Phase Modulation for Discrete-time Wiener Phase Noise Channels with Oversampling at High SNR

Hassan Ghozlan  
Department of Electrical Engineering  
University of Southern California  
Los Angeles, CA 90089 USA  
ghozlan@usc.edu

Gerhard Kramer  
Institute for Communications Engineering  
Technische Universität München  
80333 Munich, Germany  
gerhard.kramer@tum.de

Abstract—A discrete-time Wiener phase noise channel model is introduced in which multiple samples are available at the output for every input symbol. A lower bound on the capacity is developed. At high signal-to-noise ratio (SNR), if the number of samples per symbol grows with the square root of the SNR, the capacity pre-log is at least 3/4. This is strictly greater than the capacity pre-log of the Wiener phase noise channel with only one sample per symbol, which is 1/2. It is shown that amplitude modulation achieves a pre-log of 1/2 while phase modulation achieves a pre-log of at least 1/4.

I. INTRODUCTION

Communication systems often suffer from phase noise due to the instability of oscillators [1]. The characteristics of the phase noise process vary by application. In systems with phase tracking devices, such as phase-locked loops (PLL), the residual phase noise follows a stationary ARMA process [6]–[11]. In Digital Video Broadcasting DVB-S2, an example of a satellite communication system, the phase noise process is modeled by a Wiener process [4]. In fiber-optic communication, the phase noise in laser oscillators is modeled by a Wiener process [4]. For discrete-time phase noise channels with a stationary and ergodic phase noise process (whose entropy rate is finite), Lapidot showed that the capacity grows logarithmically with the signal-to-noise ratio (SNR) with a pre-log factor equal to 1/2 at high SNR [5]. The two cases of Wiener phase noise and auto-regressive-moving-average (ARMA) phase noise fall into this class. At finite SNR, numerical methods exist for computing (bounds on) the information rate for Wiener and ARMA phase noise [6]–[11].

In [12]–[15], continuous-time phase noise channels are studied. Continuous-time white phase noise is considered in [12]. In [12] and [13], a discrete-time phase noise channel is developed by discretizing a continuous-time Wiener phase noise channel by oversampling the output of an integrator-and-dump filter at the receiver. It was shown in [12] that, at high SNR, the information rate grows logarithmically with SNR with a pre-log factor equal to 1/2 when the number of samples per symbol grows with the square root of the SNR. This result was established by employing amplitude modulation only. It was shown in [13] through numerical simulations that oversampling improves the information rate for Phase Shift Keying (PSK) modulation (see Fig. 5 in [13]). The question of whether phase modulation can increase the pre-log factor at high SNR is left open.

We study in this paper a discrete-time channel model similar to [8]–[11], namely one without amplitude noise that would arise due to filtering before sampling [12]–[15]. We do this as a first step towards addressing the more complex continuous-time model. Our approach is similar to [12], [13] in that we consider oversampling receivers, where the oversampling rate increases with the square root of the SNR to achieve the maximum pre-log of 1/2 for amplitude modulation. However, as will show, we achieve an additional pre-log of 1/4 by using only 2 samples per symbol.

The paper is organized as follows. In Section II, the discrete-time model of [12] for the Wiener phase noise channel with oversampling is described and a simplified discrete-time channel model is introduced. A lower bound on capacity of the simplified channel is derived in Section III and the paper is concluded with Section IV.

II. DISCRETE-TIME MODEL

We use the following notation: $j = \sqrt{-1}$, $\ast$ denotes the complex conjugate, $\delta_\pi$ is the Dirac delta function, $\lceil \cdot \rceil$ is the ceiling operator. We use $X^k$ to denote $(X_1, X_2, \ldots, X_k)$. We describe the discrete-time model developed in [12]. Let $X^k$ be the input symbols. For every input symbol, there are $L$ output samples. The $k$-th output sample is

$$Y^\text{full}_k = X_{\lceil k/L \rceil} \Delta e^{j\Theta_k} F_k + N_k$$

where $k = 1, \ldots, nL$ and $\Delta = 1/L$. The process $\{N_k\}$ is an independent and identically distributed (i.i.d.) circularly-symmetric complex Gaussian process with mean 0 and $\mathbb{E}[|N_k|^2] = \sigma_N^2 \Delta$ while the process $\{\Theta_k\}$ is the discrete-time Wiener process

$$\Theta_{k+1} = \Theta_k + W_k$$

where $\Theta_k$ is uniform on $[-\pi, \pi]$ and $\{W_k\}$ is an i.i.d. real Gaussian process with mean 0 and $\mathbb{E}[|W_k|^2] = \sigma_W^2 = 2\pi \beta \Delta$. Moreover, $\{W_k\}$ is independent of $\{N_k\}$. The random variable
We decompose the mutual information using the chain rule into two parts:

\[ I(X^n; Y^n) = I(X^n_A; Y^n) + I(\Phi_A^n; Y^n | X^n_A). \]  

The first term represents the contribution of the amplitude modulation while the second term represents the contribution of the phase modulation. First, we analyze the amplitude modulation term. We have

\[ I(X^n_A; Y^n) \overset{(a)}{=} \sum_{k=1}^{n} I(X_{A,k}; Y^n | X_A^{k-1}) \]

\[ \overset{(b)}{\geq} \sum_{k=1}^{n} I(X_{A,k}; V_k | X_A^{k-1}) \]  

where \( V_k \) is a deterministic function of \((Y^n, X_A^{k-1})\). Step (a) follows from the chain rule of mutual information and (b) follows from the data processing inequality. We choose

\[ V_k = \sum_{\ell=1}^{L} |Y_{(k-1)L+\ell}|^2. \]

When \( X^n \) is i.i.d., the pair \((X_{A,k}, V_k)\) with \( V_k \) defined in (15) is independent of \( X_A^{k-1} \) and therefore

\[ I(X_{A,k}; V_k | X_A^{k-1}) = I(X_{A,k}; V_k). \]  

By using the auxiliary-channel lower bound theorem in [16, Sec. VI], we have

\[ I(X_{A,k}; V_k) \geq \mathbb{E}[\log Q_{V \mid X_A}(V_k | X_A)] - \mathbb{E}[\log Q_{V,k}(V_k)] \]  

where \( Q_{V \mid X_A}(v | x_A) \) is an arbitrary auxiliary channel and

\[ Q_{V,k}(v) \equiv \mathbb{E}[p_{X_{A,k}}(x_A)Q_{V \mid X_A}(v | x_A)dx_A] \]  

where \( p_{X_{A,k}}(\cdot) \) is the true distribution of \( X_{A,k} \), i.e., \( Q_{V}(\cdot) \) is the output distribution obtained by connecting the true input source to the auxiliary channel. We choose the auxiliary channel

\[ Q_{V \mid X_A}(v | x_A) = \frac{1}{\sqrt{4\pi^2 x_A^2 \Delta^2 \sigma_N^2}} \exp \left( -\frac{(v - x_A^2 \Delta - \sigma_N^2)^2}{4x_A^2 \Delta^2 \sigma_N^2} \right). \]  

Following steps similar to those in [12], it can be shown that if \( X_A^n \) is i.i.d. with \(|X_k|^2\) distributed according to \( p_{X_k} \) for \( k = 1, \ldots, n \) where

\[ p_{X_k}(|x|^2) = \begin{cases} \frac{2}{\pi} \exp \left(1 - \frac{2|x|^2}{\sigma^2} \right), & |x|^2 \geq P/2 \\ 0, & \text{otherwise} \end{cases} \]  

then

\[ \lim_{\text{SNR} \to \infty} I(X_A^n; Y^n) - \frac{1}{2} \log \text{SNR} \geq -2 - \frac{1}{2} \log(8\pi) \]  

where

\[ I(X_A^n; Y^n) \equiv \lim_{n \to \infty} \frac{1}{n} I(X_A^n; Y^n). \]
Next, we turn our attention to the contribution of the phase modulation. By using the chain rule, we have

\[
I(\Phi^n_k; Y^n | X^n_A) = \sum_{k=1}^{n} I(\Phi_X; Y^n | X^n_A, \Phi^{k-1}_X)
\]

\[
\geq \sum_{k=2}^{n} I(\Phi_X; Y^n | X^n_A, \Phi^{k-1}_X)
\]

\[
\geq \frac{I(\Phi_X; Y^n | X^n_A, \Phi^{k-1}_X)}{n} \sum_{k=2}^{n} (1 + \hat{Z}_{k-1})
\]

\[
= \frac{I(\Phi_X; Y^n | X^n_A, \Phi^{k-1}_X)}{n} \sum_{k=2}^{n} (1 + \hat{Z}_{k-1})
\]

where \(\hat{Y}_k\) is a deterministic function of \((Y^n, X^n_A, \Phi^{k-1}_X)\). Inequality (a) follows from the non-negativity of mutual information and (b) follows from the data processing inequality.

At high SNR, we use some intuition to choose a reasonable processing of \((Y^n, X^n_A, \Phi^{k-1}_X)\) for decoding \(\Phi_X\):

1) Since only the past inputs \(X^{k-1}\) are available, the future outputs \(Y^{n+k}\) are not very useful for estimating \(\Theta_{k-1}\).

2) Since \(\{\Theta_k\}\) is a first-order Markov process, the most recent past input symbol \(X_{k-1}\) and the most recent output sample \(Y_{k-1}\) are the most useful for estimating \(\Theta_{k-1}\). A simple estimator is

\[
e^{-j\Theta_{k-1}} \equiv \frac{Y_{k+1}L}{X_{k+1}\Delta} = e^{-j\Theta_{k-1}} \frac{N_{k+1}L}{X_{k+1}\Delta}
\]

3) Given the current input amplitude \(|X_k|\) and the estimate of \(\Theta_{k-1}\), the first sample \(Y_{k+1}\) in \(Y_k\) is the most useful for decoding \(\Phi_X\) because the following samples become increasingly corrupted by the phase noise. We scale \(Y_{k+1}\) to normalize the variance of the additive noise and write

\[
Y_{(k+1)L+1} = \left|X_k\right| e^{j\Phi_{X,k}} + \tilde{N}_k
e^{j\Theta_{k-1}L+1}
\]

where

\[
\tilde{N}_k \equiv \frac{N_{k+1}L + e^{-j\Theta_{k-1}L+1}}{\sqrt{\Delta}}.
\]

To summarize, we choose

\[
\hat{Y}_k = \left|X_k\right| e^{j\Phi_{X,k}} + \tilde{N}_k \left(1 + \hat{Z}_{k-1}\right) e^{jW_{k+1}}
\]

where \(\hat{N}_k\) and \(\hat{Z}_{k-1}\) are statistically independent and

\[
\hat{N}_k \sim N_c(0, 1)
\]

\[
\hat{Z}_{k-1} \sim \frac{1}{\left|X_k\right|^2 \Delta}
\]

which means that, conditioned on \(\{X_{k-1} = |x_{k-1}|\}, \hat{Z}_{k-1}\) is a Gaussian random variable with mean 0 and variance \(1/\left|X_{k-1}\right|^2 \Delta\). Moreover, \(W_{k+1}L+1\) is statistically independent of \(\hat{N}_k\) and \(\hat{Z}_{k-1}\). The choice of \(\hat{Y}_k\) in (24) implies that

\[
I(\Phi_X; Y_k | X_{A,k}, X_{k-1})
\]

\[
Q_{\hat{Y}_k} = \angle \hat{Y}_k
\]

where \(I_0(\cdot)\) is the zeroth-order modified Bessel function of the first kind and \(\alpha > 0\). This distribution is known as Tikhonov (or von Mises) distribution [17]. Furthermore, define

\[
Q_{\hat{Y}_k} = \angle \hat{Y}_k
\]

\[
\exp(\alpha \cos(\hat{\Phi}_{X,k} - \hat{\Phi}_{X,k}) - \frac{2 \pi I_0(\alpha)}{\sqrt{\text{SNR}}})
\]

where the last equality holds because \(X_1, \ldots, X_n\) are statistically independent and \(\Phi_{X,k}\) is independent of \(X_{A,k}\) with a uniform distribution on \([-\pi, \pi]\). We have

\[
I(\Phi_X; Y_k | X_{A,k}, X_{k-1})
\]

\[
\geq I(\Phi_X; Y_k | X_{A,k}, X_{k-1})
\]

\[
\geq \frac{\text{E} Q_{\hat{Y}_k} = \angle \hat{Y}_k}{\text{E} Q_{\hat{Y}_k} = \angle \hat{Y}_k}
\]

\[
\alpha \text{E} \left[\cos(\hat{\Phi}_{X,k} - \hat{\Phi}_{X,k})\right]
\]

\[
\geq \frac{1}{2} \alpha - \alpha + \text{E} \left[\cos(\hat{\Phi}_{X,k} - \hat{\Phi}_{X,k})\right]
\]

\[
\geq \frac{1}{2} \alpha - \frac{\sigma^2_{\Phi}}{2} - \frac{4 \alpha}{\text{SNR}}
\]

\[
\text{E} \left[\cos(\hat{\Phi}_{X,k} - \hat{\Phi}_{X,k})\right] \geq 1 - \frac{\sigma^2_{\Phi}}{2} - \frac{4 \alpha}{\text{SNR}}
\]

where (a) follows from the data processing inequality, (b) follows by extending the result of the auxiliary-channel lower bound theorem in [16, Sec. VI], (c) follows from (33) and (34), (d) follows from [18, Lemma 2]

\[
I_0(\cdot) \leq \sqrt{\frac{\pi}{2} \sqrt{\text{SNR}} \leq \frac{\pi}{2}}
\]

and (d) holds because

\[
\text{E} \left[\cos(\hat{\Phi}_{X,k} - \hat{\Phi}_{X,k})\right] \geq 1 - \frac{\sigma^2_{\Phi}}{2} - \frac{4 \alpha}{\text{SNR}}
\]

where the proof is omitted.

\[
\tilde{\Phi}_{X,k} \equiv \angle \hat{Y}_k
\]
for $\text{SNR}\Delta > 2$. It follows from (23), (32) and (36) that
\[
\frac{1}{n} I(\Phi_X^n; Y^n | X_A^n) \geq \frac{n-1}{n} \left[ \frac{1}{2} \log \alpha - \alpha \pi \beta \Delta - \frac{4\alpha}{\text{SNR}\Delta} \right].
\] (39)

Hence, we have
\[
I(\Phi_X^n; Y^n | X_A^n) \equiv \lim_{n \to \infty} \frac{1}{n} I(\Phi_X^n; Y^n | X_A^n) \geq \frac{1}{2} \log \alpha - \alpha \pi \beta \Delta - \frac{4\alpha}{\text{SNR}\Delta}.
\] (40)

Suppose $L$ grows with $\text{SNR}$ such that
\[
L = \left\lfloor \beta \sqrt{\text{SNR}} \right\rfloor.
\] (42)
Since $\Delta = 1/L$, we have
\[
\lim_{\text{SNR} \to \infty} \text{SNR}^2 \Delta^2 = \frac{1}{\beta^2}.
\] (43)
Therefore, by setting $\alpha = \text{SNR}\Delta$ and taking the limit of $\text{SNR}$ tending to infinity, we have
\[
\lim_{\text{SNR} \to \infty} I(\Phi_X^n; Y^n | X_A^n) - \frac{1}{4} \log \text{SNR} \geq \log \frac{1}{\beta} - \frac{\pi}{\beta} - 4.
\] (44)

The last equation implies that the phase modulation contributes $1/4$ to the pre-log of the information rate when oversampling is employed. It follows from (10), (22), (40) and (15) that
\[
I(X; Y) = I(X_A; Y) + I(\Phi_X; Y | X_A)
\] (45)
Combining (21) and (44) yields (11).

It is worth pointing out that the phase modulation pre-log of $1/4$ requires only 2 samples per symbol for which the time resolution, $1/\Delta$, grows as the square root of the SNR. It is interesting to contemplate whether another receiver, e.g., a non-coherent receiver, can achieve the maximum amplitude modulation pre-log of $1/2$ but requires only 1 sample per symbol. If so, one would need only 3 samples per symbol to achieve a pre-log of $3/4$.

IV. Conclusion

We studied a discrete-time model of a Wiener phase noise channel with oversampling. We showed that, at high SNR, the capacity grows logarithmically with SNR with a pre-log of at least $3/4$ if the number of samples per symbol grows with the square root of the SNR. It was found that amplitude modulation and phase modulation can achieve pre-log factors of $1/2$ and $1/4$, respectively. In fact, the phase modulation pre-log of $1/4$ requires only 2 samples per symbol.

Acknowledgment

H. Ghozlan was supported by a USC Annenberg Fellowship and NSF Grant CCF-09-05235. G. Kramer was supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research.

REFERENCES

[1] A. Demir, A. Mehrotra, and J. Roychowdhury. Phase noise in oscillators: a unifying theory and numerical methods for characterization. IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications, 47(5):655–674, 2000.

[2] A.J. Viterbi. Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques. Proc. IEEE, 51(12):1737–1753, 1963.

[3] E. Casini, R. D. Gaudenzi, and A. Ginesi. DVB-S2 modem algorithms design and performance over typical satellite channels. Int. J. Satell. Commun. Network, 22(3):281–318, 2004.

[4] R.W. Tkach and A.R. Chraplyvy. Phase noise and linewidth in an InGaAsP DFB laser. J. Lightwave Tech., 4(11):1711–1716, 1986.

[5] A. Lapidoth. Capacity bounds via duality: A phase noise example. In Proc. 2nd Asian-Euro. Workshop on Inf. Theory, pages 58–61, 2002.

[6] J. Daauwels and H.-A. Loeliger. Computation of information rates by particle methods. IEEE Trans. Inf. Theory, 54(1):406–409, Jan. 2008.

[7] A. Barbieri and G. Colavolpe. On the information rate and repeat-accumulate code design for phase noise channels. IEEE Trans. Comm., 59(12):3223–3228, Dec. 2011.

[8] L. Barletta, M. Magarini, and A. Spalvieri. Estimate of information rates of discrete-time first-order Markov phase noise channels. IEEE Phot. Techn. Lett., 23(21):1582–1584, Nov. 2011.

[9] L. Barletta, M. Magarini, and A. Spalvieri. The information rate transferred through the discrete-time Wiener’s phase noise channel. J. Lightwave Tech., 30(10):1480–1486, May 2012.

[10] L. Barletta, M. Magarini, and A. Spalvieri. A new lower bound below the information rate of Wiener phase noise channel based on Kalman carrier recovery. Opt. Express, 20(23):25471–25477, Nov. 2012.

[11] L. Barletta, M. Magarini, and A. Spalvieri. Tight upper and lower bounds to the information rate of the phase noise channel. IEEE Int. Symp. Inf. Theory, pages 2284–2288, 2013.

[12] G. Kramer and H. Ghozlan. On Wiener phase noise channels at high signal-to-noise ratio. IEEE Int. Symp. Inf. Theory, pages 2279–2283, 2013.

[13] H. Ghozlan and G. Kramer. Multi-sample receivers increase information rates for wiener phase noise channels. Global Telecom. Conf. (GLOBECOM), pages 1919–1924, 2013.

[14] L. Barletta and G. Kramer. On continuous-time white phase noise channels. IEEE Int. Symp. Inf. Theory, 2014.

[15] L. Barletta and G. Kramer. Signal-to-noise ratio penalties for continuous-time phase noise channels. Int. Conf. on Cognitive Rad. Oriented Wireless Net. (CROWN), 2014.

[16] H. Ghozlan and G. Kramer. Interference focusing for mitigating cross-phase modulation in a simplified optical fiber model. IEEE Int. Symp. Inf. Theory, pages 2033–2037, 2010.