Lower Bound on the Capacity of Continuous-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 continuous-time Wiener phase noise channel with an integrate-and-dump multi-sample receiver is studied. A lower bound to the capacity with an average input power constraint is derived, and a high signal-to-noise ratio (SNR) analysis is performed. The proposed lower bound suggests that the capacity pre-log depends on the oversampling factor, and amplitude and phase modulation do not equally contribute to capacity 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*. 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.

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 statistic of certain lasers used in fiber-optic communications [1]. As the sampled output of the filter matched to the transmit filter does not always represent a sufficient statistic [2], [3], oversampling does help in achieving higher rates over the continuous-time channel [4]–[6].

To simplify the analysis, some works assume a modified channel model where the filtered phase noise does not consider amplitude fading, and thus derive numerical and analytical bounds [7]–[10].

The aim of this paper is to give a capacity lower bound without any simplifying assumption on the statistic of filtered phase noise. Specifically, we extend the existing results for amplitude modulation, partly published in [5], and present new results for phase modulation.

Notation: Capital letters denote random variables or random processes. The notation $X_m^n = (X_m, X_{m+1}, \dots, X_n)$ 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.

Given a complex random variable X, we use the notation $|X|$ and $\angle X$ to denote the amplitude and the phase of X, respectively. The binary operator $\oplus$ denotes summation modulo $[-\pi, \pi)$.

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

II. System model

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

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

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

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

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.

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

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

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

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

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

where

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

$x^\star$ 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

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

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

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

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

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

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

$$\begin{aligned} Y_n &= X_n \frac{1}{\Delta} \int_{(n-1)\Delta}^{n\Delta} e^{j\Theta(t)}\,\mathrm{d}t + W_n \\ &= X_n\, e^{j\Theta((n-1)\Delta)} \frac{1}{\Delta} \int_{(n-1)\Delta}^{n\Delta} e^{j(\Theta(t)-\Theta((n-1)\Delta))}\,\mathrm{d}t + W_n \\ &\overset{\mathcal{D}}{=} X_n\, e^{j\Theta_n} \frac{1}{\Delta} \int_0^{\Delta} e^{j\gamma\sqrt{\Delta}B_n(t/\Delta)}\,\mathrm{d}t + W_n \qquad (10) \\ &\overset{(a)}{=} X_n\, e^{j\Theta_n} \int_0^1 e^{j\gamma\sqrt{\Delta}B_n(t)}\,\mathrm{d}t + W_n \\ &= X_n\, e^{j\Theta_n} F_n + W_n, \qquad (11) \end{aligned}$$

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

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

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, 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

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

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 iid with zero mean and variance $\mathsf{E}\left[|X_n|^2\right]$, and the average power constraint is

$$\begin{aligned} \mathsf{E}\left[\frac{1}{T}\int_0^T |X(t)|^2\,\mathrm{d}t\right] &= \frac{1}{ML\Delta}\sum_{n=1}^{ML} \mathsf{E}\left[|X_n|^2\right] \\ &= \frac{\mathsf{E}\left[|X_n|^2\right]}{\Delta} \le \mathcal{P}. \end{aligned} \tag{14}$$

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:

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

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).

III. Lower bound on capacity

We compute a lower bound to the capacity of the continuous-time Wiener phase noise channel (13)-(15). For notational convenience, we use the following indexing for $i = 1, \ldots, L$ and $k = 1, \ldots, M$:

$$Y_{(k-1)L+i} = X_k\, e^{j\Theta_{(k-1)L+i}} F_{(k-1)L+i} + W_{(k-1)L+i}, \tag{16}$$

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

$$\mathsf{C}\left(\mathsf{SNR}\right) = \lim_{M\to\infty} \frac{1}{M} \sup I\left(X_1^M ; \mathbf{Y}_1^M\right) \tag{17}$$

where the supremum is taken among the distributions of X_1^M, with the iid assumption on the X_k's, such that the average power constraint (14) is satisfied.

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

$$\begin{aligned} \frac{1}{M} I\left(X_1^M ; \mathbf{Y}_1^M\right) &= \frac{1}{M}\sum_{k=1}^{M} I\left(X_k ; \mathbf{Y}_1^M \,\middle|\, X_1^{k-1}\right) \\ &\overset{(a)}{=} \frac{1}{M}\sum_{k=1}^{M} I\left(|X_k| ; \mathbf{Y}_1^M \,\middle|\, X_1^{k-1}\right) + I\left(\angle X_k ; \mathbf{Y}_1^M \,\middle|\, X_1^{k-1}, |X_k|\right) \\ &\overset{(b)}{\ge} \frac{1}{M}\sum_{k=1}^{M} I\left(|X_k|^2 ; ||\mathbf{Y}_k||^2\right) + I\left(\angle X_k ; \mathbf{Y}_{k-1}^k \,\middle|\, X_{k-1}, |X_k|\right) \\ &= \underbrace{I\left(|X_1|^2 ; ||\mathbf{Y}_1||^2\right)}_{I_{||}} + \underbrace{I\left(\angle X_1 ; \mathbf{Y}_0^1 \,\middle|\, X_0, |X_1|\right)}_{I_{\angle}} \end{aligned} \tag{18}$$

where step (a) follows by polar decomposition of X_k, step (b) holds by a data processing inequality, by reversibility of the map $x \mapsto x^2$ for non-negative reals, and because X_1^{k-1} is independent of $(X_k, \mathbf{Y}_k)$. Finally, the last equality follows by stationarity of the processes.

A. Amplitude Modulation

By choosing a specific input distribution that satisfies the average power constraint we always get a lower bound on the mutual information. Specifically, we choose the input distribution as a generalized version of [5, Eq. (37)]:

$$p_{|X_k|^2}(x) = \begin{cases} \frac{1}{\lambda}\exp\left(-\frac{x-\Delta^{-s}}{\lambda}\right) & x \geq \Delta^{-s} \\ 0 & \text{elsewhere} \end{cases} \tag{19}$$

where $\lambda = \mathsf{SNR}\Delta - \Delta^{-s} > 0$ with $s > 0$. Note that with this choice the average power constraint is satisfied with equality, *i.e.*, $\mathsf{E}\left[|X_k|^2\right] = \mathsf{SNR}\Delta$.

Similar to the method used in [11, Eq. (35)] [5], we give here a lower bound to the first term on the right hand side (RHS) of (18) in the form

$$I_{||} \geq \mathsf{E}\left[-\ln q_V(V)\right] - \mathsf{E}\left[-\ln q_{V|\,|X_1|^2}(V|\,|X_1|^2)\right] \tag{20}$$

where $V = ||\mathbf{Y}_1||^2$ and

$$q_V(v) = \int_0^\infty p_{|X_1|^2}(x) q_{V|\,|X_1|^2}(v|x)\,\mathrm{d}x. \tag{21}$$

Specifically, we choose the *auxiliary channel* distribution as a generalized version of [5, Eq. (31)]:

$$q_{V|\,|X_1|^2}(v|x) = \frac{1}{\sqrt{\pi\nu x}}\exp\left(-\frac{(v - L(1+x\mathsf{E}[G]))^2}{\nu x}\right) \tag{22}$$

where $G = ||\mathbf{F}_1||^2/L$ and $\nu > 0$, for which we have [12, App. C]

$$\begin{aligned} \mathsf{E}\left[-\ln q_{V|\,|X_1|^2}(V|\,|X_1|^2)\right] &= \frac{1}{2}\ln(\pi\nu) + \frac{1}{2}\mathsf{E}\left[\ln(|X_1|^2)\right] \\ &+ \frac{L}{\nu}\left(\frac{\mathsf{E}\left[|X_1|^2\right]}{\Delta}\mathsf{Var}[G] + 2\mathsf{E}[G] + \mathsf{E}\left[\frac{1}{|X_1|^2}\right]\right) \\ &\leq \frac{1}{2}\ln(\pi\nu\lambda) + \frac{\Delta^{-s}}{2\lambda} + \frac{L}{\nu}\left(\mathsf{SNR}\cdot\mathsf{Var}[G] + 2 + \Delta^s\right) \end{aligned} \tag{23}$$

where the inequality is due to $\mathsf{E}[G] \leq 1$, $\mathsf{E}\left[|X_1|^2\right] \leq \mathsf{SNR}\Delta$, the bound $\mathsf{E}\left[|X_k|^{-2}\right] \leq \Delta^s$ which follows from the support of $|X_k|^2$, and

$$\begin{aligned} \mathsf{E}\left[\ln|X_1|^2\right] &= \int_{\Delta^{-s}}^\infty \frac{1}{\lambda}\exp\left(-\frac{x-\Delta^{-s}}{\lambda}\right)\ln(x)\,\mathrm{d}x \\ &= \ln\lambda + \int_{\Delta^{-s}/\lambda}^\infty \exp\left(-\left(u - \frac{\Delta^{-s}}{\lambda}\right)\right)\ln(u)\,\mathrm{d}u \\ &\leq \ln\lambda + \frac{\Delta^{-s}}{\lambda}. \end{aligned} \tag{24}$$

By substituting (22) and (19) into (21), and by following similar steps to those of [5], we get

$$\mathsf{E}\left[-\ln q_V(V)\right] \geq -\frac{\Delta^{-s}}{\lambda} + \frac{1}{2}\ln(L^2\mu^2\lambda^2 + \lambda\nu). \tag{25}$$

By putting together (23) and (25) we obtain

$$\begin{aligned} I_{||} \geq &-\frac{3\Delta^{-s}}{2\lambda} + \frac{1}{2}\ln(L^2\mu^2\lambda^2 + \lambda\nu) - \frac{1}{2}\ln(\pi\nu\lambda) \\ &- \frac{L}{\nu}\left(\mathsf{SNR}\cdot\mathsf{Var}[G] + 2 + \Delta^s\right). \end{aligned} \tag{26}$$

In the limit of large time resolution we have

$$\lim_{\Delta\to 0}\frac{\mathsf{Var}[G]}{\Delta^3} = \frac{\gamma^2}{45}. \tag{27}$$

Now we let the time resolution grow as a power of the SNR, *i.e.*, $\Delta^{-1} = \lceil\mathsf{SNR}^\alpha\rceil$, and the parameter $\nu = \rho\Delta^{-\beta}$, with $\rho > 0$. By using (27) in (26), in order to find a tight bound in the interval $1/3 \leq \alpha \leq 1$ we need to satisfy the conditions $\alpha < 1/(s+1)$ and $\beta \geq 1$ [12, App. C]. The lower bound is maximized by choosing $\beta = 1$ and $\rho = 4$:

$$\lim_{\mathsf{SNR}\to\infty}\left\{I_{||} - \frac{1}{2}\ln(\mathsf{SNR})\right\} \geq -\frac{1}{2}\ln(4\pi e). \tag{28}$$

For $0 < \alpha < 1/3$ we need to satisfy the conditions $\alpha < 1/(s+1)$ and $\alpha \geq 1/(\beta+2)$, and the best lower bound is obtained by choosing $\beta = \alpha^{-1} - 2$ and $\rho = 2\gamma^2/45$:

$$\lim_{\mathsf{SNR}\to\infty}\left\{I_{||} - \frac{3\alpha}{2}\ln(\mathsf{SNR})\right\} \geq -\frac{1}{2}\ln\left(\frac{2\pi\gamma^2 e}{45}\right). \tag{29}$$

B. Phase Modulation

The second term in the RHS of (18) can be lower-bounded as follows

$$\begin{aligned} I\left(\angle X_1; \mathbf{Y}_0^1 \,\middle|\, X_0, |X_1|\right) &\stackrel{(a)}{\geq} I\left(\angle X_1; \Phi \,|\, X_0, |X_1|\right) \\ &\stackrel{(b)}{\geq} \mathsf{E}\left[-\ln q_{\Phi|X_0,|X_1|}(\Phi|X_0,|X_1|)\right] - \mathsf{E}\left[-\ln q_{\Phi|X_0^1}(\Phi|X_0^1)\right] \end{aligned} \tag{30}$$

where step (a) is due to a data processing inequality with

$$\begin{aligned} \Phi &= \angle(Y_1(Y_0 e^{-j\angle X_0})^\star) \\ &= \angle X_1 \oplus \angle(|X_1|F_1 + W_1) \oplus \angle(|X_0|F_0^\star e^{jN_0} + W_0^\star), \end{aligned} \tag{31}$$

and (b) follows by choosing the auxiliary channel

$$q_{\Phi|X_0^1}(\phi|x_0^1) = \frac{\exp(\zeta\cos(\phi - \angle x_1))}{2\pi I_0(\zeta)} \tag{32}$$

where $I_0(\cdot)$ is the zero-th order modified Bessel function of the first kind, and ζ is a positive real number. Since we assume an uniform input phase distribution, the output distribution is also uniform:

$$q_{\Phi|X_0,|X_1|}(\phi|x_0,|x_1|) = \int_0^{2\pi} q_{\Phi|X_0^1}(\phi|x_0^1)\frac{1}{2\pi}\,\mathrm{d}\angle x_1 = \frac{1}{2\pi}. \tag{33}$$

Using (32), the second term in the RHS of (30) can be upper-bounded as follows for any $\Delta \leq \bar{\Delta} < \infty$:

$$\begin{aligned} \mathsf{E}\left[-\ln q_{\Phi|X_0^1}(\Phi|X_0^1)\right] &= \ln(2\pi I_0(\zeta)) - \zeta\mathsf{E}\left[\cos(\Phi - \angle X_1)\right] \\ &\leq \ln(\pi\sqrt{\pi}) + \frac{1}{2}\ln\left(\frac{1}{\zeta}\right) + \zeta\rho \\ &= \frac{1}{2}\ln\left(2\pi^3 e\rho\right) \end{aligned} \tag{34}$$

where the inequality is due to $I_0(\zeta) \le \sqrt{\pi}/2 \cdot e^{\zeta}/\sqrt{\zeta}$ derived in [13, Lemma 2], and from the result of Appendix A with

$$\rho = 1 - \mathsf{E}\left[F_0 e^{-jN_0}\right] \mathsf{E}\left[F_1\right] + 2e^{-3\gamma^2\Delta/8} \mathsf{E}\left[|X_1|^{-2}\right] K_{\bar{\Delta}} \quad (35)$$

where $K_{\bar{\Delta}} > 1$ is a finite number[1]. The last step in (34) is obtained by choosing $\zeta = (2\rho)^{-1}$.

In the limit of large time resolution we have

$$\lim_{\Delta\to 0} \left\{ \frac{\rho}{\Delta} - 2K_{\bar{\Delta}}\Delta^{s-1} \right\} \le \frac{2}{3}\gamma^2 \quad (36)$$

where the inequality follows from the bound $\mathsf{E}\left[|X_1|^{-2}\right] \le \Delta^s$. Choosing $s = 1$ and putting together (30) and (33)-(36) we get

$$\lim_{\Delta\to 0} \left\{ I_{\angle} + \frac{1}{2}\ln(\Delta) \right\} \ge \frac{1}{2}\ln\left(\frac{3}{\pi e(\gamma^2 + 3K_{\bar{\Delta}})} \right), \quad (37)$$

and letting the time resolution grow as a power of the SNR, *i.e.*, $\Delta^{-1} = \lceil \mathsf{SNR}^{\alpha} \rceil$, for $0 < \alpha \le 1/2$ we have

$$\lim_{\mathsf{SNR}\to\infty} \left\{ I_{\angle} - \frac{\alpha}{2}\ln(\mathsf{SNR}) \right\} \ge \frac{1}{2}\ln\left(\frac{3}{\pi e(\gamma^2 + 3K_{\bar{\Delta}})} \right). \quad (38)$$

For $\alpha > 1/2$ we obtain looser bounds than for the case with $\alpha = 1/2$. Since oversampling with a growth factor $\overline{\alpha}$ contains all the cases with $\alpha < \overline{\alpha}$, for the interval $\alpha > 1/2$ we can use the bound valid for $\alpha = 1/2$.

IV. Discussion

As a byproduct of (28), (29), and (38), a lower bound to the capacity pre-log is

$$\lim_{\mathsf{SNR}\to\infty} \frac{\mathsf{C}(\mathsf{SNR})}{\ln(\mathsf{SNR})} \ge \begin{cases} 2\alpha & 0 < \alpha \le 1/3 \\ (1+\alpha)/2 & 1/3 \le \alpha \le 1/2 \\ 3/4 & 1/2 \le \alpha < 1. \end{cases} \quad (39)$$

Figure 1 shows the lower bounds on the capacity pre-log versus the parameter α, as reported by (39). The contributions of amplitude and phase modulation are also shown separately: Amplitude modulation reaches full degrees of freedom by sampling more than $\sqrt[3]{\mathsf{SNR}}$ samples per symbol, while phase modulation achieves at least half of the available degrees of freedom by using a time resolution that scales as $1/\sqrt{\mathsf{SNR}}$.

The input distribution that achieves the capacity lower bound is uniform in phase and the square amplitude is distributed as a shifted exponential (19). The statistic used for detecting $|X_k|$ is $||\mathbf{Y}_k||$, and the one used for detecting $\angle X_k$ is $\angle\left(Y_{(k-1)L+1}\left(Y_{(k-1)L}e^{-j\angle X_{k-1}}\right)^{\star}\right)$.

V. Conclusions

We have derived a lower bound to the capacity of continuous-time Wiener phase noise channels with an average transmit power constraint. As a byproduct, we have obtained a lower 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 as high as that reported in (39) can be achieved.

[1] For example, choosing $\gamma^2\Delta = 0.01$ gives $K_{\bar{\Delta}} = 8.1353$. See [12, App. E] for a detailed derivation.

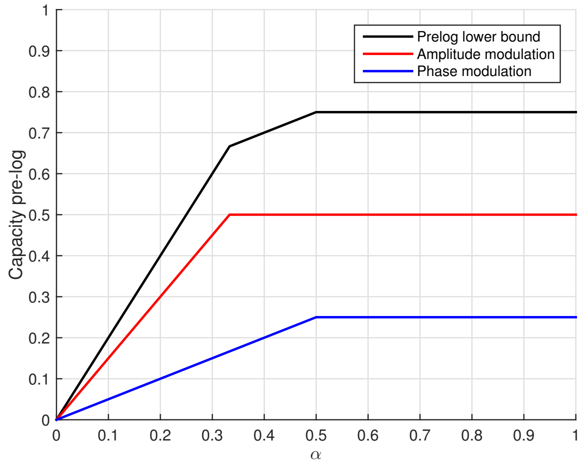

Fig. 1. Capacity pre-log lower bounds as a function of α at high SNR. The oversampling factor L is $L = \lceil \mathsf{SNR}^{\alpha} \rceil$.

Appendix A
A lower bound to $\mathsf{E}\left[\cos(\Phi - \angle X_1)\right]$

The expectation can be simplified as follows:

$$\begin{aligned} &\mathsf{E}\left[\cos(\Phi - \angle X_1)\right] \\ &\overset{(a)}{=} \mathsf{E}\left[\cos(\angle(|X_1|F_1 + W_1) - \angle(|X_0|F_0 e^{-jN_0} + W_0))\right] \\ &\overset{(b)}{=} \mathsf{E}\left[\cos(\angle(|X_1|F_1 + W_1))\right] \mathsf{E}\left[\cos(\angle(|X_0|F_0 e^{-jN_0} + W_0))\right] \\ &\quad + \mathsf{E}\left[\sin(\angle(|X_1|F_1 + W_1))\right] \mathsf{E}\left[\sin(\angle(|X_0|F_0 e^{-jN_0} + W_0))\right] \\ &= \mathsf{E}\left[\cos(\angle(|X_1|F_1 + W_1))\right] \mathsf{E}\left[\cos(\angle(|X_0|F_0 e^{-jN_0} + W_0))\right] \end{aligned} \quad (40)$$

where step (a) is due to (31), step (b) to the addition formula for cosine and independence of random variables, and the last step follows because $\mathsf{E}\left[\sin(\angle(|X_1|F_1 + W_1))\right] = 0$ as F_1 and W_1 have symmetric pdfs with respect to the real axis.

The first expectation on the RHS of (40) can be written as

$$\begin{aligned} \mathsf{E}\left[\Re\{e^{j\angle(|X_1|F_1+W_1)}\}\right] &= \mathsf{E}\left[\Re\{e^{j\angle F_1} e^{j\angle(|X_1 F_1| + W_1)}\}\right] \\ &= \mathsf{E}\left[\cos(\angle F_1)\cos(\angle(|X_1 F_1| + W_1))\right] \end{aligned} \quad (41)$$

where the first step is due to the circular symmetry of W_1, and the second step because of the symmetric pdfs of F_1 and W_1. A lower bound to (41) is given by

$$\begin{aligned} &\mathsf{E}\left[\Re\{e^{j\angle(|X_1|F_1+W_1)}\}\right] \overset{(a)}{\ge} \mathsf{E}\left[\Re\{F_1\}\cos(\angle(|X_1 F_1| + W_1))\right] \\ &\overset{(b)}{\ge} \mathsf{E}\left[\Re\{F_1\}1(\Re\{F_1\} \ge 0)\left(1 - \frac{1}{|X_1 F_1|^2}\right)\right] \\ &\quad + \mathsf{E}\left[\Re\{F_1\}1(\Re\{F_1\} < 0)\right] \end{aligned}$$

$$\stackrel{(c)}{\geq} \mathsf{E}\left[\Re\{F_1\} - 1(\Re\{F_1\} \geq 0)\frac{1}{|X_1F_1|^2}\right]$$
$$\stackrel{(d)}{\geq} \mathsf{E}\left[\Re\{F_1\} - \frac{1}{|X_1F_1|^2}\right]$$
$$\geq \frac{2}{\gamma^2\Delta}\left(1 - e^{-\gamma^2\Delta/2}\right) - \mathsf{E}\left[\frac{1}{|X_1|^2}\right] K_{\bar{\Delta}} \qquad (42)$$

where step (a) holds because $|F_1| \leq 1$, (b) follows by $\cos(x) \leq 1$ and by [12, App. D] [14, Lemma 2]

$$\mathsf{E}\left[\cos(\angle(\rho + W_1))\right] \geq 1 - \frac{1}{\rho^2}, \qquad \rho > 0, \qquad (43)$$

step (c) because $\Re\{F_1\} \leq 1$, step (d) is obtained by subtracting $\mathsf{E}\left[1(\Re\{F_1\} < 0)\,|X_1F_1|^{-2}\right]$, and the final inequality uses $\mathsf{E}\left[|F_1|^{-2}\right] \leq K_{\bar{\Delta}}$ for a finite suitable $\bar{\Delta}$ [12, App. E].

Following an analogous derivation used for finding (42), for the second factor on the RHS of (40) we have

$$\mathsf{E}\left[\Re\{e^{j\angle(|X_0|F_0e^{-jN_0}+W_0)}\}\right] \geq \mathsf{E}\left[\Re\{F_0e^{-jN_0}\} - \frac{1}{|X_0F_0|^2}\right]$$
$$\geq \sqrt{\frac{2\pi}{\gamma^2\Delta}}\,\mathrm{erf}\left(\sqrt{\frac{\gamma^2\Delta}{8}}\right)e^{-3\gamma^2\Delta/8} - \mathsf{E}\left[\frac{1}{|X_0|^2}\right]K_{\bar{\Delta}} \quad (44)$$

where $\mathrm{erf}(\cdot)$ is the error function, and the closed form for $\mathsf{E}\left[F_0e^{-jN_0}\right]$ is provided in Appendix B. Using (42) and (44) into (40), with $\Re\{\mathsf{E}\left[F_1\right]\} \leq 1$, $\Re\{\mathsf{E}\left[F_0e^{-jN_0}\right]\} \leq e^{-3\gamma^2\Delta/8}$, $\mathsf{E}\left[|X_0|^{-2}\right] \geq 0$, and $\mathsf{E}\left[|F_0|^{-2}\right] \geq 0$, the final result is

$$\mathsf{E}\left[\cos(\Phi - \angle X_1)\right] \geq \mathsf{E}\left[F_0e^{-jN_0}\right]\mathsf{E}\left[F_1\right]$$
$$- 2e^{-3\gamma^2\Delta/8}\mathsf{E}\left[\frac{1}{|X_1|^2}\right]K_{\bar{\Delta}}. \qquad (45)$$

Appendix B
Evaluation of $\mathsf{E}\left[F_0e^{-jN_0}\right]$

Knowing that $N_0 = \sigma\int_0^1 B(\tau)\,\mathrm{d}\tau$ with $\sigma = \gamma\sqrt{\Delta}$, we compute

$$\mathsf{Var}\left[\sigma B(t) - N_0\right] = \sigma^2\mathsf{Var}\left[B(t)\right] + \mathsf{Var}\left[N_0\right] - 2\sigma\mathsf{E}\left[B(t)N_0\right]$$
$$= \sigma^2(t+1) - 2\sigma^2\int_0^1 \mathsf{E}\left[B(t)B(\tau)\right]\mathrm{d}\tau$$
$$= \sigma^2(t^2 - t + 1) \qquad (46)$$

where the last step follows from the property of Wiener processes $\mathsf{E}\left[B(t)B(\tau)\right] = \min\{t, \tau\}$. Thus we have

$$\mathsf{E}\left[F_0e^{-jN_0}\right] = \int_0^1 \mathsf{E}\left[e^{j(\sigma B(t) - N_0)}\right]\mathrm{d}t$$
$$\stackrel{(a)}{=} \int_0^1 e^{-\mathsf{Var}[\sigma B(t) - N_0]/2}\,\mathrm{d}t$$
$$= \sqrt{\frac{2\pi}{\sigma^2}}e^{-\frac{3}{8}\sigma^2}\,\mathrm{erf}\left(\sqrt{\frac{\sigma^2}{8}}\right) \qquad (47)$$

where in step (a) we used the characteristic function of a Gaussian random variable, and in the last step we used (46).

Acknowledgment

L. Barletta was supported by the Technische Universität München - Institute for Advanced Study, funded by the German Excellence Initiative. G. Kramer was supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research.

References

[1] 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.

[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] ——, “Signal-to-noise ratio penalties for continuous-time phase noise channels,” in *Int. Conf. on Cognitive Radio Oriented Wirel. Networks (CROWNCOM)*, June 2014, pp. 232–235.

[4] 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.

[5] 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.

[6] ——, “Multi-sample receivers increase information rates for Wiener phase noise channels,” in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2013, pp. 1897–1902.

[7] L. Barletta, M. Magarini, and A. Spalvieri, “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.

[8] 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.

[9] H. Ghozlan and G. Kramer, “Phase modulation for discrete-time Wiener phase noise channels with oversampling at high SNR,” in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2014.

[10] L. Barletta and G. Kramer, “Upper bound on the capacity of discrete-time Wiener phase noise channels,” in *Accepted for presentation at IEEE Inf. Theory Workshop (ITW)*. [Online]. Available: http://arxiv.org/pdf/1411.0390.pdf

[11] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, “Simulation-based computation of information rates for channels with memory,” *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug 2006.

[12] L. Barletta and G. Kramer, “Lower bound on the capacity of continuous-time Wiener phase noise channels,” in *arXiv preprint*, 2015. [Online]. Available: http://arxiv.org/pdf/1503.03223v1.pdf

[13] H. Ghozlan and G. Kramer, “Interference focusing for mitigating cross-phase modulation in a simplified optical fiber model,” in *IEEE Int. Symp. Inf. Theory (ISIT)*, June 2010, pp. 2033–2037.

[14] ——, “Models and information rates for Wiener phase noise channels,” in *arXiv preprint*, 2015. [Online]. Available: http://arxiv.org/pdf/1503.03130.pdf