality oscillators, if the continuous-time waveform is processed by long filters at the receiver side. This is the case, for example, when the symbol time is very long, as happens when using orthogonal frequency division multiplexing. A study of the signal-to-noise ratio (SNR) penalty induced by filtering of a white phase noise process has been recently done in [2], where it is shown that the best projection receiver suffers an SNR loss that depends on the phase noise statistics.

Typically, the phase noise generated by oscillators is a random process with memory, and this makes the analysis of the capacity challenging. The phase noise is usually modeled as a Wiener process, as it turns out to be accurate in describing the phase noise statistics of certain lasers used in fiber-optic communications [3], and of free-running microwave oscillators [1]. Tight numerical bounds on the information rate of discrete-time phase noise channels with memory are given in [4]–[7], while analytical results on single-user Wiener phase noise channels are given in [8]–[12] where it is shown that even weak phase noise becomes the limiting factor at high SNR.

In [11] an achievable rate region for the discrete-time Wiener phase noise channel with an integrate-and-dump over-sampling receiver was derived. For the same channel and receive filter, in this paper we develop an upper bound to the capacity and characterize the pre-log[1] at high SNR.

The paper is organized as follows. The system model for the continuous-time channel is described in Sec. II, along with a simplification that leads to the discrete-time model under consideration. The upper bound to the capacity is derived in Sec. III, and the results are discussed in Sec. IV. Conclusions are drawn is Sec. V.

Notation: Capital letters denote random variables or random processes. The notation $X_m^n = (X_m, X_{m+1}, \ldots, X_n)$ with $n \geq m$ is used for random vectors. With $\mathcal{N}(0, \sigma^2)$ we denote the probability distribution of a real Gaussian random variable with zero mean and variance σ^2. The symbol $\stackrel{\mathcal{D}}{=}$ means equality in distribution.

The symbol $\widetilde{X}$ denotes the reduction of X modulo $[-\pi, \pi)$, and the binary operator $\oplus$ denotes summation modulo $[-\pi, \pi)$.

Given a complex random variable X, we use the notation $X_{||} = |X|$ and $X_{\angle} = \angle X$ to denote the amplitude and phase of X, respectively.

The operators $\mathsf{E}[\cdot]$, $\mathsf{h}(\cdot)$, and $I(\cdot\,;\cdot)$ denote expectation, differential entropy, and mutual information, respectively.

The function $\log(x)$ denotes the natural logarithm of x.

II. SYSTEM MODEL

In this Section we describe how to obtain a discrete-time version of the continuous-time channel, and we point out the main assumption that leads to the simplified model analyzed in Sec. III.

The output of a continuous-time phase noise channel can be written as

$$Y(t) = X(t)e^{j\Theta(t)} + W(t), \qquad 0 \leq t \leq T \qquad (1)$$

where $j = \sqrt{-1}$, $X(\cdot)$ is the data-bearing input waveform, and $W(\cdot)$ is a circularly symmetric complex white Gaussian noise. The phase process is given by

$$\Theta(t) = \Theta(0) + \gamma\sqrt{T}B(t/T), \qquad 0 \leq t \leq T, \qquad (2)$$

where $B(\cdot)$ is a standard Wiener process, *i.e.*, a process characterized by the following properties:

- $B(0) = 0$,
- for any $1 \geq t > s \geq 0$, $B(t) - B(s) \sim \mathcal{N}(0, t - s)$ is independent of the sigma algebra generated by $\{B(u) : u \leq s\}$,
- $B(\cdot)$ has continuous sample paths almost surely.

One can think of the Wiener phase process as an accumulation of white noise:

$$\Theta(t) = \Theta(0) + \gamma\int_0^t B'(\tau)\,\mathrm{d}\tau, \qquad 0 \leq t \leq T, \qquad (3)$$

where $B'(\cdot)$ is a standard white Gaussian noise process.

[1] The factor in the capacity high-SNR expansion in front of $\log(\mathsf{SNR})$.

A. Signals and Signal Space

Suppose $X(\cdot)$ is in the set $\mathcal{L}^2[0,T]$ of finite-energy signals in the interval $[0,T]$. Let $\{\phi_m(t)\}_{m=1}^\infty$ be an orthonormal basis of $\mathcal{L}^2[0,T]$. We may write

$$X(t) = \sum_{m=1}^\infty X_m\, \phi_m(t), \quad W(t) = \sum_{m=1}^\infty W_m\, \phi_m(t) \quad (4)$$

where

$$X_m = \int_0^T X(t)\, \phi_m(t)^\star \, \mathrm{d}t, \quad (5)$$

$x^\star$ is the complex conjugate of x, and the $\{W_m\}_{m=1}^\infty$ are independent and identically distributed (iid), complex-valued, circularly symmetric, Gaussian random variables with zero mean and unit variance.

The projection of the received signal onto the $n-$th basis function is

$$Y_n = \int_0^T Y(t)\, \phi_n(t)^\star \, \mathrm{d}t \quad (6)$$

$$= \sum_{m=1}^\infty X_m \int_0^T \phi_m(t)\, \phi_n(t)^\star\, e^{j\Theta(t)}\, \mathrm{d}t + W_n \quad (7)$$

$$= \sum_{m=1}^\infty X_m\, \Phi_{mn} + W_n. \quad (8)$$

The set of equations given by (8) for $n = 1, 2, \ldots$ can be interpreted as the output of an infinite-dimensional multiple-input multiple-output channel, whose fading channel matrix is $\Phi = [\Phi_{mn}]$.

B. Receivers with Finite Time Resolution

Consider a receiver whose time resolution is limited to Δ seconds, in the sense that every projection must include at least a Δ-second interval. More precisely, we set $ML\Delta = T$, where M is the number of independent symbols transmitted in $[0,T]$ and L is the oversampling factor, *i.e.*, the number of samples per symbol. The integrate-and-dump receiver with resolution time Δ uses the basis functions

$$\phi_m(t) = \begin{cases} 1/\sqrt{\Delta}, & t \in [(m-1)\Delta, m\Delta) \\ 0, & \text{elsewhere.} \end{cases} \quad (9)$$

for $m = 1, \ldots, ML$. With the choice (9), the fading channel matrix Φ is diagonal and the channel's output for $n = 1, \ldots, ML$ is

$$Y_n = X_n \frac{1}{\Delta} \int_{(n-1)\Delta}^{n\Delta} e^{j\Theta(t)}\, \mathrm{d}t + W_n$$

$$= X_n\, e^{j\Theta((n-1)\Delta)} \frac{1}{\Delta} \int_{(n-1)\Delta}^{n\Delta} e^{j(\Theta(t) - \Theta((n-1)\Delta))}\, \mathrm{d}t + W_n$$

$$\overset{\mathcal{D}}{=} X_n\, e^{j\Theta_n} \frac{1}{\Delta} \int_0^\Delta e^{j\gamma\sqrt{\Delta}B_n(t/\Delta)}\, \mathrm{d}t + W_n \quad (10)$$

$$\overset{(a)}{=} X_n\, e^{j\Theta_n} \int_0^1 e^{j\gamma\sqrt{\Delta}B_n(t)}\, \mathrm{d}t + W_n$$

$$= X_n\, e^{j\Theta_n} F_n + W_n, \quad (11)$$

where we have used the notation $\Theta_n = \Theta((n-1)\Delta)$ and $F_n = \int_0^1 e^{j\gamma\sqrt{\Delta}B_n(t)}\, \mathrm{d}t$. In (10) we have used (2), the property $B(t/T) - B((n-1)\Delta/T) \overset{\mathcal{D}}{=} B(t/T - (n-1)\Delta/T)$, the substitution

$$\begin{cases} t \leftarrow t - (n-1)/\Delta \\ B_n(t/T) \leftarrow B(t/T - (n-1)\Delta/T), \end{cases} \quad (12)$$

and the property $\sqrt{T}B_n(t/T) \overset{\mathcal{D}}{=} \sqrt{\Delta}B_n(t/\Delta)$. Finally, in step (a) we have used the substitution $t \leftarrow t/\Delta$.

Since the oversampling factor is L, and the basis functions are square in time domain, we have $X_{kL+1} = X_{kL+2} = \ldots = X_{kL+L}$ for $k = 0, \ldots, M-1$, and we can write the model (11) as

$$Y_n = X_{\lceil n/L \rceil L}\, e^{j\Theta_n} F_n + W_n \quad (13)$$

for $n = 1, \ldots, ML$.

The vectors X_1^{ML}, F_1^{ML}, and W_1^{ML} are independent of each other. The variables $\{X_{kL}\}_{k=1}^M$ are chosen as identically distributed with zero mean and variance $\mathsf{E}\left[|X_n|^2\right]$, and the average power constraint is

$$\mathsf{E}\left[\frac{1}{T} \int_0^T |X(t)|^2\, \mathrm{d}t\right] = \frac{1}{ML\Delta} \sum_{n=1}^{ML} \mathsf{E}\left[|X_n|^2\right]$$

$$= \frac{\mathsf{E}\left[|X_n|^2\right]}{\Delta} \leq \mathcal{P}. \quad (14)$$

Since we set the power spectral density of W to 1, the power $\mathcal{P}$ is also the SNR, *i.e.*, $\mathsf{SNR} = \mathcal{P}$.

Using (3), the variables Θ_1^{ML} follow a discrete-time Wiener process:

$$\Theta_n = \Theta_{n-1} + N_{n-1}, \qquad n = 1, \ldots, ML, \quad (15)$$

where the N_n's are iid Gaussian variables with zero mean and variance $\gamma^2\Delta$. The fading variables F_n's are complex-valued and iid, and F_n is independent of Θ_1^n. In other words, F_n is correlated only to N_n, and is independent of the vector $(N_1^{n-1}, N_{n+1}^{ML})$.

Note that for any finite Δ, or equivalently for any finite oversampling factor L, the vector Y_1^{ML} does not represent a sufficient statistic for the detection of X given Y in the model (1). In other words, the finite time resolution receiver is generally suboptimal.

In this paper we study a simplified model, where the fading variables F_1^{ML} are all one, *i.e.*, we have

$$Y_n = X_{\lceil n/L \rceil L}\, e^{j\Theta_n} + W_n, \qquad n = 1, \ldots, ML. \quad (16)$$

This is a commonly-studied model, *e.g.*, see [11], [13], and it is referred to as the discrete-time Wiener phase noise channel. The complete model (13) is harder to analyze than the model (16), because in the former the dependency between F_n and N_n must be addressed. On the other hand, if the oversampling factor L grows unbounded, then each random variable F_n converges to 1; this suggests that the analysis of the model (16) can give insights into the analysis of model (13) for receivers with high time resolution.

III. UPPER BOUND ON CAPACITY

We compute an upper bound to the capacity of the discrete-time Wiener phase noise channel (16). For notational convenience, we use the following indexing for $i = 1, \ldots, L$ and $k = 1, \ldots, M$:

$$Y_{(k-1)L+i} = X_k \, e^{j\Theta_{(k-1)L+i}} + W_{(k-1)L+i}, \qquad (17)$$

and we group the output samples associated with X_k in the vector $\mathbf{Y}_k = Y_{(k-1)L+1}^{(k-1)L+L}$.

The capacity is defined as

$$\mathsf{C}(\mathsf{SNR}) = \lim_{M \to \infty} \frac{1}{M} \sup I\left(X_1^M; \mathbf{Y}_1^M\right) \qquad (18)$$

where the supremum is taken among the distributions of X_1^M such that the average power constraint (14) is satisfied.

The mutual information rate can be upper-bounded as follows:

$$\frac{1}{M} I\left(X_1^M; \mathbf{Y}_1^M\right) = \frac{1}{M} \sum_{k=1}^{M} I\left(X_1^M; \mathbf{Y}_k \,\middle|\, \mathbf{Y}_1^{k-1}\right)$$

$$\overset{(a)}{\leq} \frac{1}{M} \sum_{k=1}^{M} I\left(X_1^M; \mathbf{Y}_k \,\middle|\, \mathbf{Y}_1^{k-1}, \widetilde{\Theta}_{(k-1)L}\right)$$

$$\overset{(b)}{\leq} \frac{1}{M} \sum_{k=1}^{M} I\left(X_k; \mathbf{Y}_k \,\middle|\, \widetilde{\Theta}_{(k-1)L}\right)$$

$$\overset{(c)}{=} I\left(X_1; \mathbf{Y}_1 \,\middle|\, \widetilde{\Theta}_0\right)$$

$$\overset{(d)}{=} I\left(X_{\|,1}; \mathbf{Y}_1 \,\middle|\, \widetilde{\Theta}_0\right) + I\left(X_{\angle,1}; \mathbf{Y}_1 \,\middle|\, \widetilde{\Theta}_0, X_{\|,1}\right) \quad (19)$$

where step (a) holds by a data processing inequality and because $\widetilde{\Theta}_{(k-1)L}$ is independent of X_1^M, (b) because $\mathbf{Y}_k$ is conditionally independent of $(\mathbf{Y}_1^{k-1}, X_1^{k-1}, X_{k+1}^M)$ given $(X_k, \widetilde{\Theta}_{(k-1)L})$, (c) follows by stationarity of the processes, and (d) by polar decomposition of X_1 and the chain rule.

For the *amplitude channel*, *i.e.*, the first term in the right-hand side (RHS) of (19), we have

$$I\left(X_{\|,1}; \mathbf{Y}_1 \,\middle|\, \widetilde{\Theta}_0\right) \overset{(a)}{\leq} I\left(X_{\|,1}; \{X_1 e^{j\widetilde{\Theta}_i} + W_i\}_{i=1}^{L} \,\middle|\, \widetilde{\Theta}_0^L\right)$$

$$\overset{(b)}{=} I\left(X_{\|,1}; \{X_1 + W_i\}_{i=1}^{L} \,\middle|\, \widetilde{\Theta}_0^L\right)$$

$$\overset{(c)}{=} I\left(X_{\|,1}; \{X_1 + W_i\}_{i=1}^{L}\right)$$

$$\overset{(d)}{=} I\left(X_{\|,1}; \left|X_1 + \frac{1}{L}\sum_{i=1}^{L} W_i\right|\right)$$

$$\overset{(e)}{\leq} \frac{1}{2}\log(1 + \mathsf{SNR}) - \frac{\log(2)}{2} + o(1) \quad (20)$$

where (a) holds by a data processing inequality and because $\widetilde{\Theta}_0^L$ is independent of X_1, (b) holds due to the circular symmetry of the W_i's, (c) because $\widetilde{\Theta}_0^L$ is independent of any other quantity, (d) because the processed variable $\left|X_1 + L^{-1}\sum_{i=1}^{L} W_i\right|$ is a sufficient statistic for the detection of $X_{\|,1}$, and (e) is an upper bound to the capacity of a non-coherent channel under an average power constraint [8,

Eq. (16)] where $o(1)$ represents a function independent of L that vanishes for $\mathsf{SNR} \to \infty$.

For the *phase channel*, *i.e.*, the second term on the RHS of (19), we have

$$I\left(X_{\angle,1}; \mathbf{Y}_1 \,\middle|\, \widetilde{\Theta}_0, X_{\|,1}\right)$$

$$= I\left(X_{\angle,1}; \{X_1 e^{j\widetilde{\Theta}_i} + W_i\}_{i=1}^{L} \,\middle|\, \widetilde{\Theta}_0, X_{\|,1}\right)$$

$$\overset{(a)}{\leq} I\left(X_{\angle,1}; \{X_1 e^{j\widetilde{\Theta}_i}\}_{i=1}^{L} \,\middle|\, \widetilde{\Theta}_0, X_{\|,1}\right)$$

$$\overset{(b)}{=} I\left(X_{\angle,1}; \{X_{\angle,1} \oplus \widetilde{\Theta}_i\}_{i=1}^{L} \,\middle|\, \widetilde{\Theta}_0, X_{\|,1}\right)$$

$$\overset{(c)}{\leq} I\left(X_{\angle,1}; \{X_{\angle,1} \oplus \widetilde{\Theta}_i\}_{i=1}^{L} \,\middle|\, \widetilde{\Theta}_0\right)$$

$$\overset{(d)}{=} I\left(X_{\angle,1}; X_{\angle,1} \oplus \widetilde{N}_0, \widetilde{N}_1^{L-1} \,\middle|\, \widetilde{\Theta}_0\right)$$

$$\overset{(e)}{=} I\left(X_{\angle,1}; X_{\angle,1} \oplus \widetilde{N}_0\right)$$

$$= \mathsf{h}\left(X_{\angle,1} \oplus \widetilde{N}_0\right) - \mathsf{h}\left(\widetilde{N}_0\right)$$

$$\overset{(f)}{\leq} \log(2\pi) - \mathsf{h}\left(\widetilde{N}_0\right) \quad (21)$$

$$\leq \frac{1}{2}\log\left(\frac{2\pi}{e\gamma^2}\right) - \frac{1}{2}\log(\Delta) - g(\Delta) \quad (22)$$

where in step (a) we bound by the information extracted by a genie-aided receiver that knows the additive noise W_1^L, (b) is obtained by deleting the amplitude contribution of $X_{\|,1}$, (c) holds because $X_{\|,1} \multimap (X_{\angle,1}, \widetilde{\Theta}_0) \multimap \{X_{\angle,1} \oplus \widetilde{\Theta}_i\}_{i=1}^{L}$ forms a Markov chain, (d) is obtained by applying reversible transformations, (e) holds because the random variables $(\widetilde{N}_1^{L-1}, \widetilde{\Theta}_0)$ are independent of any other quantity, (f) holds by choosing a uniform distribution in $[0, 2\pi)$ for $X_{\angle,1}$, and the last inequality is derived in the Appendix with

$$g(\Delta) = \frac{1}{2}\mathrm{erf}\left(\frac{\pi}{\sqrt{2\gamma^2\Delta}}\right)$$

$$- \frac{e^{-\frac{\pi^2}{2\gamma^2\Delta}}}{\sqrt{2\pi\gamma^2\Delta}}\left(\pi + \frac{4(\pi + \gamma^2\Delta/\pi)}{1 - e^{-\frac{\pi^2}{\gamma^2\Delta}}}\right) - \frac{1}{2}. \quad (23)$$

Suppose the oversampling factor grows as a power of the SNR, *i.e.*, $L = \Delta^{-1} = \lceil\mathsf{SNR}^\alpha\rceil$ for $0 < \alpha \leq 1$. Inserting (20) and (22) into (19) and using (18) yields

$$\mathsf{C}(\mathsf{SNR}) \leq \frac{1}{2}\log(1 + \mathsf{SNR}) - \frac{\log(2)}{2} + \frac{1}{2}\log\left(\frac{2\pi}{e\gamma^2}\right)$$

$$+ \frac{1}{2}\log(\lceil\mathsf{SNR}^\alpha\rceil) + o(1), \quad (24)$$

which for large SNR gives

$$\limsup_{\mathsf{SNR}\to\infty}\left\{\mathsf{C}(\mathsf{SNR}) - \frac{1+\alpha}{2}\log(\mathsf{SNR}) - \frac{1}{2}\log\left(\frac{\pi}{e\gamma^2}\right)\right\} \leq 0. \quad (25)$$

IV. DISCUSSION

As a byproduct of (25), an upper bound to the capacity pre-log is

$$\lim_{\mathsf{SNR}\to\infty}\frac{\mathsf{C}(\mathsf{SNR})}{\log(\mathsf{SNR})} \leq \frac{1+\alpha}{2}. \quad (26)$$
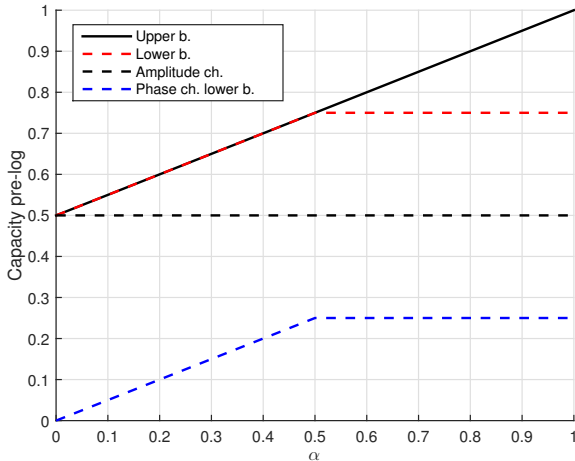
Fig. 1. Capacity pre-log bounds as a function of α for high SNR. The oversampling factor L is $L = \lceil \mathsf{SNR}^\alpha \rceil$.

As shown in the previous Section, a pre-log of $1/2$ comes from the amplitude channel, while a contribution of $\alpha/2$ comes from the phase channel. For example, if no oversampling is used ($L = 1$), one can let α go to zero and obtain just the degrees of freedom provided by the amplitude channel, *i.e.*, $1/2$. This means that, without oversampling, the Wiener phase noise channel has the same degrees of freedom of the non-coherent channel. This is in accordance with the result given in [8].

If the oversampling factor grows as $\sqrt{\mathsf{SNR}}$, for $\alpha = 1/2$, then a pre-log higher than $3/4$ can not be achieved. Indeed, a pre-log as high as $3/4$ can be achieved with the processing described in [11]: the amplitude channel contributes with pre-log $1/2$ by using the statistic

$$V_k = \sum_{i=1}^{L} |Y_{(k-1)L+i}|^2 \tag{27}$$

to detect $X_{k,||}$, and the phase channel contributes with pre-log $1/4$ by using the processing

$$\angle \widetilde{Y}_k = \angle \left(Y_{(k-1)L+1} \left(\frac{Y_{(k-1)L}}{X_{k-1}} \right)^\star \right) \tag{28}$$

to detect $X_{k,\angle}$.

Figure 1 plots the known upper and lower bounds to the capacity pre-log at high SNR. The upper bound is the result of this paper, expressed in (26). The lower bound is based on results derived in [11]. More specifically:

- the lower bound for the amplitude channel, shown as the dashed black line, was derived independent of the growth rate of the oversampling factor;
- for the phase channel it was shown how to achieve pre-log $1/4$ for $\alpha = 1/2$, hence the same pre-log can be achieved for $1/2 \leq \alpha \leq 1$. It is not difficult to use the results of [11] to extend the lower bound in the range $0 \leq \alpha < 1/2$. It turns out that the achievable pre-log linearly increases from 0 to $1/4$.

From the Figure, the capacity pre-log is exactly known in the range $0 \leq \alpha \leq 1/2$, where the upper and lower bounds agree. The upper bound derived in this paper does not rule out the possibility to achieve pre-log 1 if the oversampling factor grows faster than $\sqrt{\mathsf{SNR}}$.

Consider the case of receivers without oversampling ($L = 1$ or $\alpha \to 0$). The analysis of Sec. III shows that the simplified model (16) has capacity pre-log $1/2$, while the general discrete-time model that accounts also for the amplitude fading F_n (13) has a $\log(\log(\mathsf{SNR}))$ behavior at high SNR [14], *i.e.*, pre-log 0. This means that, at least in the case $L = 1$, the simplified model is not a good approximation of the complete model.

V. CONCLUSIONS

We have derived an upper bound to the capacity of discrete-time Wiener phase noise channels. As a byproduct, we have obtained an upper bound to the capacity pre-log at high SNR that depends on the growth rate of the oversampling factor used at the receiver. If the oversampling factor grows proportionally to SNR^α, then a capacity pre-log higher than $(1 + \alpha)/2$ can not be achieved.

Previous results on a lower bound to the capacity pre-log allow to state that the capacity at high SNR is exactly $(1 + \alpha)/2 \log(\mathsf{SNR})$ for $0 \leq \alpha \leq 1/2$.

APPENDIX

A LOWER BOUND TO $\mathsf{h}\left(\widetilde{N}_0\right)$

The probability density function of $\widetilde{N}_0$ is

$$p_{\widetilde{N}_0}(x) = \sum_{k=-\infty}^{\infty} p_{N_0}(x + 2\pi k) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\gamma^2\Delta}} e^{-\frac{(x+2\pi k)^2}{2\gamma^2\Delta}} \tag{29}$$

for $-\pi \leq x < \pi$ and zero elsewhere, and can be upper-bounded as follows for $-\pi \leq x < \pi$:

$$p_{\widetilde{N}_0}(x) \overset{(a)}{\leq} \frac{1}{\sqrt{2\pi\gamma^2\Delta}} \left(e^{-\frac{x^2}{2\gamma^2\Delta}} + 2\sum_{k=0}^{\infty} e^{-\frac{\pi^2(2k+1)^2}{2\gamma^2\Delta}} \right)$$

$$\overset{(b)}{\leq} \frac{1}{\sqrt{2\pi\gamma^2\Delta}} \left(e^{-\frac{x^2}{2\gamma^2\Delta}} + 2\sum_{k=0}^{\infty} e^{-\frac{\pi^2(2k+1)}{2\gamma^2\Delta}} \right)$$

$$= \frac{1}{\sqrt{2\pi\gamma^2\Delta}} \left(e^{-\frac{x^2}{2\gamma^2\Delta}} + 2\frac{e^{-\frac{\pi^2}{2\gamma^2\Delta}}}{1 - e^{-\frac{\pi^2}{\gamma^2\Delta}}} \right) \tag{30}$$

where step (a) follows by using $x \geq -\pi$ for the terms with $k \geq 1$ and $x < \pi$ for the terms with $k \leq -1$. Inequality (b) holds because $(2k+1)^2 \geq 2k+1$ for $k \geq 0$. The differential

entropy of $\widetilde{N}_0$ can be lower-bounded as follows:

$$\mathsf{h}\left(\widetilde{N}_0\right) = \mathsf{E}\left[-\log\left(p_{\widetilde{N}_0}(\widetilde{N}_0)\right)\right]$$

$$\overset{(a)}{\geq} \frac{1}{2}\log(2\pi\gamma^2\Delta) + \frac{\mathsf{E}\left[\widetilde{N}_0^2\right]}{2\gamma^2\Delta} - \mathsf{E}\left[\log\left(1 + \frac{2e^{\frac{\widetilde{N}_0^2-\pi^2}{2\gamma^2\Delta}}}{1-e^{-\frac{\pi^2}{\gamma^2\Delta}}}\right)\right]$$

$$\geq \frac{1}{2}\log(2\pi\gamma^2\Delta) + \frac{\mathsf{E}\left[\widetilde{N}_0^2\right]}{2\gamma^2\Delta} - \frac{2e^{-\frac{\pi^2}{2\gamma^2\Delta}}}{1-e^{-\frac{\pi^2}{\gamma^2\Delta}}}\mathsf{E}\left[e^{\frac{\widetilde{N}_0^2}{2\gamma^2\Delta}}\right] \quad (31)$$

where inequality (a) is due to (30) and the monotonicity of the logarithm, and the last inequality is due to $\log(1+x) \leq x$. A lower bound to the second moment of $\widetilde{N}_0$ is:

$$\mathsf{E}\left[\widetilde{N}_0^2\right] = \sum_{k=-\infty}^{\infty}\int_{-\pi}^{\pi} x^2 p_{N_0}(x+2\pi k)\,\mathrm{d}x$$

$$\geq \int_{-\pi}^{\pi} x^2 p_{N_0}(x)\,\mathrm{d}x$$

$$= \gamma^2\Delta\,\mathrm{erf}\left(\frac{\pi}{\sqrt{2\gamma^2\Delta}}\right) - \sqrt{2\pi\gamma^2\Delta}\,e^{-\frac{\pi^2}{2\gamma^2\Delta}} \quad (32)$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t \quad (33)$$

is the error function. Since all the terms of the summation are positive, the inequality follows by considering only the term for $k=0$. An upper bound to the last expectation on the RHS of (31) is

$$\mathsf{E}\left[e^{\frac{\widetilde{N}_0^2}{2\gamma^2\Delta}}\right] = \frac{1}{\sqrt{2\pi\gamma^2\Delta}}\sum_{k=-\infty}^{\infty}\int_{-\pi}^{\pi} e^{-\frac{(x+2\pi k)^2-x^2}{2\gamma^2\Delta}}\,\mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi\gamma^2\Delta}}\sum_{k=-\infty}^{\infty}\frac{\gamma^2\Delta}{\pi k}e^{-\frac{2\pi^2k^2}{\gamma^2\Delta}}\sinh\left(\frac{2\pi^2 k}{\gamma^2\Delta}\right)$$

$$= \frac{2}{\sqrt{2\pi\gamma^2\Delta}}\left(\pi + \sum_{k=1}^{\infty}\frac{\gamma^2\Delta}{\pi k}e^{-\frac{2\pi^2k^2}{\gamma^2\Delta}}\sinh\left(\frac{2\pi^2 k}{\gamma^2\Delta}\right)\right)$$

$$\leq \frac{2}{\sqrt{2\pi\gamma^2\Delta}}\left(\pi + \sum_{k=1}^{\infty}\frac{\gamma^2\Delta}{\pi}e^{-\frac{2\pi^2k^2}{\gamma^2\Delta}}\sinh\left(\frac{2\pi^2 k}{\gamma^2\Delta}\right)\right)$$

$$= \frac{2}{\sqrt{2\pi\gamma^2\Delta}}\left(\pi + \frac{\gamma^2\Delta}{\pi}\right). \quad (34)$$

The bound (31) with inequalities (32) and (34) give

$$\mathsf{h}\left(\widetilde{N}_0\right) \geq \frac{1}{2}\log(2\pi e\gamma^2\Delta) + \frac{1}{2}\,\mathrm{erf}\left(\frac{\pi}{\sqrt{2\gamma^2\Delta}}\right)$$

$$- \frac{e^{-\frac{\pi^2}{2\gamma^2\Delta}}}{\sqrt{2\pi\gamma^2\Delta}}\left(\pi + \frac{4(\pi+\gamma^2\Delta/\pi)}{1-e^{-\frac{\pi^2}{\gamma^2\Delta}}}\right) - \frac{1}{2}$$

$$= \frac{1}{2}\log(2\pi e\gamma^2\Delta) + g(\Delta). \quad (35)$$

Also, note that $\mathsf{h}\left(\widetilde{N}_0\right) \leq \mathsf{h}\left(N_0\right) = 1/2 \cdot \log(2\pi e\gamma^2\Delta)$, so we have that the bound (35) is tight for small Δ, *i.e.*,

$$\lim_{\Delta\to 0} g(\Delta) = 0. \quad (36)$$

REFERENCES

[1] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 47, no. 5, pp. 655–674, May 2000.
[2] L. Barletta and G. Kramer, "On continuous-time white phase noise channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, June 2014, pp. 2426–2429.
[3] G. Foschini and G. Vannucci, "Characterizing filtered light waves corrupted by phase noise," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1437–1448, Nov 1988.
[4] L. Barletta, M. Magarini, and A. Spalvieri, "Estimate of information rates of discrete-time first-order Markov phase noise channels," *IEEE Photon. Technol. Lett.*, vol. 23, no. 21, pp. 1582–1584, 2011.
[5] ——, "The information rate transferred through the discrete-time Wiener's phase noise channel," *J. Lightwave Technol.*, vol. 30, no. 10, pp. 1480–1486, 2012.
[6] ——, "Tight upper and lower bounds to the information rate of the phase noise channel," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 2284–2288.
[7] L. Barletta, M. Magarini, S. Pecorino, and A. Spalvieri, "Upper and lower bounds to the information rate transferred through first-order Markov channels with free-running continuous state," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3834–3844, July 2014.
[8] A. Lapidoth, "On phase noise channels at high SNR," in *IEEE Inf. Theory Workshop*, 2002, pp. 1–4.
[9] A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channels," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3223–3228, 2011.
[10] H. Ghozlan and G. Kramer, "On Wiener phase noise channels at high signal-to-noise ratio," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 2279–2283.
[11] ——, "Phase modulation for discrete-time Wiener phase noise channels with oversampling at high SNR," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2014.
[12] L. Barletta and G. Kramer, "Signal-to-noise ratio penalties for continuous-time phase noise channels," in *Int. Conf. on Cognitive Radio Oriented Wirel. Networks (CROWNCOM'14)*, June 2014, pp. 232–235.
[13] M. Martalò, C. Tripodi, and R. Raheli, "On the information rate of phase noise-limited communications," in *Inf. Theory and Appl. Workshop (ITA)*, Feb 2013, pp. 1–7.
[14] A. Lapidoth and S. 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, 2003.