Cyclic Water-Filling for Additive Wide-Sense Cyclostationary Gaussian Noise Channel

Byung Wook Han and Joon Ho Cho
Department of Electronic and Electrical Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Gyeongbuk 790-784, Korea
Email: {hanbw,jcho}@postech.ac.kr

***Abstract*—In this paper, we derive the capacity of a continuous-time band-limited channel corrupted by an additive wide-sense cyclostationary (WSCS) proper-complex Gaussian noise, and find the jointly optimal signaling and reception scheme to asymptotically achieve the capacity. Unlike a wide-sense stationary (WSS) noise, a WSCS noise has a structure that can be exploited in the noise suppression. Such a noise is often encountered when co-channel interference generated by linearly modulated codeword symbols from a Gaussian codebook is present in the frequency band of interest. The problem is formulated as a variational problem to find the optimal correlation function of the channel input, and the solution is obtained in the frequency domain by using the vectorized Fourier transform (VFT) and the matrix-valued PSD (MV-PSD) technique. It turns out that the optimal solution distributes the transmit power over the frequency band in a water-filling-like way that results in a WSCS, not WSS, signal as the channel input. Numerical results show that this cyclic water-filling (CWF) scheme performs significantly better than the ordinary water-filling scheme that ignores the spectral correlation of the WSCS noise.**

I. Introduction

Consider a channel model depicted in Fig. 1. This is the complex baseband equivalent model of an additive proper-complex wide-sense cyclostationary (WSCS) Gaussian noise channel examined in this paper. The channel input is $X(t)$ and the output is $Y(t)$. The key questions to be answered in this paper are

- What is the capacity of this channel?
- What are the mathematical tools needed to formulate the problem?
- What is the optimal signaling scheme that achieves this capacity?

The admissible input $X(t)$ is a complex-valued second-order random process with average power constraint P, i.e.,

$$\lim_{\epsilon\to\infty}\frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\mathbb{E}\{|X(t)|^2\}dt\le P. \tag{1}$$

This input signal passes through a frequency-selective channel modeled by a linear-time invariant (LTI) filter with impulse response $h(t)$ and the filter output is corrupted by an additive noise $n(t)$. The LTI filter has a piecewise smooth frequency response $H(\xi)$ and it is strictly band-limited to $\xi\in[-B,B)$.[1]

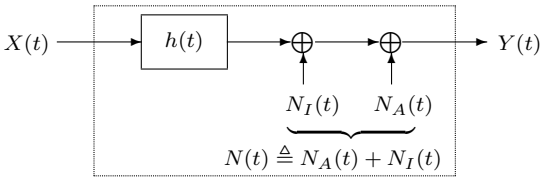

Fig. 1. Complex baseband equivalent model of additive WSCS Gaussian noise channel.

The additive noise $N(t)$ consists of a proper-complex white Gaussian noise $N_A(t)$ and a proper-complex zero-mean WSCS Gaussian random process $N_I(t)$ with cycle period $T_0>0$. The white noise models ambient noise and the WSCS random process models interference. It is assumed that the interference $N_I(t)$ is also band-limited to $\xi\in[-B,B)$. Thus, the overall additive noise $N(t)$ is also a proper-complex zero-mean WSCS Gaussian random process satisfying $\mathbb{E}\{N(t)\}=0,\forall t$, $\mathbb{E}\{N(t_1)N(t_2)\}=0,\forall t_1,t_2$, and

$$r_N(t_1,t_2)\triangleq\mathbb{E}\{N(t_1)N(t_2)^*\} \tag{2a}$$
$$=r_N(t_1+T_0,t_2+T_0),\forall t_1,t_2, \tag{2b}$$

where the superscript * denotes complex conjugation. Since the white noise outside the frequency band $\xi\in[-B,B)$ is irrelevant, we assume that $N(t)$ is band-limited to $\xi\in[-B,B)$ without loss of generality.

The objectives, in other words, are to find the maximum throughput of an optimal channel encoder and decoder pair that operates over this channel with an arbitrarily small error rate, subject to the average power constraint (1), and to find such an encoder/decoder pair. Without the interference component, the answers are well known [1], [2]. However, no answer is currently available with the WSCS interference.

Why do we consider such a channel, by the way? It is because we often encounter co-channel interference in wired and wireless digital communications and such a man-made signal in many cases exhibits cyclostationarity. In particular,

This work was supported in part by the Ministry of Knowledge Economy under the grant IITA-2008-C1090-0801-0037 for the BrOMA-ITRC@POSTECH supervised by the IITA and in part by the Ministry of Education, Science and Technology under the KOSEF grant No. R01-2008-000-11265-0.

[1] In this paper, ξ is used for frequency in Herz (Hz), while f is reserved for frequency offset from the origin $\xi=0$.

if a communicator that occupies the entire or a part of the frequency band of interest employs linear modulation, which is one of the most popular digital modulation schemes, the second-order statistical property of the interference is best captured by a WSCS random process model rather than by a wide-sense stationary (WSS) model [3]. Of course, the use of an ideal sinc pulse as the symbol waveform makes the WSCS signal degenerate to a WSS signal, because any symbol waveform with excess bandwidth less than or equal to zero makes the linearly modulated signal WSS [4]. However, a Nyquist theorem states that, to remove intersymbol interference (ISI), a non-negative excess bandwidth is necessary. Moreover, the sinc pulse that is the only Nyquist pulse with zero excess bandwidth is not an absolutely integrable function, which prevents it from being used in a practical implementation of a linear modulator. Therefore, the use of symbol waveform with a positive excess bandwidth is prevalent in the system design, which justifies our WSCS assumption.[2] Moreover, if the data symbols of the linear modulation is from a Gaussian codebook, which is the optimal codebook for an AWGN channel, the interference becomes a WSCS Gaussian random process. Precisely speaking, the use of a complex Gaussian codebook results in a strictly cyclostationary proper-complex zero-mean Gaussian random process. In this paper, we simply call this process as a WSCS Gaussian random process.

The capacity of this channel is well known [1] for real baseband channel with a zero-mean WSS real Gaussian noise. The extension is straightforward by combining the results in [1] and [8] to complex baseband channel with a proper-complex zero-mean WSS Gaussian $N(t)$. Resorting to the capacity of a parallel Gaussian channel, we can divide the frequency band into equal-length subintervals and invoke the piecewise constant approximation of the channel and the PSD of the noise. For this, it suffices to have a piecewise smooth frequency response of the channel and a piecewise smooth PSD of the noise. Then, the capacity of each subinterval can be achieved by using a linear modulating codeword symbols from a complex Gaussian codebook with the sinc pulse, where the code rate of the codebook and the transmit power of the linearly modulated signal are determined by a water-filling algorithm. The sinc pulse used in each subinterval can be approximated by any square-root Nyquist pulse with a very small excess bandwidth. Such a design with sharp transition in the frequency domain is possible because the pulse now has very low frequency components only. By letting the number of subintervals tend to infinity, the approximation loss vanishes and the problem is formulated in the frequency domain as

Problem 1:

$$\underset{R_X(\xi)}{\text{maximize}} \int_{-B}^{B} \log_2 \left(1 + \frac{|H(\xi)|^2 R_X(\xi)}{R_N(\xi)}\right) d\xi \tag{3a}$$

$$\text{subject to} \int_{-B}^{B} R_X(\xi) d\xi \leq P, \tag{3b}$$

where $R_X(\xi)$ is the PSD of the transmitted WSS signal. The solution to this optimization problem can be obtained by a water-filling algorithm as

$$C_{\text{WSS}} = \int_{-B}^{B} \log_2 \left(1 + \frac{\left[\nu_{\text{opt}} - \dfrac{R_N(\xi)}{|H(\xi)|^2}\right]^+}{\dfrac{R_N(\xi)}{|H(\xi)|^2}}\right) d\xi, \tag{4a}$$

where ν_{opt} is the unique solution to

$$\int_{-B}^{B} \left[\nu_{\text{opt}} - \frac{R_N(\xi)}{|H(\xi)|^2}\right]^+ d\xi = P \tag{4b}$$

and $[x]^+ \triangleq (x + |x|)/2$ denotes the positive part of x.

How can we formulate an optimization problem such as *Problem 1* for the complex baseband channel in Fig. 1 with a proper-complex zero-mean WSCS Gaussian noise? To handle the frequency limitation, it may be necessary to formulate the problem in the frequency domain. However, it is obvious that the power constraint (1) can no longer be written in the frequency domain as (3b) because (1) reduces to (3b) only when the noise is WSS [9]. Since we do not know yet whether the optimal transmit signal is WSS or not, we definitely need different mathematical tools to formulate the problem in the frequency domain from the ordinary ones of the Fourier transform and the PSD.

It turns out that the tools needed to formulate the problem in a mathematically tractable form are the vectorized Fourier transform (VFT) and the matrix-valued PSD (MV-PSD). These tools have been used to design cyclic Wiener filters and linear receivers in a cyclostationary noise [10]. They are also used to find jointly optimal linear transmitter (Tx) and receiver (Rx) pairs, where the mean-squared error (MSE) at the output of the linear receiver is the objective to be minimized and the transmit waveform for linear modulation is found [11]–[18]. However, no information-theoretic result has been derived yet using these tools. We will first summarize the definitions and main results from the non-information theoretic joint Tx-Rx optimization problems [16] and [19]. Then, the problem will be formulated and the solution will be derived using the definitions and the results. Interestingly, the problem described in the frequency domain resembles the capacity computation problem for multi-input multi-output (MIMO) channels. Thus, once the problem is formulated the solution is readily obtained by mostly following the well-known procedure of singular-value decomposition of the channel matrix.

The application of the solution to a selfish and a selfless overlay design is considered, where the WSCS interference is now the signal from a legacy system operating in the frequency band of interest and the system to be designed is an overlay

[2]It is noteworthy that a short-code code-division multiple-access (CDMA) transmitter can be viewed as a linear modulator with a very large excess bandwidth in the symbol waveform level and that a long-code CDMA transmitter can be viewed as a linear modulator with a small excess bandwidth in the chip pulse level [5]. For example, in the WCDMA standard, the square-root raised cosine (SRRC) pulse with excess bandwidth 0.22 is employed as the chip pulse [6]. In addition, an orthogonal frequency division multiplexing (OFDM) transmitter can also be viewed as a linear modulator having a symbol waveform with a positive excess bandwidth [7].

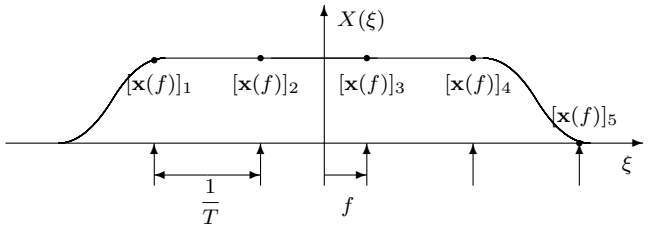

Fig. 2. Frequency-domain sampling of the ordinary Fourier transform $X(\xi)$ of a deterministic signal $x(t)$ with offset f to construct the VFT $\mathbf{x}(f)$ with reference rate $1/T$.

system sharing the frequency band with the legacy system. The application to orthogonal multiple-access communications is also considered. It is shown that the multiple-access scheme performs the same as the optimal FDMA scheme even with overlapping spectra.

The rest of this paper is organized as follows. In Section II, the preliminaries are provided that summarize the definitions and non-information theoretic results in the related previous work. In Section III, the problem is formulated and the solution is provided. Discussions and numerical results are provided in Section IV, and concluding remarks are offered in Section V.

II. Preliminaries

In this section, we first summarize the definitions that are needed to formulate and solve the information and non-information theoretic joint Tx-Rx optimization problems in additive WSCS noise. Then, we summarize the non-information theoretic results in [16] and [19]. We also find some new insights into the non-information theoretic results.

A. Definitions and Examples

In this subsection, we provide the definitions of the VFT and the MV-PSD. These definitions are modified versions of the ordinary definitions of the VFT and the MV-PSD in [16]–[18]. We also provide some examples to enhance the understanding of these definitions.

To begin with, we define the notion of excess bandwidth. This definition is a generalized version of the ordinary definition of the excess bandwidth [16, *Definition 4*] for linearly modulated signals.

Definition 1: Given a bandwidth and reference rate pair $(B, 1/T)$, the excess bandwidth $\beta(B, 1/T)$, simply denoted by β, is defined by the relation

$$BT = \frac{1+\beta}{2}. \tag{5}$$

Note that the excess bandwidth is not necessarily non-negative.[3] If applied to the linear modulation with a SRRC waveform and the reference rate equal to the symbol rate, the excess bandwidth corresponds to the roll-off factor. In what follows, we simply call reference rate as rate, whenever no confusion arises.

[3]For a system design with a negative excess bandwidth, see [20].

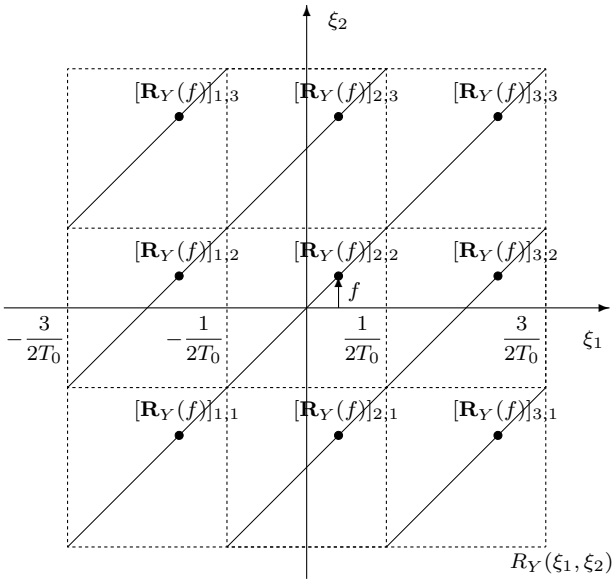

Fig. 3. 2-D frequency-domain sampling of the double Fourier transform $R_Y(\xi_1, \xi_2)$ with offset f to construct the MV-PSD $\mathbf{R}_Y(f)$ with bandwidth-rate pair $(3/(2T_0), 1/T_0)$ of a cyclostationary random process $Y(t)$ with cycle period T_0.

Definition 2: Given a bandwidth-rate pair $(B, 1/T)$, the VFT $\mathbf{x}(f)$ of a deterministic signal $x(t)$ is defined as a vector-valued function of $f \in [-1/(2T), 1/(2T))$, whose kth entry is given by

$$[\mathbf{x}(f)]_k \triangleq X\left(f + \frac{k-L+1}{T}\right) \tag{6}$$

for $k = 1, 2, ..., 2L+1$, where $X(\xi)$ is the ordinary Fourier transform of $x(t)$ and the integer $L \triangleq \lceil \beta/2 \rceil$ determines the length $(2L+1)$ of the VFT.

Fig. 2 illustrates how to construct a VFT from an ordinary Fourier transform. After forming a row vector by sampling the ordinary Fourier transform with sampling period $1/T$ and offset f, we can obtain the VFT in a column vector form at offset f by rotating the row vector 90 degrees clockwise.

Definition 3: Given a bandwidth-rate pair $(B, 1/T)$, the MV-PSD $\mathbf{R}_Y(f)$ of a WSCS random process $Y(t)$ with fundamental cycle period $T_0 = T/K$ for some positive integer K is defined as a matrix-valued function of $f \in [-1/(2T), 1/(2T))$, whose (k,l)th entry is given by

$$[\mathbf{R}_Y(f)]_{k,l} \triangleq R_Y^{(k-l)}\left(f + \frac{l-L-1}{T}\right) \tag{7}$$

for $k, l = 1, 2, ..., 2L+1$, where $R_Y^{(k)}(\xi)$ is the Fourier transform of $r_Y^{(k)}(t)$ such that

$$r_Y(t_1, t_2) \triangleq \mathbb{E}\{Y(t_1)Y(t_2)^*\} = \sum_{k=-\infty}^{\infty} r_Y^{(k)}(t_1 - t_2) e^{\mathrm{j}\frac{2\pi k}{T} t_1} \tag{8}$$

for all t_1, t_2, and the integer $L \triangleq \lceil \beta/2 \rceil$ determines the size $(2L+1)$-by-$(2L+1)$ of the MV-PSD.

Since $T = KT_0$ is also a cycle period of a random process satisfying $r_Y(t_1, t_2) = r_Y(t_1 + T_0, t_2 + T_0), \forall t_1, t_2$, by using the results in [21, p. 242] and [22], respectively, we can show that the series representation (8) of $r_Y(t_1, t_2)$ is always possible and that the double Fourier transform $R_Y(\xi_1, \xi_2)$ of $r_Y(t_1, t_2)$ can be rewritten as

$$R_Y(\xi_1, \xi_2) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_Y(t_1, t_2) e^{-\mathrm{j}2\pi(\xi_1 t_1 - \xi_2 t_2)} dt_1 dt_2 \quad (9a)$$

$$= \sum_{k=-\infty}^{\infty} R_Y^{(k)}\left(\xi_1 - \frac{k}{T}\right)\delta\left(\xi_1 - \xi_2 - \frac{k}{T}\right) \quad (9b)$$

where $\delta(\cdot)$ denotes the Dirac delta function. Fig. 3 illustrates how to construct an MV-PSD from the the double Fourier transform that consists of impulse fences. Similar to the construction of the VFT, we can obtain the MV-PSD at offset f by rotating the matrix, which is formed by 2-D sampling the double Fourier transform with sampling period $1/T$ and offset f, 90 degrees clockwise. The major difference from the VFT is that the sampled value is the height of the impulse fence at the sampling point, not the value of the double Fourier transform, which is either infinity or zero.

Example 1: (a) Any WSS process is also a WSCS process. It can be shown that the MV-PSD of a WSS process is always a diagonal matrix whose diagonal entries are the entries of the VFT of the autocorrelation function. (b) Any linearly modulated signal with a proper-complex zero-mean WSS random data sequence is a WSCS process with fundamental cycle period equal to the symbol period. It can be shown that the MV-PSD of such a process $Z(t)$ with reference rate $1/T$ equal to the symbol transmission rate is given by

$$\mathbf{R}_Z(f) = \frac{M(fT)}{T}\mathbf{s}(f)\mathbf{s}(f)^{\mathcal{H}}, \quad (10)$$

where $M(\xi)$ is the PSD of the WSS data sequence, $\mathbf{s}(f)$ is the VFT of the symbol waveform $s(t)$, and the superscript $^{\mathcal{H}}$ denotes Hermitian transposition.

When we design a band-limited signal $x(t)$ or a WSCS random process $y(t)$, not every entry in $\mathbf{x}(f)$ and $\mathbf{R}_Y(f)$ are free variables if $B < 1/(2T) + L/T$. Thus, the following refinements are useful, which differentiates the VFT and the MV-PSD from their prototypes used in [11]–[15].

Definition 4: (a) Given a bandwidth-rate pair $(B, 1/T)$, the effective VFT of $x(t)$ is defined as a variable-dimensional function of f that is obtained after removing the first entry of $\mathbf{x}(f)$ for

$$-\frac{1}{2T} \le f < -\frac{1+\beta}{2T} + \frac{L}{T} \quad (11a)$$

and the last entry for

$$\frac{1+\beta}{2T} - \frac{L}{T} \le f < \frac{1}{2T}. \quad (11b)$$

(b) The effective MV-PSD of $Y(t)$ is also similarly defined by removing the first row and column of $\mathbf{R}_Y(f)$ for f satisfying the condition (11a) and and the last row and column for f satisfying the condition (11b).

For simplicity, we do not introduce a new notation for the effective VFT and MV-PSD. In what follows, all the VFTs and the MV-PSDs are effective ones, unless otherwise specified.

Example 2: Let $s(t)$ be a SRRC pulse with roll-off factor $0 < \beta \le 1$, energy T, and bandwidth $(1+\beta)/(2T)$. Then, given a bandwidth-rate pair $((1+\beta)/(2T), 1/T)$, the effective VFT $\mathbf{s}(f)$ is given by

$$\mathbf{s}(f) = \begin{cases} 1, & \text{for } -\frac{1-\beta}{2T} \le f < \frac{1-\beta}{2T} \\ \begin{bmatrix} \sqrt{1-c(f)} \\ \sqrt{c(f)} \end{bmatrix}, & \text{for } \frac{1-\beta}{2T} \le f < \frac{1}{2T} \\ \begin{bmatrix} \sqrt{c(f)} \\ \sqrt{1-c(f)} \end{bmatrix}, & \text{for } -\frac{1}{2T} \le f < -\frac{1-\beta}{2T}, \end{cases} \quad (12)$$

where $c(f)$ is defined as

$$c(f) = \frac{1}{2} + \frac{1}{2}\cos\left[\frac{2\pi T}{\beta}\left(|f| - \frac{1-\beta}{2T}\right)\right]. \quad (13)$$

Note that the non-effective VFT is a 3-by-1 vector for all f, while the effective VFT is a scalar for $-(1-\beta)/(2T) \le f < (1-\beta)/(2T)$ and a 2-by-1 vector elsewhere.

Definition 5: Given a bandwidth-rate pair $(B, 1/T)$, the degree of freedom $\mathcal{N}(f)$ as a function of $f \in [-1/(2T), 1/(2T))$ is defined as the length of the effective VFT at f.

It is shown in [17] that $\mathcal{N}(f)$ is given by

$$\mathcal{N}(f) = \begin{cases} 1 + \lceil\beta\rceil, & \text{for } |f| < \frac{1+\beta-\lceil\beta\rceil}{2T} \\ \lceil\beta\rceil, & \text{otherwise} \end{cases} \quad (14a)$$

for even $\lceil\beta\rceil$, and

$$\mathcal{N}(f) = \begin{cases} \lceil\beta\rceil, & \text{for } |f| < \frac{\lceil\beta\rceil-\beta}{2T} \\ 1 + \lceil\beta\rceil, & \text{otherwise} \end{cases} \quad (14b)$$

for odd $\lceil\beta\rceil$, where β is the excess bandwidth. An interesting observation is made in [17] that the integral of $\mathcal{N}(f)$ always equals $2B$ whether $\lceil\beta\rceil$ is odd or even, i.e.,

$$\int_{-\frac{1}{2T}}^{\frac{1}{2T}} \mathcal{N}(f) df = \frac{1+\beta}{T} = 2B. \quad (15)$$

Note that $2B$ is the bandwidth of the frequency band of interest in the passband. Thus, $\mathcal{N}(f)$ can be interpreted as the bandwidth density at offset f.

B. Summary of Non-Information Theoretic Results

In this subsection, we provide a summary of the non-information theoretic results from [16] and [19]. Fig. 4 shows the system model, where the data sequence $(b[k])_k$ is modeled as a discrete-time WSS random process with auto-correlation function and PSD given by

$$m[k] \triangleq \mathbb{E}\{b[k+l]b[l]^*\} \text{ and } M(\xi) \triangleq \sum_{k=-\infty}^{\infty} m[k] e^{-\mathrm{j}2\pi\xi k}, \quad (16)$$

respectively, and the noise $N(t) = N_A(t) + N_I(t)$ is a WSCS random process with cycle period $T_0 = T/K$ for some positive integer K.

The objective is to find the optimal Tx and Rx waveforms $s(t)$ and $w(t)$ for linear modulation and linear demodulation

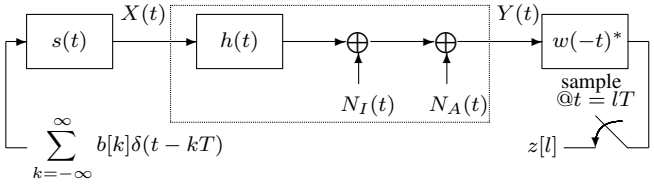

Fig. 4. Complex baseband equivalent model of additive WSCS Gaussian noise channel with linear modulator and demodulator.

with symbol rate $1/T$ that jointly minimizes the MSE at the output of the linear receiver, i.e., the optimization problem to be solved is given by

Problem 2:

$$\underset{s(t),w(t)}{\text{minimize}} \ \mathbb{E}\{|b[l]-z[l]|^2\} \tag{17a}$$

$$\text{subject to } \mathbb{E}\left\{\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\Big|\sum_{k=-\infty}^{\infty} b[k]s(t-kT)\Big|^2 dt\right\} \le P, \tag{17b}$$

where the average power constraint (1) reduces to (17b) because the power $\mathbb{E}\{|X(t)|^2\}$ as a function of t is periodic with period T.

Of course, there are implicit bandwidth constraints additional to the power constraint (17b). Define the frequency band observable by the receiver as the receive band $\mathcal{W}_R$ and the frequency band over which the transmitter can emit signal power as the transmit band $\mathcal{W}_T$. Then, there are three cases worth considering in the system design: [19]

1) Both $\mathcal{W}_T$ and $\mathcal{W}_R$ are given.
2) Only $\mathcal{W}_R$ is given.
3) Only $\mathcal{W}_T$ is given.

Note that the optimal Tx does not allocate any signal power in the complement $(\mathcal{W}_R)^c$ of $\mathcal{W}_R$ because the signal component in $(\mathcal{W}_R)^c$ is not observed by the Rx and, consequently, it cannot contribute to reducing the MSE. Thus, without loss of optimality, we re-define the transmit band as $\mathcal{W}_T \cap \mathcal{W}_R$ in case 1), and we set $\mathcal{W}_T = \mathcal{W}_R$ in case 2).

In case 3), in order not to waste any transmit power, $\mathcal{W}_R$ must be chosen to include $\mathcal{W}_T$. Moreover, it is potentially suboptimal not to include a frequency band having an observable that has non-zero correlation with the observable inside $\mathcal{W}_R$. This is because the non-zero correlation can be exploited in reducing the MSE. Thus, we set $W_{R,0}$ $(\le W_{T,0})$ and $W_{R,1}$ $(\ge W_{T,1})$, respectively, as the maximum and the minimum values of frequencies such that the observable inside $\mathcal{W}_R$ has no correlated observable outside $\mathcal{W}_R$, where $W_{R,0}$ and $W_{R,1}$ are assumed finite.

In all three cases, the transmit band is now contained in the receive band. So, it is convenient to choose the center frequency of $\mathcal{W}_R$ as the reference for complex baseband representation and the baseband bandwidth of $\mathcal{W}_R$ as the bandwidth B for the vectorization.

Using the VFT and the MV-PSD with bandwidth-rate pair $(B, 1/T)$, we can convert *Problem 2* as [16], [19]

$$\underset{\mathbf{s}(f),\mathbf{w}(f)}{\text{minimize}} \int_{\mathcal{F}} TM(fT) + \mathbf{w}(f)^{\mathcal{H}}\mathbf{R}_Y(f)\mathbf{w}(f) - 2\Re\big\{M(fT)\mathbf{w}(f)^{\mathcal{H}}\mathbf{H}(f)\mathbf{s}(f)\big\}df \tag{18a}$$

$$\text{subject to} \int_{\mathcal{F}} \frac{M(fT)}{T}\mathbf{s}(f)^{\mathcal{H}}\mathbf{s}(f)df \le P, \tag{18b}$$

$$\mathbf{G}(f)^{\mathcal{H}}\mathbf{s}(f) = \mathbf{0}, \forall f \in \mathcal{F}, \tag{18c}$$

which is now described in the frequency domain, where $\mathcal{F}$ is the Nyquist interval, with respect to the reference rate $1/T$, defined as

$$\mathcal{F} \triangleq \left\{f : -\frac{1}{2T} \le f < \frac{1}{2T}\right\}, \tag{19}$$

$\mathbf{s}(f)$ and $\mathbf{w}(f)$ are the VFTs of the Tx and the Rx waveforms $s(t)$ and $w(t)$, respectively, $\mathbf{R}_Y(f)$ is the MV-PSD of the channel output $y(t)$ given by

$$\mathbf{R}_Y(f) = \mathbf{R}_N(f) + \frac{M(fT)}{T}\mathbf{H}(f)\mathbf{s}(f)\mathbf{s}(f)^{\mathcal{H}}\mathbf{H}(f)^{\mathcal{H}}, \tag{20}$$

and $\mathbf{H}(f)$ is the channel matrix given by

$$\mathbf{H}(f) = \text{diag}\big(\mathbf{h}(f)\big) \tag{21}$$

with $\mathbf{h}(f)$ being the VFT of the channel impulse response $h(t)$.

The orthogonality constraint (18c) is required when the transmit band in the complex baseband is a proper subset of the receive band in the complex baseband. This is because $s(t)$ must have no signal energy over some frequency components in $\xi \in [-B, B)$. Let the receive band in the complex baseband be $\mathcal{B}_R \triangleq [-B, B)$, the transmit band in the complex baseband be $\mathcal{B}_T \subset [-B, B)$, and the band over which $H(\xi) \neq 0$ be $\mathcal{B}_H$. Then, the blocking matrix $\mathbf{G}(f)^{\mathcal{H}}$ is defined as

$$\mathbf{G}(f) \triangleq [\mathbf{g}_{-L}(f)\ \mathbf{g}_{-L}(f)\ \cdots\ \mathbf{g}_L(f)], \tag{22}$$

where $\mathbf{g}_l(f)$, for $l = -L, -L+1, ..., L$, is the VFT of a time-function $g_l(t)$ whose Fourier transform $G_l(\xi)$ is defined as

$$G_l(\xi) = \begin{cases} 1, & \text{for } \xi \in \left[\frac{l}{T} - \frac{1}{2T}, \frac{l}{T} + \frac{1}{2T}\right) \setminus \big(\mathcal{B}_T \cap \mathcal{B}_H\big), \\ 0, & \text{elsewhere.} \end{cases} \tag{23}$$

The orthogonality constraint is also required when we consider overlaying a cognitive radio that does not induce any interference to the legacy receiver, under the assumption that the modulation and demodulation parameters of the legacy user are fixed. In this case, the VFT of the convolution between the impulse response of the legacy receiver and the channel response from the overlay Tx to the legacy Rx needs to be augmented to the matrix $\mathbf{G}(f)$.

Example 3: [19, *Example 1*] Suppose that $|H(\xi)|$, $\mathcal{B}_R$, and $\mathcal{B}_T$ are given as illustrated in Fig. 5. Then, $\mathbf{g}_l(f)$ are given by

$$\mathbf{g}_{-1}(f) = \begin{cases} [1\ 0\ 0]^{\mathcal{H}}, & \text{for } -\frac{1}{2T} \le f < -\xi_0 + \frac{2}{2T}, \\ [0\ 0\ 0]^{\mathcal{H}}, & \text{for } -\xi_0 + \frac{2}{2T} \le f < \frac{1}{2T}, \end{cases} \tag{24a}$$

$$\mathbf{g}_0(f) = [0\ 0\ 0]^{\mathcal{H}}, \text{ for } -\tfrac{1}{2T} \le f < \tfrac{1}{2T}, \tag{24b}$$

and

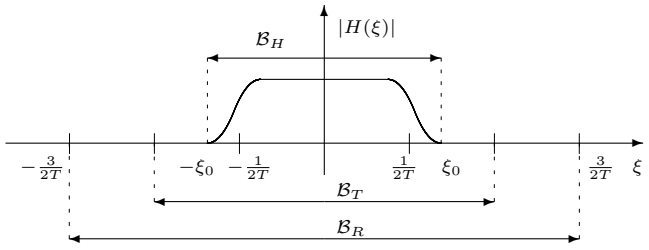

Fig. 5. $|H(\xi)|$, $\mathcal{B}_R$, $\mathcal{B}_T$, and $\mathcal{B}_H$ for *Example 3*.

$$\mathbf{g}_1(f) = \begin{cases} [0\ 0\ 0]^{\mathcal{H}}, & \text{for } -\frac{1}{2T} \le f < \xi_0 - \frac{2}{2T}, \\ [0\ 0\ 1]^{\mathcal{H}}, & \text{for } \xi_0 - \frac{2}{2T} \le f < \frac{1}{2T}. \end{cases} \quad (24c)$$

From the results in [4], an interesting observation can be made for the cases with $K \neq 1$, i.e., $T = KT_0 > T_0$: If a cyclostationary interference $N_I(t)$ has a rank-one MV-PSD for bandwidth-rate pair $(B, 1/T_0)$, i.e.,

$$\mathbf{R}_{N_I}(f) = \frac{1}{T_0}\mathbf{q}(f)\mathbf{q}(f)^{\mathcal{H}} \quad (25)$$

where $\mathbf{q}(f)$ is the VFT of a time-function $q(t)$ for bandwidth-rate pair $(B, 1/T_0)$, then the MV-PSD of $N_I(t)$ for bandwidth-rate pair $(B, 1/T)$ is given by

$$\mathbf{R}_{N_I}(f) = \frac{1}{T}[\mathbf{q}(f)\mathbf{q}(f)^{\mathcal{H}}] \odot \mathbf{E}(K, f) \quad (26)$$

where $\mathbf{q}(f)$ is the VFT of $q(t)$ for bandwidth-rate pair $(B, 1/T)$, $\odot$ denotes the entry-by-entry Hadamard product, and $\mathbf{E}(K, f)$ is the masking matrix defined as

$$\mathbf{E}(K, f) \triangleq \sum_{k=0}^{K-1} \begin{bmatrix} e^{-\mathrm{j}\frac{2\pi k}{K}\cdot 1} \\ e^{-\mathrm{j}\frac{2\pi k}{K}\cdot 2} \\ \cdots \\ e^{-\mathrm{j}\frac{2\pi k}{K}\cdot \mathcal{N}(f)} \end{bmatrix} \begin{bmatrix} e^{-\mathrm{j}\frac{2\pi k}{K}\cdot 1} \\ e^{-\mathrm{j}\frac{2\pi k}{K}\cdot 2} \\ \cdots \\ e^{-\mathrm{j}\frac{2\pi k}{K}\cdot \mathcal{N}(f)} \end{bmatrix}^{\mathcal{H}} \quad (27)$$

which is the sum of K rank-one matrices. It can be shown that [4, Eqs. (16) and (26)]

$$[\mathbf{E}(K, f)]_{k,l} = \begin{cases} K, & \text{for mod}(k-l, K) = 0 \\ 0, & \text{elsewhere,} \end{cases} \quad (28)$$

with mod$(\cdot\,,\cdot)$ denoting modular operation, so that the matrix is called the masking matrix. Note that, by the masking matrix $\mathbf{E}(K, f)$, a rank one matrix becomes a rank $\min(K, \mathcal{N}(f))$ matrix when the reference rate is reduced from $1/T_0$ to $1/T = 1/(KT_0)$.

This rank multiplication can be explained in terms of the multi-stream decomposition [23]. If $N_I(t)$ is a linearly modulated signal with waveform $q(t)$ and uncorrelated symbols with symbol rate T_0, then it can also be viewed as K linearly modulated signals with waveforms $q(t), q(t-T_0), ...,$ and $q(t-(K-1)T_0)$, respectively, and uncorrelated symbols with a common symbol rate $1/T = 1/(KT_0)$. This multi-stream decomposition of a high-rate data stream into multiple low-rate streams enable us to use (10) to show (25) becomes (26) when the reference rate is reduced from $1/T_0$ to $1/T = 1/(KT_0)$.

The following theorems provide the solution to *Problem 2* and the necessary and sufficient condition for the existence of the solution. To proceed, we define the projection matrix $\mathbf{P}(f)$ as

$$\begin{aligned} \mathbf{P}(f) &\triangleq \mathbf{I}_{\mathcal{N}(f)} - \mathbf{G}(f)\big(\mathbf{G}(f)^{\mathcal{H}}\mathbf{G}(f)\big)^{\dagger}\mathbf{G}(f)^{\mathcal{H}}, \\ &= \mathbf{I}_{\mathcal{N}(f)} - \mathbf{G}(f)\mathbf{G}(f)^{\dagger}, \end{aligned} \quad (29)$$

where $\mathbf{I}_{\mathcal{N}(f)}$ denotes the $\mathcal{N}(f)$-by-$\mathcal{N}(f)$ identity matrix and † denotes Moore-Penrose generalized inverse. We also define the support of $M(fT)$ as

$$\mathcal{F}_M \triangleq \{f \in \mathcal{F} : M(fT) \neq 0\}, \quad (30a)$$

and the support of $\lambda_{\max}(f)$ as

$$\mathcal{F}_{\lambda_{\max}} \triangleq \{f \in \mathcal{F} : \lambda_{\max}(f) \neq 0\} \quad (30b)$$

where $\lambda_{\max}(f)$ is the largest eigenvalue of $\mathbf{P}(f)\mathbf{C}(f)$ with $\mathbf{C}(f)$ defined as

$$\mathbf{C}(f) \triangleq \mathbf{H}(f)^{\mathcal{H}}\mathbf{R}_N(f)^{-1}\mathbf{H}(f). \quad (31)$$

Theorem 1: [19, *Theorem 1*] The VFTs $\mathbf{s}_{\text{opt}}(f)$ and $\mathbf{w}_{\text{opt}}(f)$ of the jointly optimal transmit waveform $s_{\text{opt}}(t)$ and the receive waveform $w_{\text{opt}}(t)$ as the solutions to *Problem 2* are given, respectively, by

$$\mathbf{s}_{\text{opt}}(f) = \begin{cases} \sqrt{a_{\text{opt}}(f)}\mathbf{v}_{\max}(f)e^{\mathrm{j}\theta(f)}, & \forall f \in \mathcal{F}_M, \\ \text{arbitrary}, & \forall f \in (\mathcal{F}_M)^c \end{cases} \quad (32)$$

and

$$\mathbf{w}_{\text{opt}}(f) = \frac{M(fT)\mathbf{R}_N(f)^{-1}\mathbf{H}(f)\mathbf{s}_{\text{opt}}(f)}{1 + \frac{M(fT)}{T}\lambda_{\max}(f)a_{\text{opt}}(f)}, \forall f \in \mathcal{F}_M, \quad (33)$$

where $\mathbf{v}_{\max}(f)$ is the normalized eigenvector corresponding to the largest eigenvalue $\lambda_{\max}(f)$ of $\mathbf{P}(f)\mathbf{C}(f)$, and $\theta(f)$ is arbitrary. The optimal energy allotment $a_{\text{opt}}(f)$ is given by

$$a_{\text{opt}}(f) = \left[\nu_{\text{opt}} - \sqrt{\frac{T}{M(fT)\lambda_{\max}(f)}}\right]^{+}\sqrt{\frac{T}{M(fT)\lambda_{\max}(f)}}, \quad (34)$$

for $f \in \mathcal{F}_M \cap \mathcal{F}_{\lambda_{\max}}$ with $\nu_{\text{opt}} > 0$ being the unique solution to

$$\int_{\mathcal{F}_M \cap \mathcal{F}_{\lambda_{\max}}} \left[\nu_{\text{opt}} - \sqrt{\frac{T}{M(fT)\lambda_{\max}(f)}}\right]^{+}\sqrt{\frac{M(fT)}{T\lambda_{\max}(f)}}df = P, \quad (35)$$

and $a_{\text{opt}}(f) = 0$ for $f \in \mathcal{F}_M \cap (\mathcal{F}_{\lambda_{\max}})^c$. This solution leads to the minimum MSE (MMSE) given by

$$\text{MMSE} = \int_{\mathcal{F}_M} \frac{TM(fT)}{1 + \frac{M(fT)}{T}\lambda_{\max}(f)a_{\text{opt}}(f)}df. \quad (36)$$

Proof: See the proof of [19, *Theorem 1*]. ■

Theorem 2: [19, *Theorem 2*] A non-trivial optimal solution exists iff the length of the set

$$\Big\{f \in \mathcal{F}_M : \text{Dim}\big(\text{null space of } \mathbf{G}(f)^{\mathcal{H}}\big) > 0\Big\} \quad (37)$$

is greater than zero, where Dim$(\cdot)$ denotes the dimension of a subspace, where a solution is trivial if it does not reduce the MSE at all, i.e., $\mathbb{E}\{|b[l] - z[l]|^2\} = \mathbb{E}\{|b[l]|^2\} = \int_{\mathcal{F}_M} TM(fT)df$.

Proof: See the proof of [19, *Theorem 2*]. ■

In the next section, using the definition and the results summarized in this section, we formulate the optimization problem to find the capacity of the WSCS Gaussian noise channel and derive the solution.

III. Capacity of WSCS Gaussian Noise Channel

In this section, the capacity of the WSCS Gaussian noise channel is derived. The following proposition provides the optimization problem to be solved to find the capacity.

Proposition 1: The capacity of the WSCS Gaussian noise channel is the optimum value of the problem given by

Problem 3:

$$\underset{(\mathbf{R}_X(f))_f}{\text{maximize}} \int_{\mathcal{F}} \log_2 \det\left\{\mathbf{I}_{\mathcal{N}(f)} + \tilde{\mathbf{H}}(f)\mathbf{R}_X(f)\tilde{\mathbf{H}}(f)^{\mathcal{H}}\right\} df \tag{38a}$$

$$\text{subject to} \int_{\mathcal{F}} \text{tr}\{\mathbf{R}_X(f)\} df = P, \tag{38b}$$

where $\mathbf{R}_X(f)$ is the matrix-valued PSD of a proper-complex zero-mean Gaussian random process $X(t)$, $\tilde{\mathbf{H}}(f) \triangleq \mathbf{R}_N(f)^{-\frac{1}{2}}\mathbf{H}(f)$, and all the VFTs and the MV-PSD have the reference rate $1/T = 1/T_0$.

Proof: Omitted. ∎

To accommodate the cases with orthogonality constraints such as the transmit band and the zero interference constraints considered in Section II-B, we modify the problem as follows.

Proposition 2: The orthogonality constraints can be written as

$$\mathbf{g}_l(f)^{\mathcal{H}}\mathbf{R}_X(f)\mathbf{g}_l(f) = 0, \forall l \in \mathcal{L}, \forall f \in \mathcal{F}, \tag{39}$$

for some VFTs $\mathbf{g}_l(f)$ and some index set $\mathcal{L}$. Then, the optimization problem to find the capacity under the additional constraints (39) is given the same as *Problem 3* except that $\tilde{\mathbf{H}}(f)$ is now re-defined as

$$\tilde{\mathbf{H}}(f) \triangleq \mathbf{R}_N(f)^{-\frac{1}{2}}\mathbf{H}(f)\mathbf{P}(f) \tag{40}$$

where $\mathbf{P}(f)$ is the projection matrix defined as (29) with $\mathbf{G}(f)$ similarly defined as (22).

Proof: Omitted. ∎

Since the integrand of the objective function in *Problem 3* resembles the objective function of the well-known MIMO channel capacity problem [24], the solution can be readily obtained except that the optimal power distribution needs to be found over all the orthogonal channels and all the offsets $f \in \mathcal{F}$.

Theorem 3: The capacity of the WSCS Gaussian noise channel as the optimum value to *Problem 3* is given by

$$C_{\text{WSCS}} = \int_{\mathcal{F}} \sum_{n=1}^{\mathcal{N}(f)} \log_2\left(1 + [\gamma_n(f)^2\nu_{\text{opt}} - 1]^+\right) df, \tag{41a}$$

where ν_{opt} is the unique solution to

$$\int_{\mathcal{F}} \sum_{n=1}^{\mathcal{N}(f)} \left[\nu_{\text{opt}} - \frac{1}{\gamma_n(f)^2}\right]^+ df = P \tag{41b}$$

with $\gamma_n(f)$ defined as the nth singular value of the matrix $\tilde{\mathbf{H}}(f)$.

Proof: Omitted. ∎

Let $X_{\text{opt}}(t)$ be the transmitted signal that achieves the capacity of the channel. Then, the following theorem states that the proper-complex zero-mean random process is cyclostationary with the same cycle period as the interference $N(t)$.

Theorem 4: $X_{\text{opt}}(t)$ is WSCS with cycle period T.

Proof: A detailed proof is omitted. Instead, we provide an outline: It can be shown that

$$r_X(t,s) \triangleq \int_{\mathcal{F}} \mathbf{a}_f(t)^{\mathcal{H}}\mathbf{R}_X(f)\mathbf{a}_f(s) df \tag{42a}$$

where

$$\mathbf{a}_f(t) \triangleq \begin{bmatrix} e^{-\text{j}2\pi\left(\frac{1}{T}+f\right)t} \\ e^{-\text{j}2\pi\left(\frac{2}{T}+f\right)t} \\ \vdots \\ e^{-\text{j}2\pi\left(\frac{\mathcal{N}(f)}{T}+f\right)t} \end{bmatrix}. \tag{43}$$

Thus, the integrand in (42a) satisfies

$$\mathbf{a}_f(t)^{\mathcal{H}}\mathbf{R}_X(f)\mathbf{a}_f(s) \tag{44a}$$

$$= \sum_{k,l} ([\mathbf{a}_f(t)]_k)^* [\mathbf{R}_X(f)]_{k,l} [\mathbf{a}_f(s)]_l \tag{44b}$$

$$= \sum_{k,l} [\mathbf{R}_X(f)]_{k,l} e^{\text{j}2\pi\left\{\left(\frac{k}{T}+f\right)t - \left(\frac{l}{T}+f\right)s\right\}} \tag{44c}$$

$$= \sum_{k,l} [\mathbf{R}_X(f)]_{k,l} e^{\text{j}2\pi\left\{\left(\frac{k}{T}+f\right)(t+T) - \left(\frac{l}{T}+f\right)(s+T)\right\}} \tag{44d}$$

$$= \mathbf{a}_f(t+T)^{\mathcal{H}}\mathbf{R}_X(f)\mathbf{a}_f(s+T), \forall t, s, \tag{44e}$$

which implies $r_X(t,s) = r_X(t+T, s+T), \forall t, s$. Therefore, the conclusion follows. ∎

Since the resultant optimal transmit signal $X_{\text{opt}}(t)$ is cyclostationary, we call the water-filling described in *Theorem 3* the cyclic water-filling.

IV. Discussions and Numerical Results

In this section, we provide discussions and numerical results related to the previous section. For simplicity, all channels are assumed flat.

A. Application to Selfish Overlay

The first application of the capacity result is to design an overlay system that maximizes its own rate regardless of the interference it causes to the legacy system. In particular, we consider the cases where the legacy system employs linear modulation with a square-root Nyquist pulse of roll-off factor $0 < \beta \leq 1$, symbol rate $1/T_0 = W/(1+\beta)$, symbols from a Gaussian codebook, and transmission power P_1. The overlay system does not necessarily employ linear modulation, but has an average power of P_2.

Fig. 6 compares the spectral efficiency of the WSCS Gaussian noise channel obtained by using the cyclic water-filling (WF) and the ordinary WF. The interference-to-signal power ratio (ISR) P_1/P_2 is 0 dB and the excess bandwidth of the legacy signal is $\beta = 0, 0.25, 0.5$, and 1. The spectral efficiency is shown as a function of the signal-to-noise ratio $P_2/(N_0W)$. As shown in the figure, the cyclic WF significantly outperforms the ordinary WF.

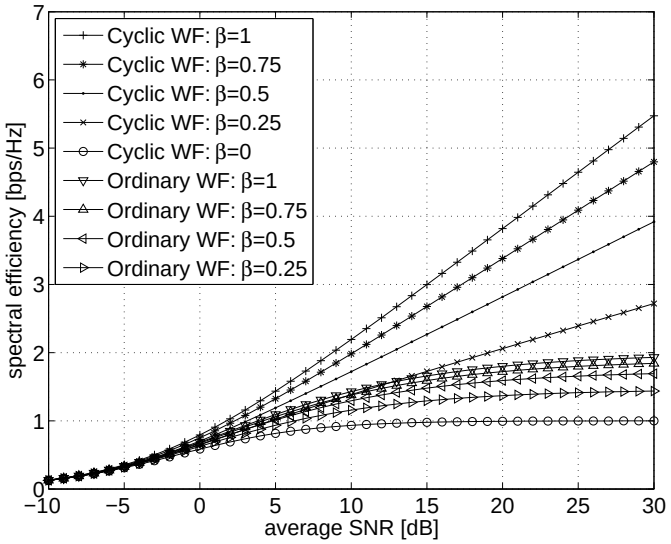

Fig. 6. Comparison of cyclic water-filling and ordinary water-filling.

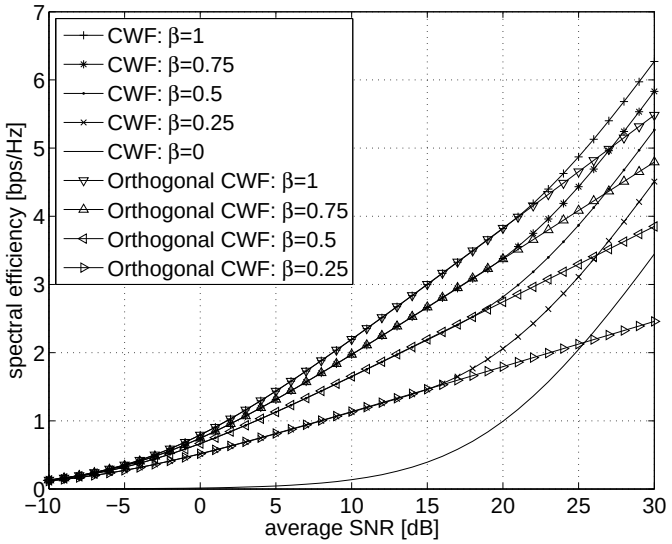

Fig. 8. Comparison of non-orthogonal and orthogonal CWF for INR 20 dB.

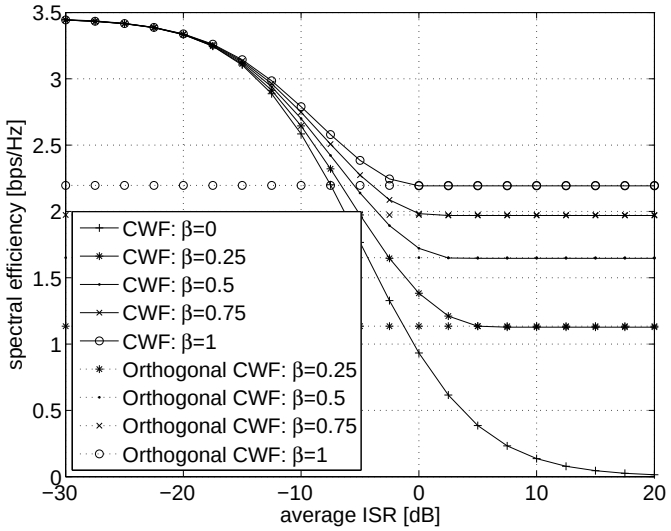

Fig. 7. Spectral efficiency of WSCS Gaussian noise channel as a function of ISR for SNR $P_2/(N_0W) = 10$ dB.

B. Application to Selfless Overlay

The second application of the capacity result is to design an overlay system that maximizes its own rate without any performance degradation of the legacy system. The same signal model is adopted for the legacy system. The overlay system is assumed to have the knowledge on the symbol timing of the legacy receiver. This assumption is viable, for example, when an orthogonal overlay system is designed for a bent-pipe satellite broadcasting system. In such a system, the multiple legacy and overlay receivers can be regarded as collocated receivers, and the terrestrial station that transmits the overlay signal may monitor the downlink signal to acquire an accurate channel estimation.

Fig. 7 shows the spectral efficiency of the WSCS Gaussian noise channel as a function of the ISR, when the cyclic and the orthogonal cyclic WFs are used. The SNR is 10 dB and the excess bandwidth of the legacy signal is $\beta = 0.25, 0.5, 0.75,$ and 1. An interesting observation can be made from Fig. 7 that, if $\beta > 0$, there is an ISR value over which the spectral efficiency of the cyclic WF (CWF) scheme no longer decreases and equals to that of the orthogonal CWF, even though the interference power increases. This phenomenon occurs because even the selfish overlay signal utilizes only the subspace that is orthogonal to that of the legacy signal when the ISR is greater than or equal to a certain level. Thus, it observes an effectively interference-free frequency band of a reduced bandwidth. The same phenomenon is observed in Fig. 8, where non-orthogonal and orthogonal CWF schemes are compared, where the interference-to-noise power ratio (INR) $P_1/(N_0W) = 20$ dB.

C. Application to Orthogonal Multiple-Access Communications

The next application is to orthogonal multiple-access communications. When the first user employs linear modulation with a square-root Nyquist pulse with excess bandwidth $\beta \in [0, \infty)$, it can be shown that the second user orthogonal to the first user has the effective bandwidth $W\beta/(1+\beta)$ out of W. Since the throughput of the first user is given by

$$C_1 = \frac{W}{1+\beta} \log_2 \left(1 + \frac{P_1(1+\beta)}{N_0 W}\right) \tag{45}$$

which is the capacity of an FDMA user that occupies $W/(1+\beta)$ of W and that of the overlay user is given by

$$C_2 = \frac{W\beta}{1+\beta} \log_2 \left(1 + \frac{P_2}{N_0 \frac{W\beta}{1+\beta}}\right), \tag{46}$$

regardless of the choice of the reference rate $1/T = 1/(KT_0)$, the first and the second users act like optimal FDMA users [2].

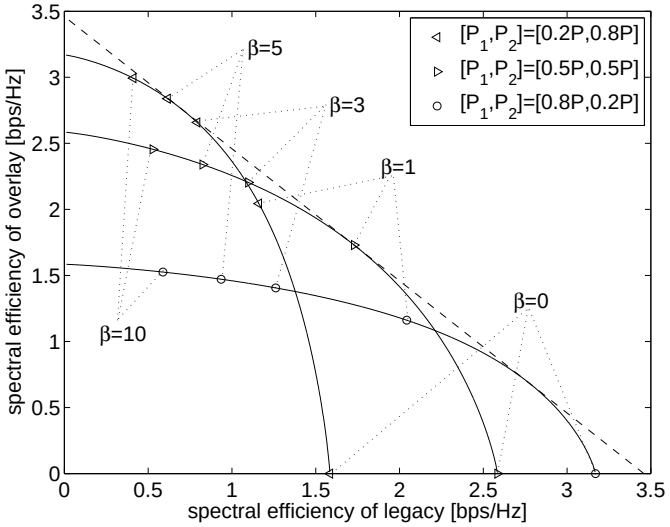

Fig. 9. Achievable rate pair of legacy and orthogonal overlay for roll-off factor $\beta \in [0, \infty)$. The total SNR is 10 dB.

Fig. 9 shows the normalized rate pairs of the users obtained by using the orthogonal CWF. Note that all the boundary points of the optimal FDMA rate region are achievable. Extension to multiple-user cases is straightforward.

V. Conclusions

In this paper, we have formulated an optimization problem to find the capacity of a proper-complex zero-mean WSCS Gaussian noise channel and derived the solution. It is shown that the optimal transmit signal is also a proper-complex zero-mean WSCS Gaussian random process with the same cycle frequency as the noise. Thus, the water-filling procedure to find the optimal solution is named the cycle water-filling. The solution accommodates orthogonality constraints that may be imposed to limit the transmit band or to make the transmitted signal induce no interference at linear receivers operating in the frequency band of interest. It is shown that the cyclic water-filling significantly outperforms the ordinary water-filling that ignores the spectral correlation of the WSCS noise. Applications to the design of selfish and selfless overlay systems and to the orthogonal multiple-access communications are considered. It is shown that the FDMA rate-pairs are achievable by the orthogonal cyclic water-filling.

References

[1] R. G. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[3] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw Hill, 2001.

[4] J. H. Cho, Y. K. Jeong, and J. S. Lehnert, "Linear suppression of wideband data-like interference in DS/SS communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 5, pp. 973–985, May 2005.

[5] J. H. Cho and J. S. Lehnert, "Blind adaptive multiuser detection for DS/SSMA communications with generalized random spreading," *IEEE Trans. Commun.*, vol. 49, no. 6, pp. 1082–1091, June 2001.

[6] ETSI SMG2, *ETSI UMTS Terrestrial Radio Access (UTRA) ITU-R RTT Candidate Submission*, May/June 1998.

[7] Q. Zhang, B. W. Han, J. H. Cho, and S. Wei, "PAPR performance of IDFT-based uncoded OFDM signals with null subcarriers and transmit filtering," in *Proc. IEEE International Conf. on Commun. 2006*, Istanbul, Turkey, 11-15 June 2006.

[8] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[9] P. Z. Peebles Jr., *Probability, Random Variables, and Random Signal Principles*, 4th ed. New York: McGraw-Hill, 2000.

[10] W. A. Gardner, Ed.,*Cyclostationarity in Communications and Signal Processing*. New York: IEEE Press, 1994.

[11] G. Golden, J. Mazo, and J. Salz, "Transmitter design for data transmission in the presence of a data-like interferer," AT&T Bell Laboratories Released Paper No. BL92-00517, May 7, 1992.

[12] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[13] G. Golden, J. Mazo, and J. Salz, "Transmitter design for data transmission in the presence of a data-like interferer," *IEEE Trans. Commun.*, vol. 43, no. 2-4, pp. 837–850 , Feb.-Apr. 1995.

[14] P. S. Kumar and S. Roy, "Optimization for crosstalk suppression with noncoordinating users," *IEEE Trans. Commun.*, vol. 44, no. 7, pp. 894–905, Jul. 1996.

[15] H. S. Mir and S. Roy, "Optimum transmitter/receiver design for a narrowband overlay in noncoordinated subscriber lines," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 992–998, June 2004.

[16] J. H. Cho, "Joint transmitter and receiver optimization in additive cyclostationary noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 14, pp. 3396–3405, Dec. 2004.

[17] J. H. Cho and W. Gao, "Continuous-time equivalents of Welch bound equality sequences," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3176-3185, Sep. 2005.

[18] J. H. Cho, "Multi-user constrained water-pouring for continuous-time overloaded Gaussian multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1437–1459, Apr. 2008.

[19] Y. H. Yun and J. H. Cho, "An optimal orthogonal overlay in cyclostationary legacy signal," *IEEE Trans. Comm.*, under review.

[20] J. H. Cho and J. S. Lehnert, "System design for DS/SSMA communications with negative chip excess bandwidth," in *Proc. IEEE Military Commun. Conf.*, vol. 2, McLean, VA, Oct. 2001, pp. 1050–1054.

[21] L. E. Franks, "Polyperiodic linear filtering," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. New York, NY: IEEE Press, 1994, pp. 240–266.

[22] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 4–14, Jan. 1975.

[23] J. H. Yeo, Y. G. Yoo, and J. H. Cho, "Power spectral density of block-transmitted signals for SC-FDE using cyclic prefix, zero padding, and unique word", in *Proc. Joint Conference on Communication and Information 2008 (JCCI'08)*, Cheju, Korea, 2008.

[24] A. Goldsmith, *Wireless Communications*. New York: Cambridge Univ. Press, 2005.