Scientific
Research

International Journal of

Communications, Network and System Sciences

www.scirp.org/journal/ijcns/

9 771913 371006   09

Scientific Research

TABLE OF CONTENTS

Volume 3   Number 9   **September 2010**

International Journal of Communications, Network and System Sciences (IJCNS)

Journal Information

SUBSCRIPTIONS

SERVICES

COPYRIGHT

PRODUCTION INFORMATION

Scientific
Research

A Virtual Channel-Based Approach to Compensation of I/Q Imbalances in MIMO-OFDM Systems

Shichuan Ma, Deborah D. Duran, Hamid Sharif, Yaoqing (Lamar) Yang
Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Omaha, USA
E-mail: *sma@huskers.unl.edu, dduran@mail.unomaha.edu, yyang3@unl.edu, hsharif@unl.edu*

Abstract

Multiple-input multiple-output (MIMO)-orthogonal frequency-division multiplexing (OFDM) scheme has been considered as the most promising physical-layer architecture for the future wireless systems to provide high-speed communications. However, the performance of the MIMO-OFDM system may be degraded by in-phase/quadrature-phase (I/Q) imbalances caused by component imperfections in the analog front-ends of the transceivers. I/Q imbalances result in inter-carrier interference (ICI) in OFDM systems and cause inaccurate estimate of the channel state information (CSI), which is essential for diversity combining at the MIMO receiver. In this paper, we propose a novel approach to analyzing a MIMO-OFDM wireless communication system with I/Q imbalances over multi-path fading channels. A virtual channel is proposed as the combination of multi-path fading channel effects and I/Q imbalances at the transmitter and receiver. Based on this new approach, the effects of the channel and I/Q imbalances can be jointly estimated, and the influence of channel estimation error due to I/Q imbalances can be greatly reduced. An optimal minimal mean square error (MMSE) estimator and a low-complexity least square (LS) estimator are employed to estimate the joint coefficients of the virtual channel, which are then used to equalize the distorted signals. System performance is theoretically analyzed and verified by simulation experiments under different system configurations. The results show that the proposed method can significantly improve system performance that is close to the ideal case in which I/Q are balanced and the channel state information is known at the receiver.

1. Introduction

Due to the rapid growth of broadband wireless applications, the next generation of wireless communication systems poses major challenges for efficient exploitation of the available spectral resources. Among the existing techniques, the combination of orthogonal frequency-division multiplexing (OFDM) and multiple-input multiple-output (MIMO) has been widely considered the most promising approach for building future wireless transmission systems [1]. OFDM, as a popular modulation scheme, converts a frequency-selective fading channel into a parallel collection of flat-fading sub-channels, enabling high data-rate transmissions [2-4]. On the other hand, by deploying multiple antennas at both ends of the transmitter (TX) and receiver (RX), MIMO architectures are capable of combating channel fading by taking advantage of the spatial diversity and/or efficient in enhancing system capacity by employing spatial multiplexing [5,6]. The MIMO-OFDM technique has attracted significant research interest in recent years [7-10] and has been standardized in a number of communication systems [11-13].

Although OFDM presents numerous advantages, it suffers from performance degradations due to the hardware component flaws in the analog front-ends of the transceivers, including phase noise [14], carrier frequency offset [15], and in-phase/quadrature-phase (I/Q) imbalance [16,17]. The imbalance between the in-phase and the quadrature-phase branches is a major factor in performance degradation. When the received radio-frequency (RF) signal is down-converted to baseband, the analog front-end imperfections cause imbalances between the I and Q branches, which introduces inter-carrier interference (ICI) and frequency-dependent distortion to the received data. This leads to a decrease in the operating signal-to-noise

ratios (SNRs) and low data rates. Although several methods were proposed to overcome ICI [18-20], I/Q imbalances cannot be handled using these methods. The I/Q imbalances become more severe when a direct-conversion receiver (DCR) [21] or lower intermediate frequency (IF) [22] is utilized. Furthermore, the effect of the I/Q imbalances on the system is mixed with the signal distortion from the fading channel, making it difficult to compensate. Consequently, estimation and equalization of the I/Q imbalances from the received data are critical in OFDM-based systems. Frequency-independent and frequency-dependent I/Q imbalance models are reported in [16] and [17], respectively. Based on these models, the effects of I/Q imbalances are studied in OFDM systems [23-28] and in MIMO-OFDM systems [29-33].

In [29], the input-output relation of a MIMO-OFDM system with frequency-independent I/Q imbalances at the receiver is derived. Based on this result, an adaptive method is introduced to compensate for the received data. In [31], the effect of frequency-dependent I/Q imbalances on MIMO-OFDM system is studied, using a pilot-based compensation scheme. In both studies, however, the received signals are diversity combined using inaccurate channel state information (CSI) estimated under I/Q imbalances, leading to system performance degradation. We have proposed a method to deal with this problem in [33]. By jointly estimating and compensating for multipath fading channels and I/Q imbalances, this method can effectively mitigate the I/Q effects.

In this paper, a novel approach is proposed to analyze MIMO-OFDM wireless communication systems with I/Q imbalances over multi-path fading channels. A *virtual channel* is proposed to bypass the channel estimation under I/Q influence. Based on this approach, the TX and RX I/Q imbalances are treated as parts of the fading channels. The effects of both the fading channels and I/Q imbalances on the system can be jointly estimated before diversity combining and are then employed for diversity combining and signal compensation. A minimum mean square error (MMSE) estimator and a least square (LS) estimator are used to estimate the joint coefficients of the virtual channel. A signal compensation approach based on a zero-forcing algorithm is also provided. The system performance is theoretically analyzed, and bit error rate (BER) is expressed in closed-form. Extensive simulation results verified that the I/Q imbalances at both transmitter and receiver sides can be effectively mitigated by using the virtual channel approach.

The rest of this paper is organized as follows. In the following section, a frequency-dependent I/Q imbalances model and an Alamouti scheme-based MIMO-OFDM model with transmitter and receiver I/Q imbalances are described. The input-output relation is derived in the frequency domain. In Section 3, the MMSE and LS estimators of the joint coefficients of the virtual channel are developed. In Section 4, the BER of the system is theoreti-

cally analyzed. Simulation results are given in Section 5. Finally, Section 6 concludes the paper.

Notations: A small (capital) letter represents a variable in time (frequency) domain, and a bold small (capital) letter represents a vector (matrix). The superscripts *, T, H represent the conjugate, the transpose, and the Hermitian transpose operations, respectively. $|\cdot|$ denotes absolute value operation, $\|\cdot\|$ denotes Frobenius norm, and $\varepsilon\{\cdot\}$ represents expectation. $Q(\cdot)$ represents the Q-function. $\otimes$ denotes convolution.

2. System Model

2.1. I/Q Imbalance Model

A frequency-dependent I/Q model at the receiver side is shown in **Figure 1**. As described in [17], the I/Q imbalances arise from two effects. One is the effect of the quadrature demodulator, which is frequency-independent and determined by I/Q amplitude imbalance g_{RX} and I/Q phase imbalance θ_{RX}; another is the effect of branch components, which is frequency-dependent and modeled as two filters with frequency response $L_{I,RX}(f)$ and $L_{Q,RX}(f)$. If the I/Q branches are perfect, then $g_{RX} = 1$, $\theta_{RX} = 0$, and $L_{I,RX}(f) = L_{Q,RX}(f) \equiv 1$.

Assuming $r_{BB}(t)$ is the baseband equivalent signal of the received signal $r_{RX}(t)$, the down-converted signal $z_{RX}(t)$ can be written as

$$z_{RX}(t) = g_{1,RX}(t) \otimes r_{BB}(t) + g_{2,RX}(t) \otimes r_{BB}^*(t) \quad (1)$$

where

$$
\begin{aligned}
g_{1,RX}(t) &= \left[l_{I,RX}(t) + l_{Q,RX}(t) g_{RX} e^{-j\theta_{RX}} \right] / 2 \\
g_{2,RX}(t) &= \left[l_{I,RX}(t) - l_{Q,RX}(t) g_{RX} e^{j\theta_{RX}} \right] / 2
\end{aligned} \quad (2)
$$

$l_{I,RX}(t)$ and $l_{Q,RX}(t)$ are the time-domain representations of $L_{I,RX}(f)$ and $L_{Q,RX}(f)$, respectively. From the point of view of the frequency domain, the expression in (1) can be written as

$$Z_{RX}(f) = G_{1,RX}(f) R_{BB}(f) + G_{2,RX}(f) R_{BB}^*(-f) \quad (3)$$
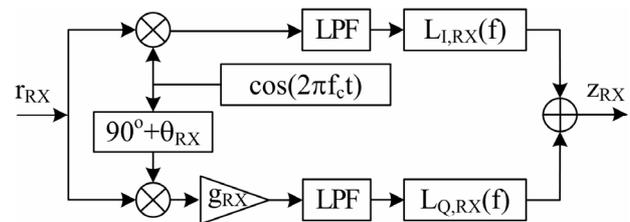
Figure 1. Block diagram of frequency-dependent I/Q imbalance at a receiver.

where $R_{BB}(f)$ is the Fourier transform of $r_{BB}(t)$, and

$$
\begin{aligned}
G_{1,RX}(f) &= \left[L_{I,RX}(f) + L_{Q,RX}(f)g_{RX}e^{-j\theta_{RX}} \right]/2 \\
G_{2,RX}(f) &= \left[L_{I,RX}(f) - L_{Q,RX}(f)g_{RX}e^{j\theta_{RX}} \right]/2
\end{aligned} \tag{4}
$$

In (3), the term $R_{BB}^{*}(-f)$ represents the inter-carrier interference (ICI) projected from the mirror frequency to the signal frequency. This is called image projection, a major problem caused by I/Q imbalances in signal demodulations.

Similarly, the relation between the signal R_{TX} and the transmitted baseband signal Z_{TX} with I/Q mismatch can be written as

$$
Z_{TX}(f) = G_{1,TX}(f)R_{TX}(f) + G_{2,TX}(f)R_{TX}^{*}(-f) \tag{5}
$$

where

$$
\begin{aligned}
G_{1,TX}(f) &= \left[L_{I,TX}(f) + L_{Q,TX}(f)g_{TX}e^{j\theta_{TX}} \right]/2 \\
G_{2,TX}(f) &= \left[L_{I,TX}(f) - L_{Q,TX}(f)g_{TX}e^{-j\theta_{TX}} \right]/2
\end{aligned} \tag{6}
$$

where g_{TX} denotes the I/Q amplitude imbalance at TX, θ_{TX} represents the I/Q phase imbalance at TX, and $L_{I,TX}(f)$ and $L_{Q,TX}(f)$ indicate the non-linear frequency characteristics of the I and Q branches at TX.

2.2. MIMO-OFDM Model with I/Q Imbalances

A block diagram of a MIMO-OFDM wireless communication system with Alamouti diversity scheme and frequency-dependent TX and RX I/Q imbalances is shown in **Figure 2**. In this system, there are two transmit antennas and N_r (≥ 1) receive antennas. All signals are represented in the form of space-time coded (STC) blocks in the frequency domain. For example, $S_{i,1} \mid S_{i,2}$ denotes an STC block data of $N \times 2$ matrix, where $i = 1, 2, \cdots$ is the STC block index, and N is the number of the used OFDM sub-carriers. Vector $\mathbf{s}_{i,j} = [S_{i,j}(-N/2)\cdots S_{i,j}(-1)\, S_{i,j}(1)\cdots S_{i,j}(N/2)]^{T}$ is an OFDM symbol to be transmitted over the system at the jth, $j \in \{1, 2\}$ time slot of the ith STC block, where $S_{i,j}(k)$ denotes the symbol at the kth, $k \in \{-N/2, \cdots, -1, 1, \cdots, N/2\}$ sub-carrier with average symbol energy $E_s/2$. Vector $S_{i,1}$ is transmitted followed by vector $S_{i,2}$. For simplicity, the block index i is omitted in **Figure 2**. $H_{n,m}$ denotes the channel frequency response between the nth transmit antenna and the mth receive antenna, where $n \in \{1, 2\}$ and $m \in \{1, 2, \cdots, N_r\}$. The TX I/Q imbalance parameters are denoted by $G_{1,TX}^{(n)}$ and $G_{2,TX}^{(n)}$, and the RX I/Q imbalance parameters are denoted by $G_{1,RX}^{(m)}$ and $G_{2,RX}^{(m)}$, where again, $n \in \{1, 2\}$ and $m \in \{1, 2, \cdots, N_r\}$.

Assume two consecutive data symbols, $S_{i,1}(k)$ and $S_{i,2}(k)$, to be transmitted over the kth sub-carrier. Based
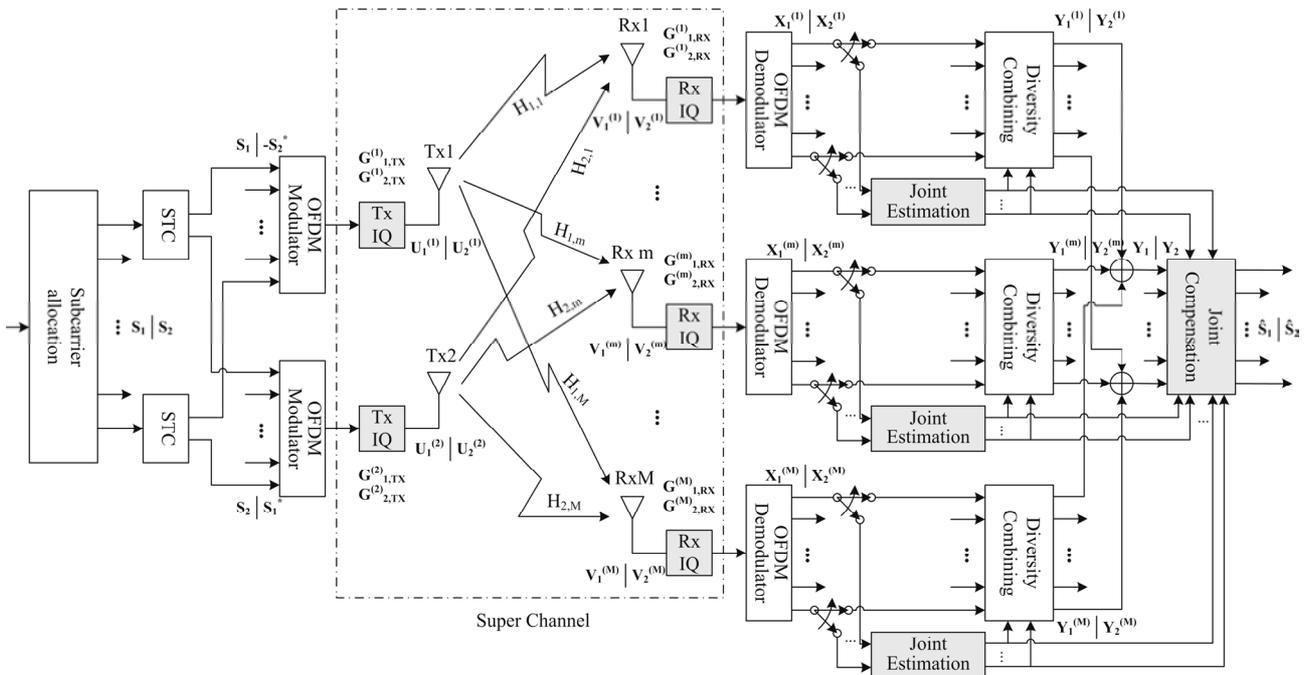
Figure 2. Block diagram of a $2 \times N_r$ MIMO-OFDM wireless communication system with transmitter and receiver I/Q imbalances. The virtual channel is illustrated in the dashed block.

on the Alamouti scheme, $S_{i,1}(k)$ and $-S_{i,2}^*(k)$ are distributed into the kth sub-carrier data stream of the first OFDM modulator, while $S_{i,2}(k)$ and $S_{i,1}^*(k)$ are distributed into the kth sub-carrier data stream of the second OFDM modulator. Data streams at all sub-carriers are then processed through OFDM modulation, including operations of padding zeros, inverse fast Fourier transform (IFFT), and adding cyclic prefix (CP). Because the IFFT operation only transforms the signal from frequency domain to time domain, the signals (viewed in the frequency domain) are not changed after OFDM modulation. Therefore, after OFDM modulation, the two consecutive data symbols at the kth sub-carrier of the first transmitter in the frequency domain are still $S_{i,1}(k)$ and $-S_{i,2}^*(k)$.

The OFDM-modulated signals are then distorted by TX I/Q imbalances. According to (5), the signals to be transmitted via the first transmit antenna are given by

$$U_{i,1}^{(1)}(k) = G_{1,TX}^{(1)}(k)S_{i,1}(k) + G_{2,TX}^{(1)}(k)S_{i,1}^*(-k)$$
$$U_{i,2}^{(1)}(k) = -G_{1,TX}^{(1)}(k)S_{i,2}^*(k) - G_{2,TX}^{(1)}(k)S_{i,2}(-k) \tag{7}$$

Similarly, the signals to be transmitted via the second transmit antenna are

$$U_{i,1}^{(2)}(k) = G_{1,TX}^{(2)}(k)S_{i,2}(k) + G_{2,TX}^{(2)}(k)S_{i,2}^*(-k)$$
$$U_{i,2}^{(2)}(k) = G_{1,TX}^{(2)}(k)S_{i,1}^*(k) + G_{2,TX}^{(2)}(k)S_{i,1}(-k) \tag{8}$$

The signals are then transmitted over a multi-path fading channel. The received signals at the mth receive antenna are given as

$$V_{i,1}^{(m)}(k) = H_{1,m}(k)U_{i,1}^{(1)}(k) + H_{2,m}(k)U_{i,1}^{(2)}(k)$$
$$V_{i,2}^{(m)}(k) = H_{1,m}(k)U_{i,2}^{(1)}(k) + H_{2,m}(k)U_{i,2}^{(2)}(k) \tag{9}$$

The received signals at the mth receive antenna are further corrupted by I/Q imbalances in the mth receiver as follows:

$$W_{i,1}^{(m)}(k) = G_{1,RX}^{(m)}(k)V_{i,1}^{(m)}(k) + G_{2,RX}^{(m)}(k)V_{i,1}^{*(m)}(-k)$$
$$W_{i,2}^{(m)}(k) = G_{1,RX}^{(m)}(k)V_{i,2}^{(m)}(k) + G_{2,RX}^{(m)}(k)V_{i,2}^{*(m)}(-k) \tag{10}$$

Assuming the noise in the system is additive white Gaussian noise (AWGN), then

$$X_{i,1}^{(m)}(k) = W_{i,1}^{(m)}(k) + N_{i,1}^{(m)}(k)$$
$$X_{i,2}^{(m)}(k) = W_{i,2}^{(m)}(k) + N_{i,2}^{(m)}(k) \tag{11}$$

where $N_{i,j}^{(m)}(k)$ is independently identically distributed (i.i.d.) complex zero-mean Gaussian noise with variance N_0.

Combining (7), (8), (9), (10) and (11), the received data symbols at the mth receive antenna before diversity combining can be written as

$$X_{i,1}^{(m)}(k) = A^{(m)}(k)S_{i,1}(k) + B^{(m)}(k)S_{i,1}^*(-k) +$$
$$C^{(m)}(k)S_{i,2}(k) + D^{(m)}(k)S_{i,2}^*(-k) + N_{i,1}^{(m)}(k)$$
$$X_{i,2}^{(m)}(k) = C^{(m)}(k)S_{i,1}^*(k) + D^{(m)}(k)S_{i,1}(-k) -$$
$$A^{(m)}(k)S_{i,2}^*(k) - B^{(m)}(k)S_{i,2}(-k) + N_{i,2}^{(m)}(k)$$
$$\tag{12}$$

where $A^{(m)}(k)$, $B^{(m)}(k)$, $C^{(m)}(k)$, and $D^{(m)}(k)$ are defined as

$$\begin{aligned} A^{(m)}(k) \quad & G_{1,RX}^{(m)}(k)G_{1,TX}^{(1)}(k)H_{1,m}(k) + \\ & G_{2,RX}^{(m)}(k)G_{2,TX}^{*(1)}(-k)H_{1,m}^*(-k) \\ B^{(m)}(k) \quad & G_{1,RX}^{(m)}(k)G_{2,TX}^{(1)}(k)H_{1,m}(k) + \\ & G_{2,RX}^{(m)}(k)G_{1,TX}^{*(1)}(-k)H_{1,m}^*(-k) \\ C^{(m)}(k) \quad & G_{1,RX}^{(m)}(k)G_{1,TX}^{(2)}(k)H_{2,m}(k) + \\ & G_{2,RX}^{(m)}(k)G_{2,TX}^{*(2)}(-k)H_{2,m}^*(-k) \\ D^{(m)}(k) \quad & G_{1,RX}^{(m)}(k)G_{2,TX}^{(2)}(k)H_{2,m}(k) + \\ & G_{2,RX}^{(m)}(k)G_{1,TX}^{*(2)}(-k)H_{2,m}^*(-k) \end{aligned} \tag{13}$$

From (12) and (13), it is observed that the channel frequency response ($H_{n,m}$) and the I/Q effects ($G_{1,TX}^{(n)}$, $G_{2,TX}^{(n)}$, $G_{1,RX}^{(m)}$ and $G_{2,RX}^{(m)}$) are mixed together and produce the coefficients $A^{(m)}(k)$, $B^{(m)}(k)$, $C^{(m)}(k)$, and $D^{(m)}(k)$. If we treat the I/Q effects as parts of the fading channels, we can model the dashed block in **Figure 2** as a virtual channel with joint coefficients $A^{(m)}(k)$, $B^{(m)}(k)$, $C^{(m)}(k)$, and $D^{(m)}(k)$. Moreover, $A^{(m)}(k)$ and $C^{(m)}(k)$ are critical for data decoding, while the presence of $B^{(m)}(k)$ and $D^{(m)}(k)$ may introduce ICI. If the I/Q branches are perfect, then $A^{(m)}(k) = H_{1,m}(k)$, $B^{(m)}(k) = 0$, $C^{(m)}(k) = H_{2,m}(k)$, and $D^{(m)}(k) = 0$. Furthermore, the received data symbols at the kth sub-carrier are distorted by the mixture effects of channel and I/Q imbalances, and interference is introduced by other data symbols within the STC block at the kth and the mirrored -kth sub-carriers.

To compensate for the received signals, the joint coefficients should be estimated. The estimation methods are described in the next section. Now, assuming the accurate estimate of the joint coefficients are obtained, the received data can be combined to achieve the transmit diversity as

$$Y_{i,1}^{(m)}(k) = A^{*(m)}(k)X_{i,1}^{(m)}(k) + C^{(m)}(k)X_{i,2}^{*(m)}(k)$$
$$Y_{i,2}^{(m)}(k) = C^{*(m)}(k)X_{i,1}^{(m)}(k) - A^{(m)}(k)X_{i,2}^{*(m)}(k) \tag{14}$$

It should be noted that the combining method is slightly different from the Alamouti scheme, where channel state information is employed. In our scheme, the joint coefficients $A^{(m)}(k)$ and $C^{(m)}(k)$ are used. If the I/Q branches are perfect, (14) is reduced to the standard Alamouti diversity combining scheme as substituting (12) and (13) into (14), we obtain

3. Estimators and Signal Compensation

In this section, we describe two training sequence-based estimators to show how to estimate the joint coefficients of the virtual channel. The first is a minimal mean square error (MMSE) estimator. Although MMSE estimator is an optimal linear detector, it requires some priori knowledge of the estimated variables and is computationally intensive. An alternative is a least square (LS) estimator, which may significantly reduce the computational complexity at the expense of negligible BER degradation. The estimated joint coefficients can be used to perform diversity combining as in (14) and to compensate for the received signals as described at the end of this section.

3.1. MMSE Estimator

Assume total N_{tr} blocks of training sequences are used in this system (That is $2N_{tr}$ OFDM symbols). Let $P_{i,j}(k)$, $j \in \{1,2\}$ denote the jth training symbol in the ith block at the kth sub-carrier. According to (12), the input-output relation for one block can be written as

$$\mathbf{u}_i^{(m)}(k) = \mathbf{P}_i(k)\mathbf{v}^{(m)}(k) + \mathbf{n}_i^{(m)}(k) \qquad (22)$$

where

$$\mathbf{u}_i^{(m)}(k) = \left[X_{i,1}^{(m)}(k) \ X_{i,2}^{(m)}(k) \right]^T \qquad (23)$$

$$\mathbf{P}_i(k) = \begin{bmatrix} P_{i,1}(k) & P_{i,1}^*(-k) & P_{i,2}(k) & P_{i,2}^*(-k) \\ -P_{i,2}^*(k) & -P_{i,2}(-k) & P_{i,1}^*(k) & P_{i,1}(-k) \end{bmatrix} \qquad (24)$$

$$\mathbf{v}_i^{(m)}(k) = \left[A^{(m)}(k) \ B^{(m)}(k) \ C^{(m)}(k) \ D^{(m)}(k) \right]^T \qquad (25)$$

$$\mathbf{n}_i^{(m)}(k) = \left[N_{i,1}^{(m)}(k) \ N_{i,2}^{(m)}(k) \right]^T \qquad (26)$$

For the total T blocks, the input-output relation is written as

$$\mathbf{u}^{(m)}(k) = \mathbf{P}(k)\mathbf{v}^{(m)}(k) + \mathbf{n}^{(m)}(k) \qquad (27)$$

where

$$\mathbf{u}^{(m)}(k) = \left[\mathbf{u}_1^{(m)}(k)^T \ \mathbf{u}_2^{(m)}(k)^T \cdots \mathbf{u}_{N_{tr}}^{(m)}(k)^T \right]^T \qquad (28)$$

$$\mathbf{P}(k) = \left[\mathbf{P}_1(k)^T \ \mathbf{P}_2(k)^T \cdots \mathbf{P}_{N_{tr}}(k)^T \right] \qquad (29)$$

$$\mathbf{n}^{(m)}(k) = \left[\mathbf{n}_1^{(m)}(k)^T \ \mathbf{n}_2^{(m)}(k)^T \cdots \mathbf{n}_{N_{tr}}^{(m)}(k)^T \right]^T \qquad (30)$$

The MMSE estimate of $\mathbf{v}$ is given as [34]

$$\hat{\mathbf{v}}^{(m)}(k) = \mathbf{R}_{vu}\mathbf{R}_u^{-1}\mathbf{u}^{(m)}(k) \qquad (31)$$

where

$$\mathbf{R}_{vu} = E\{\mathbf{v}^{(m)}(k)\mathbf{u}^{(m)}(k)^H\} = \mathbf{R}_v\mathbf{P}^H(k) \qquad (32)$$

$$\mathbf{R}_u = E\{\mathbf{u}^{(m)}(k)\mathbf{u}^{(m)}(k)^H\} = \mathbf{P}(k)\mathbf{R}_v\mathbf{P}^H(k) + \mathbf{R}_n \qquad (33)$$

$$\mathbf{R}_n = E\{\mathbf{n}^{(m)}(k)\mathbf{n}^{(m)}(k)^H\} = N_0\mathbf{I}_{2T} \qquad (34)$$

$$\mathbf{R}_v = E\{\mathbf{v}^{(m)}(k)\mathbf{v}^{(m)}(k)^H\} \qquad (35)$$

$$Y_{i,1}^{(m)}(k) = H_{1,m}^*(k)X_{i,1}^{(m)}(k) + H_{2,m}(k)X_{i,2}^{*(m)}(k)$$
$$Y_{i,2}^{(m)}(k) = H_{2,m}^*(k)X_{i,1}^{(m)}(k) - H_{1,m}(k)X_{i,2}^{*(m)}(k) \qquad (15)$$

$$Y_{i,1}^{(m)}(k) = Q_1^{(m)}(k)S_{i,1}(k) + Q_2^{(m)}(k)S_{i,1}^*(-k)$$
$$\qquad + Q_3^{(m)}(k)S_{i,2}^*(-k) + \tilde{N}_{i,1}^{(m)}(k)$$
$$Y_{i,2}^{(m)}(k) = Q_1^{(m)}(k)S_{i,2}(k) + Q_2^{*(m)}(k)S_{i,2}^*(-k)$$
$$\qquad - Q_3^{*(m)}(k)S_{i,1}^*(-k) + \tilde{N}_{i,2}^{(m)}(k) \qquad (16)$$

where

$$Q_1^{(m)}(k) \quad |A^{(m)}(k)|^2 + |C^{(m)}(k)|^2$$
$$Q_2^{(m)}(k) \quad A^{*(m)}(k)B^{(m)}(k) + C^{(m)}(k)D^{*(m)}(k) \qquad (17)$$
$$Q_3^{(m)}(k) \quad A^{*(m)}(k)D^{(m)}(k) - C^{(m)}(k)B^{*(m)}(k)$$

and

$$\tilde{N}_{i,1}^{(m)}(k) = A^{*(m)}(k)n_{i,1}^{(m)}(k) + C^{(m)}(k)n_{i,2}^{*(m)}(k)$$
$$\tilde{N}_{i,1}^{(m)}(k) = C^{*(m)}(k)n_{i,1}^{(m)}(k) - A^{(m)}(k)n_{i,2}^{*(m)}(k) \qquad (18)$$

Finally, the received data symbols at multiple receive antennas are combined to obtain receiver diversity, and the raw data symbols are given by

$$Y_{i,1}(k) = \sum_{m=1}^{N_r} \left[Y_{i,1}^{(m)}(k) \right] =$$
$$\underbrace{Q_1(k)S_{i,1}(k)}_{\text{signal}} + \underbrace{Q_2(k)S_{i,1}^*(-k) + Q_3^{(m)}(k)S_{i,2}^*(-k)}_{\text{intercarrier interference}} + \underbrace{\tilde{N}_{i,1}(k)}_{\text{noise}}$$

$$Y_{i,2}(k) = \sum_{m=1}^{N_r} \left[Y_{i,2}^{(m)}(k) \right] =$$
$$\underbrace{Q_1(k)S_{i,2}(k)}_{\text{signal}} + \underbrace{Q_2(k)S_{i,2}^*(-k) - Q_3^{*(m)}(k)S_{i,1}^*(-k)}_{\text{intercarrier interference}} + \underbrace{\tilde{N}_{i,2}(k)}_{\text{noise}}$$
$$\qquad (19)$$

where

$$Q_1(k) = \sum_{m=1}^{N_r} \left[Q_1^{(m)}(k) \right]$$
$$Q_2(k) = \sum_{m=1}^{N_r} \left[Q_2^{(m)}(k) \right] \qquad (20)$$
$$Q_3(k) = \sum_{m=1}^{N_r} \left[Q_3^{(m)}(k) \right]$$

are defined as the combined coefficients, and

$$\tilde{N}_{i,1}(k) = \sum_{m=1}^{N_r} \left[\tilde{N}_{i,1}^{(m)}(k) \right]$$
$$\tilde{N}_{i,2}(k) = \sum_{m=1}^{N_r} \left[\tilde{N}_{i,2}^{(m)}(k) \right] \qquad (21)$$

are the combined noise.

From (19), it can be seen that there is ICI in the signals after receiver combining, which is caused by the image projections from the mirrored frequency. To improve system performance, it is necessary to compensate for the received signals.

3.2. LS Estimator

Although MMSE estimator is optimal, it suffers from high computational complexity and requires the knowledge of $\mathbf{R}_v$, which must be estimated using a large amount of transmission data. A simple but effective method is least square estimator. To further reduce the computational complexity, we design a special training pattern to avoid matrix inversion operation. In order to utilize this special pattern, training sequences must be transmitted in groups of two blocks. Let s, s, s, and s^* be the four consecutive training symbols within the ith and the $(i+1)$th STC blocks at the kth sub-carrier, where $s = p(1+j)$ is a complex number with identical real and imaginary parts p. The corresponding training symbols at the $-k$th sub-carrier are also s, s, s, and s^*. According to (12), the received data within the ith and the $(i+1)$th STC blocks at the kth sub-carrier can be represented in matrix form as

$$
\begin{bmatrix} X_{i,1}^{(m)}(k) \\ X_{i,2}^{(m)}(k) \\ X_{i+1,1}^{(m)}(k) \\ X_{i+1,2}^{(m)}(k) \end{bmatrix} = \begin{bmatrix} s & s^* & s & s^* \\ -s^* & -s & s^* & s \\ s & s^* & s & s \\ -s & -s^* & s^* & s \end{bmatrix} \mathbf{v}^{(m)}(k) + \begin{bmatrix} N_{i,1}^{(m)}(k) \\ N_{i,2}^{(m)}(k) \\ N_{i+1,1}^{(m)}(k) \\ N_{i+1,2}^{(m)}(k) \end{bmatrix}
$$

(36)

Thus, the LS estimate of the joint coefficients is given as [35]

$$
\begin{aligned}
\hat{\mathbf{v}}^{(m)}(k) &= \begin{bmatrix} s & s^* & s & s^* \\ -s^* & -s & s^* & s \\ s & s^* & s & s \\ -s & -s^* & s^* & s \end{bmatrix}^{-1} \begin{bmatrix} X_{i,1}^{(m)}(k) \\ X_{i,2}^{(m)}(k) \\ X_{i+1,1}^{(m)}(k) \\ X_{i+1,2}^{(m)}(k) \end{bmatrix} \\
&= \frac{1}{4p} \begin{bmatrix} 0 & -1-j & 1 & j \\ 0 & -1+j & 1 & -j \\ 1-j & 0 & j & 1 \\ 1+j & 0 & -j & 1 \end{bmatrix} \begin{bmatrix} X_{i,1}^{(m)}(k) \\ X_{i,2}^{(m)}(k) \\ X_{i+1,1}^{(m)}(k) \\ X_{i+1,2}^{(m)}(k) \end{bmatrix}
\end{aligned}
$$

(37)

The estimates of the virtual channel coefficients from different training sequence groups can be averaged to obtain more accurate estimate.

3.3. Signal Compensation Approach

Based on the MMSE estimate given by (31) or LS estimate given by (37) of the joint coefficients, the estimate of the combined coefficients, $\hat{Q}_1(k)$, $\hat{Q}_2(k)$, and $\hat{Q}_3(k)$, can be calculated according to (17) and (20), and then can be used to equalize the raw data symbols. According to (19), the raw data symbols at the kth and the $(-k)$th sub-carriers within the ith STC block can be written in matrix form as (38). Thus, the zero-forcing estimates of the transmitted data symbols are given by (39).

$$
\begin{bmatrix} Y_{i,1}(k) \\ Y_{i,1}^*(-k) \\ Y_{i,2}(k) \\ Y_{i,2}^*(-k) \end{bmatrix} =
$$

$$
\begin{bmatrix} \hat{Q}_1(k) & \hat{Q}_2(k) & 0 & \hat{Q}_3(k) \\ \hat{Q}_2^*(-k) & \hat{Q}_1^*(-k) & \hat{Q}_3^*(-k) & 0 \\ 0 & -\hat{Q}_3^*(-k) & \hat{Q}_1(k) & \hat{Q}_2^*(k) \\ -\hat{Q}_3(-k) & 0 & \hat{Q}_2(-k) & \hat{Q}_1^*(-k) \end{bmatrix} \begin{bmatrix} S_1(k) \\ S_1^*(-k) \\ S_2(k) \\ S_2^*(-k) \end{bmatrix} + \begin{bmatrix} \tilde{N}_{i,1}(k) \\ \tilde{N}_{i,1}(-k) \\ \tilde{N}_{i,2}(k) \\ \tilde{N}_{i,2}(-k) \end{bmatrix}
$$

(38)

$$
\begin{bmatrix} \hat{S}_1(k) \\ \hat{S}_1^*(-k) \\ \hat{S}_2(k) \\ \hat{S}_2^*(-k) \end{bmatrix} = \begin{bmatrix} \hat{Q}_1(k) & \hat{Q}_2(k) & 0 & \hat{Q}_3(k) \\ \hat{Q}_2^*(-k) & \hat{Q}_1^*(-k) & \hat{Q}_3^*(-k) & 0 \\ 0 & -\hat{Q}_3^*(-k) & \hat{Q}_1(k) & \hat{Q}_2^*(k) \\ -\hat{Q}_3(-k) & 0 & \hat{Q}_2(-k) & \hat{Q}_1^*(-k) \end{bmatrix}^{-1} \begin{bmatrix} Y_{i,1}(k) \\ Y_{i,1}^*(-k) \\ Y_{i,2}(k) \\ Y_{i,2}^*(-k) \end{bmatrix}
$$

(39)

4. Performance Analysis

According to (19), the received raw data symbols are contaminated by noise and interference from the signals at the mirrored sub-carriers. If the ICI can be successfully canceled by the proposed algorithm, the post-processing SNR at the kth sub-carrier, $\eta(k)$, can be calculated as

$$
\begin{aligned}
\eta(k) &= \frac{\varepsilon\{|Q_1(k)S_{i,1}(k)|^2\}}{\varepsilon\{|\tilde{n}_{i,1}(k)|^2\}} = \frac{\varepsilon\{|Q_1(k)S_{i,2}(k)|^2\}}{\varepsilon\{|\tilde{n}_{i,2}(k)|^2\}} \\
&= \sum_{m=1}^{N_r}\left[|A^{(m)}(k)|^2 + |C^{(m)}(k)|^2\right]\frac{E_s/2}{N_0} \\
&= \frac{1}{2}\sum_{m=1}^{N_r}\left[|A^{(m)}(k)|^2 + |C^{(m)}(k)|^2\right]\rho
\end{aligned}
$$

(40)

where $E_s/2$ is the average transmit energy per symbol period per antenna and $\rho = E_s/N_0$ can be interpreted as the average SNR for the single-input single-output scheme.

This post-processing SNR is determined by the joint effects of channels and I/Q imbalances. For classical i.i.d. channels [5] and perfect I/Q characteristics, the post-processing SNR becomes

$$
\eta(k) = N_r\rho
$$

(41)

This shows that the system with perfect I/Q over i.i.d. channel can achieve an array gain of N_r.

In OFDM systems, each sub-carrier can be treated as a frequency-flat channel. Assuming optimum detection at the receiver, the corresponding symbol error rate for rectangular M-ary QAM is given by [36]

$$P_s(k) = 1 - \left(1 - 2\left(1 - \frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{3\eta(k)}{M-1}}\right)\right)^2 \quad (42)$$

Assuming that only one single bit is changed for each erroneous symbol, the equivalent bit error rate for rectangular M-ary QAM is approximated as [37]

$$P_b(k) \approx \frac{1}{log_2(M)}P_s(k) \quad (43)$$

5. Simulation Results

To evaluate the proposed virtual channel idea, the two estimators, and the signal compensation approach, we used MATLAB to simulate an OFDM-based $2 \times N_r$ MIMO system with frequency-dependent TX and RX I/Q imbalances. The size of fast Fourier transform (FFT) is 128, the number of used sub-carriers is 96, and the length of CP is 32. The multi-path channel is modeled by six independent complex taps with a power delay profile of a 3 dB decay per tap. It should be noted that the actual channel length can be estimated [38]. The simulation bandwidth is set to 20 MHz, leading to a 156.25 KHz sub-carrier spacing and a maximum $0.3 \mu s$ excess delay. Sixty-four quadrature amplitude modulations (64QAM) are used.

The non-linear frequency characteristics of the I and Q branches, $L_I(f)$ and $L_Q(f)$, are modeled as two first-order finite impulse response (FIR) filters. A form of parameters $\{g, \theta, [a_I, b_I], [a_Q, b_Q]\}$ is used to describe the I/Q imbalance, where $[a_I, b_I]$ and $[a_Q, b_Q]$ are the coefficients of the FIR filters for the I and Q branches. I/Q parameters of $\{1.03, 3, [0.01, 0.9], [0.8, 0.02]\}$ and $\{[0.8, 0.02], [0.01, 0.9]\}$ are used for the two transmitters, respectively. For simplification of simulation, all receivers use I/Q parameters of $\{1.05, 3, [0.8, 0.02], [0.9, 0.01]\}$. Because the estimation of the joint coefficients is performed separately in each receiver, the same I/Q parameters for different I/Q models still lead to generalization of the simulation results. It should be noted that the I/Q imbalance parameters are chosen to be worst case in order to evaluate the robustness of the proposed approach.

A frame-by-frame transmission scheme is employed in the simulation. The multi-path channel is independently generated for each frame. One frame consists of N_{tr} STC blocks of training symbols followed by 50 STC blocks of data symbols. A total of 5,000 frames (288 Mbits) are simulated for each scenario at a given SNR.

Figure 3 shows the frequency response of two of the FIR filters used for simulating the frequency-dependent characteristics of I/Q imbalances. For both filters, the variations of the amplitude are within 0.5 dB, but the average gains are differentiated by 1 dB. It should be noted that the signals after TX I/Q distortions must be normalized in the simulations to compensate for the energy lost due to attenuations of the filters. While the variation of the phase

response for the first filter is small, the variation is very large for the second filter. The amplitude and phase differences of the filters are suitable to simulate the frequency-dependent characteristics of I/Q imbalances.

Figure 4 shows the typical constellations of one frame of symbols generated in an ideal channel environment (without multi-path fading and noise). **Figures 4** (a) and (b) show the constellations of the raw data symbols under frequency-independent and frequency-dependent I/Q imbalances, respectively. It is observed that the constellation in **Figure 4** (b) becomes nearly random compared with **Figure 4** (a) due to the effect of frequency-de pendent I/Q imbalances. **Figure 4** (c) shows the constellation of signals recovered by the proposed LS algorithm, which shows that the proposed algorithm can perfectly compensate frequency-independent and frequency-dependent I/Q imbalances.

The BER performances of the proposed estimators and compensation approach are shown in **Figures 5-10**. To make better comparisons, a series of simulations under different scenarios was conducted, including a scenario of perfect I/Q and CSI known at receivers termed as "ideal case", a scenario with frequency-independent I/Q imbalances termed as "indep-IQ", and a scenario with frequency-dependent I/Q imbalances termed as "dep-IQ". The legend term of "NoComp" means that I/Q imbalances are present but no compensation scheme is applied (the CSI is estimated under I/Q imbalances), "MMSE (N)" means that the MMSE estimator with N STC blocks of training sequences (the total $2 \times N$ OFDM
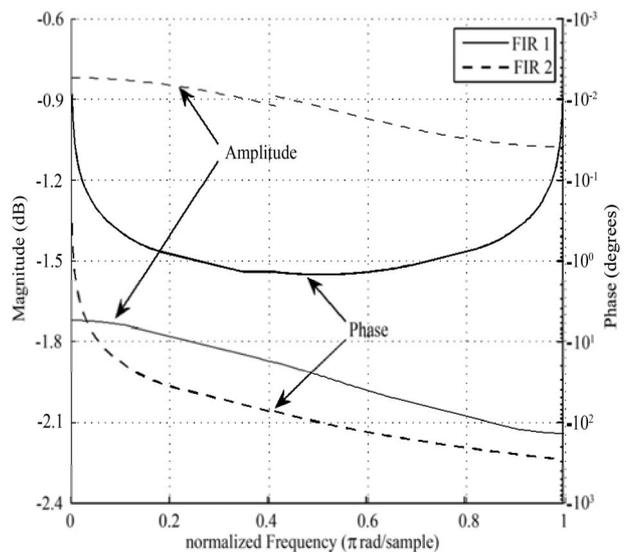
Figure 3. Frequency response of the FIR filters used for simulating the frequency-dependent characteristics of I/Q imbalances; coefficients of FIR 1 are [0.8, 0.02], and coefficients for FIR 2 are [0.01, 0.9].

(a)   (b)   (c)
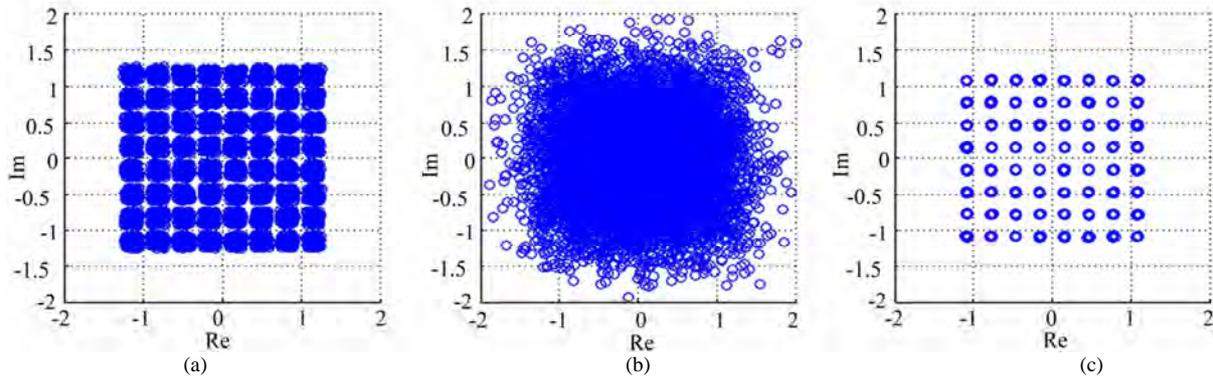
Figure 4. Constellations of signals in a frame generated by a 2 × 1 MIMO-OFDM system under ideal channel environment. (a) raw data constellations under frequency-independent I/Q imbalances; (b) raw data constellations under frequency-dependent I/Q imbalances; (c) constellations of the compensated data under frequency-dependent I/Q imbalances.

symbols) is used, and "LS(N)" means that the LS estimator with N STC blocks of training sequence is used to estimate the joint coefficients. For both MMSE and
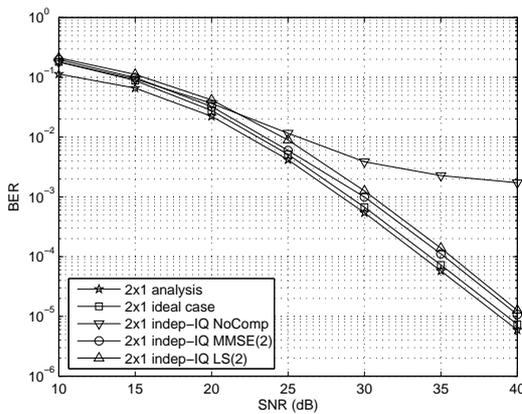
Figure 5. Performance of a 2 × 1 MIMO-OFDM system with TX and RX frequency-independent I/Q imbalances over multi-path fading channel.
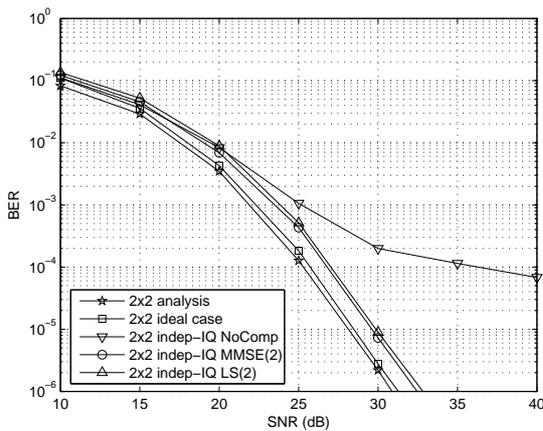
LS estimators, the proposed compensation approach is applied to compensate the received raw data symbols. Furthermore, the BER is theoretically calculated according to (43) and is shown in the following figures with legend term "analysis".

Figure 5 and **Figure 6** show the performance of a 2×1 and a 2×2 systems with frequency-independent I/Q imbalances, respectively. It is observed from the "No-comp" curves that the I/Q imbalances significantly degrade the system performance, resulting in high error floor. These results agree with the constellation analysis mentioned above. With the proposed estimators and signal compensation approach, the I/Q distortion can be effectively mitigated and the system performance is significantly improved. By using two STC blocks of training sequences, the system performance resulting from both MMSE and LS estimators is close to the ideal case. The performance degradation is less than 1 dB. Although LS estimator performs a little worse than MMSE estimator, the low computational complexity makes the LS estimator more competitive. By assuming perfect estimation of the

Figure 6. Performance of a 2 × 2 MIMO-OFDM system with TX and RX frequency-independent I/Q imbalances over multi-path fading channel.
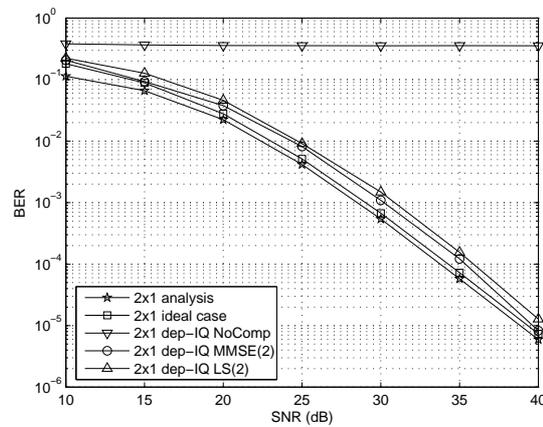
Figure 7. Performance of a 2 × 1 MIMO-OFDM system with TX and RX frequency-dependent I/Q imbalances over multi-path fading channel.
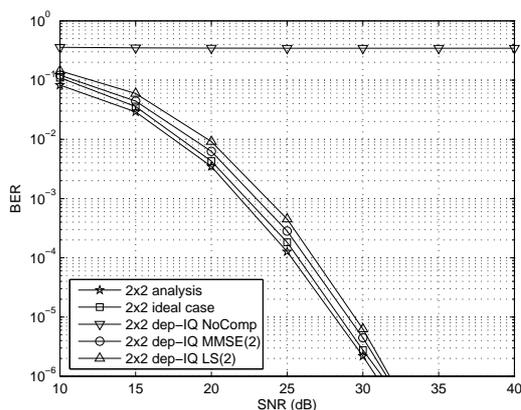
Figure 8. Performance of a 2 × 2 MIMO-OFDM system with TX and RX frequency-dependent I/Q imbalances over multi-path fading channel.
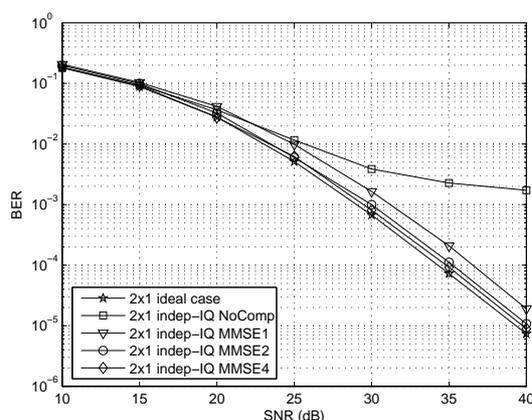
Figure 9. Comparison of the performances of the MMSE estimator with different lengths of training symbols for the 2 × 1 MIMO-OFDM system with TX and RX frequency-independent I/Q imbalances over multi-path fading channel.
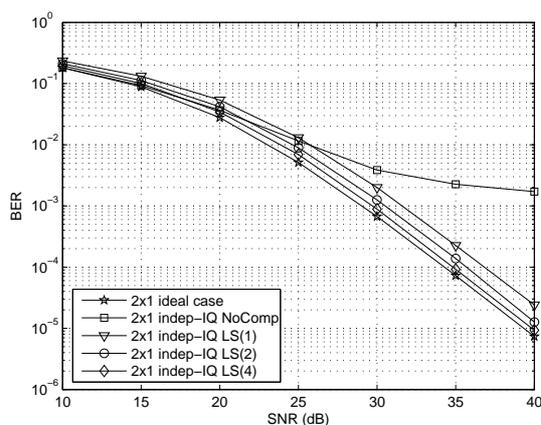
Figure 10. Comparison of the performances of the LS estimator with different lengths of training symbols for the 2 × 1 MIMO-OFDM system with TX and RX frequency-independent I/Q imbalances over multi-path fading channel.

joint coefficients and compensation of the received signal, it is reasonable that the analysis results are slightly better than the ideal case.

The performance of the systems with frequency-dependent I/Q imbalances are shown in **Figure 7** and **Figure 8** for the 2×1 and 2×2 scenarios, respectively. The frequency dependent characteristic of the I/Q imbalances causes fatal influence on the MIMO-OFDM systems. Without compensation, the BER remains one half regardless of the SNR. Our proposed approach can also successfully combat the worst fading effect caused by frequency-dependent I/Q imbalances, resulting in good performance that is close to the ideal case. This significant performance improvement has not been reported by other literatures.

An intuitive sense is that a longer training sequence could result in better performance. To demonstrate this point, we compare the system performances with different lengths of training symbols in **Figure 9** and **Figure 10** for MMSE and LS estimators, respectively. It is observed that BER decreases with the increase of the training symbol numbers. When four blocks of training sequences are used, the performance loss compared to the ideal case is negligible.

6. Conclusions

In this paper, we introduce a new virtual channel concept to analyze the I/Q imbalances in a MIMO-OFDM wireless communication system over multi-path fading channels. The input-output relation is derived in frequency domain, which incorporates the effect of I/Q imbalances with multi-path fading channels. The integrated effect can be modeled by the joint coefficients of the virtual channel. By using this approach, inaccurate estimation of the channel state information can be avoided at the diversity combining stage. The joint coefficients are estimated by using the proposed MMSE and LS estimators, and are used to compensate the received signals. Simulation results show that the proposed approach can effectively mitigate the TX and RX I/Q imbalances in MIMO-OFDM systems. Although only a two transmit antenna scheme is illustrated in this paper, our proposed approach can be easily extended to any number of transmit antenna configurations by using orthogonal space-time block coding.

7. Acknowledgements

8. References

[1] M. Jang and L. Hanzo, "Multiuser MIMO-OFDM for Next-Generation Wireless Systems," *Proceedings of the IEEE*, Vol. 95, No. 7, July 2007, pp. 1430-1469.

[2] R. Chang and R. Gibby, "A Theoretical Study of Performance of an Orthogonal Multiplexing Data Transmission Scheme," *IEEE Transactions on Communications Technology*, Vol. 16, No. 4, August 1968, pp. 529-540.

[3] Y. Wu and W. Y. Zou, "Orthogonal Frequency Division Multiplexing: A Multi-Carrier Modulation Scheme," *IEEE Transactions on Consumer Electronics*, Vol. 41, No. 3, August 1995, pp. 392-399.

[4] L. Hanzo, M. Muinster, B. Choi and T. Keller, "OFDM and MC-CDMA for Broadband Multi-User Communications," Wlans and Broadcasting, Wiley, 2003.

[5] A. Paulraj, R. Nabar and D. Gore, "Introduction to Space-Time Wireless Communication*s*," Cambridge University Press, Cambridge, 2003.

[6] Y. Yang, G. Xu and H. Ling, "An Experimental Investigation of Wideband MIMO Channel Characteristics Based on Outdoor Non-Los Measurements at 1.8 GHz," *IEEE Transactions on Antennas Propagation*, Vol. 54, November, 2006, pp. 3274-3284.

[7] H. Sampath, S. Talwar, J. Tellado, V. Erceg and A. J. Paulraj, "A Fourth-Generation MIMO-OFDM Broadband Wireless System: Design, Performance and Field Trial Results," *IEEE Communications Magazine*, Vol. 40, No. 9, September 2002, pp. 143-149.

[8] G. L. Stüber, J. R. Barry, S. W. McLaughlin, Y. G. Li, M. A. Ingram and T. G. Pratt, "Broadband MIMO-OFDM Wireless Communications," *Proceedings of the IEEE*, Vol. 92, No. 2, February 2004, pp. 271-294.

[9] Y. Li, J. H. Winters and N. R. Sollenberger, "MIMO-OFDM for Wireless Communications: Signal Detection with Enhanced Channel Estimation," *IEEE Transactions on Communications*, Vol. 50, No. 9, September, 2002, pp. 1471-1477.

[10] M. Shin, H. Lee and C. Lee, "Enhanced Channel-Estimation Technique for MIMO-OFDM Systems," *IEEE Communications Letters*, Vol. 53, No. 1, January 2004, pp. 262-265.

[11] IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access System, October 2004.

[12] IEEE Candidate Standard 802.11n: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2004.

[13] 3GPP Technical Specification Group Radio Access Network, Overall Description of E-UTRA/E-UTRAN, 2007.

[14] S. R. Herlekar, C. Zhang, H.-C. Wu, A. Srivastava and Y. Wu, "OFDM Performance Analysis in the Phase Noise Arising from the Hot-Carrier Effect," *IEEE Transactions on Consumer Electronics*, Vol. 52, No. 3, August 2006, pp. 757-765.

[15] J. Lei and T.-S. Ng, "A Consistent OFDM Carrier Frequency Offset Estimator Based on Distinctively Spaced Pilot Tones," *IEEE Transactions on Wireless Communications*, Vol. 3, No. 2, March 2004, pp. 588-599.

[16] C. L. Liu, "Impacts of I/Q Imbalance on QPSK-OFDM-QAM Detection," *IEEE Transactions on Consumer Electronics*, Vol. 44, No. 3, August 1998, pp. 984-989.

[17] M. Valkama, M. Renfors and V. Koivunen, "Compensation of Frequency-Selective I/Q Imbalances in Wideband Receivers: Models and Algorithms," *Proceedings of the 3rd IEEE Workshop on Signal Processing Advances in Wireless Communications*, Taiwan, March 2001, pp. 42-45.

[18] H.-C. Wu and Y. Wu, "Distributive Pilot Arrangement Based on Modified M-sequences for OFDM Inter-Carrier Interference Estimation," *IEEE Transactions on Wireless Communications*, Vol. 6, No. 5, May 2007, pp. 1605-1609.

[19] X. Huang and H.-C. Wu, "Robust and Efficient Inter-Carrier Interference Mitigation for OFDM Systems in Time-Varying Fading Channels," *IEEE Transactions on Vehicular Technology*, Vol. 56, No. 5, September 2007, pp. 2517-2528.

[20] H.-C. Wu, X. Huang, Y. Wu and X. Wang, "Theoretical Studies and Efficient Algorithm of Semi-Blind ICI Equalization for OFDM," *IEEE Transactions on Wireless Communications*, Vol. 7, No. 10, October 2008, pp. 3791-3798.

[21] A. A. Abidi, "Direct-Conversion Radio Transceivers for Digital Communications," *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 12, December 1995, pp. 1399-1410.

[22] J. Crols and M. Steyaert, "Low-IF Topologies for High-Performance Analog Front Ends of Fully Integrated Receivers," *IEEE Transactions on Circuits System*, Vol. 45, No. 3, March 1998, pp. 269-282.

[23] A. Tarighat, R. Bagheri and A. H. Sayed, "Compensation Schemes and Performance Analysis of IQ Imbalances in OFDM Receivers," *IEEE Transactions on Signal Processing*, Vol. 53, No. 8, August 2005, pp. 3257-3268.

[24] J. K. Hwang, W. M. Chen, Y. L. Chiu, S. J. Lee and J. L. Lin, "Adaptive Baseband Compensation for I/Q Imbalance and Time-Varying Channel in OFDM System," *Intelligent Signal Processing and Communications*, Japan, December 2006, pp. 638 – 641.

[25] J. Tubbax, B. Come, L. van der Perre, S. Donnay, M. Engles and C. Desset, "Joint Compensation of IQ Imbalance and Phase Noise," *IEEE Semiannual on Vehicular Technology Conference*, Vol. 3, April 2003, pp. 1605-1609.

[26] D. Tandur and M. Moonen, "Joint Compensation of OFDM Frequency-Selective Transmitter and Receiver IQ Imbalance," *EURASIP Journal on Wireless Communications and Networking*, Vol. 2007, No. 68563, May 2007, pp. 1-10.

[27] L. Anttila, M. Valkama and M. Renfors, "Frequency-Selective I/Q Mismatch Calibration of Wideband Direct-Conversion Transmitters," *IEEE Transactions on Circuits System*, Vol. 55, No. 4, April 2008, pp. 359-363.

[28] S. Ma, D. Duran and Y. Yang, "Estimation and Compen-

sation of Frequency-Dependent I/Q Imbalances in OFDM Systems over Fading Channels," *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing*, Beijing, September 2009.

[29] A. Tarighat, W. M. Younis and A. H. Sayed, "Adaptive MIMO OFDM Receivers: Implementation Impairments and Complexity Issues," *IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, Yokohama, August 2004.

[30] Y. Zou, M. Valkama and M. Renfors, "Digital Compensation of I/Q Imbalance Effects in Space-Time Coded Transmit Diversity Systems," *IEEE Transactions on Signal Process,* Vol. 56, No. 6, June 2008, pp. 2496-2508.

[31] Y. Zou, M. Valkama and M. Renfors, "Analysis and Compensation of Transmitter and Receiver I/Q Imbalances in Space-Time Coded Multi-Antenna OFDM Systems," *EURASIP Journal on Wireless Communications and Networking*, No. 391025, January 2008, pp. 1-10.

[32] D. Tandur and M. Moonen, "Compensation of RF Impairments in MIMO OFDM Systems," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, April 2008, pp. 3097-3100.

[33] S. Ma, D. Duran, H. Sharif and Y. Yang, "An Adaptive Approach to Estimation and Compensation of Frequency-Dependent I/Q Imbalances in MIMO-OFDM Systems," *Proceedings of IEEE Global Communications Conference*, Honolulu, 2009.

[34] A. H. Sayed, Adaptive Filters, Wiley-IEEE Press, New Jersey, 2008.

[35] D. Manolakis, V. K. Ingle and S. M. Kogon, "Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing," MA: Artech House Publishers, Boston, 2005.

[36] J. G. Proakis, "Digital Communications," 4th edition, Mc-Graw-Hill, New York, 2000.

[37] R. L. Peterson, R. E. Ziemer and D. E. Borth, "Introduction to Spread-Spectrum Communications," 3rd edition, Prentice Hall, New Jersey, 1995.

[38] X. Wang, H. C. Wu, S. Y. Chang, Y. Wu and J. Y. Chouinard, "Efficient Non-Pilot-Aided Channel Length Estimation for Digital Broadcasting Receivers," *IEEE Transactions on Broadcasting*, Vol. 55, No. 3, September 2009, pp. 633-641.

Scientific
Research

A New Method for Improving Robustness of Registered Fingerprint Data Using the Fractional Fourier Transform

Reiko Iwai, Hiroyuki Yoshimura
Graduate School of Engineering, Chiba University, Chiba, Japan
E-mail: reiko@tu.chiba-u.ac.jp, yoshimura@faculty.chiba-u.jp
Received June 21, 2010; revised July 29, 2010; accepted August 30, 2010

Abstract

Inspired by related studies, a new data processing method in fingerprint authentication using the fractional Fourier transform (FRT) was proposed for registered fingerprint data. In this proposal, protection of personal information was also taken into account. We applied the FRT instead of the conventional Fourier transform (FT) which has been used as one of the representative fingerprint authentication algorithm. Our method solved the problem of current registration method and the robustness was verified. In this study, a modeled fingerprint image instead of the original raw fingerprint images was analyzed in detail to make the characteristic clear. As one dimensional (1D) modeled fingerprint image, we used the finite rectangular wave which is regarded as the simplification of the grayscale distribution in an arbitrary scanned line of the raw fingerprint images. As a result, it was clarified that the data processed by the FRT provide higher safety than those processed by the FT, because it is difficult to specify the orders from the intensity distribution of FRTs (the intensity FRTs) when the combination of the various FRT's order at every scanned line is used.

1. Introduction

The fingerprint images are indispensable and easy to use the information to identify individuals. It is often-used for logging into a PC, access control, as well as diligence and indolence management in an office. Registration methods of the fingerprint images are classified into three major categories. One method is to register the whole fingerprints images as two dimensional (2D) data without modification. The others include to register a priori extracted features of the images, such as minutiae templates [1], and to register spatial frequency data transformed from 1D data extracted from the original 2D image in a specific direction [2]. In the former two methods, there exists the problem that unfair use is possible when the information leaks out. For the security reasons, the third method might be more preferable. However, it also has following problems: 1) the registered data can easily be decoded to the original 1D data by the inverse FT (IFT), when the registered data are generated using the conventional FT; 2) additional processing time is necessary for any trials to solve the issue 1).

We focused on the application of the FRT [3] by generalizing the FT to solve these problems, where the FRT's order can be set arbitrarily. Because of this unique feature of the FRT, it has capability of the encryption if the FRT's order is not revealed. The FRT of the 1D image out of a 2D fingerprint image includes such features as less computational complexity to retrieve the original, if needed, and impossibility to be decoded to the corresponding raw data by unauthorized persons who have no knowledge on the FRT's order used. Therefore, even if the registered fingerprint information leaked out from an identification system, security would be guaranteed. Since the FRT will be processed by an optical system incorporating a lens and a laser light source to conduct the necessary process physically from scanning the fingerprint images to generating the FRT image [4]. Therefore, it could drastically reduce the time needed for individual identification.

In the present study, taking into account such future application, the intensity FRTs are considered in terms of any possibility of unlawful reveal of the hidden FRT's orders used to register the information corresponding to the fingerprint images. Namely, we compare the intensity distribution data obtained by the FT and FRT. In the comparison, the actual data are simplified to be the finite-length rectangular waveforms because the analytical results could be explained easily. The purpose of this analysis is to demonstrate that unauthorized third persons cannot retrieve

the original data from the registered data in FRT. In this paper, we analyze from the following two perspectives: 1) behavior of the peak value of the cross-correlation function between the finite rectangular wave and the intensity FRT, and 2) behavior of the peak value of the cross-correlation function between the finite rectangular wave and the intensity inverse FRT (IFRT). These analyses allow us to show the difference between the intensity FRT, intensity IFRT and the original image quantitatively. This fact means that we cannot identify the original fingerprint image as the difference becomes greater and greater. In addition, we verify the robustness from the following two viewpoints: 1) behavior of the second-maximum value of auto-correlation function of the intensity FRT, and 2) behavior of the peak value of the cross-correlation function between the finite rectangular wave and the intensity IFRT (the FRT's and IFRT's orders are different cases). First, we show that the FRT's order cannot be identified from the characteristics of the intensity FRT. Next, we show the intensity FRT cannot be decoded to the corresponding original data by using the conventional IFT by unauthorized persons who have no knowledge on the FRT's order used.

2. A New Data Processing Method in Fingerprint Authentication by Use of the Intensity FRT

2.1. Definition of the Fractional Fourier Transform (FRT)

The FRT is the generalization of a conventional FT. The FRT of 1D input data $u(x)$ is defined [3, 5] as

$$u_p(x_p) = F^{(p)}\left[u(x)\right] = \int u(x)\exp[i\pi(x_p^2 + x^2)/s^2 \tan\phi]$$

$$\times \exp[-2i\pi x_p x / s^2 \sin\phi]dx, \qquad (1)$$

where a constant factor has been dropped; $\phi = p\pi/2$, where p is the FRT's order; s is a constant. In particular, in the optical FRT, s is called a scale parameter expressed in terms of $s = \sqrt{\lambda f_s}$ where λ is the wavelength and f_s is an arbitrarily fixed focal length. In this paper, the value of s was fixed at 1.0.

When p takes a value of $4n+1$, n being any integer, the FRT corresponds to the conventional FT. The intensity distribution of the FRT, $I_p(x_p)$, is obtained by calculating $|u_p(x_p)|^2$. In addition, $u_p(x_p)$ can be decoded to $u(x)$ by the IFRT with the order $-p$ as follows:

$$u(x) = F^{(-p)}\left[u_p(x_p)\right]. \qquad (2)$$

In our study, we call p in Equation (2) the IFRT's order. "Disfrft.m" [6] was used in our numerical calculation of the FRT.

2.2. Modeling Waveform Pattern of the Fingerprint

Although the grayscale levels are composed of intermediate values between 0 and 255 at the actual scanned lines in the case of 2D black and white image of 8 bits, in order to highlight the FRT as our new method together with its feasibility, a finite rectangular wave is assumed to be the simplification of the grayscale distribution of the fingerprint image as shown in **Figure 1**. Horizontal axis is intentionally made of 1024 (2^{10}) pixels for the results of the FT and the FRT to be smoothly illustrated. In the present paper we show the result only at the one scanned line especially. However, we premise the application of the FRT to the 2D original fingerprint image which has multiple lines with random FRT's orders. In addition, the FRT's orders can be used as arbitrary real numbers.

2.3. Application of the FRT

The algorithm of the FRT has been intensively studied [7-9]. Alternatively, the FRT was also applied to the fake finger detection [10].

In the present paper we apply the FRT to the 1D finite rectangular wave data shown in **Figure 1** as a model of fingerprint images. Basically, the FRT with the order p is applied to the finite rectangular wave in Equation (1). The FRT with the order p can be decoded to the finite rectangular wave by the IFRT with the same order p as already explained in Equation (2).

Figure 2 demonstrates the results of the FRTs in comparison with the conventional FT (*i.e.*, the FRT with p=1.0). Namely, **Figure 2(a)** shows the result of the FT as the amplitude distribution at the upper portion and the phase distribution at the lower portion. **Figures 2(b)** and
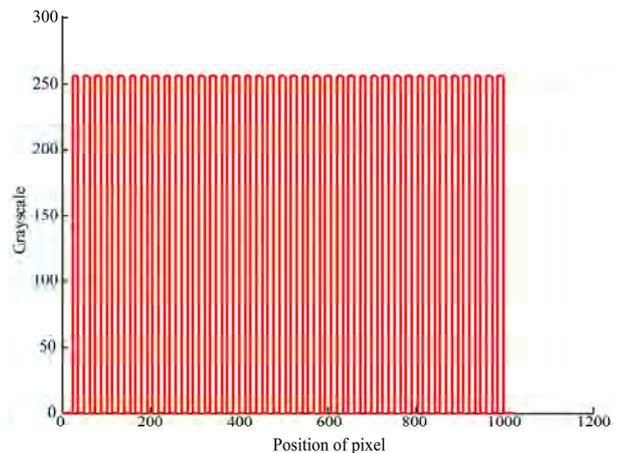

Figure 1. The finite rectangular wave as a modeled fingerprint image.

(a)

(b)

(c)
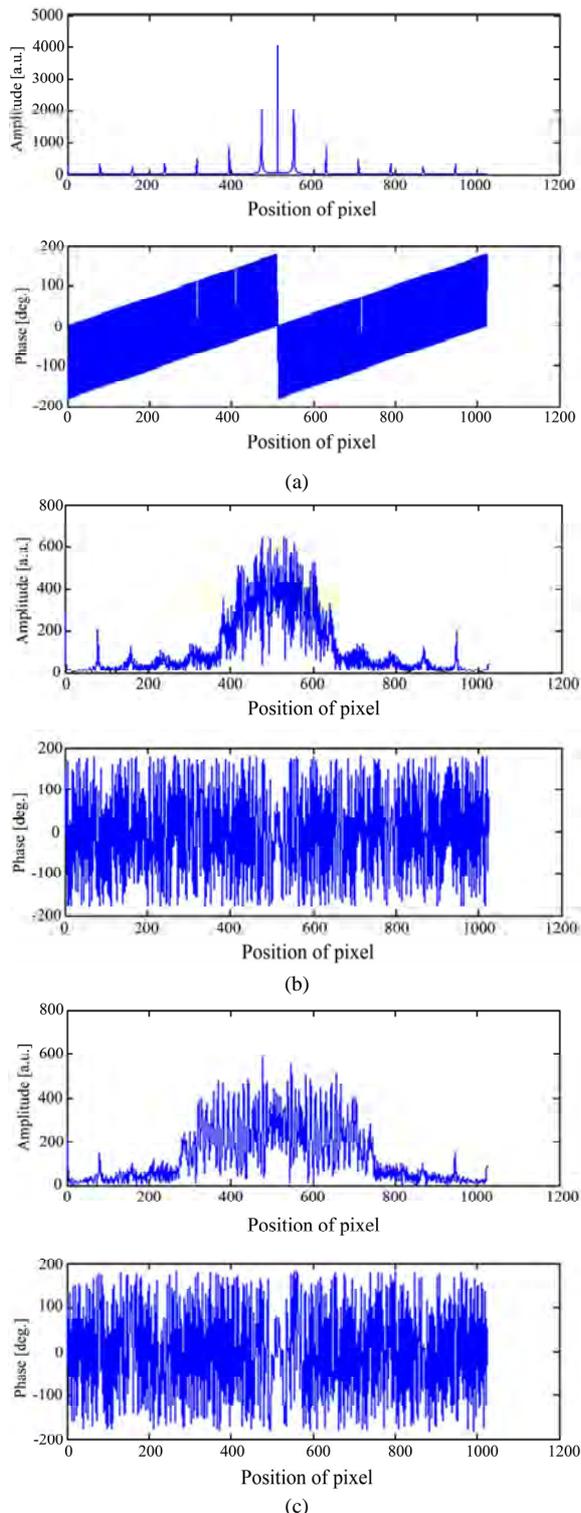
Figure 2. Examples of the amplitude and phase distributions of the FRTs applied to the finite rectangular wave, when ps = (a) 1.0, (b) 0.9 and (c) 0.8.

2(c) are the results of the FRTs with ps = 0.9 and 0.8, respectively. As a result, the peak values of the ampli-

tude distributions in **Figures 2(a)**, **2(b)** and **2(c)** are 4.04×10^3, 6.59×10^2 and 5.80×10^2, respectively.

It is found that the peak value of the amplitude distribution falls remarkably and the width of spread increases when the value of the FRT's order p decreases. It is also found that there is little difference in phase distributions between **Figures 2(b)** and **2(c)**. In the case of FT shown in **Figure 2(a),** the order p can be identified through the waveform of the amplitude and phase distributions. However, in the case of FRT, the order p might not be identified through them. In particular, it is difficult to identify the FRT's orders ps through the waveforms of the phase distributions shown in **Figures 2(b) and 2(c)**. Therefore, this fact led us the new method safer than the conventional method using the FT, because the FRT's order has highly-confidential in the applied FRT condition.

In this way, we focused on the intensity distribution of the FRT from a viewpoint of the security of individual information, because the intensity FRT may not be completely decoded to the original fingerprint image by the IFRT. **Figure 3** depicts the intensity FT of **Figure 2(a)** and the intensity FRT of **Figure 2(b).** The peak values of the intensity distributions in **Figures 3(a) and 3(b)** are 1.63×10^7 and 4.34×10^5, respectively. It is found from the comparison between **Figures 2** and **3** that the peak value of the wave pattern of the intensity distribution is very high. In the next section, we compare the registered information of the intensity FT and the intensity FRT by changing the FRT's order p.

3. Cross-Correlation Properties between the Intensity FRT and the Finite Rectangular Wave

First, we focused on and analyzed the peak value of the normalized cross-correlation function between the finite rectangular wave and the intensity FRT by changing the FRT's order p. The peak value quantitatively indicates the waveform difference between the two.

Figure 4 depicts the normalized cross-correlation function between the finite rectangular wave shown in **Figure 1** and the intensity FRT shown in **Figure 3**. The peak values of the normalized cross-correlation functions in **Figures 4(a)** and **4(b)** are 0.0522 and 0.312, respectively. It is found that, in the case of $p = 1.0$, the waveform is similar to the finite rectangular wave, though the peak value is very small. On the other hand, in the case of $p = 0.9$, the waveform is not similar to the finite rectangular wave, though the peak value is higher than that in the case of $p = 1.0$. However significant peak cannot be seen in both of the cases.
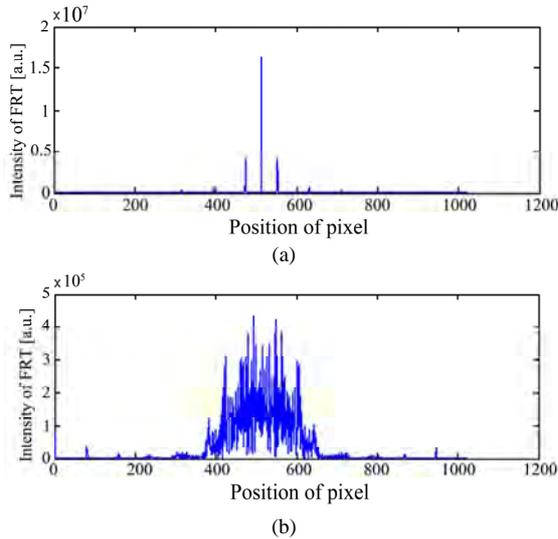
Figure 3. The intensity distributions of the FRTs of the finite rectangular wave shown in Figure 1, when *p*s = (a) 1.0 and (b) 0.9.
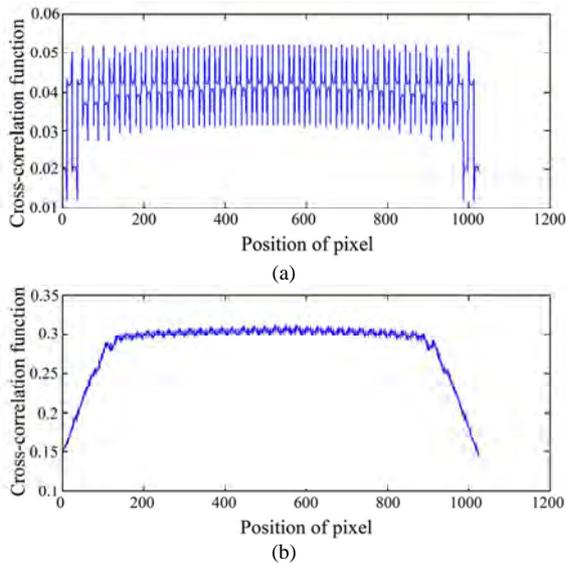
Figure 4. The normalized cross-correlation functions between the finite rectangular wave shown in Figure 1 and the intensity FRT shown in Figure 3, when *p*s = (a) 1.0 and (b) 0.9.

Figure 5 illustrates the peak value of the normalized cross-correlation function by changing the FRT's order *p* from 0.1 to 1.0 by 0.1. As a result, it is understood that the waveform difference between the finite rectangular wave and the registered intensity distribution of the FRT becomes larger with increasing FRT's order *p*. In particular, the intensity FRT with *p*=0.1 is proximate to the finite rectangular wave. However, basically the original fingerprint image is 2D and consisted of multiple scanned lines. Therefore, it is not a problem in our method because
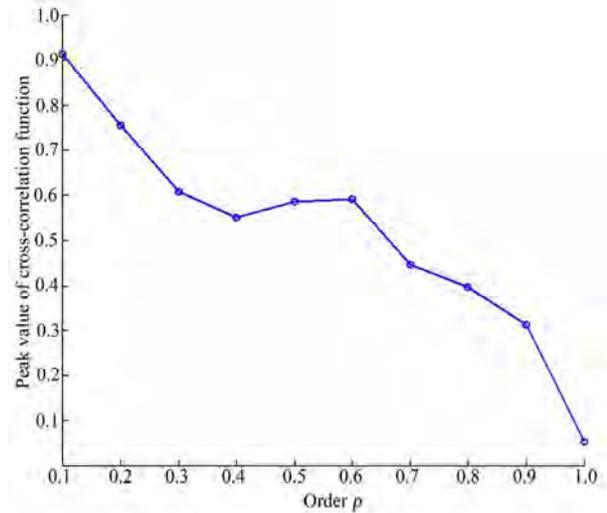
Figure 5. Behavior of the peak value of the cross-correlation function between the intensity FRT and the finite rectangular wave on the FRT's order *p*.

the combination of the various FRT's orders *p*s would be applied in the 2D fingerprint image and the values of the low FRT's order *p* would not be used.

4. Cross-Correlation Properties between the Intensity IFRT and the Finite Rectangular Wave

Next, we focused on and analyzed the peak value of normalized cross-correlation function between the finite rectangular wave shown in **Figure 1** and the intensity distribution of the IFRT (the intensity IFRT) of intensity FRT shown in **Figure 3**, in order to investigate the difficulty to retrieve the original fingerprint image.

Figure 6 depicts examples of the intensity IFRT when the FRT's and IFRT's orders *p*s are 1.0 and 0.9, respectively. From this figure, it is understood that the intensity FRT cannot be decoded to the finite rectangular wave **Figure 7** shows the normalize cross-correlation function between the finite rectangular wave shown in **Figure 1** and the intensity IFRT shown in **Figure 6**. We numerically calculated the normalized cross-correlation function to analyze the difference between them quantitatively. As a result, the peak values of the normalized cross-correlation functions in **Figures 7(a)** and **7(b)** are 0.869 and 0.197, respectively, and the peak value when the order *p* = 0.9 is much smaller than that when the order *p* = 1.0.

The analytical result of the every FRT's order is shown in **Figure 8**. It is found that the intensity FRT can be decoded to closely the finite rectangular wave in the cases of FRT's orders *p*s = 0.1 and 1.0. The peak

(a)

(b)
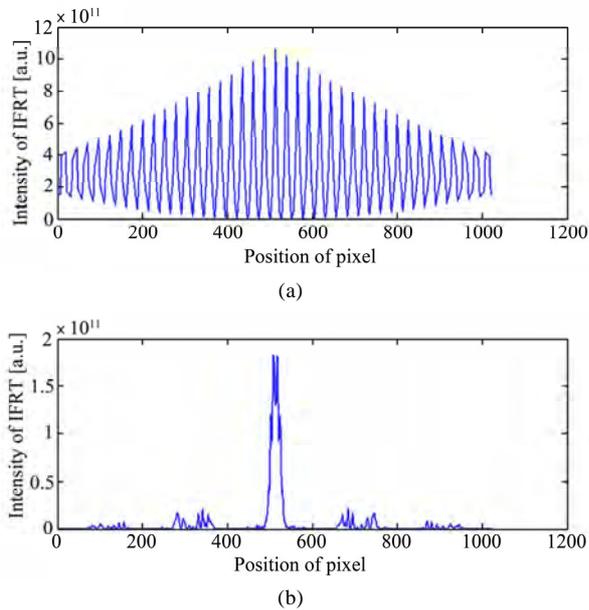
Figure 6. Examples of the intensity IFRT of the intensity FRT shown in Figure 3, when the FRT's and IFRT's orders *p*s = (a) 1.0 and (b) 0.9.

(a)

(b)
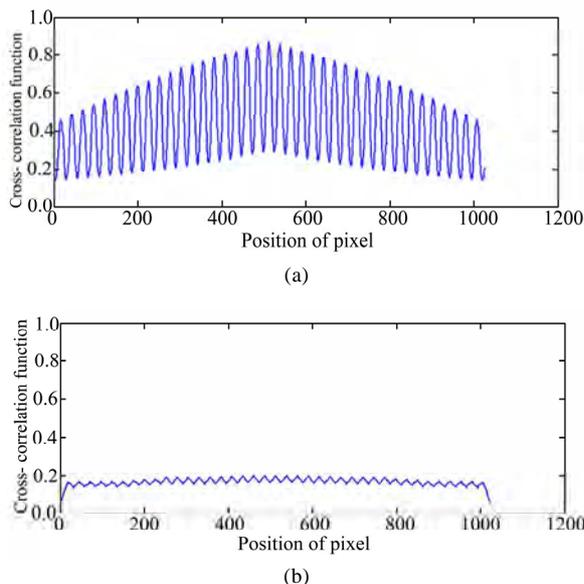
Figure 7. The normalized cross-correlation function between the finite rectangular wave shown in Figure 1 and the intensity IFRT shown in Figure 6, when the FRT's and IFRT's orders *p*s = (a) 1.0 and (b) 0.9.

value of the normalized cross-correlation function when the order *p*=1.0 becomes drastically high as shown in **Figure 8**. It is understood from **Figure 6 (a)** that the intensity FRT in the case of FRT's order *p*=1.0 can be decoded to closely the finite rectangular wave. However, it may not be decoded to the finite rectangular wave in the cases of the other FRT's orders *p*s. As
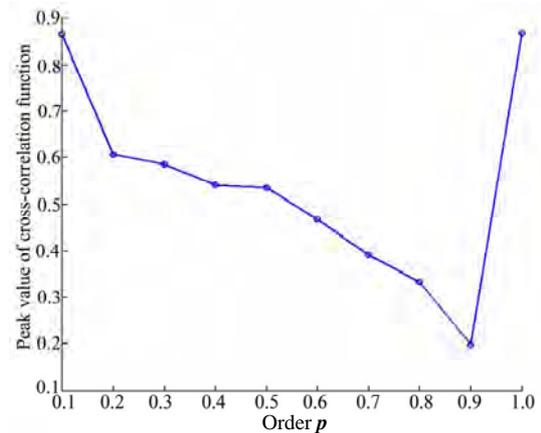

Figure 8. Behavior of the peak value of the normalized cross-correlation functions between the intensity IFRT and the finite rectangular wave when the IFRT's order *p* is the same as the FRT's order *p*.

described previously in Section 3, in our method, the combination of the various FRT's orders *p*s would be applied and the FRT's order *p* = 1.0 and the low FRT's order *p* would not be used. Therefore, it is clarified that the data processed by the FRT would provide higher safety than the cases processed only by the FT.

5. Verification of Robustness

Finally, we verify the robustness of our proposed method from the following two viewpoints: 1) behavior of the second-maximum value of auto-correlation function of the intensity FRT, and 2) behavior of the peak value of the cross-correlation function between the finite rectangular wave and the intensity IFRT (the FRT's and IFRT's orders are different cases).

First, we show the FRT's order cannot be identified from the characteristics of the intensity FRT. Next, we show the intensity FRT cannot be decoded to the finite rectangular wave by using the conventional IFT by unauthorized persons who have no knowledge on the FRT's orders used.

5.1. The Order Dependency on the Second-Maximum Values of the Auto-Correlation Function of the Intensity FRT

First, we focused on and analyzed the second-maximum value of the auto-correlation function of the intensity distribution of the FRT by changing the order *p*, because the peak values of the auto-correlation functions are all 1 and are not distinguishable.

Figure 9 depicts examples of the auto-correlation function of the intensity distribution of the FRT when the FRT's orders *p*s = 1.0 and 0.9 shown in **Figure 3**.

As a result, the second-maximum values in **Figures 9(a)** and **9(b)** are 0.501 and 0.390, respectively.

 Figure 10 illustrates the FRT's order dependency on the second-maximum value of the auto-correlation function by changing the order p. The second maximum peak values for the FRT's orders from 0.1 to 1.0 by 0.1 are 0.180, 0.223, 0.191, 0.179, 0.145, 0.225, 0.299, 0.359, 0.390 and 0.501, respectively. In particular, the second-maximum peak value is the smallest when the FRT's order $p = 0.5$. The reason can be considered that the intensity FRT equally includes both of the characteristics of the spatial information and spatial frequency information of the finite rectangular wave when the FRT's order $p = 0.5$.

 As a result, it is shown that the FRT's order cannot be found from the characteristics of the intensity FRT, because we cannot see noticeable difference in the second-maximum values in any cases of FRT's order p.

(a)

(b)
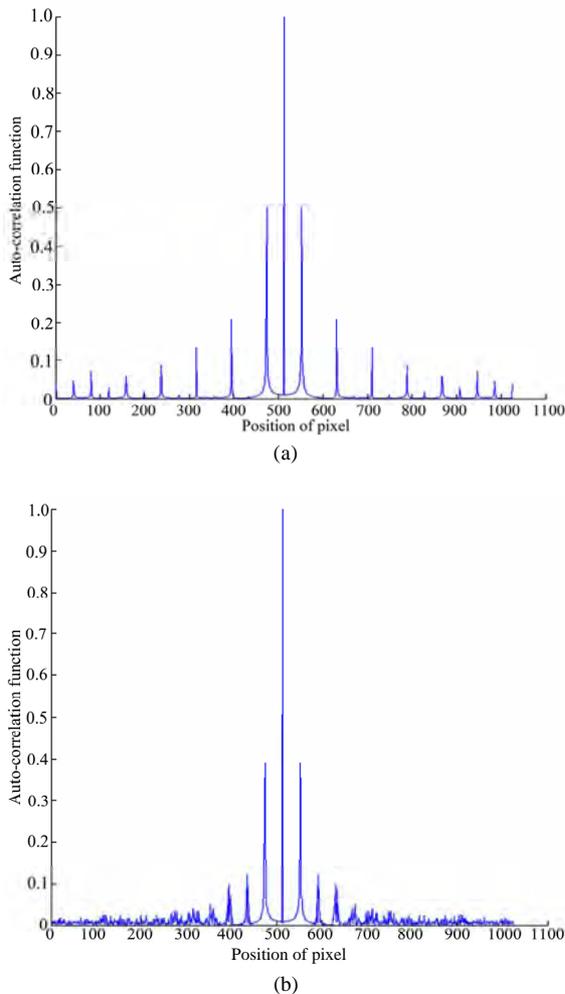
Figure 9. The normalized auto-correlation functions of the intensity FRTs shown in Figure 3, when ps = (a) 1.0 and (b) 0.9.
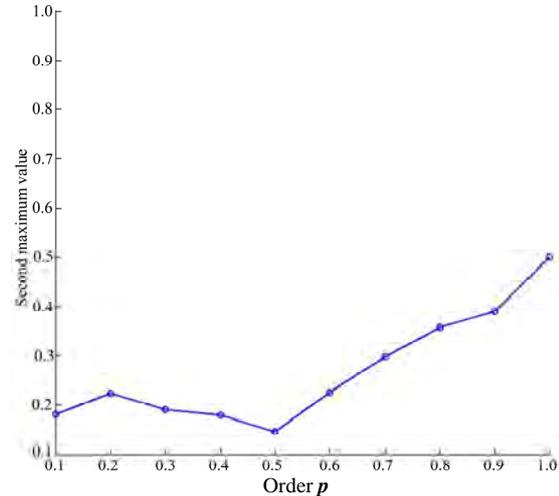
Figure 10. Dependence of the second-maximum value of the normalized auto-correlation functions on the FRT's order p.

5.2. Cross-Correlation Properties between the Intensity IFRT and the Finite Rectangular Wave (The FRT's and the IFRT's Orders are Different Cases)

Next, we focused on and analyzed the peak values of normalized cross-correlation function between the finite rectangular wave and the intensity IFRT of intensity FRT, when the FRT's and IFRT's orders are different from each other. In particular, we fixed the IFRT's order at 1.0 which corresponds to the conventional IFT.

 Figure 11 depicts examples of the intensity IFRT of intensity FRT when the FRT's and IFRT's orders ps are (a) 0.9 and 1.0 and (b) 0.8 and 1.0, respectively. As a result, it is understood that the waveforms are completely different from the finite rectangular wave. **Figure 12** shows the normalized cross-correlation function between the finite rectangular wave shown in **Figure 1** and the intensity IFRT shown in **Figure 11**. As a result, the peak values in **Figures 11(a) and 11(b)** are 0.135 and 0.0579, respectively, and the peak values are very small for both of the FRT's orders ps=0.9 and 0.8.

 The analytical result of every FRT's order is shown in **Figure 13**. The peak values of the FRT's orders from 0.1 to 0.9 by 0.1 are 0.0563, 0.0592, 0.0587, 0.0624, 0.0579, 0.0690, 0.0884, 0.107 and 0.135, respectively, and fully small. The peak value decreases with a decrease in the FRT's order p when p takes a value from 0.9 to 0.5 by 0.1. However, the peak value is almost the same value when p takes a value from 0.5 to 0.1 by 0.1. The reason can be considered that the retrieving effect to the finite rectangular wave by the FT is very small when $p=0.5$ or less because the intensity FRT has little characteristics of the spatial frequency information of the finite rectangular wave. As a result, it is found that the intensity IFRT
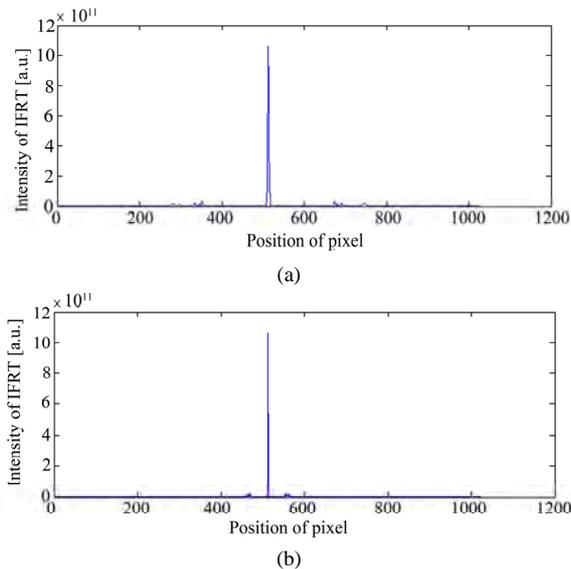
Figure 11. Examples of the intensity IFRT of the intensity FRT shown in Figure 3, when the FRT's and IFRT's orders *p*s = (a) 0.9 and 1.0 and (b) 0.8 and 1.0, respectively.
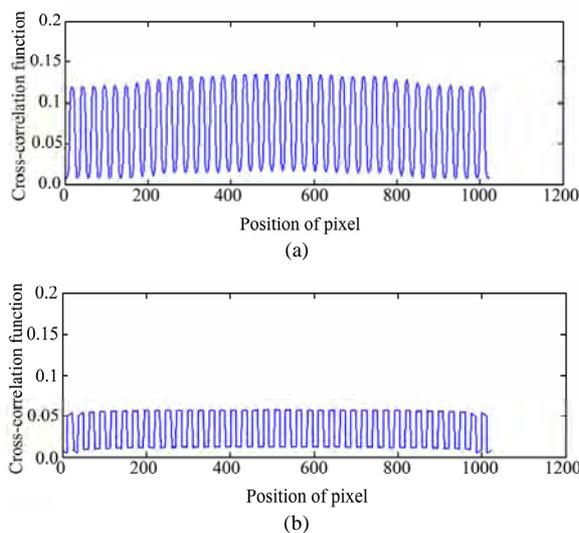

Figure 12. The normalized cross-correlation function between the finite rectangular wave shown in Figure 1 and the intensity IFRT shown in Figure 11, when the FRT's and IFRT's orders *p*s = (a) 0.9 and 1.0 and (b) 0.8 and 1.0, respectively.

cannot be decoded to the finite rectangular wave in all cases of the FRT's orders *p*s.

Moreover, it is understood by comparison between **Figures 8** and **13** that the intensity FRT may not be decoded to the finite rectangular wave for all FRT's orders *p*s when the FRT's and IFRT's orders *p*s are different. Therefore, it is clarified that the data processed by the FRT would provide higher safety than the cases processed only by the FT.
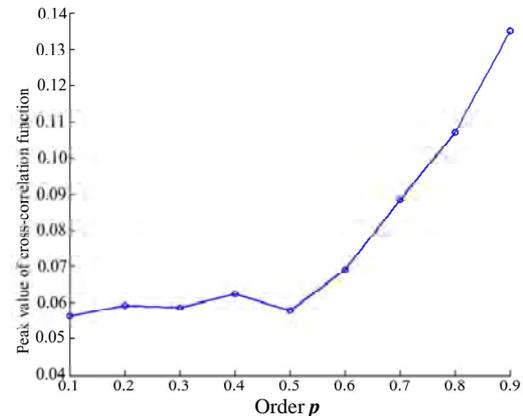
Figure 13. Behavior of the peak value of the normalized cross-correlation function between the finite rectangular wave and the intensity IFRT, when the IFRT's order is fixed at 1.0 and the FRT's orders *p*s are different from 1.0.

From the above-mentioned results about the robustness of image processed by our method and the results shown in **Figure 5** in Section 3 and **Figure 8** in Section 4, we can say that the most suitable single value of the FRT's order *p* is 0.9, though the combination of the various *p* at every scanned line of the raw fingerprint image is significant in our method.

6. Conclusions

In the present study, we proposed a new data processing method for registering the fingerprint image by using the FRT. Moreover, the robustness was examined. In this study, our new method was analyzed in detail by using a modeled fingerprint image instead of the original raw fingerprint image to make the characteristic clear. As a result, it was found that our proposed method can register the fingerprint related data that cannot be easily decoded to the corresponding original fingerprint data by unauthorized persons. In addition, our new method was performed by simple numerical calculation. Furthermore, it was understood that our new method has high robustness and security in the combination of the various FRT's orders. As a further study, we would analyze the accuracy of our method for dust and sebum regarded as noise and hurts regarded as the lost part.

7. References

[1] D. Maltoni, D. Maio, A. K. Jain and S. Prabhakar, "Handbook of Fingerprint Recognition," Springer, New York, 2003.

[2] H. Takeuchi, T. Umezaki, N. Matsumoto and K. Hirabayashi, "Evaluation of Low-Quality Images and Imaging Enhancement Methods for Fingerprint Verification,"

Electronics and Communications in Japan, Vol. 90, No. 10, 2007, pp. 40-53.

[3]   H. M. Ozaktas, Z. Zalevsky and M. A. Kutay, "The Fractional Fourier Transform," John Wiley & Sons, New Jersey, 2001.

[4]   A. W. Lohmann, "Image Rotation, Wigner Rotation, and the Fractional Fourier Transform," *Journal of the Optical Society of America A*, Vol. 10, No. 10, 1993, pp. 2181-2186.

[5]   A Bultheel and H. E. Martinez Sulbaran, "Computation of the Fractional Fourier Transform," *Applied and Computational Harmonic Analysis*, Vol. 16, No. 3, 2004, pp. 182-202.

[6]   A.Bultheel and H. M. Sulbaran, "Computation of the Fractional Fourier Transform". http://nalag.cs. kuleuven. be/research/software/FRFT/

[7]   F. J. Marinho and L. M. Bernardo, "Numerical Calculation of Fractional Fourier Transforms with a Single Fast-Fourier-Transform Algorithm," *Journal of Optical Society of America A*, Vol. 15, No. 8, 1998, pp. 2111-2116.

[8]   X. Yang, Q. Tan, X. Wei, Y. Xiang, Y. Yan and G. Jin, "Improved Fast Fractional-Fourier Transform Algorithm," *Journal of Optical Society of America A*, Vol. 21, No. 9, 2004, pp. 1677-1681.

[9]   D. H. Bailey and P. N. Swarztrauber, "The Fractional Fourier Transform and Applications," *SIAM Review*, Vol. 33, No. 3, 1991, pp. 389-404.

[10] H. Lee, H. Maeng and Y. Bae, "Fake Finger Detection Using the Fractional Fourier Transform," *Lecture Notes in Computer Science*, Heidelberg, Vol. 5707, 2009, pp. 318-324.

◆◆ Scientific
◆◆ Research

Statistical Approach to Mitigating 3G Interference to GPS in 3G Handset

Taher AlSharabati, Yinchao Chen
University of South Carolina Columbia, South Carolina, USA
E-mail: *sharabat@email.sc.edu, CHENYIN@cec.sc.edu*

Abstract

In this paper, we show how statistical decision theory can be used to solve real life product problems. Global Positioning System (GPS) performance in a mobile handset degrades whenever it is simultaneously used with a 3G data or voice service. This degradation is due to the 3G transmitter interference. Mitigation methods to interference in GPS have been proposed. However, most of these methods depend on hardware and signal processing or just hardware solutions. In some cases, it maybe difficult to implement these hardware methods, especially in mobile handsets due to its small size, printed circuit board (PCB) layout issues and the added cost. In this paper, a novel signal processing statistical algorithm approach is proposed to mitigate the 3G interference to GPS. This statistical approach utilizes the knowledge of the statistical characteristics and distributions of both the GPS signal and noise. Then the method utilizes the probabilities to make a statistical decision to remove the effect of noise. This method does not require room on the PCB of the mobile handset and therefore no layout challenges arise. In addition, cost is minimized and the product turn in cycle is shortened. This paper offers theoretical as well as practical insight to the GPS operation during 3G call inside the mobile phone.

1. Introduction

Global Positioning System (GPS) plays a major role in our daily lives. Its applications extend from naval, aerospace, first time emergency responders and location based-finding and directions in cell phones to name a few.

The existence of GPS in cell phones presents its own challenges since it co-exists with other services such as 3G WCDMA and GSM carriers [1].

The spatial distance or proximity of these services and subsystems and their components inside the cell phone is very small. Especially since the mobile handset is being packed with as many services as possible with the smallest form factor. This distance is small enough to the point where it does not present enough attenuation to the 3G transmitted Radio Frequency (RF) signals toward other receivers inside the handset. This, in turn, will present itself as interference to the receivers of the services that exist in the handset especially the GPS receiver. The effect of the distance on the attenuation or path loss of the transmitted signal strength is governed by the equation:

$$A = 32.4 + 10\log(rf)^2 \qquad (1)$$

where A is the attenuation in dB, f is the frequency of the transmitted signal and r is the distance between the transmitter antenna and the GPS receiver.

The spectrum separation or proximity between the frequency bands of the mobile handset wireless services is another source of interference. The wireless services are allocated frequency bands for their transmitters and their receivers. The closer the frequency of a transmitter to the frequency of a receiver, the more of interference the frequency will present. This is due to the non ideal characteristics of the band pass filter of the transmitter [2]. As a matter of fact, some filter designs exhibit "fly back", which is worse insertion loss characteristics especially towards the band edges of the filter. These band edges could be the operating frequencies of another service like the GPS. For example, the GPS receive frequency (called L1 band) is 1575.42 MHz, while the DCS band lowest transmit frequency (or channel) in GSM is

1710 MHz and that of WCDMA band IV is 1712.4 MHz. The frequency separation is less than 135 MHz.

It should be mentioned that architecture and system engineers do a thorough interference lineup analysis and calculations to assess the effects of different scenarios of the effects of interference of each band to itself and others with the given components of the subsystems before or on the onset of any transceiver chipset deployment. The effects of interference, if not mitigated, on receivers and the GPS receiver in particular are severe. These effects could range from degradation in user position accuracy or loss of satellite acquisition. This situation, when it happens to a receiver in a product like mobile phones, is called "de sense" *i.e.* the receiver is "de sensed". In [1], the GPS receiver was de sensed by 5.5 dB. Shown in **Figure 1** is a bar graph of the GPS satellites where eight of them have been acquired based on an acquisition metric of 2.5, where the horizontal axis is the satellite number and the vertical axis is the metric value, (Permission from [3]) when there is no 3G interference added to the data collected in [3]. Any satellite with an acquisition metric of 2.5 and above is considered acquired.

2. Analysis

In the following we present an analysis and examples for calculating two of the important parameters of the GPS receiver, namely; the average signal strength and the root mean squared (RMS) code tracking error.

2.1. Calculating the Average Signal Strength

The average carrier strength (C_{av}) (or C_s) of the acquired satellites can be calculated based on Equation (9) in [4]. Equation (9) is repeated here for clarity purposes for the case $1 < Db < \pi$ and assuming that $\sigma^2 = 0.0332$ (σ^2 is the normalized GPS code tracking error variance in units of $\sec^2$ [5], D is the two sided GPS correlator spacing and

$$b = \beta_r * T_c \qquad (2)$$

b is the normalized bandwidth, assuming $\beta_r = 6MHz$, T_c is the chip period):

From this equation, the carrier to noise power ratio $\frac{C_s}{N_o}$ is:

$$\frac{C_s}{N_o} \mathrm{dB} = 21.67 dB \cdot$$

Assuming system noise figure (NF) of 2dB and thermal noise power (N_o) of -174dB/Hz. From:

$$C_{av} = N_o + NF + C_s/N_o \qquad (3)$$

$C_{av} \sim = -150$ dBm. This is the average signal strength or carrier power of one of the acquired satellites. **Figure 2** shows a plot of the average carrier power of the acquired satellites of **Figure 1**. The horizontal axis is the satellite number, while the vertical axis is the average carrier signal strength in dBm. Note how the values correspond to the metric values in **Figure 1**.

2.2. Calculating the RMS Code Error

One of GPS measurement accuracy parameters is the root mean squared (RMS) code tracking error in units of meters. Let's denote the RMS code tracking error as E then [5],

$$E = \sigma * c \qquad (4)$$

where σ is the standard deviation in units of chips [6] and c is the speed of light. Each chip period is approximately $1\mu s$. For example, for the above case where $\sigma^2 = 0.0332$, $E = 54.66$ m. This is the code tracking error. **Figure 3** displays the RMS code tracking error (in meters) of the acquired satellites. It is evident from this plot and that of **Figure 2** how the average signal power affects this parameter. The stronger the power is; the less the error is and therefore the accurate the user position is.

3. Procedure

3.1. GPS Data and 3G Noise Generation

Being inside the phone, the GPS receiver suffers the most interference from the 3G transmitters inside the User Equipment (UE) while they are transmitting. From this point on, the 3G transmitted signals will be referred to as 3G noise because they present themselves as noise to the GPS receiver. The GPS data used in this paper are data collected in [3], which is data output of the ADC in that platform. The 3G noise will be simulated on the baseband level by using the WCDMA Quadrature Phase Shift Keying (QPSK) modulation scheme which is:

$$\left(\frac{\sigma_\tau^2}{T_c}\right) \cong \frac{B_L(1-0.5B_LT)}{2\frac{C_s}{N_o}}\left[\frac{1}{b}+\frac{b}{\pi-1}(D-\frac{1}{b})^2\right]\left[1+\frac{2}{T\frac{C_s}{N_o}(2-D)}\right], \quad 1 < Db < \pi \quad (9) \text{ in [4]}$$
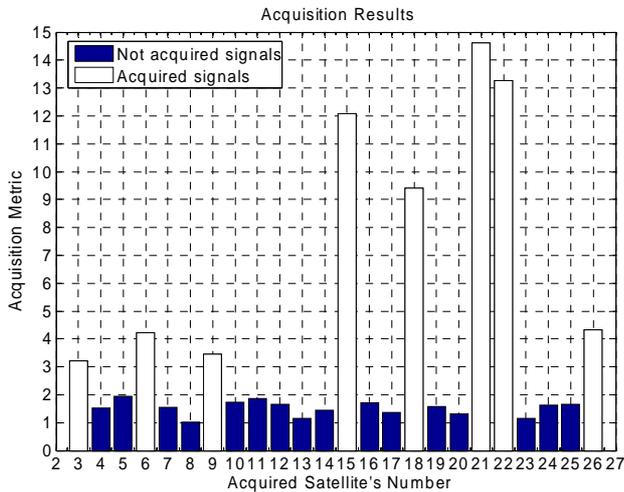
Figure 1. Number of acquired satellites when 3G noise is not present.
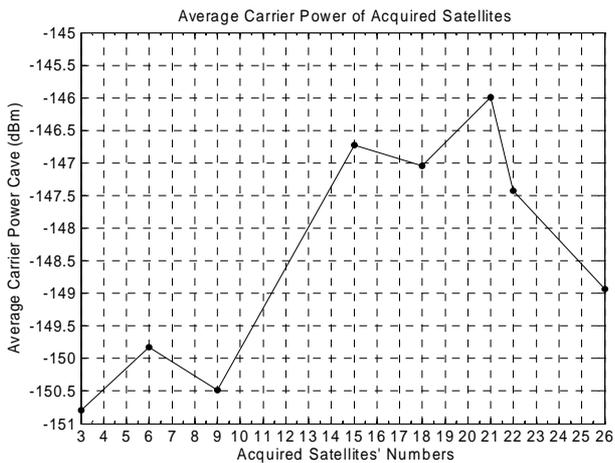
Figure 2. Plot of the average power of the acquired satellites.
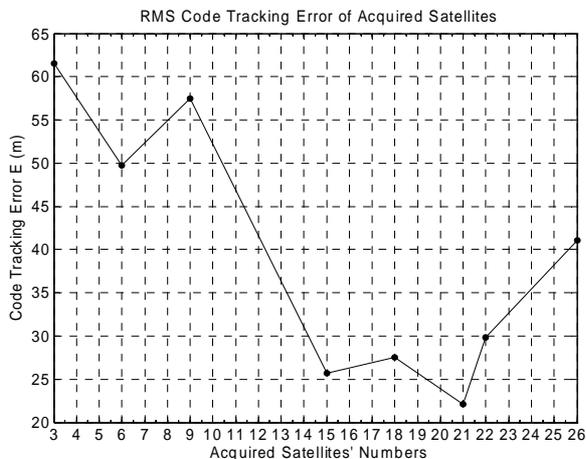
Figure 3. Plot of RMS code tracking error of the acquired satellites.

$$V_{3G}(t) = \frac{Vrms}{\sqrt{2}}(d_I(t)\cos(w_{IF}t + \frac{\pi}{4}) + d_Q(t)\sin(w_{IF}t + \frac{\pi}{4}))$$

$$d_I(t),\ d_Q(t) = \pm 1$$

(5)

where $d_I(t)$, $d_Q(t)$ are the in phase and quadrature phase baseband components of the data streams (which are random and normally distributed with zero mean and variance of one) and w_{IF} is the intermediate frequency of the down converted signal. The 3G noise is sampled at the rate of the sampling frequency of the ADC in [3]. Then the 3G and GPS data are added together to form the GPS plus interference signal.

3.2. Determining 3G Noise Level

This task will take into consideration the total system architecture components of both the 3G transmit (3G TX) section and GPS receive section. **Figure 4** shows a block diagram of a typical GPS-3G coexistence scenario [1]. The GPS section could be the hardware platform in [3]. Going from right to left; the heart of the GPS (GPS RX) receiver is the GPS Radio Frequency Integrated Chip (RFIC). Most of GPS RFICs have on board cross functional sections that span from RF to baseband. The SiGe 4120L [8] is one of SiGe's GPS line of products. It has a low noise amplifier (LNA) on board and it outputs serial IQ data. So, its functionality span from RF to bits. A bandpass filter is used to filter out any signal outside the designed bandwidth of the GPS RX. At the forefront is the GPS antenna and antenna matching circuitry to bring the antenna impedance to a 50-ohm system. On the interfering 3G TX side, the TX-RX 3G antenna accommodates both the WCDMA and GSM carrier bands. Then, the WCDMA/GSM Power Amplifier (PA) along with the duplexer.

Based on typical values of 3GPP WCDMA transmitter noise into GPS, values of 3G noise into GPS [9] (–140 dBm/Hz), duplexer insertion loss and spatial isolation, a value of V_{rms} (3G noise) in Equation (5) was calculated. This value came to be ~0.28 mV. In these calculations (for illustration purposes only), the spatial isolation is assumed to be 15 dB, the total receive line up gain is ~100 dB in 50 ohm system. First the power (P_{dB}) of the interferer was calculated and then it was converted to voltage based on a 50 ohm system:

$$P_{dB} = \text{noise into GPS} + \text{Lineup GaindB} +$$
$$\text{IsolationdB} = -55\ \text{dB}$$

(6)

$$p_{dBw} = pdB - 30$$

(7)

$$P_{dBw} = 10 \times \log(V_{rms}^2 / 50)$$

(8)

$$\therefore V_{rms} = 0.28\ mV$$
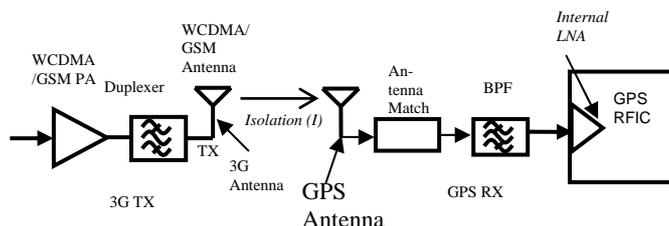
Figure 4. Typical co-existence 3G-GPS scenario.

4. Approach to the Algorithm

Previous published work has focused on suggesting either a combination of signal processing and hardware or just hardware solutions to mitigating the interference. In [7], antenna null steering was suggested to minimize Wide Band (WB) RF interference (RFI). While this method works in products other than cell phones, it is difficult to implement in the cell phone because it requires complex implementation of antenna design, RF front end electronics, feedback and signal processing. This method can not be implemented in the cell phone due to its complexity and immaturity for mobile phone environment. Other methods are proposed to prevent pulsed RFI [7]. While the methods just mentioned aimed at preventing interference from occurring, they do not mitigate the effects of interference after it happens. These methods do not offset the effects of errors generated once the interference is mixed along the line up and down converted to baseband with the desired GPS signal. Instead, a statistical method is proposed to mitigate these errors.

The GPS data output from the ADC is Gaussian distributed with zero mean and therefore their probability distribution function is normal and is in the form of:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma}} \qquad (9)$$

Figure 5 shows that distribution and weights in bar graph of the values and therefore it gives the probability of each value output from the Analog to Digital Converter (ADC). The red trace is the Normal distribution fit of the data. The horizontal axis are the values output from the ADC, the vertical axis are their weights. This figure does not literally give the probability values, but it shows the weights of these values.

5. Algorithm and Results

5.1. Mitigation Algorithm Development

The receiver tries to make a decision or best guess about a symbol s_i given that **r** was received. In other words, the receiver tries to decide the probability of **s** given that **r** was received; $p(\mathbf{s}|\mathbf{r})$ based on the test statistic:

$$T = p(r \mid s_i) p(s_i) \qquad (10)$$

Since the data follow the Gaussian distribution, then for N = 2 [10], where N is the number of dimensions of the signal space:

$$p(r \mid s_i) = \frac{e^{-\frac{(r-s_i)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \qquad (11)$$

and (10) becomes;

$$p(r \mid s_i)p(s_i) = \frac{e^{-\frac{(r-s_i)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} p(s_i) \qquad (12)$$

$p(s_i)$ is the probability of the noise free quantized output GPS signal from the ADC and σ is the standard deviation of the 3G noise. The novelty of this approach is as follows: since the Coarse Acquisition (C/A) code of the satellites is comprised of 1's and –1's (–1 for 0), the prototype messages can be reduced to two levels, namely; 1 and –1 and therefore calculating the probabilities in **Figure 5** reduces to calculating the probabilities of the negative values of the signal and the probabilities of the positive values of the signal. We just need to recover the right phase of the signal. The decision reduces down to deciding whether a 1 or a –1 is being sent. In this case (12) generates two values:

$$p(1 \mid s_i)p(s_i) = \frac{e^{-\frac{(r-s_i)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} p(s_i > 0) \qquad (13)$$

and

$$p(-1 \mid s_i) p(s_i) = \frac{e^{-\frac{(r-s_i)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} p(s_i \leq 0) \qquad (14)$$

The decision goes to the one with the higher statistic.

5.2. Results

To determine user position, the GPS receiver has to acquire at least four satellites. **Figure 6** shows the results of acquisition of GPS data plus 3G noise when no detection algorithm is used [2], namely PRN#: 15, 18, 21. From the results, only three satellites were acquired, not enough to determine user position. While after using the detection algorithm, five satellites (**Figure 7**) are acquired, namely PRN#: 6, 15, 18, 21, and 22. This will

help the GPS receiver in calculating the user position even if the user is using the mobile handset in either a data call or a voice call.

Based on Equations (2), (3) and (10) in [4], values of the average carrier power (C_{av}) were approximated and a comparison plot for three cases was obtained. The first case corresponds to the acquired satellites in **Figure 1** where the results are for the raw data where no 3G noise was added to the data and therefore no detection algorithm was used. The second case corresponds to the ac-

quired satellites in **Figure 6** where 3G noise was added to the data and no detection and estimation algorithm was used. The third case corresponds to the acquired satellites in **Figure 7** where 3G noise was added to the data and the detection and estimation algorithm was used. **Figure 8** shows such a comparison. It is evident from this plot how the proposed detection algorithm and estimation not only recovers the satellites lost due to noise, but it improves the average detected power of those satellites.

By using (4), a comparison was made for the code
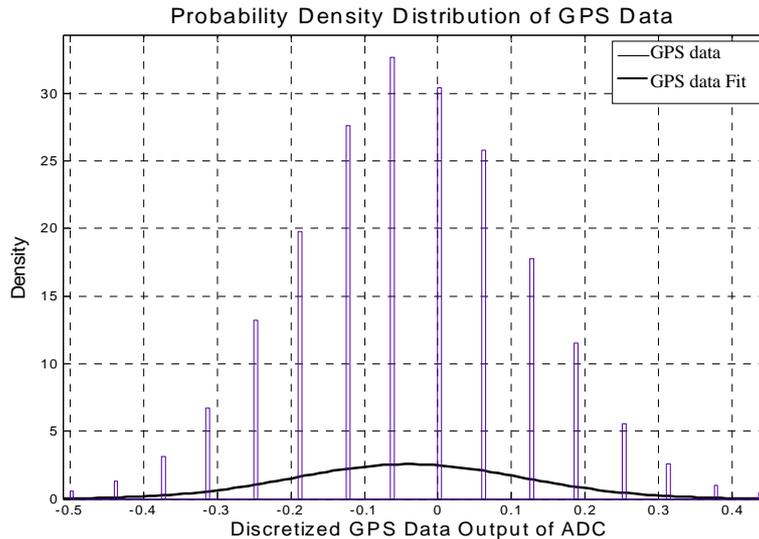
Figure 5. Distribution of GPS data from the ADC in [3].
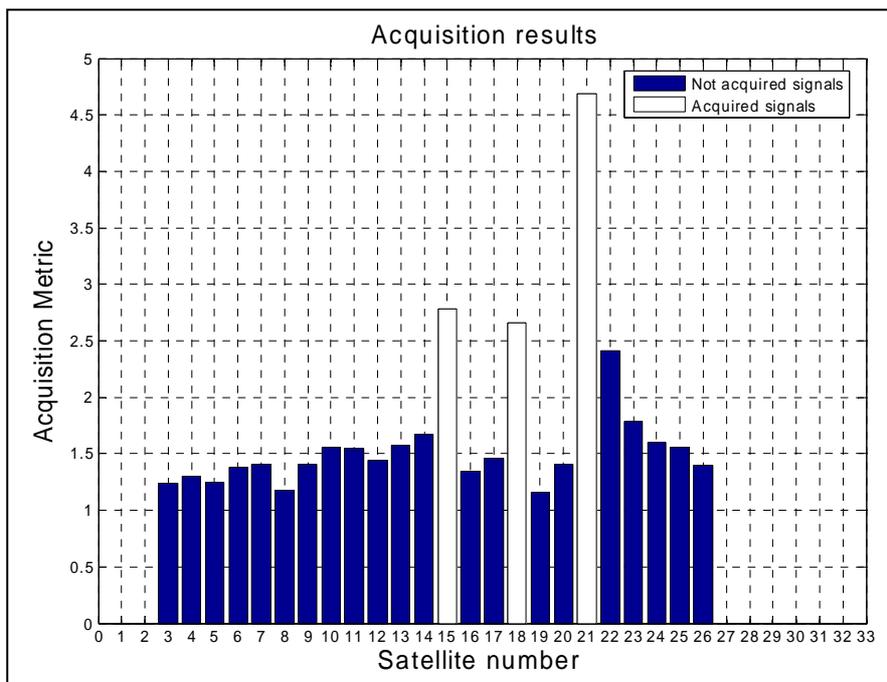
Figure 6. Results of acquisition when no detection used.

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　IJCNS
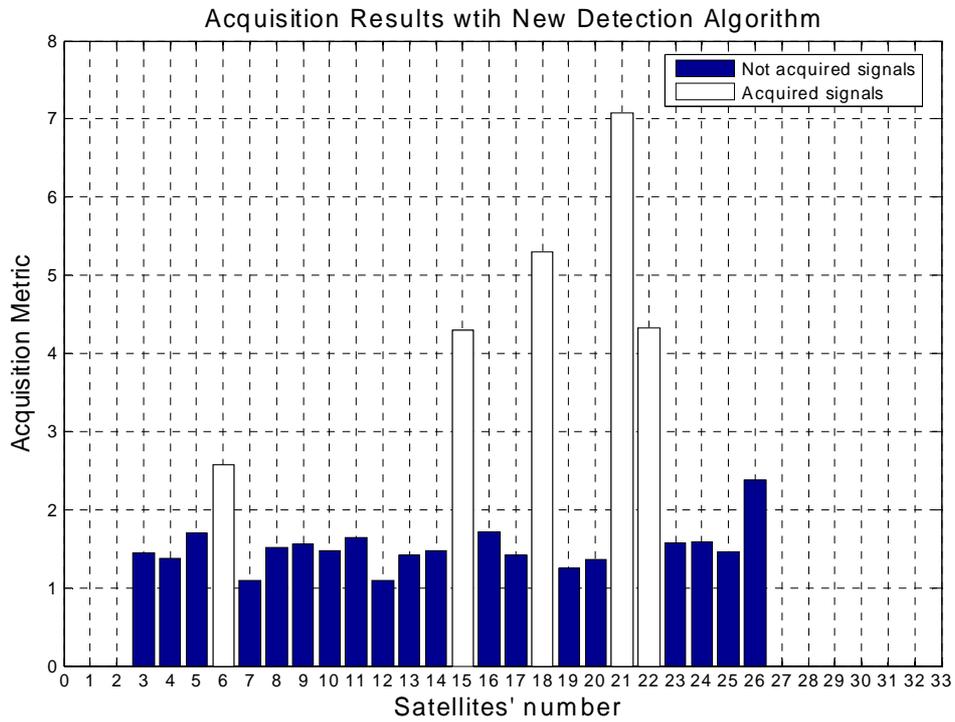
Figure 7. Results of acquisition when using detection algorithm.
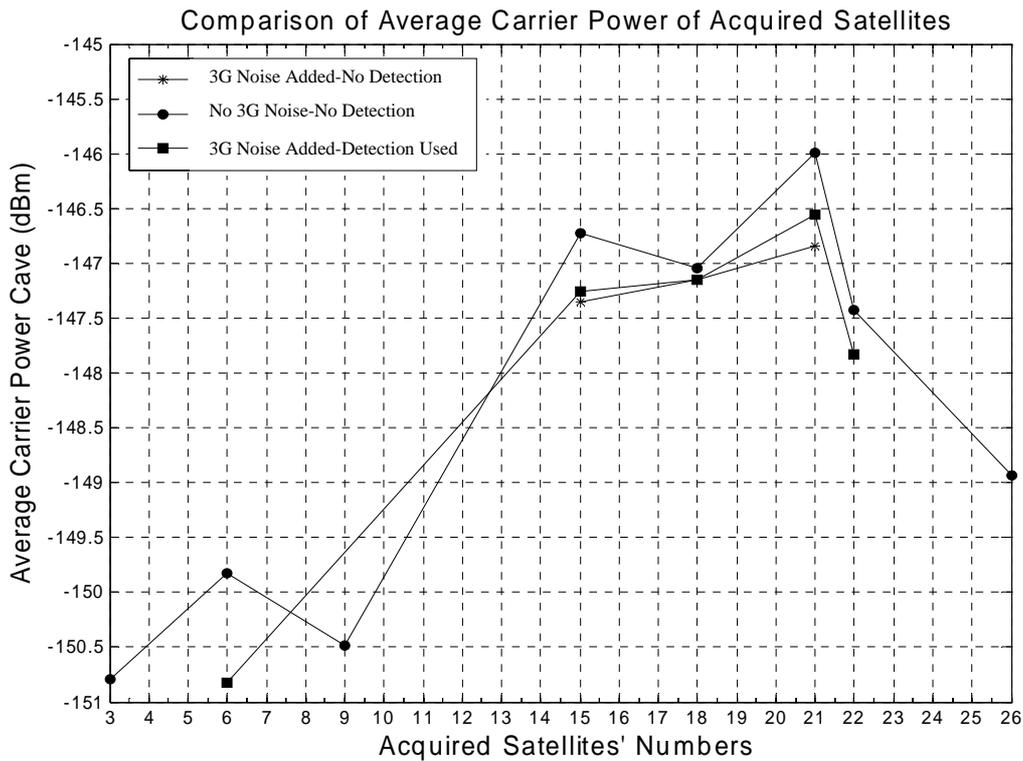
Figure 8. Comparison of average carrier power of acquired satellites when (1) No 3G noise added. (2) 3G noise added-no detection algorithm used. (3) 3G noise added- detection algorithm used.
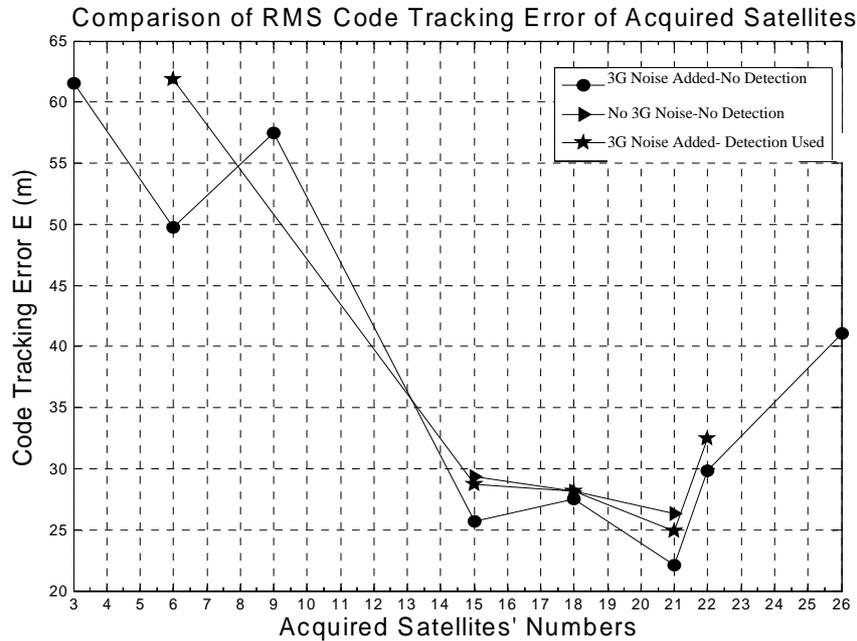
Figure 9. Comparison of RMS code tracking error of acquired satellites when (1) No 3G noise added. (2) 3G noise added-no detection algorithm used. (3) 3G noise added- detection algorithm used.

tracking error for the same three cases as shown in **Figure 9.** The standard deviation was calculated for each tracked satellite. It is seen how the proposed algorithm reduces the error by showing less numbers in meters. This means that the user position is more accurate for the satellites acquired with noise and detection used than under noise without detection algorithm. It should be mentioned that the situation in **Figure 6** is very optimistic. In many cases, the GPS receiver loses track of all satellites and hence no GPS fix can be obtained. The suggested algorithm guarantees to recover most if not all the satellites that are tracked as if the 3G noise is not present.

6. Conclusions

A cross functional effort has been displayed in this paper. The effort spanned from presenting architectural analysis and calculations for a 3G-GPS RF TX-RX system focusing on the interference related aspects and sources, to estimating the average power of the received signal, to laying out the algorithm analysis and derivation of the mitigation scheme. We showed how the mitigation scheme can recover the acquisition of the satellites that were lost due to the interference. Also, the average carrier power and RMS code tracking error of the recovered satellites are improved and therefore the user position accuracy is improved even though the mobile handset is in a 3G data or voice service when using the GPS service. This algorithm could be implemented at the phasing stage of a mobile handset development and manufacturing.

7. References

[1] T. Sharabati and Y. Chen, "Mitigating 3G Carrier Interference to GPS Due to Co-Existence in 3G Handset," *IEEE NAECON*, 2009, pp. 86-91.

[2] T. Sharabati, "Proposal for Mitigating 3G Interferrence to GPS," University of South Carolina, South Carolina, 2009.

[3] K. Borre, D. M. Akos, N. Bertelsen, P. Rinder and S. H. Jensen, "A Software Defined GPS and Galileo Receiver Birkhauser," Birkhauser, Boston, 2007.

[4] W. B. John and R. K. Kevin, "Extended Theory of Early-Late Code Tracking for A Band limited GPS Receiver," *Navigation, Journal of the Institute of Navigation*, Vol. 47, No. 3, 2000, pp. 211-226.

[5] W. B. John, "Effect of Narrowband Interference on GPS Code Tracking Accuracy," ION NTM, 26-28 January 2000, pp. 16-27.

[6] B. Michael and A. J. V. Dierendonck, "GPS Receiver Architecture and Measurements," *Proceedins of the IEEE*, Vol. 87, No. 1, 1999, pp. 48-64.

[7] P. Ward, "Interference Heads Up," *GPS World*, Vol. 19, No. 6, June 2008, pp. 64-73.

[8] S. Semiconductor, "SE4120L GNSS Receiver IC," June 2008.

[9] A. Technologies, "UTMS1700/2100(1710-1755MHz) and UMTS1700 (1750-1785MHz)," *Avago Technologies*, AV02-1907EN, 2009.

[10] W. Couch Leon II, "Digital and Analog Communication Systems," 3rd edition, Collier Macmillan, New York, 1990.

Scientific Research

Fuzzy Integral Based Information Fusion for Water Quality Monitoring Using Remote Sensing Data

Huibin Wang[1], Tanghuai Fan[2], Aiye Shi[1], Fengchen Huang[1], Huimin Wang[3]
[1]*College of Computer and Information Engineering, Hohai University, Nanjing, China*
[2]*Department of Computer Science and Technology, Nanchang Institute of Technology, Nanchang, China*
[3]*State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China*
E-mail: *hhuwhb@gmail.com*

Abstract

To improve the monitoring precision of lake chlorophyll a (Chl-a), this paper presents a fusion method based on Choquet Fuzzy Integral (CFI) to estimate the Chl-a concentration. A group of BPNN models are designed. The output of multiple BPNN model is fused by the CFI. Meanwhile, to resolve the over-fitting problem caused by a small number of training sets, we design an algorithm that fully considers neighbor sampling information. A classification experiment of the Chl-a concentration of the Taihu Lake is conducted. The result shows that, the proposed approach is superior to the classification using a single neural network classifier, and the CFI fusion method has higher identification accuracy.

1. Introduction

The reason of eutrophication is that the excess nutrients accumulation leading to high biological productivity (such as algae, aquatic plants, reeds, fish, plankton) in the water, which includes natural factors and human factors. Eutrophication can cause deterioration of water quality and change the water ecology, and in severe cases can cause the original function of the lake losing, change the lake's ecological environment. Therefore, Lake eutrophication monitoring [1] is an important part of water environment monitoring. However, it is difficult to identify the lake eutrophication distribution information exactly for relying solely on the ground sensors, or by way of point source measurement. It needs to take the advantages of remote sensing, using remote sensing fusion manner through the eutrophication monitoring.

Remote sensing water quality monitoring is to get the distribution of water quality by establishing the qualitative or quantitative relationship between water spectral characteristics and water quality parameters, and the migration and changes of water quality by comparing the information in different time. Lake water quality inversion through combining remote sensing data with the measured data of ground monitoring is the main form of

the present research.

But most inversion models are just single model, only the inversion precision of the monitoring point itself is analyzed, and the inversion precision of the non-monitoring points which are adjacent to the monitoring point have not been fully considered. Therefore, the precision of inversion and robustness are very poor, which makes a great affection to the result of information analysis.

One of the research challenges is how to create an appropriate inversion model for water quality parameters inversion of different regions and different characteristics. Conventional inversion methods include empirical method, semi-empirical method and analysis method [2]. These methods are essentially achieving the water quality inversion via the establishment of linear regression model. However, since the non-linear relationship between remote sensing image data and water quality parameters, the estimation using linear regression and the small complementary between the data will lead to a poor accuracy. Therefore, many scholars conduct water quality inversion by nonlinear regression method. Non-linear regression method can be both explicit and implicit. The parameters of explicit one are often difficult to determine, while the parameters of implicit one avoid the above problems. The implicit regression models commonly used

are neural network regression model [3,4] and genetic algorithm regression model [5]. To improve the accuracy of water quality inversion, some scholars establish the water quality inversion model using other methods combined with the neural network method [6].

However, all of the existing results establish the inverse model by using the internal relations between water quality spectral characteristics and remote sensing image data. In addition, the existing inversion models only analyze the accuracy of the local monitoring point without considering the water quality inversion accuracy of neighboring non-monitoring points.

Eutrophication evaluation commonly uses chlorophyll a (Chl-a) as an indicator [7], the concentration of Chl-a is an important parameter to measure the degree of eutrophication. The method using TM remote sensing data to inverse Chl-a concentration can often be divided into two categories: one is linear estimation model, and the other is non-linear estimation model. The former algorithm is less complex, while the inversion precision of the Chl-a concentration is lower. The latter algorithm is more complex, while the inversion precision of the Chl-a concentration is higher.

This paper focuses on the water environment monitoring information fusion classification of remote sensing combined with ground monitoring data, the monitoring data is classified by a group of water quality classifiers, and the multiple output results of a group of water quality classifiers are fused based on fuzzy integral fusion model to get the Chl-a concentration estimates. This is based on TM remote sensing image information extraction, combining with the monitoring information from ground sensors, to optimize information processing.

2. The Proposed Fuzzy Integral Model for Water Quality Monitoring

Using remote sensing for water quality monitoring and water quality status recognition, essentially, is an uncertain problem. Because it is usually hard to determine the regression model of water quality parameters between remote sensing and ground monitoring, which requires a lot of testing. Meanwhile, the water environmental information is strongly related and has many complex factors. Since it is difficult to extract the thematic information of water quality parameters and status, the monitoring information from only one class of remote image is incomplete, and the monitoring is not very robust. To improve the robustness of water quality recognition, an improved method is used. First, the data fusion process using multi-spectral, radar and other remote sensing image data are conducted, then, the water quality based on fusion results is recognized. However, most recognition

applications are based on a single inversion model structure which causes lower identification accuracy.

This paper designs a new fusion model based on Choquet Fuzzy Integral (CFI) (showed in **Figure 1**) to estimate the Chl-a concentration of Taihu Lake.

This model uses a group of inversion models, firstly, multiple ground monitoring data and remote sensing data n are input into the various inversion models (water quality classification) n, the output is the initial Chl-a concentration classification estimation, then the CFI is used to fuse the outputs which come from different inversion models to get the final result of Chl-a concentration classification estimation. Water quality classification can be composed of semi-empirical methods, neural network, SVM and other models.

The fuzzy integral based fusion model is adopted, mainly considering the fact that the fusion algorithm of water quality classification needs not only the objective results provided by inversion models, but also the importance of inverse models in the fusion process. The Choquet Fuzzy Integral (CFI) can obtain the importance of inverse models by defining the fuzzy measure. Accordingly, the fusion process transforms to the generalized Lebesgue integral of objective results provided by inversion models related to the importance of the corresponding inverse models. As the importance of inversion models has been taken into account, the CFI fusion method can make the results more objective and accurate. In the implementation process, the importance of inversion models is obtained from the remote sensing data of monitoring points and the training samples which consist of Chl-a concentration. The value of importance level is between region [0,1]. The closer to 1 the value is, the more important inversion model is. On the other hand, the smaller value indicates the less importance of the inversion model. Meanwhile, CFI fully considers the impact of various factors. The inversion model which has great impact on one category, but has a small importance level can also affect the fusion results by fuzzy integral.
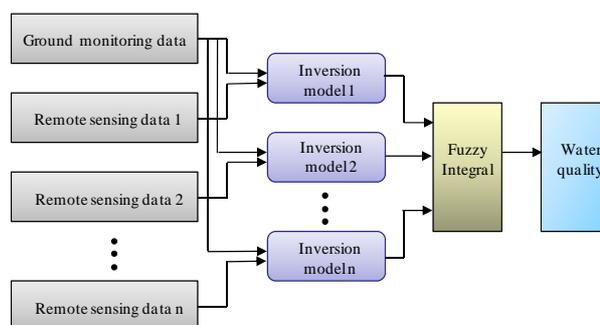

Figure 1. Water quality monitoring model based on fuzzy integral.

3. NN Based Water Quality Classification

Artificial Neural Network (ANN) is an effective non-linear approximation method. We select the back- propagation neural network (BPNN) model as a water quality classifier. Although the BP neural network has some inherent shortcomings, it still has a good generalization performance since it's a global approximation algorithm. Usually, the transfer function used by neurons of BP neural network is sigmoid type differentiable function, which can achieve any nonlinear mapping between the input and output. In this method, all of the inversion models use BP neural network.

This research uses multi-layer BPNN, which is defined as follows: if the network is shown as 3-3-4, that is, there are 3 nodes in the input layer, one hidden layer and also 3 nodes in the hidden layer, and 4 nodes in the output layer, and so the others. The input variables are data from band 1 to band 3 of the Land sat TM image, while the output ones are inversion of the Chl-a concentration category. Each of the BPNN used has three input nodes which have the same value, while it has four output nodes. But it is different for the numbers of hidden layer and its nodes in each BPNN. In addition, each BPNN uses different learning rules, in order to ensure the independence of the BPNN.

The approach uses 11 training sets (through Chl-a concentration category collected on field and the corresponding values of TM radiation). So far, there is no general rule to select the appropriate number of training set, but the more the number and decidability of training set can be, the more promotion neural network has. In general, the number of training set must be twice more than the number of free variables. In the research, as the number of available training set is relatively small, there may be "over-fitting" problem during the network training. To solve that, we consider that it should utilize the near information of each sampling site, shown in **Figure 2**.

Assume that each monitoring point and its four nearest neighbor of Chl-a concentration are in the same category.

(●Representative sampling site, ○Representative neighbor site,

between sampling site and neighbor site about 30 meters)
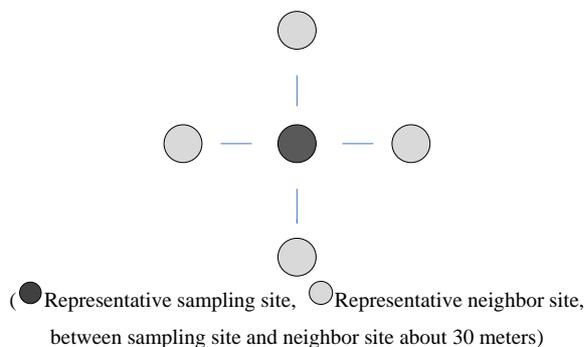
Figure 2. Sampling site and four neighbor sites.

This is because the spatial resolution of the Landsat TM images from band 1 to band 3 is 30m, and correspondingly the distance between its monitoring point of its ground with the same name and its four nearest neighbor point is also 30 meters (Landsat TM images have been registered with the coordinate of ground monitoring point). Due to the short distance, this assumption is reasonable. In the research we select 11 monitoring points, and the nearest 4 points in the neighborhood of each monitoring points, which 55 pairs of training set are composed of. The Chl-a concentration category of 8 monitoring points and their corresponding 4 points in the neighborhood, and the value of radiation corresponding to TM images with the same name are used as the training set (that is, 40 training sets). The Chl-a concentration category of other 3 monitoring points and their corresponding values of TM radiation are used as a test set. We select training set considering its representation in the entire lake region.

The traditional interpolation method of the neural network may result in poor fitting or over-fitting of data. Poor fitting will result in more training error, while the over-fitting would lead to much larger inspection error. Therefore, in the training network, we must make appropriate adjustments to the network's size, in order to make training error similar to testing error.

Network input and output vectors have been normalized to between 0 and 1 so that the input and output maintain a certain dynamic range when the network is training.

4. Fuzzy Integral Based Fusion Method

CFI fuzzy integral is a nonlinear function with the fuzzy measure. The distinction between Fuzzy integral and other examples is that it takes the objective evidence provided by the various sources and the desired value of the sources' subsets into account.

The two most commonly-used fuzzy integrals are Sugeno [9] fuzzy integrals and Choquet [10] fuzzy integrals. The Sugeno fuzzy integral is the nonlinear function defined on the fuzzy measure, and it eliminates the effect of the secondary factors. Compared with the weighted average, it enhances the effect of the main factors, but completely ignores the secondary factors. Choquet fuzzy integral takes various factors into account, in order to avoid the defects of Sugeno fuzzy integral. The application of the Choquet fuzzy integral in multi-source information fusion has been emphasized and widely used [11]. As this paper selects the water quality by fuzzy integral, considering the interaction of the different water quality grades, this article carries out the information fusion by Choquet fuzzy integral.

Fuzzy integral can be interpreted as a fuzzy expectation, or the maximum degree of consistency between two

opposite trends or between objective evidence with the expectation. In this paper, the fuzzy integral is understood as a vague expectation that it expects the high recognition accuracy of the water quality classes by means of the fuzzy integral calculation.

The basic principle of fuzzy integral [12] is as follows:

Suppose that S is a random set, $P(S)$ is the power set of S, If Set Function g meets following conditions:

1) $g(\phi) = 0$, $g(S) = 1$

2) $g(A) \leq g(B), if A \subset B, and A, B \in P(S)$

3) $f A_i \in P(S), and \left\{A_i\right\}_{i=1}^{\infty}$ is monotonous, so

$$\lim_{i \to \infty} g(A_i) = g\left(\lim_{i \to \infty} A_i\right) \quad (1)$$
$$\gamma = 2$$

Then g is one of fuzzy measures on $P(S)$.

According to the above definition, Sugeno introduced g_λ fuzzy measure. It satisfies the following conditions:

$$ig(A \cup B) = g(A) + g(B) + \lambda \, g(A)g(B) \quad (2)$$

To any $\lambda > -1$, $A, B \in P(S)$ and $A \cap B = \phi$, Suppose that S is a set formed by information source, Where, $S = \left\{s_1, s_2, \cdots, s_m\right\}$, And written as $g_i = g_\lambda \left(\{s_i\}\right)$, $g_i \left(i = 1, 2, \cdots, m\right)$ is called as fuzzy-density.

Suppose $A = \left\{s_1, s_2, \cdots, s_m\right\} \subseteq X$, so $g_\lambda(A)$ is defined as:

$$g_\lambda(A) = \sum_{i=1}^{i=m} g_i + \lambda \sum_{i=1}^{m-1} \sum_{j=j+1}^{m} g_i g_j + \cdots + \lambda^{m-1} g_1 g_2 \cdots g_m \quad (3)$$

When $\lambda \neq 0$, according to the property of Fuzzy measure, the value of λ is determined by the following polynomial

$$\lambda + 1 = \prod_{i=1}^{m} (1 + \lambda g_i) \quad (4)$$

For a fixed set $\{g_i\}$, $0 < g_i < 1$, there is exclusive $\lambda \in (-1, +\infty)$ and $\lambda \neq 0$, in conformity with (4).

So, the fuzzy measure is totally decided by its fuzzy-density g_λ. Murofushi and Sugeno proposed Choquet Fuzzy integral, called CFI, showing by (5):

$$\int_s h(s) dg(\bullet) = \sum [h(s_i) - h(s_{i-1})] g(A_i) \quad (5)$$
$$h(s_0) = 0$$

where $A_i = \{s_1, s_2, \cdots, s_i\}$, the order of the $h(\bullet)$ is descending, that is $0 \leq h(s_1) \leq h(s_2) \leq \cdots \leq h(s_n) \leq 1$.

In the new-proposed method, the output of BPNN is fused by CFI. Firstly, for *i*th Water quality classifier, its correct recognition rate to water qualification is g_i. So, the fuzzy measure of every class different water quality classifier can be deduced from (3) and (4).

In order to calculate CFI, we need to make certain belief function h, its value is determined by the output of BPNN. Finally these BPNN are fused. According to (5), the Fuzzy integral values of every water quality classification are calculated. Then the class corresponding to the maxima value is chosen as the correct recognition class.

5. Experimental Results and Analysis

This work is conducted at the Taihu Lake, which is in the Yangtze River Delta in China. The Taihu Lake is the third largest fresh water lake and is the typical plain type shallow lake with its average depth 1.89 m. The area of the Taihu Lake body is 2338.1 km^2. The water body volume probably is 0.47 km^3; the change coefficient of the water body is 1.18. The prevailing wind of Taihu region is southeaster in spring and summer, the concentration of Chl-a is a little high in its west and northwest region; additionally, the pollution of industry, household garbage and the circle net catches fish make the Chl-a concentration nearby the lake is comparably high, too.

5.1. Data Sampling and Processing

The data of TM is the most used multi-spectral remote sensing data in inland water quality monitoring. The unit of TM data is scene, a scenery data, which compared to the earth area, is the area of 185 km×185 km, and space resolution is 30 m. The data of every scene is decided by the satellite orbit number and central latitude, the system, using the orbit and coordinate to decide the scene center, is called Global index system.

Data usually offer users with CCT. Each unit of data (which called the pixels) records reflective luminance on the ground reflection area in each band, that are almost the same as the resolution of sensors, the quantitative series are 256. The parameters are shown in **Table 1**.

Locate direction about representative's point of research region respectively by using GPS, and measured transparency and reflective spectra of water in each point simultaneously. Meanwhile, in order to ensure the reliability of the water quality research, the ground water quality parameters acquisition time by sensors, should consistent with the TM remote sensing data acquisition time.

Due to the large surface of Taihu Lake, the width and spatial resolution of a set of data from the TM are sufficient to cover the entire region. The TM data is obtained on 1997-05-04; its pseudo color composition diagram is shown in **Figure 3**.

Table 1. The observation parameters of landsat -4/5.

Wave band	Wavelength (μm)	Calibration of spectral regions	spatial resolution (m)
1	0.45~0.52	blue	30
2	0.52~0.60	green	30
3	0.63~0.69	red	30
4	0.76~0.90	near infrared	30
5	1.55~1.75	mid-infrared	30
6	10.4~12.5	infrared	60
7	2.08~2.35	mid-infrared	30
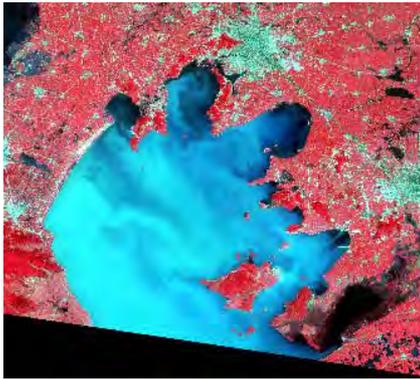
Figure 3. The 432 band Color composition diagram of Taihu Lake (on 4 May,4,1997).

Before estimating the Chl-a concentration, we should made atmospheric correction of the TM data. Here the Dark-Object methods is used, it is assumed that the radiation received by TM sensors including the water-leaving radiance and the radiation caused by atmospheric effects. If there is no atmospheric interference, then the gray value of clean water (the general location is mid-lake) in the near infrared band (the forth band) image should be set to 0, otherwise, we assume the value is n4, it indicates that the TM data suffered atmospheric interference, so the forth band value of atmospheric correction is radiation value subtract n4. Correspondingly, the 1, 2, and 3 bands then made atmospheric correction on the basis of n4 (generally according to the specific image to adjust).

Furthermore, the ground space coordinate and the image coordinate should be aligning. This article use Map-Info7.0 to align coordinate. **Figure 4** is composition diagram of the original TM image and ground monitoring point after alignment.

As shown in **Figure 5**, there is 11 monitoring points in Taihu Lake in May 1997. **Table 2** shows the Chl-a concentration value of the sampling points.

When we use TM multi-spectral remote sensing data to study lake water quality, the band of choice is very
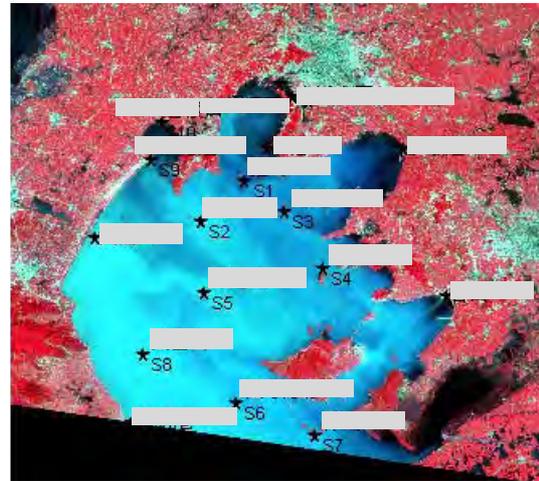
Figure 4. The effect picture of monitoring point and remote sensing image after alignment.
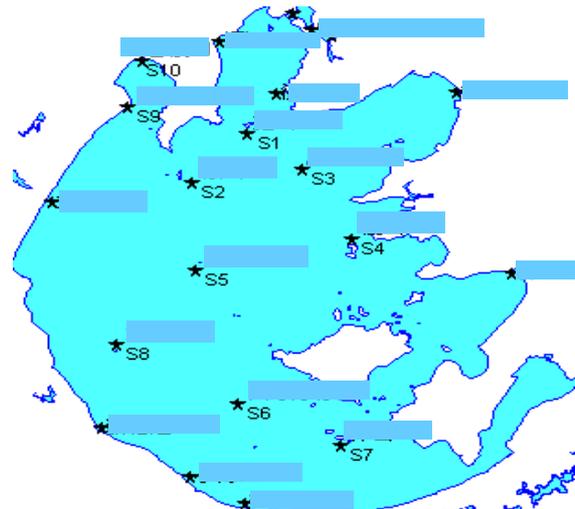
Figure 5. The distribution of monitoring points in Tai hu Lake.

Table 2. The concentration value of Chl-a of the sampling points in Taihu lake.

Sampling points	Concentration value of Chl-a (μg/L)
S1	39
S2	22
S3	16
S4	17
S5	13
S6	16
S7	17
S8	8
S9	22
S10	47
S11	16

important. Related research shows that:

In the range of 400-500 nm (corresponding to the spectral range of TM1), Since the absorption peak of Chl-a in the blue band and strong absorption of yellow substance in the range, reflectivity of water is low; but as the impact of suspended solids, Chl-a absorption peak at 440nm is not very obvious.

In the range of 510-620 nm (corresponding to the spectral range of TM2), the reflection peak is due to weak absorption of chlorophyll, carotene and scattering role of cells and suspended particles. The reflection peak is related to the composition of the pigment, the higher the concentration of chlorophyll in the water is, the higher the peak of the radiation is, and then it can be used as quantitative indicators of chlorophyll.

The low valley of 630 nm is aroused by the absorption of phycocyanin. 675 nm is another absorption peak of Chl-a. So when algal density is high, the spectral reflectance curve of water will be in the valley there.

Based on the above analysis, this chapter uses the first three-band Landsat TM to study lake water quality.

5.2. Results and Analysis

1) Experiment 1: The single SVM inversion model for Chl-a concentration in Taihu Lake Classification and Analysis of Experiments.

SVM is based on statistical learning theory. It can find the best compromise between learning ability and the model complexity based on a limited sample of the information, in order to obtain the best generalization ability (prediction accuracy). It shows many unique advantages when addressing the small sample, non-linear and high dimensional study.

The kernel function by SVM chosed is radial basis function which is commonly used, namely:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \qquad (6)$$

Parameters γ and penalty coefficient C in SVM model is determined by training the test. By selecting different parameter values several times in the training process, and reviewing the model predictions results about the training sample and the test samples, the optimal model parameters will be fond. After screening, best SVM model parameters is $\gamma = 2$ and penalty coefficient is C = 100, using 6-Fold cross-training.

Table 3 shows the category of Chl-a concentration, which are divided into four categories: C1, C2, C3 and C4. Where, C1, C2, C3 and C4 are coded as (0.1 0.1 0.1 0.9) ^{T}, (0.1 0.1 0.9 0.1) ^{T}, (0.1 0.9 0.1 0.1) ^{T}, (0.9 0.1 0.1 0.1) ^{T}, T denotes the transpose.

According to **Table 3**, all the concentration classes of monitoring points are shown in **Table 4**.

Table 5 and **Table 6** show the inversion results of train-

Table 3. Chl-a concentration category identification.

Classification	C1	C2	C3	C4
Chl-a concentration （μg/L）	≤10	≤20	≤40	>40

Table 4. Chl-a concentration classification of the sampling points.

Sampling Points	Classification
S1	C3
S2	C3
S3	C2
S4	C2
S5	C2
S6	C2
S7	C2
S8	C1
S9	C3
S10	C4
S11	C2

Table 5. SVM training set classification results.

Monitoring points	SVM classification	Correct category	The overall classification accuracy
S1	C3	C3	
S4	C2	C2	
S5	C2	C2	
S6	C2	C2	
S7	C2	C2	87.5%
S8	C2	C1	
S10	C4	C4	
S11	C2	C2	

Table 6. SVM Test Set Classification Results.

Monitoring points	SVM classification	Correct category	The overall classification accuracy
S2	C2	C3	
S3	C2	C2	33.3%
S9	C2	C3	

ing samples.

Table 5 and **Table 6** show that SVM classification accuracy on the training samples is high (87.5%), but its classification accuracy on the validation samples is low, only 33.3%. This is mainly because the type of target uniformity of the sample is insufficient (poor reparability).

In 11 samples sets, 54.55% is C2. Although the SVM is suitable for fitting a small sample data, but its generalization performance is also poor because of the poor divisibility of target category. In addition, the encoding for the target sample of SVM classifier is inferior to neural networks. Such as for the class of the target sample m, encoded is the only positive integer m, which also makes the category fitting degree of water quality parameters lower in the case of too few training samples using SVM method. The result is show as in **Figure 6**.

2) Experiment 2: We use fuzzy integral proposed in this paper combining with neural network model to compare and analyze the Chl-a concentration category in Taihu Lake. Besides, it is compared with the single neural network model.

In this research, the training sets include Chl-a concentration category of eight monitoring points: S1, S4, S5, S6, S7, S8, S10, S11 and their four neighbor points and the corresponding TM image radiance. Test set is constituted with The Chl-a concentration classification of other three monitoring points and corresponding TM image radiation.

Three BPNN used to generate the initial value of Chl-a concentration category. To ensure the independence of three BPNN, the first BPNN structure is 3-4-4, the second BPNN structure is 3-5-4, and the third BPNN structure is 3-3-4-4. The three BPNN input vectors are the value of atmospheric correction radiance in band TM1, TM2, TM3. In addition, the learning rules of each BPNN are different. The first BPNN is momentum of back propagation gradient descent; second BPNN is decreasing gradient adaptive learning rate back propagation; third using Bayesian back propagation rule. The expected output of BPNN is the Chl-a concentration category.

All of the BPNN is trained using the MATLAB software. Test set for each BPNN model results and the results of CFI are shown in **Table 7**.

As can be seen from **Table 7**, the classification obtained using the CFI model results better than the single BPNN classifier. Although each BPNN can give the correct classification to the test set, while CFI gives better classification of test results.

For sampling point S2, correct classification is given by BPNN according to the principle of selecting the maximum, However, for the reason that the confidence values of class 3 (C3) and class 4 (C4) (outputs of BPNN) are similar (confidence values of C3 and C4 are 0.534 and 0.526 respectively), it is not easy to distinguish the two classes. Moreover, by using model CFI, not only can we distinguish sampling point S3 well, but also can distinguish C3 from other categories.

Figure 7 shows the Chl-a concentration distribution produced by the model this paper proposed. As is shown

Figure 6. SVM method of inversion of the Taihu Lake water quality categories.
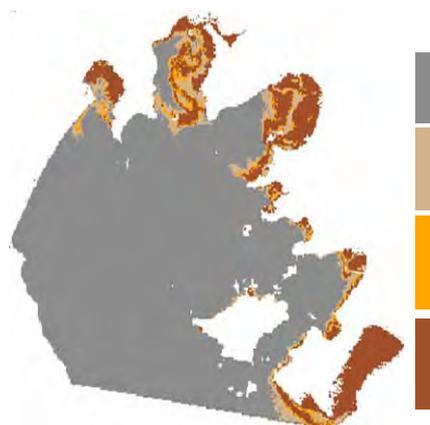
Figure 7. Chl-a concentration distribution of TaiHu Lake.

in **Figure 7**, the highest concentration of Chl-a are in the northern area, northeast corner, southwest corner, and the whole eastern coastal area of Taihu Lake, while the value of Chl-a concentration is 0-10μg/L in central southwest area.

From the comparison by **Figure 6** and **Figure 7**, there is a great difference between the water quality category of SVM inversion and the new method. When using SVM method, there are only water C2, C3 and C4 in Taihu Lake, by comparison, C1, C2, C3, C4 all exist when using the new method. From concentration of Chl-a of monitoring sites, The case of four classes of water all existing in Taihu Lake, is consistent with the actual situation. It shows that, for TM data of 1997, the category precision of water quality inversion using new method is superior to using a single inversion model (such as SVM).

6. Conclusions

For water environment monitoring based on remote

Table 7. Outputs of CFI, BPNN and results of Chl-a concentration classification.

Monitoring Sites	S2				S3				S9			
Classification	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
BPNN1	0.151	0.282	0.534	0.526	0.231	0.672	0.255	0.415	0.376	0.124	0.781	0.257
BPNN2	0.243	0.355	0.796	0.325	0.217	0.725	0.326	0.321	0.215	0.236	0.819	0.208
BPNN3	0.226	0.324	0.807	0.365	0.354	0.747	0.317	0.318	0.326	0.245	0.834	0.235
CFI	0.235	0.354	0.810	0.495	0.324	0.753	0.315	0.374	0.350	0.242	0.859	0.241

sensing, in order to improve the monitoring accuracy of lakes chlorophyll-a, the method for estimating Chl-a concentration using fuzzy integral based fusion model is proposed in the paper. The estimation of Chl-a concentration classification is obtained by fusing several outputs of retrieval models using CFI, while the classifier is composed of multiple BPNNs. Neighbor information of sampling points is fully considered to solve the problem of small number training sets. The experiments of Chl-a concentration classification show that the proposed method is superior to single neural network classifier or individual SVM classifier. In addition, the fusion model can also be extended to classify other water quality parameters.

7. Acknowledgements

8. References

[1] P. E. Greeson, "Lake Eutrophication—A Natural Process," *JAWRA Journals of the American Water Resources Association*, Vol. 5, No. 4, 2007, pp. 16-30.

[2] Q. Yin, C. Gong, D. Kuang, N. Zhou, Y. Hu, *et al.*, "Method of Satelite Remote Sensing of Lake Water Quality and Its Applications," *Journals of Infrared and Millimeter Waves* (in Chinese), Vol. 24, No. 3, 2005, pp. 198-202.

[3] G. Hanrahan, "Modelling of Pollutants in Complex Environmental Systems", ILM Publications, St. Albans, 2009.

[4] R. O. Strobl, F. Forte and L. Pennetta, "Application of Artificial Neural Networks for Classifying Lake Eutrophication Status," *Lakes and Reservoirs: Research and Management*, Vol. 12, No. 1, 2007, pp. 15-25.

[5] L. Chen, "A Study of Applying Genetic Programming to Reservoir Trophic State Evaluation Using Remote Sensor Data," *International Journals of Remote Sensing*, Vol. 24, No. 11, 2003, pp. 2265-2276.

[6] A. Shi, T. Fan, L. Xu and J. Zhou, "A Fuzzy Integral Model for Estimating Chlorophyll Concentrations in Tai Lake from Thematic Mapper Imagery," *Proceedings of IEEE International Conference on Information Acquisition*, Shandong, China, 2006, pp. 1117-1121.

[7] B. Zhang and J. L. Liu, "Evaluation Method for Lake Eutro-Phication Influence and Public Satisfaction," *Advances in Water Science*, Vol. 20, No. 5, 2009, pp. 695-700.

[8] D. E. Rumelhart, J. L. McClelland and the PDP research group, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition," MIT Press, Cambridge, 1986.

[9] M. Sugeno, "Theory of Fuzzy Integrals and Its Applications", Tokyo Institute of Technology Tokyo, Japan, 1974.

[10] J. H. Chiang, "Aggregating Membership Values by a Choquet-Fuzzy-Integral Based Operator," *Fuzzy Sets and Systems*, Vol. 114, No. 3, 2000, pp. 367-375.

[11] S. Auephanwiriyakul, J. Keller and P. Gader, "Generalized Choquet Fuzzy Integral Fusion," *Information Fusion*, Vol. 3, No. 1, 2002, pp. 69-85.

[12] L. Yang and Y. Gao, "Principle and Application of Fuzzy Mathematics," South China University of Technology Press, Guangzhou, China, 2001.

[13] H. Duan, L. Yu, B. Zhang, D. Liu, K. Song and Z. Wang, "Hyperspectral Remotesensing of Chlorophyll-a in the Chagan Lake," *Environmental Science*, Vol. 27, No. 3, 2006, pp. 503-507.

Scientific Research

Design and Implementation on the P2P VOD System Based on the RF-IPS Algorithm

Zhiyi Qu, Lili Li, Jianxu Li

School of Information Science & Engineering, Lanzhou University; Information Management Department, Lanzhou Jiaotong University bowen college, Lanzhou, China
E-mail: quzy@lzu.edu.cn, lily402@126.com

Abstract

This paper presents a P2P VOD system based on the RF-IPS algorithm, and this system consists of two parts, server-side and clients. We put emphasis on the piece selection module of server-side, and in this part we apply the RF-IPS algorithm proposed in the past. In the piece selection module, we also proposed a weighting function except the RF-IPS algorithm. Using this function we will find the best server peers effectively and quickly, it will enhance the performance of this system. We carry on experiments on RF-IPS algorithm, and the results show this algorithm is efficient. Overall, we believe that the RF-IPS algorithm is feasible to further improve system performance and provide a cost-effective P2P VOD service with acceptable user experience.

Keywords: VOD, P2P, RF-IPS, Piece Selection

1. Introduction

Video-on-Demand (VOD) has becoming one of the most important streaming service and catches many people's research interest. VOD connection is an active connection between the client and the server or the client and the client, and people can select the video resources as they pleased at any time and can do some VCR operations, such as FORWARD, BACKWARD and REWIND [1]. In the traditional VOD system, users only download video resource from the server, and the server becomes the bottleneck of the system when the users reach a certain numbers. Once the server failure, the whole system would be paralyzed. But in the mode of P2P VOD technology, video resource is uploaded or downloaded between the clients, that are peer nodes, and it is effective to avoid the bottleneck of the system. As the P2P network can provide the basic network support for the large-scale VOD applications, many of the emerging P2P VOD distribution strategies are able to provide the most basic way of data transmission in the P2P network. Described in this paper based on the RF-IPS algorithm of P2P VOD system also has a server part, but it is not the traditional sense of the server, just as a super-node has more service capabilities. The client is the peer node in the P2P network, and it can be the family PC, laptop computer and smart cell phones

and so on [2].

2. Related Work

VOD system is on-demand video system and a major application of streaming technology. It is great to challenge the traditional video services, where the users can only receive the broadcast programs passively, and can not afford the program content [3]. But somewhat differently, the VOD service system equipped with a large video library, and users can browse the video directory and choose one of the programs on demand at any time, while the system plays the corresponding broadcast program based on the users' requests. Traditional VOD system consists of server system, network system and client system.

2.1. Server System

Server system consists of three modules, service module [4], application service module and management module. The media service module is mainly used to provide users with multimedia and service information. The application service module is mainly used to accept users' requests and download a variety of user terminals, achieve the data exchange with external network and

offer users access interface. The management module is responsible for the running of the whole system and record the relevant information. In addition [5], server system also includes a Web-based management system, where administrators can manage the system via any terminals after certifications in any location.

2.2. Network System

Network system is responsible for the real-time transmission of a stream of video information, and it is a very critical part to affect the performance of continuous media web services. Including some network transmission equipments, you can use the mode of LAN or WAN [6]. Also include the QoS of the network and some application layer protocols. If only reasonably choose the available network protocols and make the network bandwidth and latency time to maintain a certain level, the video data can smoothly playback on the client node.

2.3. Client System

Client system mainly includes some user terminals, achieved by personal computers and special set-top boxes. High performance personal computer can be achieved by loading software [7]. Lower performance personal computers can be achieved by Decompression Card to complete the decoding. We can watch TV by using a special set-top box.

This system is based on the C/S mode [8], and the server connects with each user by the way of unicast. The P2P VOD technology is that the media server sends only a few media flow to several users connected firstly. Users connected later get data from the previous users but not from the server directly. In the P2P VOD system, the function of the server is no longer a traditional server but a peer with better service abilities. It is just achieved to guide the clients to download, maintain the information of the users and the video.

3. Design and Implementation of the P2P VOD System

3.1. Structure Design and Implementation

This system consists of two parts, server and client. The server is not in the traditional sense, can be understood as a super node in the P2P network, and is mainly responsible for maintaining the users' state, updating the resource list, maintaining the hot video resource, piece selection and guiding the client to download resource. The client is not in the traditional sense either, and it is the peer in the P2P network, and is responsible for uploading and downloading resource, maintaining the media buffer, and

searching and classifying some information. **Figure 1** shows the structure of this system.

3.2. Client Modules

The client in the P2P network has dual roles, is both the video provider and the video receiver. However, the function of cache management should be considered when we designed, and ensure to upload the existing resource and download and play the needed resource. The following is the proceeding and handling methods for each module, and include buffer management module, P2P network module and video playback module. The relationship among these modules is complementary as **Figure 2.** The P2P network module is achieved to the message communication and data transmission between server and client, or among the clients. The buffer management module receives the data from P2P network module and manages the data effectively.

3.3. Server Modules

The server can be divided into two parts, Index Server (IS) and Entry Server (ES). The IS part consists of four main function modules, including User Information Maintain module, Resource List Maintain module, Video and User Synchronize module and the Expired Resource Deletion module. The ES part also consists of four main function modules, including Hot Resource Maintain module, Resource Transmission module, Buffer Management module and Piece Selection module. **Figure 3** shows the processing modules between the two servers.

3.4. Piece Selection Modules

1) The function of piece selection module

The main function of piece selection module is to receive the ID information of requests peers, and calculate
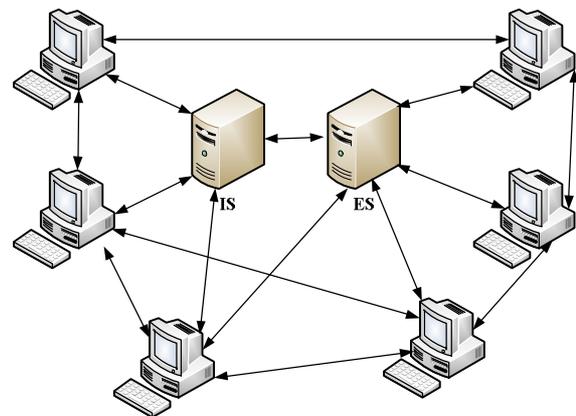

**Figure 1. The structure of the P2P VOD system.**
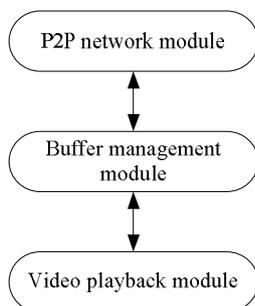
**Figure 2. The relationship among three modules.**
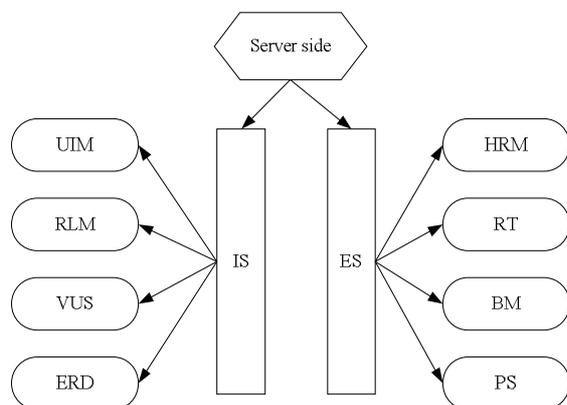
Figure 3. The processing between the two servers.

the piece ID and its download address based on the RF-IPS algorithm for this peer. In order to minimize the users' processing, these functions are both implemented on the server. In this module, we adopt the RF-IPS algorithm to choose the server peers. **Figure 4** show the main function of this module

There are a variety of peers with different performances distributed in the P2P network. When many Peers owned the video resource users needed at the same time, how to choose the server peers becomes one of the functions of this module. Specific methods are as follows:

a) Scan table Res_ID_Peers and table Res_ID_Pieces, and get the list of user information owned the piece. Table Res_ID_Peers is used to save the information of the users owned the relevant resource and table Res_ID_Pieces is used to save the information of the resource.

b) According to the weighting function, get the first numbers of users owned the maximum weight.

The weighting function is as follows:

$$f(a, b, c, d) = k_1 \times a + k_2 \times b + k_3 \times c - k_4 \times d$$

In this function, a is relevant to user's IP, b is relevant to the user's level, c is relevant to user whether to be mobile, d is relevant to the user's maximum weight, and k_1, k_2, k_3, k_4 are weights.

c) Return the piece number and the list of download address to the client.
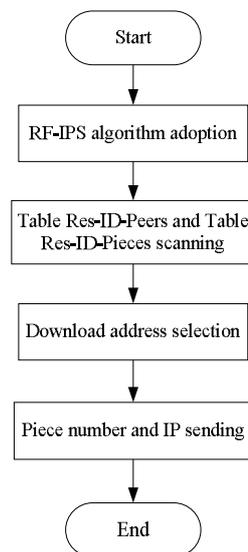
Figure 4. The main function of this module.

2) The RF-IPS algorithm

In this algorithm, we use buffer to record the states of every piece of a peer and partition it into three parts. We call each part an interval. According to the degree of urgently need for the pieces, put the pieces into intervals accordingly, and the pieces in different intervals are downloaded by using the sequential downloading strategy, RF downloading strategy, and random downloading strategy separately [9].

Figure 5 shows the partitioned buffer, where p and g are two pointers which respectively point to the current playback position and the position of data obtained. The pieces have already downloaded are in t0 and those have downloaded but not to be played are in t_1. The pieces need to be downloaded later are in t_2, which will be partitioned into three parts, —interval t_3, interval t_4, and interval t_5.

The current data in t_3 will be played immediately, so it is more urgently needed than the data in other intervals. We adopt sequential downloading strategy and set higher selection priority to the pieces in this interval at a time.

The pieces in t_4 are not so urgently needed than those in t_3, but more urgently needed than those in t_5. In order to ensure the harmonious distribution of data in the P2P network we adopt RF strategy in t_4.

Interval t_5 is the farthest away from the playback position, so in this interval we adopt random selection mechanism.

Figure 6 shows the processing procedure of the RF-IPS algorithm.

4. Experiments

The experiments we describe below were performed on a
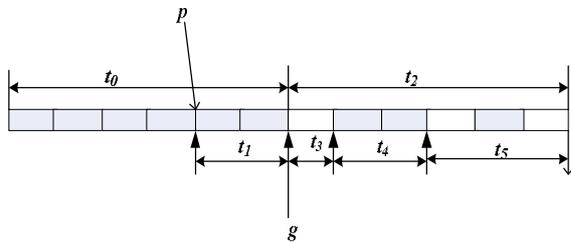
Figure 5. The buffer partitions in RF-IPS algorithm.
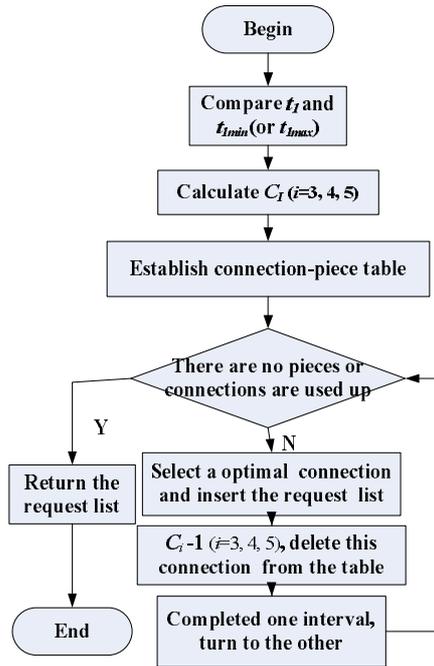
Figure 6. The processing of the RF-IPS algorithm.

50 peers of Inter(R) Pentium(R) D CPU 2.80GHz, 1GB Memory, running Red Hat Linux 3.2.3-47 with kernel 2.4.21-27 using network bandwidth of 100Mb/s in the lab of our university. One peer executed the index server to guide client-side to download, whereas other peers executed multiple instances of clients. The format of a movie is "wmv", the code rate is 756 Kb/s, the length is 120 minutes, and the piece size is 256 K.

In order to ensure the fair results for every algorithm, we designed five systems to represent five algorithms and tested them respectively at the same conditions. We present performance results of our experiments in these parts: the initial delay time and the delay time after 4 times fast-forward operations, including Played at 10 s, fast forward to 120 s, Played at 10 min, fast forward to 20 min, Played at 30 min, fast forward to 70 min and Played at 71 min, fast forward to 119 min. Table 1 presents all results of our experiments.

From **Table 1**, we obtain that the RF-IPS algorithm and sequential downloading algorithm both have the

Table 1. Experiment results.

System name	The delay time (s)				
	Initial	0.17~2(min)	10~20(min)	30~70 (min)	71~119 (min)
Sequential Downloading	7	10	85	128	146
RF	10	8	48	108	120
Bittorrent	9	8	50	80	88
Data Block scheduling	10	5	21	97	101
RF-IPS	7	3	4.5	6	7

shortest initial delay time. When played at 0.17 min, fast-forward to 2 min, the delay time of five algorithms is similar, only the RF-IPS algorithm has the shortest delay. Yet, the delay time of the five algorithms have a great difference when the time interval increased gradually, the delay time of RF-IPS is steady, while the others increase dramatically. So we come to the conclusion that the performance of our proposed algorithm is optimal and the algorithm can improve user experience.

5. Conclusions

It is a challenging job to provide VOD service on the Internet. The rapid development of P2P technology makes it become the powerful tool to build the large-scale distributed system in the WAN. The P2P VOD system becomes a hot topic in the recent distributed system area as one of the most application of P2P technology. Although the applications of P2P at the aspect of streaming have researched and implemented at some degree, their performances and results are still not ideal and mature. This paper examines the RF-IPS algorithm in the P2P VOD system, and achieves the P2P VOD system as a basis. Although this system solve the key problem in the VOD transmission system, and achieve the RF-IPS algorithm, there are many problem existed, because of limited time. With the further improvement of the network bandwidth and the enhancement of the computer performance, we believe that more and better VOD system will be broadly applied.

6. References

[1] Y. Cai, A. Natarajan and J. Wong, "On Scheduling of Peer-to-Peer Video Service," *IEEE Journals of Selected Areas in Communications*, Vol. 25, No. 1, 2007, pp. 140-145.

[2] Media Stream Website. http://www.lmtw.com.

[3] X. Liao, H. Jin, Y. Liu, L. M. Ni and D. Deng, "Anysee:

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　IJCNS

Peer-to-Peer Live Streaming," *INFOCOM*, 2006.

[4] B. Cohen "Bit Torrent Protocol Specification". http://www.bittorrent.com/

[5] PP Stream, http://www.ppstreaming.com.

[6] Y. Cai and J. Zhou, "An Overlay Subscription Network for Live Internet TV Broadcast," *IEEE Transactions on Knowledge Data English*, Vol. 18, No. 12, 2006, pp. 1711-1720.

[7] PPlive, http://www.pptv.com.

[8] A. Habia and J. Chuang, "Incentive Mechanism for Peer-to-Peer Media Streaming," 12*th IEEE International Workshop on Quality of Service*, 2004, pp. 171-180.

[9] Z. Y. Qu and L. L. Li, "A RF-IPS Algorithm for Peer-to-Peer Video-on-Demand System," 5*th Internationl Conference on Wireless Communications, Networking and Mobile Computing*, Beijing, 2009, pp. 4136-4139.

Scientific
Research

Two Slot MIMO Configuration for Cooperative Sensor Network

Ibrahim Mansour, Jamal S. Rahhal, Hasan Farahneh
Electrical Engineering Department, University of Jordan, Amman, Jordan
E-mail: rahhal@ju.edu.jo

Abstract

Sensor networks are used in various applications. Sensors acquire samples of physical data and send them to a destination node in different topologies. Multiple Input Multiple Output (MIMO) systems showed good utilization of channel characteristics. In MIMO Sensor Network, multiple signals are transmitted from the sensors and multiple sensors are used as receiving nodes. This provides each sensor multiple copies of the transmitted signal and hence, array processing techniques help in reducing the effects of noise. In this paper we devise the use of MIMO sensor network and array decision techniques to reduce the noise effect. The proposed system uses a transmission time diversity to form the MIMO system. If the number of sensors is large then groups of sensors will form the MIMO system and benefited from the diversity to reduce the required transmitted power from each sensor. Enhancing the BER reduces the required transmitted power which results in longer battery life for sensor nodes. Simulation results showed an overall gain in SNR that reaches 11 dB in some sensor network scenarios. This gain in SNR led to the opportunity of reducing the transmitted power by similar amount and hence, longer battery life is obtained.

Keywords: Wireless Sensor Networks (WSN), Cooperative Sensor Network (CSN), MIMO, Diversity

1. Introduction

Wireless Sensor Network (WSN) is defined as spatially distributed autonomous sensors to cooperatively monitor physical or environmental conditions. The development of wireless sensor networks is motivated by military applications and is used in many industrial and civilian application areas, such as environmental, pollutants, medical, vehicles, energy management, inventory control, home and building automation, homeland security and others [1-3].

A collection of sensors, actuators, controllers or other elements that communicate with each other and are able to achieve, more or less autonomously, a common goal are defined as cooperating objects. Thus, sensors and actuators form the hardware interfaces with the physical world, where the sensors retrieve information from the physical environment and the actuators modify the environment in response to appropriate commands. Controllers process the information gathered by sensors and issue the appropriate commands to the actuators, in order to achieve control objectives.

Performance of WSN is measured and optimized based on various criteria such as: capacity; bit error rate; SNR; Cross-layer Optimal Scheduling; power requirements; security and robustness.

Power consumption of WSN is an important issue, because if batteries are to be changed constantly, a lot of potential applications will be lost and widespread adoption will not occur. The power consumption must be minimized when the sensor node is designed. The power consumption can be reduced by, reduce the amount of data transmitted through the use source encoding, compression, lower the transceiver duty cycle and the repetition rate of data transmissions, reduce the frame overhead, managing power by using power-down and sleep modes. Implement an event-driven transmission strategy; only transmit data when a sensor event occurs. Turn power on to sensor only when sampling, Turn power on to signal conditioning only when an event occurs. Lower sensor sampling rate to the minimum required by the application.

Virtual MIMO has been studied in a wide range in recent years, in order to advise energy-efficient schemes, constrained by allowed physical size and battery. An individual sensor is allowed to contain only one antenna.

Previous studies showed that if these individual sensors jointly form a MIMO system, tremendous energy is saved while satisfying the required performance. However the disadvantages of the virtual MIMO are the increased complexity and the cost of multiple Radio Frequency (RF) chains.

Since wireless transceivers usually consume a major portion of battery power [3], it is critical to improve their power efficiency. Nevertheless, one of the major difficulties comes from the harsh communication environment with multipath propagation and severe fading [4,5]. Sophisticated and yet computationally efficient techniques is used for reliable and efficient signaling [6]. Moreover, optimization techniques have been used to solve problems arising in wireless networks. Achievable rate combinations were computed in [7,8]. Also, cross-layer optimizations to maximize throughput have been considered in [9].

In this work, we consider a wireless sensor network in which nodes are distributed in a certain region; each node can vary its transmission power to maintain the energy-constraint. A group of sensors may sample one physical quantity forming multi input to the transmission channel and at the Central Node (CN) receiver we employ multiple antenna elements to form the MIMO system. The use of MIMO system creates parallel channels that can be used for independent transmissions [10,11]. This will provide a promising solution to enhance the received signal quality and hence reducing the BER that leads to power saving.

2. System Description

In this paper we devise a solution to implement a MIMO system in wireless sensor networks by having a group of sensor nodes repeat transmitting the same signal that originally initiated by some sensor and another group of nodes acts as a multiple receivers. This architecture of cooperative sensor network will enhance the received signal error rate and hence, improves the network performance. The main idea is that; each sensor will transmit its own signal and repeats other sensors signals. The sensor selects the best K signals received and retransmits them once again as shown in **Figure1**.

Transmitting the same signal twice from the same node is not allowed, and hence a transmission convergence is reached.

Node 4 will receive the signal from node 1 and replicas from nodes 2, 3 and 5. Then at node 4 multiple versions of the signal produced from node 1 will arrive, each from different direction and goes different channel conditions. Node 5 will receive node 1 signal from node 1, 2 and 4. Node 2 will receive the signal from nodes 1, 4
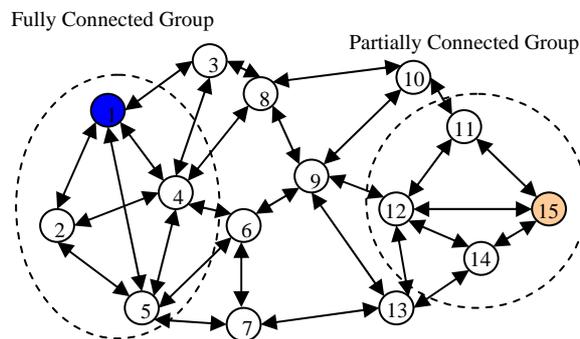

Figure 1. Cooperative sensor network as MIMO system.

and 5. Node 3 also receives node 1 signal from two paths (node 1 and node 4) but this node does not receive the signal from nodes 5 and 2, therefore, it is not in the first group.

To form MIMO system we first need to define the nodes forming each group, these nodes will have wireless connectivity among each others. This can be seen by forming the connectivity matrix as follows:

$$\Omega = \begin{bmatrix} & C_1 & C_2 & C_3 & C_4 & C_5 & \cdots & C_{N_T} \\ C_1 & 1 & 1 & 1 & 1 & 1 & & \\ C_2 & 1 & 1 & 0 & 1 & 1 & & \\ C_3 & 1 & 0 & 1 & 1 & 0 & & \\ C_4 & 1 & 1 & 1 & 1 & 1 & & \\ C_5 & 1 & 1 & 0 & 1 & 1 & & \\ \vdots & & & & & & \ddots & \\ C_{N_T} & 0 & 0 & 0 & 0 & 0 & & 1 \end{bmatrix} \quad (1)$$

When the entry $ji=1$ it means an RF channel from node C_i to node C_j is possible. A fully connected group will have all ones in its corresponding sub matrix as the case of C_1, C_2, C_4 and C_5. A partially connected group will have ones in most of its sub matrix elements as the case of C_{11}, C_{12}, C_{14} and C_{15}. A MIMO system can be constructed from a fully connected group, for example the group **G** having N points fully connected forms an N-1 × N-1 MIMO system. That means each sensor transmits its data to N-1 other sensors. This transmission will be fully available at the second transmission interval, since each sensor will repeat all other sensors signals after it receives them. If the destination node for the data is not a member of the current group, the border nodes (C_1, C_4 and C_5) (C_{11}, C_{12} and C_{14}) will be associated with other groups and hence transfer the data to the next group.

The whole network will be constructed from many groups; each group will be formed from a sub matrix that contains all ones. This can be done by omitting some rows and columns in the matrix as well as performing some permutation to create groups in the matrix. Usually,

sensors forming a subgroup are close to each others in space.

For each group the data is transferred by group cooperation and the signal from node C_i will be transmitted in the ith time slot. Each sensor will store the signals corresponding to all group members and then process the received signals to obtain the best detection. The signals R_j^n arrived at sensor j corresponding to sensor n data can be written as:

$$R_j^n = \begin{bmatrix} r_{j1}^n \\ r_{j1}^n \\ \vdots \\ r_{jN-1}^n \end{bmatrix} \quad where \ r_{ji}^n = h_{ji} x_{ni} + w_{ji} \quad (2)$$

where r_{ji}^n is the signal arrived at sensor j corresponding to sensor n from sensor i, x_{ni} is the nth sensor signal transmitted by sensor i, h_{ij} is the channel coefficient from sensor i to sensor j and w_{ji} is an *iid* $N(0, \sigma_w^2)$ white Gaussian noise. In matrix form we can write all the arrived signals at sensor j as:

$$R_j = H_j X_n + W_j \quad (3)$$

where:

$$H_j = \begin{bmatrix} h_{j1} & 0 & \cdots & 0 \\ 0 & h_{j2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{jN-1} \end{bmatrix} \quad X_n = \begin{bmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{nN-1} \end{bmatrix}$$

$$and \ W_j = \begin{bmatrix} w_{j1} \\ w_{j2} \\ \vdots \\ w_{jN-1} \end{bmatrix} \quad (4)$$

At the jth sensor a Weighted Least Square (WLS) detector can be used to recover the data for each transmitting sensor as:

$$\hat{X}_n = \left[H_j^T V_{ww}^{-1} H_j \right]^{-1} H_j^T V_{ww}^{-1} R_j \quad (5)$$

V_{ww} is the noise covariance matrix.
An example of a fully connected group from the network shown in **Figure 1** will be:

(1,2,4,5), (4,5,6), …, (11,12,14), (11,12,15). And a partially connected groups could be: (4,6,8,9), (6,7, 9, 13), …, (11,12,14,15). Data propagation is done via the boundary nodes, where the decision is made.

In the following we consider an example of a 5 nodes sensor network, the close sensors will have communication channels between each other and groups are formed to propagate data between different sensors. Each sensor will repeat the data once, and the boundary sensors will make a decision when the data is fully available for all sensors.

Example:
**Figure 2** shows a 5 nodes simple sensor network, if
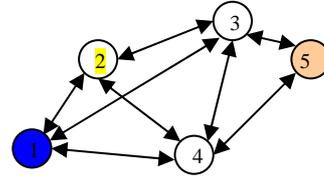
Figure 2. A simple 5 sensor example.

sensor 1 sends a packet to sensor 5, it transmits its signal first to its neighbouring sensors (2, 3 and 4). Then in the second transmitting interval sensors 2, 3 and 4 resends the signal received from sensor 1 again and after the second signalling interval sensor 2 will have the following received signals (corresponding to sensor 1) r_{21}^1, r_{23}^1 and r_{24}^1, sensor 3 will have the following received signals r_{31}^1, r_{32}^1 and r_{34}^1 and sensor 4 will have the following received signals r_{41}^1, r_{42}^1 and r_{43}^1. The combined signals at sensors 2, 3 and 4 form a 3x3 MIMO system. The received signals at sensor 2 are given by:

$$\begin{bmatrix} r_{21}^1 \\ r_{23}^1 \\ r_{24}^1 \end{bmatrix} = \begin{bmatrix} h_{21} & 0 & 0 \\ 0 & h_{23} & 0 \\ 0 & 0 & h_{24} \end{bmatrix} \begin{bmatrix} \hat{x}_{12} \\ \hat{x}_{13} \\ \hat{x}_{14} \end{bmatrix} + \begin{bmatrix} w_{21} \\ w_{23} \\ w_{24} \end{bmatrix} \quad (6)$$

where;

$$\hat{x}_{12} = x \quad \hat{x}_{13} = h_{31} x \ and \ \hat{x}_{14} = h_{41} x \quad (7)$$

x is the transmitted signal from sensor 1. We can combine equations 10 and 11 for each sensor as:

$$\begin{bmatrix} r_{21}^1 \\ r_{23}^1 \\ r_{24}^1 \end{bmatrix} = \begin{bmatrix} h_{21} & 0 & 0 \\ 0 & h_{23}h_{31} & 0 \\ 0 & 0 & h_{24}h_{41} \end{bmatrix} \begin{bmatrix} x \\ x \\ x \end{bmatrix} + \begin{bmatrix} w_{21} \\ w_{23} \\ w_{24} \end{bmatrix} \quad (8)$$

$$\begin{bmatrix} r_{31}^1 \\ r_{32}^1 \\ r_{34}^1 \end{bmatrix} = \begin{bmatrix} h_{31} & 0 & 0 \\ 0 & h_{32}h_{21} & 0 \\ 0 & 0 & h_{34}h_{41} \end{bmatrix} \begin{bmatrix} x \\ x \\ x \end{bmatrix} + \begin{bmatrix} w_{31} \\ w_{32} \\ w_{34} \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} r_{41}^1 \\ r_{42}^1 \\ r_{43}^1 \end{bmatrix} = \begin{bmatrix} h_{41} & 0 & 0 \\ 0 & h_{42}h_{21} & 0 \\ 0 & 0 & h_{43}h_{31} \end{bmatrix} \begin{bmatrix} x \\ x \\ x \end{bmatrix} + \begin{bmatrix} w_{41} \\ w_{42} \\ w_{43} \end{bmatrix} \quad (10)$$

Sensor j will decide for the received signal $\hat{x}_j$ using equations 3 and 8. The network creates a diverse transmission system. This diversity will enhance the BER performance of the over all data transmission. If a diversity order of L_d is used then the BER will be reduced exponentially by a factor of L_d [15]. In the proposed structure and for a group of N nodes, we retransmit the signal N times (lets call it power repetition $L_p=N$) and the diversity order is $L_d = (N-1)^2$. The power repetition is the cost we pay for retransmitting the signal and the gain we achieve is the reception diversity L_d. For the above

example we have $L_d = 9$ and $L_p = 4$. As N grows larger we can achieve better gain compared to one transmission scenario. The overall diversity gain we achieve using the proposed structure can be written as:

$$G_d = \frac{(N-1)^2}{N} \text{ dB} \qquad (11)$$

This means that, for a preset BER performance we can reduce the average transmitted power from each sensor by a factor related to G_d.

The performance of this network is calculated by BER and average power transmitted from each sensor. The main goal is to achieve minimum BER at minimum transmitted power from each sensor. The power constraint imposed in equation 6 makes sure that each sensor will remain under its maximum allowable transmitted power and hence, maximizes its battery life. The BER performance depends on the MIMO sub groups formulated in the network; therefore, a simulation program is used next to determine the BER performance under the power constraint.

3. Simulation Results

In this section we used a MATLAB routine to simulate different sensor networks and results was obtained at different SNR's. The simulation flow is implemented as follows:

1) Initialize the network topology, the power constraint and calculates the groups.

2) Generate random data for each sensor.

3) Transmit the first packet from each sensor, setting the second transmission to all zeroes.

4) Receive the packets at each sensor, append new data to the received packet and retransmit it again (The receiving is done by using equation 3 and the detection is done by using equation 5, then the BER performance is calculated).

5) Repeat (4) until all data is transmitted.

The MIMO part is constructed from the received second transmission from other sensors and the current received signal from the current sensor. This means that we construct the multiple output from the received signal vector arrived from other sensors and the multiple input from the transmitted signals arrived from other sensors.

Four sensor networks were simulated for 5, 10, 15 and 20 sensors. Each simulation uses 1 million runs to calculate the BER performance for the proposed system and the system without MIMO construction. **Figures 3** to **5** shows the BER vs SNR for both with and without MIMO. The system with MIMO has better BER in all cases but to calculate the overall gain in power we select the required BER and the overall gain is found as:

$$G_T = (SNR - SNR_{MIMO}) - 10\log(N_{av} - 1) \ dB \qquad (12)$$

where the last term represents the average extra power needed to be transmitted to form the MIMO system in
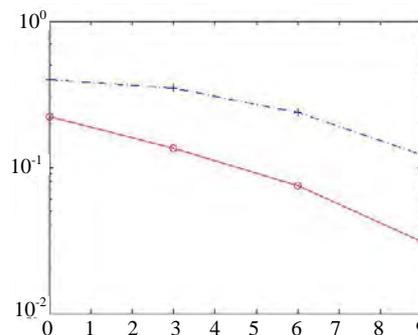
Figure 3. BER vs SNR performance for the proposed network with 5 sensors (cont. line) and one transmission network (dotted line).
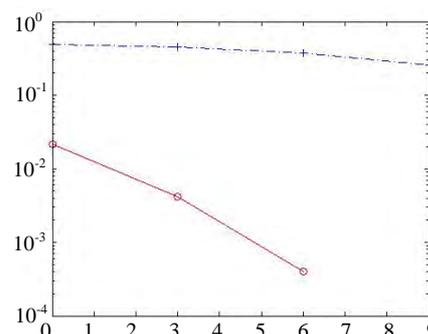
Figure 4. BER vs SNR performance for the proposed network with 10 sensors (cont. line) and one transmission network (dotted line).
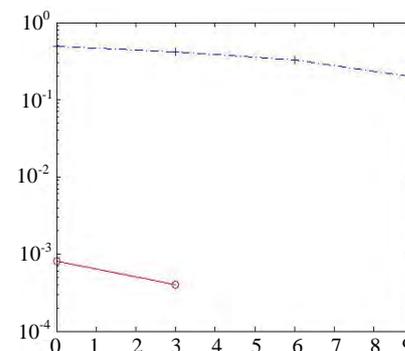
Figure 5. BER vs SNR performance for the proposed network with 15 sensors (cont. line) and one transmission network (dotted line).

the proposed solution.

As the number of nodes increases, the overall gain increases also. This can be seen in **Table 1** where we calculate the overall gain at different network sizes.

The shown results suggests that the sensor power can be reduced to smaller values even with signal repetition and still get the same BER performance as without repetition.

Tabel 1. Total Gain as a Function of Network Size.

N_T	SNR_{MIMO}	SNR	N_{av}	G_T
5	0 dB	6 dB	4	1.2 dB
10	0 dB	13 dB	5	6.0 dB
15	0 dB	19 dB	6	11.2 dB

4. Conclusions

The proposed system has showed an opportunity to enhance the wireless sensor network life by constructing a MIMO system from signal repetition emitted from each sensor. In MIMO structure we can use statistical detection techniques. This provides better signal detection and at the same time makes sure that the transmitted power from each sensor does not exceed a certain preset value. The proposed method requires more signal processing and it will delay the reception by one packet time interval.

5. References

[1] J. Liang and Q. L. Liang, "Channel Selection in Virtual MIMO Wireless Sensor Networks," *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 5, June 2009, pp. 2249-2257.

[2] S. Cui and A. Goldsmith, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE Journals of Selective Areas Communications*, Vol. 22, No. 6, pp. 1089-1098, August 2004.

[3] E. U. Biyikoglu and A. E. Ga, "On Adaptive Transmission for Energy Efficiency in Wireless Data Networks," *IEEE Transactions on Information Theory*, Vol. 50, No. 12, December 2004, pp. 3081-3094.

[4] S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, Achievable Rates and Sum Capacity of Gaussian MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, Vol. 49, No. 10, August 2003, pp. 2658-2668.

[5] S. K. Jayaweera, "Virtual MIMO-Based Cooperative Communication for Energy-Constrained Wireless Sensor Networks," *IEEE Transactions on Wireless Communications*, Vol. 5, No. 5, May 2006, pp. 984-989.

[6] Y. B. Li and X. G. Xia, "A Family of Distributed Space-Time Trellis Codes With Asynchronous Cooperative Diversity," *IEEE Transactions on Communications*, Vol. 55, No. 4, April 2007, pp. 790-800.

[7] A. D. Coso, U. Spagnolini and C. Ibars, "Cooperative Distributed MIMO Channels in Wireless Sensor Networks," *IEEE Journals of Selective Areas Communications*, Vol. 25, No. 2, February 2007, pp. 402-414.

[8] Y. Yuan, Z. He, and M. Chen, "Virtual MIMO-Based Cross-Layer Design for Wireless Sensor Networks," *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 3, May 2006, pp. 856-864.

[9] G. Thatte and U. Mitra, "Sensor Selection and Power Allocation for Distributed Estimation in Sensor Networks: Beyond the Star Topology," *IEEE Transactions on Signal Processing*, Vol. 56, No. 7, July 2008, pp. 2649-2661.

[10] S. Valentin, *et. al*, "CooperativeWireless Networking Beyond Store-and-Forward: Perspectives in PHY and MAC Design," *Wireless Personal Communications*, Vol. 48, 2009, pp. 49-68.

[11] A. B. Gershman and N. D. Sidiropoulos (Eds). "Space-Time Processing for MIMO Communications," John Wiley & Sons, New Jersey, 2005.

Scientific
Research

A New Effective and Efficient Measure of PAPR in OFDM

Ibrahim M. Hussain[1], Imran A. Tasadduq[2], Abdul Rahim Ahmad[3]

[1]*Department of Computer Engineering, Sir Syed University of Engineering and Technology, Karachi, Pakistan*
[2]*Department of Computer Engineering, Umm Al-Qura University, Makkah Al-Mukerrimah, Saudi Arabia*
[3]*Systems Design Engineering, University of Waterloo, Waterloo, Canada*
E-mail: *ibrahimh@ssuet.edu.pk, iatasadduq@uqu.edu.sa, arahim@uwaterloo.ca*

Abstract

In multi-carrier wireless OFDM communication systems, a major issue is high peaks in transmitted signals, resulting in problems such as power inefficiency. In this regard, a common practice is to transmit the signal that has the lowest Peak to Average Power Ratio (PAPR). Consequently, some efficient and accurate method of estimating the PAPR of a signal is required. Previous literature in this area suggests a strong relationship between PAPR and Power Variance (PV). As such, PV has been advocated as a good measure of PAPR. However, contrary to what is suggested in the literature, our research shows that often low values of PV do not correspond to low values of PAPR. Hence, PV does not provide a sound basis for comparing and estimating PAPR in OFDM signals. In this paper a novel, effective, and efficient measure of high peaks in OFDM signals is proposed, which is less simpler PAPR. The proposed measure, termed as Partial Power Variance (PPV), exploits the relationship among PAPR, Aperiodic Autocorrelation Co-efficient (AAC), and PV of the transmitted signal. Our results demonstrate that, in comparison to PV, Partial Power Variance is a more efficient as well as a more effective measure of PAPR. In addition, we demonstrate that the computational complexity of PPV is far less than that of PAPR.

Keywords: Aperiodic Autocorrelation Co-Efficient, OFDM, PAPR, Power Variance, Partial Power Variance

1. Introduction

Several communication systems and techniques have been used for transferring data and information reliably at high speed over wireless channel. One such technique is Orthogonal Frequency Division Multiplexing (OFDM) used for high data rate wireless transmission [1]. In OFDM, data bits are transmitted in parallel using various carriers. Although OFDM is a multi-carrier technology, it is very efficacious in mitigating the effects of multi-path delay spread over a wireless radio channel. However, a major drawback with OFDM is the high Peak-to-Average Power Ratio (PAPR) of the transmitted signal. The high PAPR mainly results from certain data sequences, such as those containing all zeros or all ones. Such OFDM signals with high peaks result in poor power efficiencies. Appropriate measures should be taken to tackle this problem. Otherwise, the high PAPR signals would substantially limit the usefulness of battery powered equipment such as portable wireless devices. In addition, these high peaks cause problems such as inter-symbol

interference (ISI) and out-of-band radiation. Transmitting high PAPR signals by increasing the operating range of the power amplifier deteriorates the power efficiency of the transmission equipment. Hence, it is imperative to reduce these peaks in the transmission signals. This issue has been addressed by several researchers [2-19]. One widely accepted method for reducing the peaks in OFDM signals is based on using Power Variance (PV) as a measure of PAPR of the signal [2-4]. The computational complexity of PAPR depends on the complexity of the IFFT block, which increases by increasing the number of subcarriers. However, Power Variance is computationally less complex than PAPR for specific range of subcarriers [2-4].

In this paper, we investigate the relationship between the PAPR and the PV of the transmitted OFDM signals. We show that a low value of PV in an OFDM sequence does not always correspond to low value of PAPR and vice versa. Therefore, it would not be generally correct to compare PAPR of different signals by comparing their corresponding PVs. We show that PV is not a good meas-

ure for PAPR and hence, our results contradict the widely accepted premise, as stated in [3,4,20-23]. In addition, the computational complexity of PV is a major drawback, which makes it a poor choice for PAPR performance measurement. Towards this end, we propose a new, effective, and computationally efficient measure, called Partial Power Variance (PPV), for estimating the PAPR. Such approximate and partial computational techniques are quite popular, as well as shown effective in various domains [24,25]. We demonstrate the efficiency of PPV through experimental results.

The rest of the paper is organized as follows. Section 2 introduces OFDM systems. Section 3 describes PAPR as a performance measure for OFDM signals. Section 4 presents various PAPR minimization techniques in OFDM. Section 5 discusses the relationship between AAC, PV and PAPR. Section 6 highlights the computational complexity issues in calculating PV and PAPR. Section 7 proposes PPV as a more efficient and less complex estimator of PAPR. Section 8 concludes the paper with some interesting future research directions.

2. Orthogonal Frequency Division Multiplexing (OFDM)

A typical OFDM system consists of a transmitter and a receiver, as shown in **Figure 1(a)** and **Figure 1(b)** respectively. Such a system works as follows: Serial stream of bits $\{b_0, b_1, b_2, \ldots\}$ are encoded using an encoder such that $b_i = 0$ or 1; for $i = 0, 1, 2, \ldots$. If T_b represents the duration of a single bit then the data rate would be $K = 1/T_b$ bits/second (bps). The serial bit stream at the output of channel encoder is fed into a serial to parallel converter

block that forms a parallel stream. This is achieved by increasing the time period of each bit from T_b to NT_b, where N represents the number of subcarriers used in the OFDM system. These bits are transmitted simultaneously to maintain the same data rate as the original rate of K bps. The number of bits entering a particular branch or subcarrier depends on the mapper (*i.e.* the digital modulation block) which is used after the serial to parallel converter. The number of bits per subcarrier is given by $L = log_2 M$ where M is the constellation size used by the mapper and depends on the modulation scheme being used. Hence the duration of a symbol per subcarrier is given by $T_s \times N$ where $T_s = L \times T_b$ represents the duration of each subcarrier symbol at the output of the serial to parallel converter which also represents the duration of a single OFDM symbol.

The output of the mapper consists of complex numbers representing the constellation points in a particular modulation scheme. These complex numbers are given by $D = \{d_0, d_1, \ldots, d_{N-1}\}$. For Quadrature Phase Shift Keying (QPSK) mapper, the values of d_k can take one of the values from $(1, -1, j, -j)$.

The Inverse Fast Fourier Transform (IFFT) block transforms the discrete complex signal into another discrete complex signal. A typical baseband signal at the output of the IFFT block is given by the following well known Inverse Discrete Fourier Transform (IDFT) Equation [2]:

$$s(q) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} d_k e^{j2\pi kq/N}; k, q = 0, 1, \ldots, N-1 \qquad (1)$$


(a)

(b)
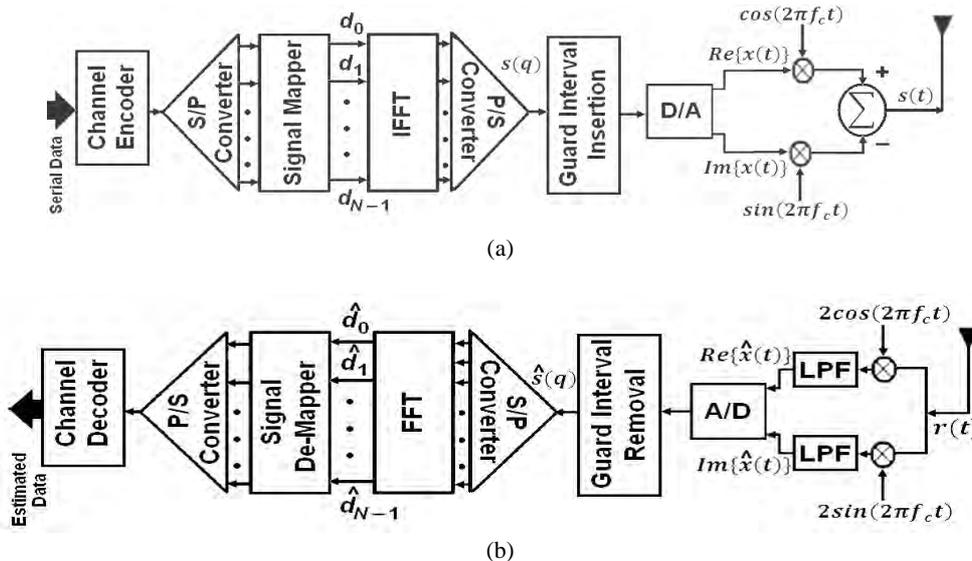
Figure 1. (a) A typical OFDM transmitter; (b) A typical OFDM receiver.

In Equation (1), k indicates the subcarrier index, q is the discrete time index, and d_k represents the complex numbers at the output of the mapper. As indicated by (1), the signal at the output of the IFFT block is the result of summation of various complex sinusoids with varying amplitudes and phases. Hence, the baseband signal given by (1) can be represented as a row vector *i.e.* S = $\{s_0, s_1, \ldots, s_{N-1}\}$. The resulting signal is converted into a serial stream using parallel to serial block after which a cyclic prefix or guard interval of length G is appended to it. The discrete-time sequence S which is input to the guard interval block is cyclically extended to form the new symbol sequence which is indicated as

$$S' = \{s_{N-G+1}, s_{N-G+2}, \ldots, s_{N-1}, s_0, s_1 \ldots, s_{N-1}\}.$$

The cyclically extended discrete-time sequence has new length of $N' = N + G - 1$. This guard interval helps in mitigating the effect of multipath fading in wireless channels. The use of guard interval results in a loss of data throughput as bandwidth is wasted on repeated data. However, in this tradeoff, the loss in data throughput is compensated by significant gains through mitigation in interference. The cyclically extended discrete-time sequence is passed through a digital-to-analog converter to form the baseband OFDM signal. Finally, the baseband OFDM signal is modulated using a carrier frequency for transmission through a wireless channel.

Transmitting a signal through a wireless channel results in convolution of the signal with the impulse response $h(q)$ of the channel. Consequently, the signal is distorted by the additive white Gaussian noise (AWGN) $n(q)$ present in the channel. The convolution between the transmitted signal and channel's impulse response is a circular convolution due to the guard interval. Thus, as seen by channel, the discrete-time sequence S' looks as if s is repeated periodically for all time.

The process of recovering the transmitted data sequence begins with the down conversion of the received signal performed by an IQ detector. The output of the IQ detector is the distorted version of the complex signal $s(t)$, indicated as $\hat{s}(t)$. The signal $\hat{s}(t)$ is passed through an analog-to-digital (A/D) converter to obtain a complex discrete signal $\hat{s}(q)$. Subsequently, the cyclic prefix is discarded and the signal becomes [2]:

$$\hat{s}_q = \sum_{m=0}^{N-1} s_{(q-m)_N} h_m ; q = 0, 1, \ldots, N-1 \qquad (2)$$

where $(q - m)_N$ represents modulo N subtraction. In vector form, $\hat{s}_q$ is represented as $\hat{S} = \{\hat{s}_0, \hat{s}_1, \ldots, \hat{s}_{N-1}\}$. After passing the received sequence through the FFT block, an estimate of transmitted complex symbols is obtained which is given by:

$$\hat{d}_k = \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} \hat{s}_q e^{-j2\pi kq/N} ; k = 0, 1, \ldots, N-1 \qquad (3)$$

After substituting (2) into (3) $\hat{d}_k$ becomes:

$$\hat{d}_k = H_k \times d_k \qquad (4)$$

here, H_k represents the transfer function component of the channel and $\hat{d}_k$ is the received subcarrier information at the output of the FFT block. The complex sequence at the output of the FFT block *i.e.* $\hat{D} = \{\hat{d}_0, \hat{d}_1, \ldots, \hat{d}_{N-1}\}$ is then passed through the signal de-mapper and parallel-to-serial converter to obtain an estimate of the encoded information. The decoder is then used to arrive at an estimate of the information transmitted.

3. Peak-to-Average Power Ratio (PAPR)

As pointed out in the previous section, the baseband OFDM signal is the result of summation of sinusoidal waves at the output of the IFFT block. At some sample points of these sinusoidal signals, constructive summations may occur, resulting in high peaks in the signal. When transmitting high peak signals through a non-linear power amplifier, distortion occurs within the transmitted signal at the output of the amplifier in the form of ISI and out-of-band radiation. Hence, the influence of high peaks is evident at the output of a non-linear power amplifier but the point of occurrence of these peaks is at the output of an IFFT block. For this reason, the non-linear amplifier is neither used for the analysis throughout the paper nor in simulations being carried out as our main concern revolves around the measurement of these high peaks at the point of occurrence.

One of the widely used measures for the power of these peaks is Peak to Average Power Ratio (PAPR) which is mathematically expressed as:

$$PAPR = \frac{\max\{P(q)\}}{P_{avg}} = \frac{\displaystyle\max_{0 \le q \le N-1} |s(q)|^2}{\dfrac{1}{N} \displaystyle\sum_{q=0}^{N-1} |s(q)|^2} \qquad (5)$$

here, $P(q)$ represents the instantaneous power and P_{avg} represents the average power of the OFDM signal. For constant envelope signals, it can be shown that $P_{avg} = N$. In order to simulate such a system, samples of OFDM signals are needed. For better PAPR estimation, over-sampling is required to capture these peaks since in normal symbol spaced sampling; some of the peaks might be missed and may result in less accurate PAPR measure. Hence oversampling (1) by a factor of J where ($J > 1$) gives a better PAPR estimation. It has been shown that $J = 4$ is sufficient to capture the peaks [5]. The peak value of an OFDM signal and the corresponding time domain signal differs from one mapper to another (*e.g.* peak in 32−Quadrature Amplitude Modulation (QAM) is different from the peak value in 32−Phase Shift Keying (PSK)).

For all phase shift keying, the maximum peak has the value of N^2 and hence a maximum PAPR of N. **Figure 2** shows all possible OFDM signals for $N = 4$ and BPSK mapper. It can be seen that the maximum normalized absolute peak in such signals is 4 as indicated in the sequences 0000, 1001, 1010 and 1111.

Since the transmitted bits are generated randomly, the transmitted OFDM signals are random in nature and therefore the envelope of an OFDM signal, as given by (1), can be considered a random variable. For large values of N, according to the central limit theorem, the expected amplitudes of OFDM signals follow a Gaussian distribution. In addition, $P(q)$ has a Chi square probability density function with two degrees of freedom [5]. The PAPR performance of a system is usually measured by the Complementary Cumulative Distribution Function (CCDF) curves, a standard way of depicting and describing PAPR related statistics. The CCDF shows the probability of an OFDM sequence exceeding a given PAPR (PAP$_0$). For PAPR, the lower bound on CCDF for a specific PAPR value (*i.e.*, PAP$_0$) is given by [7]:

$$\beta = \Pr\left(PAPR > PAP_0\right)$$

This results in the following relationship:

$$\beta \geq 1 - \left(1 - e^{-PAP_0}\right)^N \qquad (6)$$

4. Techniques for Minimizing PAPR

The objective is to minimize PAPR as much as possible so as to obtain signals with smaller peaks. Various algorithms and methods have been proposed for reducing PAPR. One simple method for reducing PAPR is direct clipping of high peaks and subsequent filtering of the signal [7,8]. In addition, various modulation schemes are used for efficient transmission of signals such as Continuous Phase Modulation (CPM) [9,10].

Recently, constellation and shaping methods have

been used to reduce PAPR. In these methods, a mapping between the original complex numbers and the finally transmitted complex numbers takes place based on an algorithm or a coding technique. One such method is the trellis shaping method using a metric which is based on the Viterbi algorithm [11]. One of the variants of this method uses a metric-based symbol predistortion algorithm resulting in some implementation complexity [12]. Another promising technique for reducing PAPR involves scrambling the incoming OFDM sequence using some rotation vectors resulting in multiple signals that represent the same original information. Among these signals, the signal with the lowest PAPR is transmitted. Examples of such methods are Selected Mapping (SLM) [13-15] and Partial Transmit Sequence (PTS) [16,17]. In SLM, shown in **Figure 3(a)**, from a single OFDM sequence of length N, U sequences are generated that represent the original information or OFDM sequence. The sequence having lowest PAPR value is transmitted. These sequences are generated by multiplying the original OFDM sequence with U different factors. These factors are given in vector form as $B^{(i)} = \left[b_0^{(i)}, b_1^{(i)}, \ldots, b_{N-1}^{(i)}\right]$ where $i = 1$ to U represents the indices of these factors. After multiplying these factors by the original OFDM sequence D, we get $X^{(i)} = \left[d_0 b_0^{(i)}, d_1 b_1^{(i)}, \ldots, d_{N-1} b_{N-1}^{(i)}\right]$. These factors are phase rotations selected appropriately such that multiplying a complex number by these factors


(a)

(b)
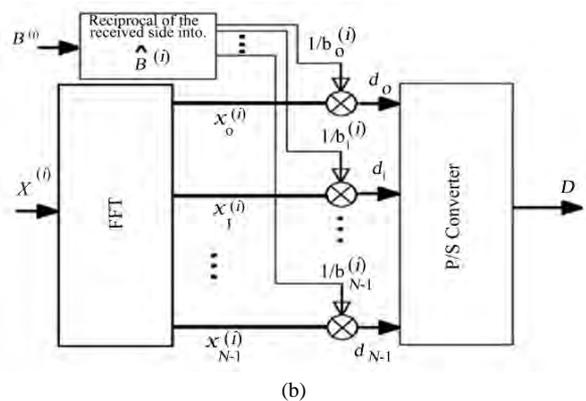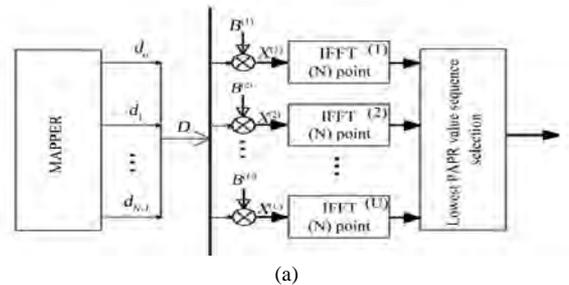
Figure 3. The SLM method (a) at the transmitter and (b) at the receiver.

Figure 2. Peak values in OFDM signals for $N = 4$ by using BPSK mapper.

results in rotation of that complex number to another complex number representing a different point in the constellation. Hence, $b_n^{(i)} = e^{-j\theta_n^{(i)}}$ where $\theta_n^{(i)} \in [0, 2\pi)$.

At the receiver end as shown in **Figure 3(b)**, the original OFDM sequence is recovered by multiplying the received sequence by the reciprocal of the vector being used at the transmitter end. Hence, in SLM, it is essential to transmit the rotational vector $B^{(i)}$ as side information to recover the original sequence.

On the other hand, in PTS method, as shown in **Figure 4**, the original OFDM sequence D is first partitioned into H disjoint sub-sequences. The length of each sub-sequence is still N but padded with zeros. Each sub-sequence is fed into a separate IFFT block of length N each. Hence, there are H number of IFFT blocks. The set of complex numbers at the output of each IFFT block is multiplied by a factor belonging to one of the rotation factors as indicated in SLM technique. These factors are optimized in such a manner that the PAPR of the combined sub-sequences are reduced. After multiplying by the factors, the complex numbers from all the IFFT blocks are added together carrier-wise, resulting in a final sequence. This final sequence has lower PAPR than the original one.

Many other proposed algorithms tackle different parameters for reducing the PAPR indirectly. One such method is the Aperiodic Autocorrelation Coefficient (AAC) of the transmitted OFDM signals in which PAPR reduction is achieved using selective scrambling of the transmitted sequence, generating a number of statistically independent sequences [2]. A Selective Function (SF) is computed for every sequence and the sequence with the lowest SF, which also corresponds to the lowest PAPR, is transmitted.

Another factor that plays an important role in reducing
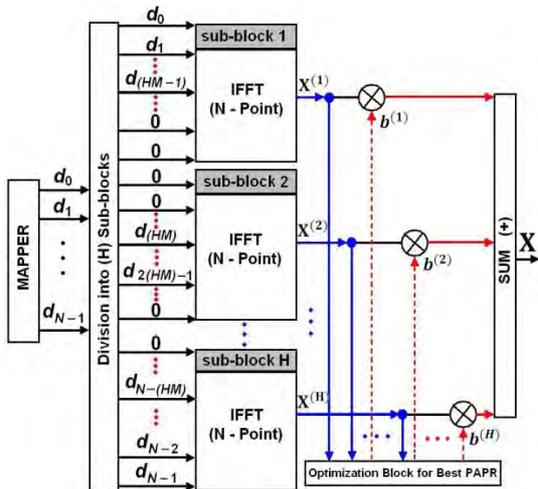
Figure 4. PTS method.

PAPR is the Power Variance (PV) of an OFDM sequence. It is indicated in the literature that there exists a strong relationship among AAC, PV and PAPR [2-4,20-23,26]. In the next section, we study this relationship in OFDM signals.

5. PV and Aperiodic Autocorrelation Coefficients

For a complex envelope signal given by (1), the instantaneous power is given as:

$$P(q) = (q)S^*(q) \tag{7}$$

where '*' denotes conjugate of a complex signal. By combining (1) and (7) we get:

$$P(q) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{p=0}^{N-1} d_k e^{\frac{j2\pi kq}{N}} d_p^* e^{\frac{-j2\pi pq}{N}} \tag{8}$$

Since the subcarriers in the OFDM signal are orthogonal, they satisfy the following condition [2]:

$$\sum_{q=0}^{N-1} \Phi(i,q)\Phi^*(j,q) = \begin{cases} 0, & i \neq j \\ N, & i = j \end{cases} \tag{9}$$

such that,

$$\Phi = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{j2\pi/N} & \cdots & e^{j2(N-1)\pi/N} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{j2(N-1)\pi/N} & \cdots & e^{j2(N-1)^2\pi/N} \end{bmatrix} \tag{10}$$

By simple manipulation of (8) and using the orthogonality property of the subcarriers given by (9), the instantaneous power can be expressed as:

$$P(q) = \frac{1}{N}\left[R_0 + \sum_{i=1}^{N-1} R_i e^{-\frac{j2\pi iq}{N}} \sum_{i=1}^{N-1} R_i^* e^{\frac{j2\pi iq}{N}} \right] \tag{11}$$

where R_i is called the ith Aperiodic Autocorrelation Co-efficient of the complex OFDM sequence D and given as [3]:

$$R_i = \sum_{k=0}^{N-1-i} d_k d_{k+i}^*; 0 \leq i \leq N-1 \tag{12}$$

Note that for $i = 0$, (12) becomes $R_0 = \sum_{k=o}^{N-1} |d_k|^2 = P_{avg}$.

in case of a constant envelope mapper, $P_{avg} = N$. By combining (11) and (12) then dividing by P_{avg}, we get the normalized instantaneous power $\gamma(q)$ given below:

$$\gamma(q) = \frac{P(q)}{P_{avg}}$$

$$\frac{1}{NP_{avg}}\left[N + \sum_{i=1}^{N-1} R_i \left\{ \cos\left(\frac{2\pi iq}{N}\right) - j\sin\left(\frac{2\pi iq}{N}\right) \right\} + \right.$$

$$\left. \sum_{i=1}^{N-1} R_i^* \left\{ \cos\left(\frac{2\pi iq}{N}\right) + j\sin\left(\frac{2\pi iq}{N}\right) \right\} \right]$$

hence, $\gamma(q)$ is given by:

$$\gamma(q) =$$

$$\frac{1}{NP_{avg}}\left[N + 2\sum_{i=1}^{N-1}\text{Re}\{R_i\}\cos\left(\frac{2\pi iq}{N}\right) - 2\sum_{i=1}^{N-1}\text{Im}\{R_i^*\}\sin\left(\frac{2\pi iq}{N}\right)\right] \quad (13)$$

In compact form, (13) can be reduced into:

$$\gamma(q) = \frac{1}{NP_{avg}}\left[R_0 + 2\sum_{i=1}^{N-1}|R_i|\cos\left(\frac{2\pi iq}{N} + \tan^{-1}\left(\frac{Img\{R_i^*\}}{\text{Re}\{R_i\}}\right)\right)\right] \quad (14)$$

Using Chebyshev polynomials, an upper bound on (14) is given by [19]:

$$\gamma(q) \le \Gamma = \frac{1}{NP_{avg}}\left[R_0 \, 2\sum_{i=1}^{N-1}|R_i|\right] \quad (15)$$

By taking the difference between the instantaneous power and the average power, *i.e.* $\Delta P(q) = P(q) - P_{avg}$, we get the Power Variance of OFDM signal using the following expression:

$$PV = \frac{1}{N}\sum_{q=0}^{N-1}\left[\Delta P(q)\right]^2 \quad$$

which can be further expanded using trigonometric properties into the following expression:

$$PV = \frac{1}{N}\left[\frac{4}{N^2}\sum_{i=1}^{N-1}\sum_{N-1}^{m=1}|R_i||R_m|\right] \times$$
$$\left[\sum_{q=1}^{N-1}\cos\left(\frac{2\pi iq}{N} + \varphi_i\right)\cos\left(\frac{2\pi mq}{N} + \varphi_m\right) - N\left(P_{avg} - 1\right)^2\right] \quad (16)$$

Hence, an upper limit on (16) becomes:

$$PV \le PV_{norm} = \sum_{i=0}^{N-1}|R_i|^2 \quad (17)$$

here, PV_{norm} is the normalized power variance. It can be observed from (11-15) that $P(q)$, $\gamma(q)$ and Γ are all functions of AAC. Low values of $|R_i|$ correspond to low instantaneous power and hence low values of Γ. Based on this assessment, some authors use (15) as a parameter for measuring and comparing PAPR values of OFDM sequences. In contrast, other authors show that PAPR analysis based on (14) and (15) gives misleading results [19]. Through an example, they show that sequences exist that have low PAPR values but high Γ values and vice versa. They conclude that Γ cannot serve as parameters for PAPR comparison of sequences and hence Γ is not a good measure for PAPR. In addition, many authors concluded that PV is also a good measure of PAPR [2-4,20-23]. However, in subsequent paragraphs, we show that this is not always the case.

Figures 5 and **6** show normalized PAPR and PV values respectively for different sets of 300 randomly generated OFDM sequences when eight subcarriers are used. Apparently, these figures suggest that PV is a good measure
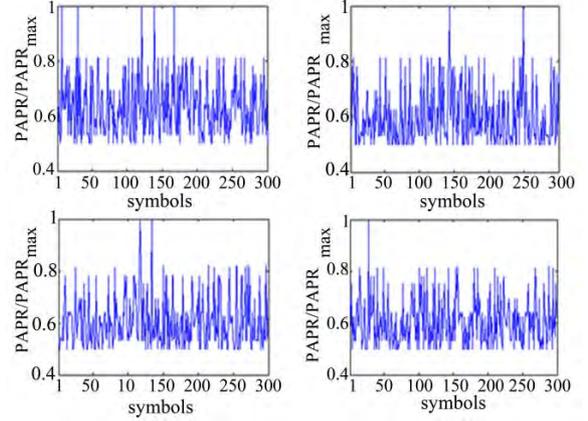

**Figure 5. Normalized PAPR for randomly generated OFDM sequences with 8 subcarriers.**
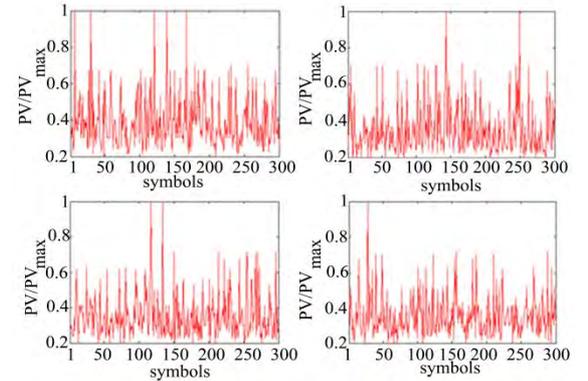
Figure 6. Normalized PV for randomly generated OFDM sequences with 8 subcarriers.

of PAPR as the pattern of both the figures for the same set of symbols is almost the same. Hence, many authors concluded that PV is a good measure for PAPR [2-4,20-23]. However, we here demonstrate that this observation is misleading because not all low PAPR sequences have low PV and vice versa. To refute this assertion, we investigate the relation between PV and PAPR given as [4,5]:

$$Q\left(\frac{PAPR - 1}{\sqrt{PV}}\right) + Q\left(\frac{1}{\sqrt{PV}}\right) = \beta \quad (18)$$

where $Q(.)$ is the complementary error function and β denotes $Pr\left(P(q) \le P(q)_{max}\right)$ which is given by (6). Figure 7 shows a plot of (18) for selective values of PAPR for 256 subcarriers with $\beta = 0.7434$. Two aspects in this **Figure 7** are noteworthy. First, for a particular value of PAPR, there are two values of PV. For instance, at PAPR = 6.5 *dB*, the values of PV are 35 and 93.37. Second, PV does not vary in proportion to PAPR. For instance, at PAPR = 6*dB*, one of the values of PV is 79.89 whereas at PAPR = 7*dB*, one of the values of PV is 44.31, *i.e.* an in-
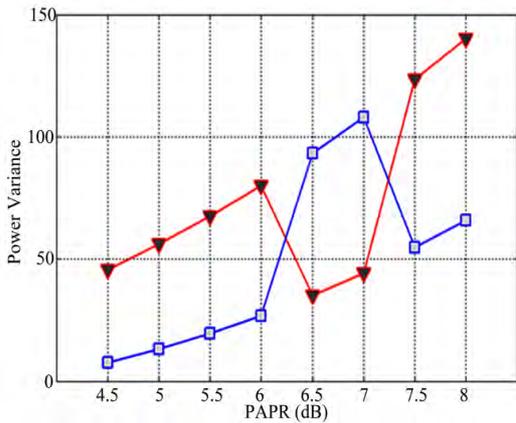
Figure 7. Relationship between PAPR and PV for β = 0.7434 with 256 subcarriers.

crease in PAPR does not always correspond to an increase in PV. It shows that PV cannot always constitute a basis for comparison of PAPR between two sequences.

Hence, PV cannot be considered a reliable measure for PAPR. In addition, to emphasize our assessment of relationship between PV and PAPR, CCDF plots are used for PAPR measure based on PV using SLM and PTS techniques.

In the first simulation experiment, CCDF curves are plotted for 50,000 randomly generated OFDM sequences using SLM technique where U = 4, 16 and 64 as shown in **Figure 8**. In this approach, U different sequences are generated from a single OFDM sequence D using randomly generated phase rotation factors. The sequences having the lowest PV and the lowest PAPR are selected for transmission. In short, the selection decision for transmitting a sequence is based on both the lowest PV and the lowest PAPR. This simulation gives two sets of CCDF curves, as shown in **Figure 8**. It is clear that the reduction in PAPR based on PV is not the same as the one based on PAPR. For instance at CCDF = 0.001, when U = 64, the PAPR reduction using a PV-based decision is approximately $2dB$ while the PAPR reduction based on PAPR itself is around $3.8dB$. A difference of almost $2dB$ is evident between the two transmission decisions. This difference in PAPR reduction is sufficient to show that PV is not a good and reliable measure for the purposes of PAPR reduction.

For the second simulation experiment, CCDF curves are plotted for 50,000 randomly generated OFDM sequences using PTS technique using 256 subcarriers and H = 16, shown in **Figure 9**. As it is the case in SLM, two curves are generated for both PV- and PAPR-based transmission decisions. A PAPR reduction of $2dB$ is achieved in case of PV-based decision whereas a reduction of $3.5dB$ is achieved in case of PAPR-based decision. Once again, the difference in PAPR performance be-
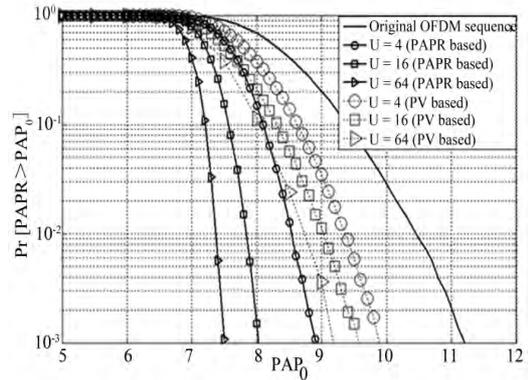

**Figure 8. CCDF curves for SLM based on PAPR and PV for various values of *U*.**
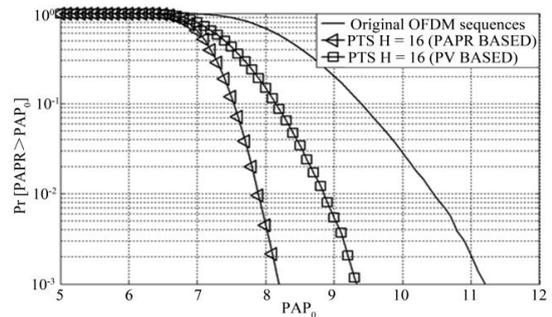
Figure 9. CCDF curves for PTS technique based on PAPR and PV for *H* = 16.

tween PV-based and PAPR-based transmission decisions is significant and obvious. All these results suggest that low values of PV do not always correspond to low PAPR values and vice versa. Consequently, PV is not a reliable measure of PAPR and it cannot be used as a parameter for OFDM sequence selection in both SLM and PTS techniques. Since OFDM sequences follow a random process, it is difficult to tell the range of PAPR values or sequences that correspond to high PV values and vice versa.

It can also be noted that for a particular number of subcarriers N, the values of β vary with PAPR values. Although theoretically PAPR, PV and β can take any value, as indicated by (18), the actual values of these three parameters are finite and specific for a given number of subcarriers. For instance, through simulation we generate all possible OFDM symbols for 16 subcarriers using BPSK mapper and plot their corresponding PAPR values against PV values as shown in **Figure 10**. The total number of possible OFDM symbols is 65,536. The maximum and minimum PAPR values are 16 and 1.7071 respectively and the corresponding PV values for these PAPR values are 1240 and 24 respectively. Note that the plot is discontinuous through PAPR axis because the
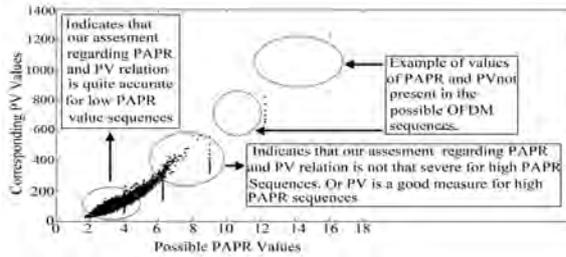
Figure 10. PV versus PAPR values for OFDM sequences with 16 subcarriers and BPSK mapper.

OFDM symbols have finite and specific values of PAPR, as mentioned earlier.

It is interesting to note that our assessment regarding the relation between PV and PAPR is more pronounced for low values of PAPR than high values of PAPR (*i.e.* more concentration of points in the lower region of PAPR). It can be seen that for a single PAPR value in the lower range (approximately from 2*dB* to 5*dB*), there are more than one corresponding PV value. Similarly, for a single value of PV (approximately from 24 to 300), there are more than one corresponding PAPR value. Again, this plot is only possible for small number of subcarriers. For high number of subcarriers, it is difficult to find the distribution of PAPR and PV values for all possible OFDM sequences.

6. Computational Complexity of PV

SLM algorithm based on PAPR comparison has moderate complexity. The main complexity is based on the computation of IFFT operation. This complexity increases as U increases. Hence, IFFT has to be computed for every sequence (*i.e.* U times) before transmitting the final selected sequence. Similarly, when evaluating PV, Equation (17) needs to be evaluated for every sequence (*i.e.* U times) before transmitting the final selected sequence. So, it would be interesting to compare the complexity associated with computing PAPR with the complexity in evaluating PV.

The complexity expressions in terms of complex additions and multiplications for evaluating IFFT, and therefore PAPR, are shown in **Table 1** [28]. Similarly it could easily be shown that the complexity for evaluating Equation (17) in terms of both complex additions and multiplications is given as:

$$Complex\ Additions = \frac{N(N-1)-2}{2}$$
$$Complex\ Multiplications = (N-1)\left(\frac{N}{2}+1\right) \quad (19)$$

Table 1. Computational complexity of various parameters.

Expression	Complex Additions	Complex Multiplications
IFFT	$JN(\log_2 JN)$	$JN(\log_2 JN)/2$
IFFT and PAPR	$JN(\log_2 JN)$	$(JN(\log_2 JN)/2)+JN$
PV	$(N(N-1)-2)/2$	$(N-1)((N/2)+1)$
PPV	$(B(2N-B-1)-2)/2$	$(B(2N-B-1)/2)+B$

Table 2. Computational complexity of various parameters for different subcarriers.

	Complex Additions				
N	PAPR ($J=1$)	PAPR ($J=4$)	PV	PPV (B)	PPV ($B=1$)
4	8	64	5	4 (2)	2
8	24	160	27	21 (4)	6
16	64	384	119	91 (8)	14
32	160	896	495	219 (8)	30
64	384	2048	2015	475 (8)	62
128	896	4608	8127	1911 (16)	126
256	2048	10240	32639	3959 (16)	254
	Complex Multiplications				
N	PAPR ($J=1$)	PAPR ($J=4$)	PV	PPV (B)	PPV ($B=1$)
4	4	32	9	7 (2)	4
8	12	80	35	26 (4)	8
16	32	192	135	100 (8)	16
32	80	448	527	228 (8)	32
64	192	1024	2079	448 (8)	64
128	448	2304	8255	1928 (16)	128
256	1024	5120	32895	3976 (16)	256

The computational complexity for both PAPR and PV for different values of N is presented in **Table. 2**. Note that, for comparison purposes, PAPR complexity is also evaluated when oversampling OFDM signals by a factor of 4. It is clear that the computational complexity of PV is lower than that of PAPR only for very small values of N (*i.e.* $N=2$ and 4). For large values of N, the computational complexity of PV is far greater than that of PAPR complexity even when $J=4$. Hence, PV-based selection becomes impractical for large values of N. This is another disadvantage of using PV as a selection criterion in SLM or PTS techniques, pronouncing the need for an efficient measure of PAPR. In the next section, we propose one such measure of PAPR

7. AAC-SLM Algorithm and Partial Power Variance (PPV)

In this section, we introduce a new parameter called Par-

tial Power Variance (PPV) that can be effective in reducing PAPR. Indeed, such Soft Computing tools and approximate, but effective and efficient, computational techniques are rapidly gaining popularity in subjective and complex application domains [24,25]. We show that PPV is not only effective but also computationally efficient than both PV and PAPR. Towards this end, we also propose an algorithm for minimizing PAPR. The proposed algorithm is modeled on SLM technique and we call it AAC-SLM. As indicated by (17), the normalized PV expression consists of $N-1$ aperiodic autocorrelation coefficients. The objective of AAC-SLM algorithm is to investigate these coefficients (*i.e.* $R_1, R_2, \ldots, R_{N-1}$) individually and find their contribution in reducing PAPR. For QPSK mapper, the proposed algorithm involves the following steps. From a single OFDM sequence, $(N-1)$ different sequences are generated by reducing $|R_i|$ of the sequence to its minimum possible values of 0 (for even i) and 1 (for odd i) where $i = 1, 2, \ldots, N-1$. These sequences represented by $Y_1, Y_2, \ldots, Y_{N-1}$ have the following property: the generated sequence Y_i is the result of minimizing $|R_i|$ of the original OFDM sequence to its minimum value (*i.e.* either 0 or 1 depending on i). Hence, for each new sequence, only its respective $|R_i|$ is minimized while other $|R_i|$'s for that sequence may not have minimum value. For example, a sequence generated by minimizing $|R_3|$ *i.e.* Y_3 may or may not have minimum values of other $|R_i|$, *i.e.* $|R_1|, |R_2|, |R_4|, \ldots$ *etc*. From these sequences, the one with the lowest PAPR is transmitted. The method for reducing the coefficients to their minimum values is given in the Appendix.

The CCDF curves for AAC-SLM algorithm using 256 subcarriers and 100,000 randomly generated OFDM sequences are shown in **Figure 11**. It is clear that a maximum reduction of 3.8 *dB* is achieved when all sequences from Y_1 through Y_{255} are included. The sequence with the lowest PAPR is selected. Further, when reducing the number of sequences to 64 (*i.e.* $Y_1, Y_2, \ldots, Y_{64}$) and selecting the one with lowest PAPR, a negligible degradation in PAPR occurs. This suggests that the higher order sequences Y_{65} through Y_{255} only have a marginal and negligible contribution in reducing PAPR. Similarly, when considering only Y_1 and Y_2, a PAPR reduction of 3.2*dB* is achieved. In fact, only by reducing the first AAC (*i.e.* Y_1 only), almost 3*dB* reduction is achieved in PAPR performance. In other words, all the generated sequences in the AAC-SLM algorithm other than Y_1 have a minimal contribution in PAPR reduction. It follows that all AAC in the PV expression have a minimal contribution in PAPR reduction except for $|R_1|$.

Based on the previous observations, **Figure 8** is re-
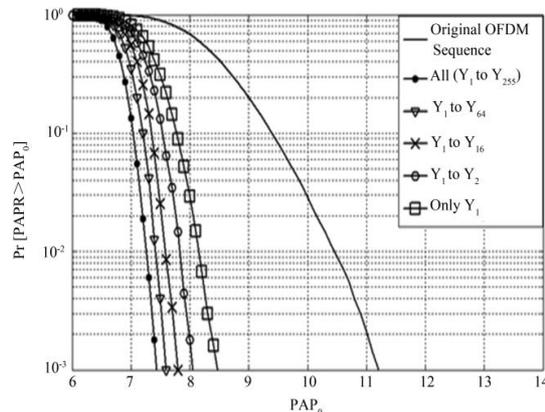

Figure 11. AAC-SLM with 256 subcarriers.
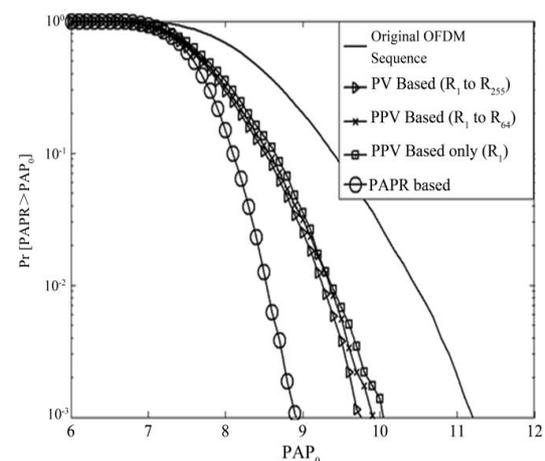
Figure 12. PPV in PAPR reduction for different values of *B* with *N* = 256.

plotted for $U = 4$, shown in **Figure 12**. Around 100,000 OFDM symbols are randomly generated to carry out this experiment. Similar to **Figure 8**, CCDF curves for both PV- and PAPR-based decision are shown in **Figure 12**. The two additional CCDF curves are a result of making the transmission decision based on PPV, where PPV is the truncated version of the PV expression given by (17). The first curve is generated when the decision is based only on the first 64 AAC ($R_1, R_2, \ldots, R_{64}$) in (17) instead of using the whole expression for PV. Hence in this case,. From **Figure 12** it is clear that a negligible degradation in PAPR performance results using this truncated expression. In fact, when a single AAC is used *i.e. PPV = $|R_1|^2$*, a degradation of only 0.2*dB* occurs from the PV-based CCDF curve. These results also support our earlier observations for the efficacy of the AAC-SLM algorithm. In other words, we can say that AAC terms in PV expression do not contribute in reducing PAPR significantly except for the first term *i.e.* R_1 that has the maximum contribution in reducing PAPR. In short, we can use PPV instead of PV while achieving almost the same

PAPR reduction.

In addition, the computational complexity of PPV is less than both PV and PAPR. To compare the computational complexities of the PV- and PPV-based decisions, we write a general expression for PPV as follows:

$$PPV_B = \sum_{i=1}^{B} |R_i|^2 \qquad (20)$$

where B indicates the number of AAC terms included in the PPV expression. Note that when $B = N - 1$, PPV expression becomes PV. In terms of B, the complexity of PPV is given below (and also shown in Table 1):

$$\left.\begin{array}{l} Complex\ Additions = \dfrac{B(2N-B-1)-2}{2} \\[2mm] Complex\ Multip. = \dfrac{B(2N-B-1)}{2} + B \end{array}\right\} \qquad (21)$$

In case of $B = 1$, the complexity reduces to $N - 2$ and N complex additions and multiplications, respectively. This shows a considerable reduction in computational complexity when PPV is used for transmission decision instead of PV or even PAPR itself. The reduction in computational complexity for different values of B is shown in **Table 2**. Once again, it can be seen that the maximum reduction in complexity is achieved when $B = 1$. For instance, in case of 256 subcarriers, the number of complex additions in PAPR (with $J = 1$) is 2048, whereas in case of PPV_1 it reduces to 254, a reduction by a factor of nearly 8. The computational complexity is decreased considerably compared with IFFT complexity. Hence PPV is faster to implement both in hardware using digital signal processing techniques and software than IFFT.

8. Conclusions

In this paper, we have established that PV is not a good measure of power efficiency as has been claimed in the literature. Our results clearly show that using PV as the power efficiency measure gives misleading results. Further, we show that PV is computationally more complex than PAPR and hence cannot be used as a power efficiency measure for OFDM. In addition, we have developed a new, effective and efficient measure for power efficiency called PPV which is computationally less complex than PAPR. The amount of reduction achieved in terms of complex additions and multiplications for a 256-subcarrier system is more than 8 times as compared to PV and 3.5 times as compared to PAPR in order to achieve the best power efficiency. In fact, based upon the flavor used for PPV, the reduction in complexity may go down as low as 40 times as compared to PAPR at a nominal degradation in power efficiency. Hence, PPV is a more useful, realistic and cost-effective measure for power

efficiency of OFDM signals. In this paper, we demonstrated the efficacy of the new measure by applying it on PTS and SLM techniques. The proposed measure can also be applied on other established algorithms to decrease the computational complexity. Further, the performance of the new measure can also be tested for more practical systems where the number of subcarriers may go beyond 2048.

In this paper the bit error rate (BER) performance of PPV is not discussed as the main objective was to investigate the PAPR performance of PPV. As a future work, it would be interesting to find out the BER performance of PPV both in AWGN and multipath fading channels.

9. References

[1] Special Issue on 4G Mobile Communications: "Toward Open Wireless Architecture, *IEEE Wireless Communications*, Vol. 11, No. 2, April, 2004.

[2] P. Van Eetvelt, G. Wade and M. Tomlinson, "Peak to Average Power Reduction for OFDM Schemes by Selective Scrambling," *Electronics Letters*, Vol. 32, No. 21, October 1996, pp. 1963-1964.

[3] H. Nikookar and R. Prasad, "Weighted Multicarrier Modulation for Peak-to-Average Power Reduction," *IEICE Transactions on Communications*, Vol. E83-B, August 2000, pp. 1396-1404.

[4] H. Nikookar and Knut Sverre Lidsheim, "Random Phase Updating Algorithm for OFDM Transmission with Low PAPR," *IEEE Transactions on Broadcasting*, Vol. 48, No. 2, June 2002, pp. 123 - 128.

[5] C. Tellambura, "Use of M-sequences for OFDM Peak-to-average Power Ratio Reduction," *IEEE Electronics Lett*ers, Vol. 33, No. 15, July 1999, pp. 1300-1301.

[6] C. Tellambura, "Computation of the Continuous-Time PAR of an OFDM Signal with BPSK Subcarriers," *IEEE Communications Letters,* Vol. 5, No. 5, May 2001, pp. 185-187.

[7] K. Y. Xue, H. W. Yang and S. L. Su, "The Clipping Noise and PAPR in the OFDM System," *Proceedings of WRI International Conference on Communications and Mobile Computing*, Vol. 1, Yunnan, January 2009, pp. 265-269.

[8] H. Gacanin and F. Adachi, "PAPR Advantage of Amplitude Clipped OFDM/TDM," *IEICE Transactions on Communications*, Vol. E91-B, No. 3, March 2008, pp. 931-934.

[9] I. A. Tasadduq and R. K. Rao, "OFDM-CPM Signals for Wireless Communications," *Canadian Journal of Electrical & Computer Engineering,* Vol. 28, No. 1, January 2003, pp. 19-25.

[10] I. A. Tasadduq and R. K. Rao, "PAPR Reduction of OFDM-CPM System Using Multi-Amplitude CPM Signals," *Proceedings of 21st Biennial Symposium on Communications,* Kingston, June 2002, pp. 225-229.

[11] H. Ochiai, "A Novel Trellis-shaping Design with Both Peak and Average Power Reduction for OFDM Systems," *IEEE Transactions on Communications*, Vol. 52, No. 11, November 2004, pp. 1916-1926.

[12] S. Sezginer and H. Sari, "Metric-Based Symbol Predistortion Techniques for Peak Power Reduction in OFDM Systems," *IEEE Transactions on Wireless Communications*, Vol. 6, No. 7, July 2007, pp. 2622-2629.

[13] S. Yang and Y. Shin, "An Adaptive SLM Scheme Based on Peak Observation for PAPR Reduction of OFDM Signals," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E91, No. 1, January 2008, pp. 422-425.

[14] A. D. S. Jayalath and C. Tellambura "SLM and PTS Peak-Power Reduction of OFDM Signals without Side Information," *IEEE Transactions on Wireless Communications,* Vol. 4, No. 5, September 2005. pp. 2006-2013.

[15] C. L. Wang and S. J. Ku, "Novel Conversion Matrices for Simplifying the IFFT Computation of an SLM-Based PAPR Reduction Scheme for OFDM Systems," *IEEE Transactions on Communications*, Vol. 57, No. 7, July 2009, pp. 1903-1907.

[16] L. Yanga, R. S. Chena, K. K. Soob and Y. M. Siub, "An Efficient Sphere Decoding Approach for PTS Assisted PAPR Reduction of OFDM Signals," *International Journal of Electronics and Communications*, Vol. 61, No. 10, November 2007, pp. 684-688.

[17] Y. Zhang, Q. Ni and H.-H. Chen, "A New Partial Transmit Sequence Scheme Using Genetic Algorithm for Peak-to-Average Power Ratio Reduction in a Multi-Carrier Code Division Multiple Access Wireless System," *International Journal of Autonomous and Adaptive Communications Systems*, Vol. 2, No. 1, March 2009, pp. 40-57.

[18] C. Tellambura, "Upper Bound on Peak Factor of *N*-multiple Carriers," *Electronic Letters,* Vol. 33, September 1997, pp. 1608-1609.

[19] N. Y. Ermolova and P. Vainikainen, "On the Relationship between Peak Factor of a Multicarrier Signal and Aperiodic Autocorrelation of the Generating Sequence," *IEEE Communications Letters*, Vol. 7, No. 3, March 2003, pp. 107-108.

[20] A. Ghassemi and T. A. Gulliver, "Low Autocorrelation Fractional PTS Subblocking for PAPR Reduction in OFDM Systems," *Proceedings of 6th Annual Conference on Communication Networks and Services Research*, Nova Scotia, May 2008, pp. 41-45.

[21] C. Y. Hsu and H. Do, "The New Peak-to-Average Power Reduction Algorithm in the OFDM System," *Wireless Personal Communications*, Vol. 41, No. 4, June 2007, pp. 517-525.

[22] E. Sun, K. Yi, B. Tian and X. Wang, "A Method for PAPR Reduction in MSE-OFDM Systems," *Proceedings of International Conference on Advanced Information Networking and Applications 2006,* Vienna, Vol. 2, April 2006, pp. 18-20.

[23] D. Wu, S. Predrag and S. Ivan, "Ternary Complementary Sets for Multiple Channel DS-UWB with Reduced Peak-to-Average Power Ratio," *Proceedings of IEEE GLOB ECO,* Texas, November 2004, pp. 3230-3234.

[24] A. R. Ahmad, O. Basir and K. Hassanein, "An Intelligent Expert Systems Approach to Layout Decision Analysis and Design Under Uncertainty," Springer-Verlag, 2008, pp. 312-365.

[25] A. R. Ahmad, "An Intelligent Expert System for Decision Analysis & Support," Ph.D. Thesis, University of Waterloo, Waterloo.

[26] I. M. Hussain and I. A Tasadduq, "PAPR Analysis in OFDM Signals Based on Power Variance," *Proceedings of Wireless Communications, Networking and Mobile Computing*, Dalian, October 2008, pp. 1-4.

[27] M. Negnevitsky, "Artificial Intelligence: A Guide to Intelligent Systems," Pearson, Sydney, 2008.

[28] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, Englewood Cliffs: Prentice Hall, 1989.

Appendix

As indicated in (17), the normalized PV expression contains $N-1$ AAC terms. AAC term can be represented by (12). In case of constant envelope mapper (*e.g.* QPSK which is used throughout our simulations), (12) can be modified into:

$$R_i = \sum_{k=0}^{N-1-i} \frac{d_k}{d_{k+i}}; \quad 0 \le i \le N-1$$

$$R_i = \left(\frac{d_0}{d_i} + \frac{d_1}{d_{i+1}}\right) + \left(\frac{d_2}{d_{i+2}} + \frac{d_3}{d_{i+3}}\right) + \cdots + \frac{d_{N-i-1}}{d_{N-1}} \quad (22)$$

Hence each expression of R_i contains $N-i$ terms of kind (d_k / d_{k+i}). Using (22) and (17) it can easily be shown that the total number of such terms in the power variance expression is $\sum_{i=1}^{N-1} N-i$. Similarly, the total number of pairs of adjacent terms in R_i as indicated by parenthesis in (22) is $N-i/2$. Now one way to reduce AAC to their minimum values (*i.e.* 0 or 1 depending on i) is to make the summation of each and every pair of terms in (22) equal to zero. Let $\left(\dfrac{d_p}{d_{p+i}} + \dfrac{d_{p+1}}{d_{p+i+1}}\right)$ be any pair in (22) where $p = 0,2,4,\ldots,N-2-i$ *(for even i)* and $p = 0,2,4,\ldots,N-3-i$ *(for odd i)*. In case of $i = 1,2,3,\ldots,N-1$, to minimize the values of all such pairs to zero, we have to multiply every second complex number (*i.e.* subcarrier) starting from d_{p+i} by a factor $g_{i,v}$. This factor is obtained as shown below:

For the first $i/2$ number of pairs:

$$\left(\frac{d_p}{g_{i,v} \times d_{p+i}} + \frac{d_{p+1}}{d_{p+i+1}}\right) = 0 \Rightarrow g_{i,v} = -\frac{d_p d_{p+1+i}}{d_{p+1} d_{p+i}}; \quad (23)$$

For rest of the pairs when i is even:

$$g_{i,v} = -\frac{\overline{d}_p d_{p+1+i}}{d_{p+1} d_{p+i}}; \quad (24)$$

For rest of the pairs when i is odd

$$g_{i,v} = -\frac{d_p d_{p+1+i}}{\overline{d}_{p+1} d_{p+i}}; \quad (25)$$

where

$$\overline{d}_p = d_p \times g_{i,v-i/2}, \overline{d}_{p+1} = d_{p+1} \times g_{i,v-i/2} \text{ and } v = \frac{p}{2}+1.$$

For example, in case of $N = 16$ subcarriers and $i = 4$, the fourth AAC becomes:

$$R_4 = \left(\frac{d_0}{d_4} + \frac{d_1}{d_5}\right) + \left(\frac{d_2}{d_6} + \frac{d_3}{d_7}\right) + \left(\frac{d_4}{d_8} + \frac{d_5}{d_9}\right) + \left(\frac{d_6}{d_{10}} + \frac{d_7}{d_{11}}\right)$$
$$+ \left(\frac{d_8}{d_{12}} + \frac{d_9}{d_{13}}\right) + \left(\frac{d_{10}}{d_{14}} + \frac{d_{11}}{d_{15}}\right)$$

This expression has 6 pairs *i.e.* $16-4/2$ and as many factors and each factor is used to minimize the corresponding pair to zero. Since i is even, then $p = 0,2,4,\ldots,10$. For instance, the third pair in R_4 is $\left(\dfrac{d_4}{d_8} + \dfrac{d_5}{d_9}\right)$ where $v = 3$ and $p = 4$. In order to find the six factors used to minimize each pair, (23) is used for the first 2 pairs *i.e.* $g_{4,1} = -\dfrac{d_0 d_5}{d_1 d_4}$ and $g_{4,2} = -\dfrac{d_2 d_7}{d_3 d_6}$. Rest of the factors are calculated using (24) as i is even. Hence, $g_{4,3} = -\dfrac{\overline{d}_4 d_9}{d_5 d_8}$, $g_{4,4} = -\dfrac{\overline{d}_6 d_{11}}{d_7 d_{10}}$, $g_{4,5} = -\dfrac{\overline{d}_8 d_{13}}{d_9 d_{11}}$ and $g_{4,6} = -\dfrac{\overline{d}_{10} d_{15}}{d_9 d_{14}}$, where $\overline{d}_4 = d_4 \times g_{4,1}$, $\overline{d}_6 = d_6 \times g_{4,2}$, $\overline{d}_8 = d_8 \times g_{4,3}$ and $\overline{d}_{10} = d_{10} \times g_{4,4}$. Further, each alternate complex number starting from d_4 (*i.e.* d_4, d_6, d_8, d_{12}, d_{14}, d_{16}) will be multiplied by $g_{4,1}$, $g_{4,2}$, $g_{4,3}$, $g_{4,4}$, $g_{4,5}$ and $g_{4,6}$ respectively. In this manner $|R_4|$ will be minimized to zero. It is clear that $\overline{d}_p$ or $\overline{d}_{p+1}$ means that before minimizing the corresponding pair to zero, updation of d_p into $\overline{d}_p$ has to take place first. In the above example, for the third pair, updation of d_4 (*i.e.* $\overline{d}_4 = d_4 \times g_{4,1}$) has to take place first after which minimization of that pair takes place. Thus, the third pair has the form $\left(\dfrac{\overline{d}_4}{d_8} + \dfrac{d_5}{d_9}\right)$ rather than $\left(\dfrac{d_4}{d_8} + \dfrac{d_5}{d_9}\right)$. Now in case of $i = 1$ *i.e.* R_1 which is an exception case, factor expression is given by (23) for the first pair and for the rest of the pairs, it is given by (24) where $\overline{d}_p = d_p \times g_{i,v-1}$. Further, each complex number starting from d_1 is going to be multiplied by the corresponding factor. In this manner, we can minimize the coefficients of any sequence to their respective minimum values using the procedure above.

◆◆ Scientific
◆◆ Research

An Efficient Noisy-ICA Based Approach to Multiuser Detection in IDMA Systems

Abdelkrim Hamza[1,2] , Salim Chitroub[1], Gérard Salut[2]

[1]*LCPTS Laboratoire de Communication et du Traitement du Signal, FEI-USTHB, Alger, Algeria*
[2]*Laboratoire d'Analyse et d'Architecture des Systèmes-CNRS;LAAS, F-31077 Toulouse, France*
E-mail: *ahamza@usthb.dz*

Abstract

Interleaved Division Multiple Access (IDMA) is a new access scheme that has been proposed in the literature to increase the capacity of wireless channels. The performance of such systems depends on the accuracy of the channel state information at the receiver. In this paper, a Noisy-Independent Component Analysis (N-ICA) based IDMA receiver for multiple access communication channels is proposed. The N-ICA component is applied as a post processor. Unlike other IDMA receivers, the proposed scheme detects and separates the transmitted symbols without channel state information tracking. The performance of the proposed technique is presented in terms of raw bit error rate (BER) without channel coding for different signal to noise ratios (SNR). Simulation results demonstrate that N-ICA post processor provides an improvement in performance in BER in loaded systems. When the system is not loaded, the proposed post processor has the same performance as conventional IDMA receiver with less iterations leading to a complexity reduction.

Keywords: IDMA, ICA, Multi-User Detection

1. Introduction

One of the challenges of next generation wireless systems is the spectral efficiency. The goal is to provide sufficient quality and capacity to the diverse and rich multimedia content that need to be transmitted. Recently proposed Interleave-Division Multiple-Access (IDMA) communication system is one of the most promising technologies for high data rate wireless networks [1]. IDMA can be regarded as a special case of Code Division Multiple Access (CDMA). In CDMA systems, users are separated using signatures or spreading codes; while in IDMA systems, distinct interleavers are employed to separate users [2].

At the receiver of IDMA system, the signal of the user of interest needs to be extracted out of the Multiple Access Interference (MAI) and the Inter Symbol Interference (ISI). Moreover, since conventional IDMA detector is sensitive to channel estimation errors [3], a good channel tracking algorithm is mandatory. This might drastically increase the overall complexity at the receiver. To overcome those drawbacks, in this paper, we propose a new blind receiver for IDMA systems. In our approach, a

noisy Independent Component Analysis (N-ICA) scheme is introduced as a post processor. Independent Component Analysis (ICA) has attracted special attentions in the wireless communication fields for interference suppression of CDMA systems [4,5].

In this paper, we propose to detect and separate the transmitted symbols without channel tracking and by including the noise in the global model; leading to the N-ICA model. We will show that our model is very suitable for symbol detection and separation in the IDMA context. In terms of complexity, as a post processor, the proposed solution starts the processing just after conventional IDMA processing. In this case, a hardware reuse is possible if an FPGA implementation is carried on for example (currently being finalized). Therefore, the proposed N-ICA block does not represent a complexity increase in the overall system.

The remainder of this paper is organized as follows. The next section is devoted to the IDMA system model. In Section 3, we detail the proposed N-ICA model for IDMA. In Section 4, an estimation algorithm is presented for N-ICA in an IDMA context. Using some evaluation criteria, computer simulation results are presented in Section 5 to

provide a comparative study. Conclusions are drawing in Section 6.

2. IDMA System Model

As shown in **Figure 1** (the upper plot is the transmitter and the lower one is the receiver), we consider an IDMA system with K users. A single path channel h_k and BPSK modulation are considered here. The *nth* bit in the sequence d_k of *k*th user is spread, generating a sequence vector denoted $c_k = [c_k(1), c_k(2), ..., c_k(J)]^T$ where J is the frame length, C is the spreading factor and the superscript T is the transpose operator. Then c_k is permuted by an interleaver π_k and at the output of the interleaver, the vector $x_k = [x_k(1), x_k(2), ..., x_k(J)]^T$ is obtained. The elements in c_k and x_k are considered as chips. The chip rate is C times higher than the bit rate. Users are distinguished mainly by their respective interleavers π_k. Each user can have its own signature sequence or all users can share the same spreading code [3]. The received signal can be modeled as:

$$r(j) = \sum_{k=1}^{K} h_k x_k(j) + n(j) \qquad j=1...J \qquad (1)$$

As illustrated in figure 1, this sub optimal receiver consists of an elementary signal estimator (ESE) and K single user a posteriori probability (APP) decoders (DEC) [1]. The multiple access and coding constraints are considered separately in the ESE and DEC. The outputs of the ESE and DEC are extrinsic log likelihood ratios (LLRs) about $\{x_k(j)\}$ defined in [2]. These LLRs are further distinguished by subscripts, $e_{ESE}(x_k(j))$ and

Figure 1. IDMA-N-ICA system model.

$e_{DEC}(x_k(j))$, depending on whether they are generated by the ESE or DEC. A global turbo-type iterative process is then applied to process the LLRs generated by the ESE and DEC, as detailed below [1,3].

$$e(x_k(j)) = \log \frac{\Pr(r|x_k(j) = +1, h_k)}{\Pr(r|x_k(j) = -1, h_k)} \forall k, j \qquad (2)$$

2.1. The Basic ESE

The Equation (1) can be rewritten as:
$$r(j) = h_k x_k(j) + \xi_k(j) \qquad (3)$$
$$\xi_k(j) = r(j) - h_k x_k(j) \qquad (4)$$
where ξ_k is the distortion including interference plus noise in r(j) with respect to user-k [3]. The mean and the variance functions are noted by E(.) and Var(.) respectively.

$\xi_k(j)$ can be approximated by a random Gaussian variable.

with mean and variance :
$$E(\xi_k(j)) = E(r(j)) - h_k E(x_k(j)) \qquad (5)$$
$$Var(\xi_k(j)) = Var(r(j)) - |h_k|^2 Var(x_k(j)) \qquad (6)$$
where

$$E(r(j)) = E\left(\sum_{k=1}^{K} h_k x_k(j)\right) + E(n(j)) \qquad (7)$$

$$Var(r(j)) = \sum_{k=1}^{K} |h_k|^2 Var(x_k(j)) + \sigma^2 \qquad (8)$$

Therefore, the log likelihood ratio [1] is given by:

$$e_{ESE}(x_k(j)) = 2.h_k(r(j) - E(\xi_k(j))) / Var(\xi_k(j)) \qquad (9)$$

2.2. The DEC Function

The Dec in our structure performs despreading operation and the extrinsic LLRs $e_{DEC}(c_k(j))$ are used to update $E(x_k(j))$ and $Var(x_k(j))$ as [3]:
$$E(x_k(j)) = \tanh \tanh e_{DEC}(x_k(j)) / 2 \qquad (10)$$
$$Var(E(x_k(j))) = 1 - E(x_k(j))^2 \qquad (11)$$
This iterative process is repeated a preset number of times.

2.3. The ESE Function for Multi-Path Channels

When we consider a multipath fading channel with L paths; the received signal is represented by:

$$r(j) = \sum_{k=1}^{K} \sum_{l=0}^{L-1} h_{k,l} x(j-l) + n(j) \forall j \qquad (12)$$

where $h_{k,l}$ are the coefficients related to user k.

Following a similar principle as that for single path we obtain Algorithm 1 below for detection in a multipath channel:

Algorithm 1: Detection in a Multi-Path Channel
1) Estimation of interference Mean

$$E(r(j)) = \sum_{k=1}^{K} \sum_{l=0}^{L-1} h_{k,l} E(x(j-l)) \qquad (13)$$

$$E\left(\xi_{k,l}(j)\right) = E(r(j+l)) - h_{k,l} E\left(x_k(j)\right) \qquad (14)$$

2) Estimation of interference variance

$$Var(r(j)) = \sum_{k=1}^{K} \sum_{l=0}^{L-1} \left|h_{k,l}\right|^2 Var\left(x(j-l)\right) + \sigma^2 \qquad (15)$$

$$Var\left(\xi_{k,l}(j)\right) = Var\left(r(j+l)\right) - \left|h_{k,l}\right|^2 Var\left(x_k(j)\right) \qquad (16)$$

3) LLR generation

$$e_{ESE}(x_k(j))_l = 2.h_{k,l} \frac{r(j+l) - E\left(\xi_{k,l}(j)\right)}{Var\left(\xi_{k,l}(j)\right)} \qquad (17)$$

$$e_{ESE}\left(x_k(j)\right) = \sum_{l=0}^{L-1} e_{ESE}(x_k(j))_l \qquad (18)$$

3. ICA and N-ICA Principle

The application of ICA consists of estimating the unknown input signals from the output signals without prior knowledge of the channel state information [6].

Let's suppose that the sources are statistically independent. This is a fundamental assumption for using ICA that is generally verified in communication systems [6].

The extraction of the sources can be done by ICA by exploiting the essential features of the sources and system [7]. In the simplest form of ICA, we observe n scalar variables $r_1, r_2, \dots r_n$ which are linear combinations of l unknown independent sources or components ICs denoted by $b_1, b_2, \dots, b_l$.

If we express the observed random variables with the vector, $r = (r_1, r_2, \dots, r_n)^T$ and the ICs variables b_j with the vector $b = (b_1, b_2, \dots, b_l)^T$ then the relationship is given by [8]:

$$r_m = Gb_m + n_m \qquad (19)$$

where r_m is the *mth* observed data vector, G is an unknown full rank mixing matrix, b_m is an unknown non Gaussian source vector and n_m is an additive Gaussian noise process.

The goal is to estimate the noise free ICs b_m using only the observations r_m and the assumption of the independence of the sources. This means that a set of vectors w_1, w_2, should be estimated such that $W = [w_1, w_2, \dots]$ is the separating matrix; therefore, the output source estimations $w_1^T . r_m, w_2^T . r_m, \dots$ *i.e.*:

$$y_m = W^T r_m \qquad (20)$$

are independent and each of them can be used to represent one of the sources.

3.1. Mathematical Representation of IDMA by N-ICA Model

In this subsection, we develop the theoretical framework and show the similarity between Noisy ICA model and IDMA system model. We focus our atten-

tion on synchronous IDMA systems for simplicity and brevity. However, the method can be extended to an asynchronous system by extending the observation interval.

After chip rate sampling *i.e. C* equal spaced samples per symbol are taken, the sampled data is processed within a window of specific size. For synchronous model, data propagated through a single path channel fall into the same window of size T_b for desired and interfering symbols.

The samples are then collected into a *C×1* vectors r_m.

$$r_m = \sum_{k=1}^{K} h_k d_{k,m} s_k + n_m \qquad (20)$$

here s_k is the *C×1* vector representation of *kth* user's interleaved signature sequence and n_m denotes the noise vector.

The last equation can be rewritten in a matrix form:

$$b_m = [d_{1,m}, d_{2,m}, \dots, d_{K,m}]$$
$$s_1 = [s_{1,1}, s_{2,1}, \dots, s_{C,1}]^T \quad C \times 1 \text{ vector}$$

$$r_m = [s_1, \cdots, s_K] \begin{bmatrix} h_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_K \end{bmatrix} \begin{bmatrix} b_{1,m} \\ \vdots \\ b_{K,m} \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_K \end{bmatrix}$$

$$r_m = [s_1 h_1, \cdots, s_K h_K] b_m + n_m \qquad (22)$$

Equation (22) can be represented in a more compact form:

$$r_m = Gb_m + n_m \qquad (23)$$

where the *C×K* matrix *G* is assumed full rank. We can see the similarity between the IDMA model of Equation (23) and the N-ICA model of Equation (19). The goal of the Noisy-ICA based IDMA detection is to recover the symbol vector b_m for each user k without knowing the parametric form of *G* which depends on the channel coefficients.

4. N-ICA Estimation Algorithm

The proposed system is a hybrid structure composed of two parts where a classical IDMA receiver is combined with a N-ICA block as shown in figure 1. Block IDMA, described in the previous section, works for a number of iterations (*it*) after which the block N-ICA takes over. The proposed N-ICA will act as a post processor attached to an IDMA receiver in the presence of noise. The aim of our N-ICA block is to avoid continuous tracking of channel state information [9]. In this section, we will derive estimation algorithms for the proposed N-ICA post processor in IDMA context.

4.1. Principal Component Analysis Processing

The Principal Component Analysis (PCA) based part of the model consists of whitening the input signals. This

step of processing is achieved by using the PCA in [10] to extract the Principal Components (PCs). This can be done for the noiseless case as follows:

$$Y = \Lambda^{-1/2} U^T G B \qquad (24)$$

where the matrix U corresponds to the Eigen vector of the data covariance matrix C and the diagonal matrix Λ that contains the related Eigen values λ:

$$\Lambda^{-1/2} = diag\ [\ \lambda_1^{-1/2}, \lambda_2^{-1/2}, \cdots, \lambda_n^{-1/2}\ ] \qquad (25)$$

This PCA processing can be extended to noisy data using bias removal technique [8]. In the regular ICA process, the covariance matrix of the noise free data r_m^{nf} can be given by:

$$C = E\left\{ r_m^{nf} (r_m^{nf})^T \right\} = GG^T \qquad (26)$$

On the other hand, the covariance matrix of the observed noisy data can be written as:

$$\Gamma = E\left\{ r_m\, r_m^T \right\} = GG^T + \sigma^2 I = C + C_n \qquad (27)$$

where σ^2 is the noise power and C_n is the diagonal noise covariance matrix. In the noise bias removal technique, the Eigen values and vectors of matrix $\Gamma - C_n$ is used for whitening instead of matrix Γ which is called quasi-whitening [10].

In fact, quasi whitening can be performed on the noisy data as follows:

$$z = (\Lambda - \sigma^2 I)^{-1/2} U^T r_m \qquad (28)$$

The covariance matrix of quasi white data is then given by:

$$E\left\{ z\, z^T \right\} = I + \sigma^2 (\Lambda - \sigma^2 I)^{-1} \qquad (29)$$

From (29), we notice that the covariance matrix is different from the identity matrix. Therefore, we have to take into account the non-whiteness of the data.

This is achieved by using the fast ICA algorithm that is presented in the next subsection.

4.2. Fast ICA Algorithm

Since only the second order statistics are used to compute the matrix, the PCA used in the first part of the model does not provide the best results. Higher order statistics of the received signals contain additional information about the non Gaussian properties of the noise.

The purpose of this work is to establish a new scheme in which the system can take into account such random deformations in the detection step. To improve the performance, the presence of the noise should be reduced to the minimum using the extracted PCs without additional prior knowledge of their statistical properties. This is the purpose of the ICA based part of the model. Therefore, the ICA model should include a noise term as well in its linear transform matrix.

The ICA approach that we present here is our contribution to take into account the noise in the ICA model. This means that the bias due to noise should be removed, or at least reduced.

Noisy-FastICA has been applied to the blind source separation and interference suppression in multiple access communications before in [11] and FastICA in [9], [10]. The Noisy-FastICA algorithm performs as follows (Algorithm 2):

Algorithm 2: Noisy-FastICA

Let k be the desired user, r_m, $m = 1, .., M$ the received block data and b denotes the estimate of the symbol b.

1) First perform PCA for dimension reduction
2) Quasi- whitened the noisy data using (28)
3) Start ICA

Let t=1 and update

$$w(t) = \Gamma^{-1} E\left\{ z_m (w(t-1)^T z_m)^3 \right\} - 3E\left\{ (w(t-1)^T z_m)^2 \right\} w(t-1)$$

where $\Gamma = I + \sigma^2 (\Lambda - \sigma^2 I)^{-1}$

Normalize $w(\text{t}): w(t) = w(t) \Big/ \sqrt{\mathbf{w}(\text{t})^T \Gamma w(t)}$

If $| w(t)^T w(t-1) | < (1 - 10^{-4})$, let $t= t+1$ and go to step 3.

4. Output the estimated desired user's bit: $\hat{b}_{k,m} = sgn$ (Z_m)

5. Simulation Results

In this section, we present the simulation results of the proposed Noisy-Independent Component Analysis (N-ICA) based Interleaved Division Multiple Access (IDMA) presented in this paper. In all simulation results, the following notations have been adopted:

It is the iterations number used in IDMA block. T represents the percentage of load rate defined by the ratio between the number of users and the spreading factor (K/C). IDMA is the conventional IDMA receiver described in Section 2.

IDMA-ICA is the hybrid structure described in [12]. IDMA-N-ICA is the proposed hybrid structure described in Section 3 using the noisy fast ICA algorithm.

To evaluate the detection and separation ability of the proposed N-ICA model, performances are presented in terms of raw Bit Error Rate (BER) before decoding for different Signal to Noise Ratios (SNR). We consider a time varying channel, BPSK modulation and Gold spreading codes of length C. Among the parameters that influence the performances are the effect of load rate and the number of iterations for IDMA block. The obtained results are presented in **Figures 2**-**6.**

In **Figures 2** and **3** we show a comparison between our proposed post processor N-ICA and the IDMA receiver for single path and multipath channel respectively.

In **Figure 4**, performances of our proposed receiver are presented for different values of τ (rate of load) and a spreading factor of 63.
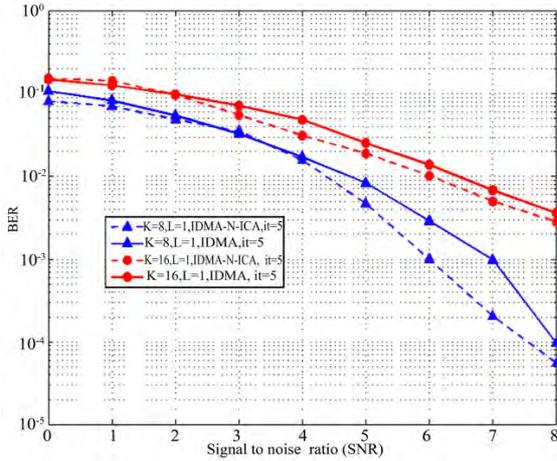
Figure 2. Performance comparison between IDMA-N-ICA and IDMA systems in single path.
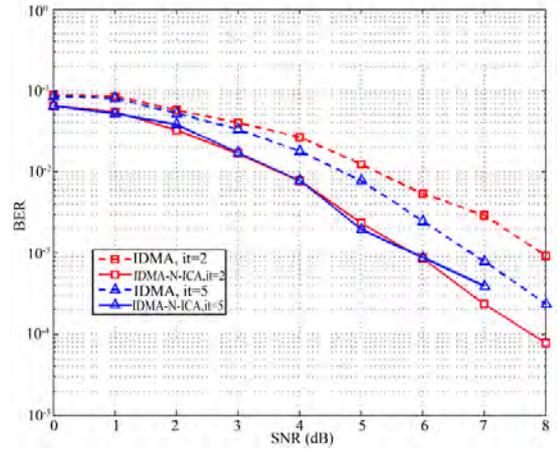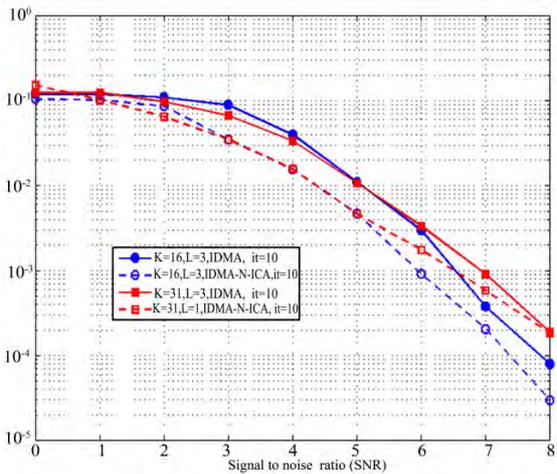

Figure 3. Performance comparison between IDMA-N-ICA and IDMA systems in single path.
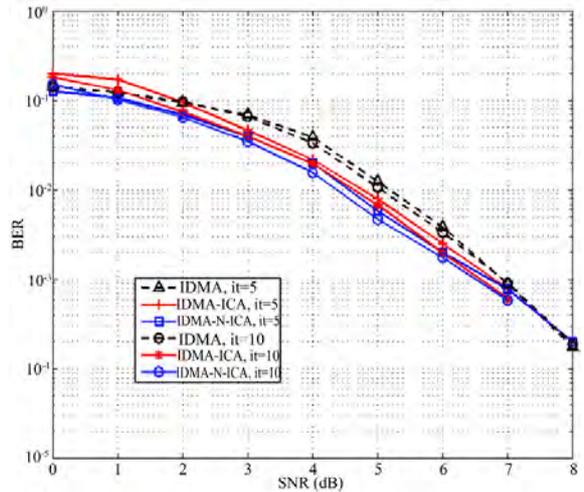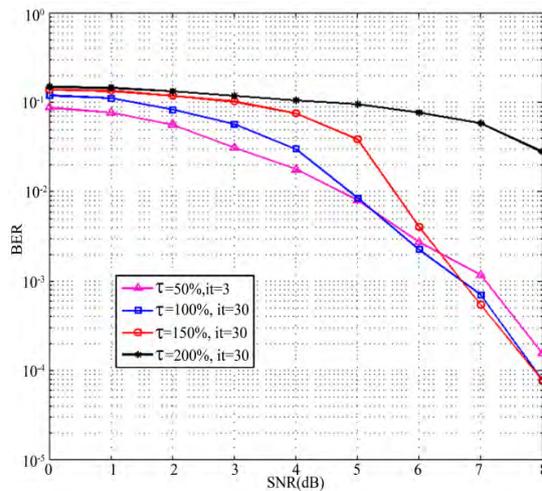

Figure 4. IDMA-N-ICA performance comparison for different rate load and C = 63.


**Figure 5. Performance comparison between IDMA-N-ICA and IDMA receiver when c = 31 and τ = 50%.**

Figure 6. Performance comparison between IDMA-N-ICA and IDMA receiver when c = 31 and τ = 100%.

We notice that our proposed scheme handles very well the MAI interferences since convergence is warranted even at very loaded systems (τ100%).

Figure 5 shows the added value of our proposed post-processor N-ICA when compared to the conventional IDMA receiver for loading rate 50% and for a spreading factor of 31. We notice that both convergence speed and better BER performances are achieved. Therefore, the proposed N-ICA approach can be employed in high loading rate in order to improve the performance of the system in terms of quality of service. Moreover, in case of low loading rate (50%), the proposed post processor allows a reduction in the number of iterations needed by the IDMA block leading to complexity reduction of the overall receiver.

In the last simulation scenario, we evaluate the added value of the noisy ICA post processor over the ICA post processor. **Figure 6** provides a comparison between

IDMA, IDMA-ICA and IDMA-N-ICA receivers when the spreading factor is 31 and the load rate is 100%. When SNR is low, N-ICA outperforms the ICA post processor. However, when SNR is high, both receivers present the same performance. These observations are expected since N-ICA takes into account the presence of noise.

It is worth noting also that both IDMA-ICA and IDMA-N-ICA receivers outperforms the conventional IDMA receiver.

6. Conclusions

In this paper, N-ICA post processor is proposed in IDMA context. N-ICA algorithm constitutes an efficient tool for symbol recovery and it offers an efficient alternative to the IDMA systems with block channel estimation.

The major contribution of this work is the application of blind detection technique in the IDMA context. The proposed algorithm has better performance compared to the IDMA receiver in loaded systems because it allows dimension reduction (PCA) which helps to reduce the amount of noise in the system.

For unloaded systems, the proposed post processor allows a complexity reduction by reducing the number of iterations needed by the IDMA block. In future work, to better analyze the complexity of the proposed scheme, FPGA implementation of IDMA and proposed post processor will be realized.

7. References

[1]   P. Li, L. Liu, K. Wu and W. K. Leung, "Interleave Division Multiple Access," *IEEE Transactions on Wireless Communication*, Vol. 5, No. 4, 2006, pp. 938-945.

[2]   I. Mahafeno, C. Langlais and C. Jego, "Reduced Complexity Iterative Multi-User Detector for IDMA (Interleave Division Multiple Access) System," *Proceedings of IEEE Global telecommunication Conference*, Vol. 1, San Francisco, 2006, pp. 530-540.

[3]   H. Schoeneich and P. Hoeher, "Semi-Blind Pilot-layer Aided Channel Estimation with Emphasis on Interleave Division Multiple Access Systems,'' *Proceedings of IEEE Global telecommunication Conference*, Vol. 3, St. Louis, 2005, pp. 110-118.

[4]   T. Ristaniemi and J. Joutsensalo, "Advanced ICA Based Receivers for DS-CDMA Systems," *Signal Processing*, Vol. 82, 2002, pp. 417-431.

[5]   M. Gupta and B. Santhanam, "ICA Based Blind Adaptive MAI Suppression in DS-CDMA Systems," *Proceedings of IEEE-DSP Workshop*, New Mexico, 2004, pp. 204-210.

[6]   J.-F. Cardoso, "Blind Signal Separation: Statistical Principles," *Proceedings of the IEEE*, Vol. 86, No.10, June 1978, pp. 2009-2025.

[7]   A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, Vol. 10, No. 3, 1999, pp. 626-634.

[8]   T. Huovinen, A. Shahed and M. Valkama, "Blind Diversity Reception and Interference Cancellation Using ICA," *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, 2007, pp. 685-688.

[9]   A. O. Dahmane, "A New MMSE Based Cascade Fliter MUD Tracking Mode in Time-Varying Channels," *Proceedings of IEEE Conference on Wireless and Mobile Computing, Networking and Communications*, Vol.5, 2005, pp. 103-107.

[10] M. E. Davies, "Identifiability Issues in Noisy ICA," *Proceedings of IEEE Signal Processing letters*, Vol. 11, No. 5, 2004, pp. 470-473.

[11] O. Ekici and A. Yongacoglu, "Application of Noisy Independent Component Analysis for CDMA Signal Separation, *Proceedings of IEEE-DSP Workshop*, Vol. 3, July 2004, pp. 3812-3816.

[12] A. Hamza, G. Salut and S. Chitroub, "Independent Component Analysis in IDMA Systems," *Proceedings of IEEE NEWCAS-TAISA*, Toulouse, July 2009, pp. 64-68.

[13] R. H. Mahadevappa, "Interference Cancellation Technique for Multi-User Communication System", PhD thesis, Northeaster University, Massachusetts, 2002.

◆◆ Scientific
◆◆ Research

Modulation Index Estimation of Frequency and Phase Modulated Signals

Geng Peng[1,2], Derong Cai[1], Zhiqiang He[1], Zhitao Huang[2]
[1]*Naval Academy of Armament, Beijing, China*
[2]*School of Electronic Science and Engineering, National University of Defense Technology, Hunan, China*
E-mail: *Andrew.higgins@csiro.au, Leonie.pearson@csiro.au, Luis.laredo@csiro.au*

Abstract

Modulation index estimation is important in the demodulation and recognition of Angle Modulation (AM) signals which include Frequency Modulation (FM) and Phase Modulation (PM) signals. In this paper, we firstly analyzed the AM signals with baseband modulation types, such as monotone, PSK, FSK, in the time and frequency domains. Then, we established a unified mathematical representation for the AM signals. Finally, we derived a blind estimation algorithm for the modulation index without using any prior knowledge. Simulation results verified the capabilities of the proposed algorithm.

Keywords: Information Processing Technology, Blind Estimation, Modulation Index, AM Signal

1. Introduction

Angle Modulation (AM), a general form of Frequency Modulation (FM) and Phase Modulation (PM), is an important nonlinear analogue modulation mode used in various communication systems. AM signal transmits information by changing (switching) carrier frequencies or phases according to the variation of the baseband signal. The modulation recognition algorithms for the FM and PM signals can be found in [1-7]. Besides, the modulation index ranges considered in these literatures are listed in **Table 1**.

It can be seen from **Table 1** that the current analogue FM/PM modulation recognition algorithms essentially use the differences in modulation indices. Once these indices are effectively estimated, the FM/PM analogue signals can be identified according to its range in the practical application. As a result, the AM signals can be demodulated blindly based on the modulation index and modulation type. Therefore, the blind estimation of the AM index is of great significance.

As far as we know, there are few researches on the modulation index estimation yet. In this paper, we analyze the AM signal with baseband signals, such as monotone, FSK, PSK, in the TF domain, and establish a unified mathematical expression to help propose an AM index estimation method, which requires no prior knowledge of the signals.

The remainder is arranged as follows: the signal model is

established in Section 2; in Section 3, the blind estimation method for the modulation index is derived; the simulation experiments and conclusions are given in Section 4 and Section 5, respectively.

2. Signal Model

AM signal is usually expressed as follows [8]:

$$s(t) = A_c \cos\left[w_c t + \varphi(t)\right] \qquad (1)$$

where A_c is a constant carrier amplitude, $w_c = 2\pi fc$, $\varphi(t)$ is the instant phase offset.

For the PM signal, the instant phase offset proportionally changes with the baseband modulation signal, namely:

$$\varphi(t) = k_p m(t) \qquad (2)$$

$$s_{\text{PM}}(t) = A_c \cos\left[w_c t + k_p m(t)\right] \qquad (3)$$

Table 1. Modulation index ranges of analogue FM/PM modulation recognition algorithms in literatures.

literatures	phase modulation index m_p of PM signals	frequency modulation index m_f of FM signals
Ref. [1]	$[-\pi, \pi]$	$(1.0, 2.0, 5.0)\times 2\pi$
Ref. [2]	$\leq \pi/2$	$5.0\times 2\pi$
Ref. [3-4]	≤ 1.5	≥ 5.0
Ref. [5-6]	$\pi/(18, 12, 9)$	$(1, 5, 10)\times 2\pi$
Ref. [7]	$m_p \gg m_f$	

where k_p is a constant phase offset, $m(t)$ is the baseband modulation signal.

For the FM signal, the change rate of the instant phase offset varies with the baseband signal proportionally, namely:

$$\varphi(t) = 2\pi k_f \int_{-\infty}^{t} m(\tau)\, d\tau \qquad (4)$$

$$s_{\text{FM}}(t) = A_c \cos\left[w_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)\, d\tau \right] \qquad (5)$$

where k_f is a constant frequency offset, which is also named as the frequency sensitivity of a frequency modulator.

3. Blind AM Index Estimation Method

In this section, we first analyze the AM signals with different baseband modulation types in the TF domain. Then, we establish a unified mathematical expression of these signals and deduce a blind estimation method for the AM index.

3.1. The TF Analysis of the AM Signals

The AM signal with monotone baseband is analyzed in the TF domain first. Then, the baseband modulation types, such as PSK, FSK, are also taken into account. At last, the uniform mathematical expression of the AM signal is obtained.

For the monotone baseband denoted as $m(t)=V_m\cos(\Omega t)$ ($\Omega = 2\pi f_m$), the $s(t)$ can be given by

$$s(t) = \begin{cases} A_c \cos\left[w_c t + m_p \cos(\Omega t) \right], & \text{PM} \\ A_c \cos\left[w_c t + m_f \sin(\Omega t) \right], & \text{FM} \end{cases} \qquad (6)$$

$$\begin{cases} m_p = k_p \cdot V_m \\ m_f = k_f \cdot V_m / f_m \end{cases} \qquad (7)$$

where m_p and m_f are the phase modulation index and the frequency modulation index respectively [8].

The PM signal can be rewritten in an exponential form as

$$s_{\text{PM}}(t) = A_c \cdot \text{Re}\left[e^{jw_c t} \cdot e^{jm_p \cos\Omega t} \right] \qquad (8)$$

where $\text{Re}[x(t)]$ is the real part of the function $x(t)$, $e^{jm_p \cos\Omega t}$ is the periodic function of Ω. Then, the Fourier expansion of the PM signal is:

$$\begin{aligned} s_{\text{PM}}(t) &= A_c \cdot \text{Re}\left[e^{jw_c t} \cdot e^{jm_p \sin(\Omega t + \pi/2)} \right] \\ &= A_c \cdot \text{Re}\left\{ \sum_{n=-\infty}^{\infty} J_n(m_p) e^{j[(w_c + n\Omega t)t]} e^{j(n\pi/2)} \right\} \end{aligned} \qquad (9)$$

where $J_n(x)$ is type I Bessel function of x with an order n.

Considering the baseband signal $m(t) = V_m\cos(\Omega t)$, FM and PM signals can be expressed in frequency domain as:

$$\begin{cases} \text{PM: } \sum_{n=-\infty}^{\infty} \left\{ \delta[f - f_0] + \delta[f + f_0] \right\} \cdot J_n(m_p)\cos(n\pi/2) \\ \text{FM: } \sum_{n=-\infty}^{\infty} \left\{ \delta[f - f_0] + \delta[f + f_0] \right\} \cdot J_n(m_f) \end{cases} \qquad (10)$$

where $f_0 = f_c + n f_m$.

Similarly, FM and PM signals with the baseband signal $m(t) = V_m\sin(\Omega t)$ can be expressed in frequency domain as:

$$\begin{cases} \text{PM: } \sum_{n=-\infty}^{\infty} \left\{ \delta[f - f_0] + \delta[f + f_0] \right\} \cdot J_n(m_p) \\ \text{FM: } \sum_{n=-\infty}^{\infty} \left\{ \delta[f - f_0] + \delta[f + f_0] \right\} \cdot J_n(m_f)\cos(n\pi/2) \end{cases} \qquad (11)$$

It can be seen from (6), (10) and (11) that the FM and PM signals with monotone baseband are essentially identical in the TF domain.

For the PSK baseband case, $m(t)$ is represented as:

$$m(t) = g(t)\cos\left[2\pi f_m t + 2\pi(i-1)/M \right], \ 0 \le t \le T_c \qquad (12)$$

where $g(t)$ is the response function of the shaping pulse, $i = 0, \cdots, M-1$, T_c is the time interval of the code chip.

For the FSK baseband case, $m(t)$ is written as:

$$\begin{aligned} m(t) &= g(t)\cos\left[\frac{\pi}{T_c}(n_c + i)t \right] \\ &= g(t)\cos\left(2\pi f_m t + \frac{i\pi}{T_c}t \right), \quad 0 \le t \le T_c \end{aligned} \qquad (13)$$

where $f_m = n_c/(2T_c)$ and n_c is an integer.

Generally, signals such as PSK and FSK take the sinusoidal signals as the carrier. The information is transmitted by varying some of the parameters of the carrier. It can be drawn from (12), (13) that both the PSK and FSK basebands are monotones with amplitude function $g(t)$ in a code period T_c. (When the function is a rectangular pulse, $g(t)=1$). To transmit information, the initial phase and/or frequency for different code intervals are different. So, it is representative to analyze the AM signal with monotone baseband in the TF domain. Accordingly, the subsequent investigation takes the monotone baseband modulation signal as an example.

To facilitate the analysis process, a unified mathematical expression of AM signals with baseband modulations such as monotone, PSK, FSK, is given as follows:

$$s(t) = A_c \cos\left[w_c t + m_p g(t)\cos(\Omega t + \theta_1) + \theta_0 \right], 0 \le t \le T_c \qquad (14)$$

where θ_0 is the initial phase of the carrier, θ_1 is the initial phase of baseband modulation signal and $g(t)$ can be a square pulse.

3.2. Derivation of the Blind Modulation Index Estimation Method

The monotone baseband modulation FM signal shown in (5) is considered here as an example. After amplitude normalization and differentiating, Equation (5) evolves to:

$$\frac{d\,s_{FM}(t)}{dt} = \frac{dA}{dt}\cos\left[w_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right]$$
$$- A\left[w_c + 2\pi k_f m(t)\right]\sin\left[w_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right]$$

(15)

where A is the amplitude of the normalized A_c, *i.e.* $A=1$, then (15) becomes:

$$\frac{d\,s_{FM}(t)}{dt} = -\left[w_c + 2\pi k_f m(t)\right]\sin\left[w_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right]$$

(16)

It can be seen from (16) that the FM signal changed into AM-FM signal after differentiation. Hilbert transformation of the differentiation of the FM signal is:

$$H\left[\frac{d\,s_{FM}(t)}{dt}\right]$$
$$= \left[w_c + 2\pi k_f m(t)\right]\cdot\exp\left\{j\left[w_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau + \pi/2\right]\right\}$$

(17)

Generally, w_c is not always greater than $2\pi k_f m(t)$, then (17) turns into:

$$H\left[\frac{d\,s_{FM}(t)}{dt}\right]$$
$$= \left|w_c + 2\pi k_f m(t)\right|\cdot\exp\left\{j\left[w_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau + \phi(t)\right]\right\}$$

(18)

$$\phi(t) = \begin{cases} \pi/2, & \text{if } w_c + 2\pi k_f m(t) \geq 0 \\ 3\pi/2, & \text{if } w_c + 2\pi k_f m(t) < 0 \end{cases}$$

(19)

The modulus of (17) is $|w_c+2\pi k_f m(t)|$, and $\phi(t)$ can be obtained by calculating the argument of (18) and relieving the illegibility. Interested readers can refer to papers [9,10] for the concrete algorithms.

When the $|w_c+2\pi k_f m(t)|$ and $\phi(t)$ are obtained, $\theta(t)$ can be expressed as:

$$\theta(t) = \left|w_c + 2\pi k_f m(t)\right|\cdot\exp\left\{\phi(t)\right\}$$
$$= w_c + 2\pi k_f m(t)$$

(20)

To achieve an expression with the modulation index, we substitute the monotone baseband modulation signal $m(t) = V_m\cos(2\pi f_m t)$ into (20), subtract $\theta(t)$ by its mean, and make a integration. The following expression can be achieved:

$$\int_0^t 2\pi k_f \left\{m(\tau) - mean\left[m(\tau)\right]\right\}d\tau$$
$$= 2\pi k_f \cdot \frac{V_m}{2\pi f_m}\sin\left(2\pi f_m\tau\right)\Big|_0^t$$

(21)

$$= m_f \cdot \sin\left(2\pi f_m t\right)$$

where m_f is obtained by calculating the amplitude deviation of the sinusoidal signal according to (21). Therefore, the blind estimation of the AM index of FM signal is accomplished.

Using a similar procedure as described in the above section, an integral form for the PM case yields:

$$\int_0^t k_p \left\{m'(\tau) - mean\left[m'(\tau)\right]\right\}d\tau = k_p V_m \cdot \cos\left(2\pi f_m\tau\right)\Big|_0^t$$
$$= m_p \cdot \left[\cos\left(2\pi f_m t\right) - 1\right]$$

(22)

where $m'(t)$ is the differentiation of $m(t)$.

Similarly, m_p can be obtained by calculating the amplitude deviation of the sinusoidal signal according to (22). Therefore, the AM index of PM signal is also achieved.

By now, a blind AM index estimation algorithm is derived. To be clearer, the main steps of this method are summarized as follows:

Step1. Judging whether the baseband modulation of the input AM signal is monotone or not. If not, estimate the code rate of the signal.

Step2. Normalizing the amplitude of the input signal and differentiating the normalized signal.

Step3. Executing the Hilbert transformation.

Step4. Calculating the modulus and solving $\phi(t)$ by removing phase ambiguity.

Step5. Obtaining $\theta(t)$ by (20) with the estimated modulus and phase.

Step6. Subtracting the mean of $\theta(t)$ and calculating the subsequent integral.

Step7. If the baseband is monotone, the modulation index is estimated according to the amplitude deviation between the adjacent integral values; otherwise, the modulation index is estimated in each code interval respectively.

Step8. Averaging previous estimated values to reduce noise influence on blind estimation of modulation index as little as possible.

4. Simulations

Experiments are executed to verify the capabilities of the proposed method. In the experiments, the blind estimation method of the modulation index is realized for the AM signal with baseband modulations, such as monotone, PSK, FSK. Conclusions are drawn according to the simulation results.

4.1. Monotone Baseband Case

The AM signal with monotone baseband is generated. The sampling frequency is 200 MHz, the carrier frequency is 70 MHz, and $V_m = 1$ V. The following six scenarios are conducted:

1) FM signal with $k_f = 50$ kHz/V, $f_m = 40$ kHz, namely m_f=1.25;

2) FM signal with $k_f = 50$ kHz/V, $f_m = 30$ kHz, namely m_f=5/3;

3) FM signal with $k_f = 40$ kHz/V, $f_m = 30$ kHz, namely m_f=4/3;

4) PM signal with $k_p = 0.4$ rad/V, $f_m = 40$ kHz, namely m_p=0.4;

5) PM signal with $k_p = 0.8$ rad/V, $f_m = 40$ kHz, namely m_p=0.8;

6) PM signal with $k_p = 0.8$ rad/V, $f_m = 20$ kHz, namely m_p=0.8.

Normalized deviation and RMSE (Root-Mean-Square Error) of blindly estimated AM index with respect to SNR (Signal-to-Noise Ratio) is shown in **Figures 1** and **2**.

(a)

(b)
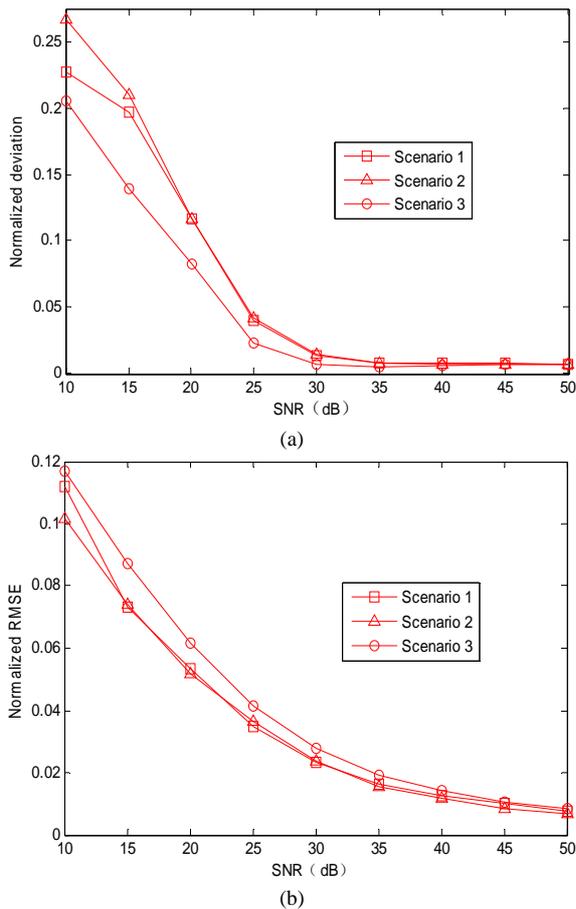
Figure 1. Blindly estimated result of AM index for FM signal with monotone baseband. (a) Normalized deviation; (b) Normalized RMSE.

(a)

(b)
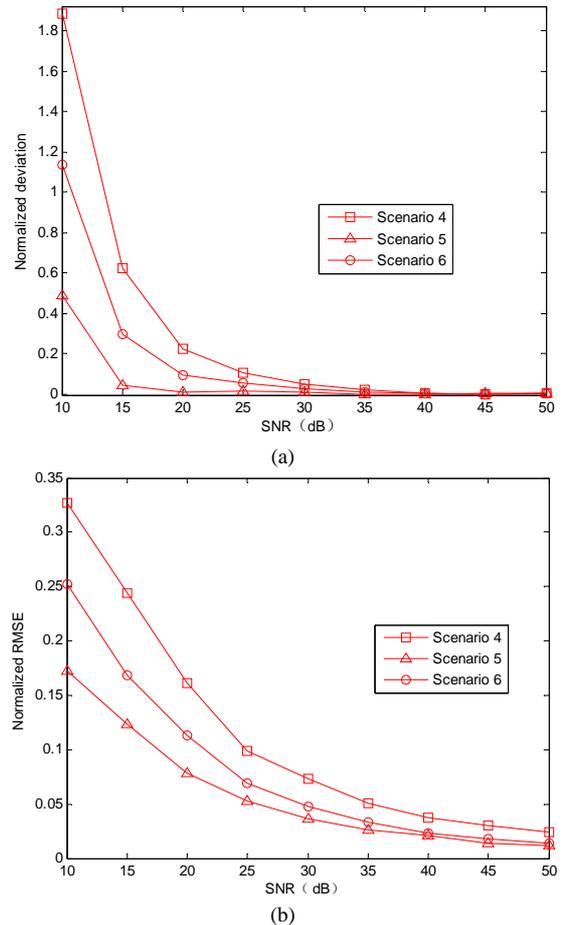
Figure 2. Blindly estimated result of AM index for PM signal with monotone baseband. (a) Normalized deviation; (b) Normalized RMSE.

The data length used by the algorithm is 1 M. The noise is additive white Gaussian noise. Simulation results are averaged from 100 Monte Carlo runs.

4.2. PSK, FSK Baseband Cases

The AM signal with BPSK baseband is generated. The sampling frequency is 200 MHz, the carrier frequency is 70 MHz, and $V_m = 1$ V. The following four simulation scenarios are conducted:

7) FM signal with $k_f = 50$ kHz/V, $f_m = 40$ kHz, $f_z = 10$ kHz, namely $m_f = 1.25$;

8) FM signal with $k_f = 50$ kHz/V, $f_m = 30$ kHz, $f_z = 10$ kHz, namely $m_f = 5/3$;

9) PM signal with $k_p = 0.8$ rad/V, $f_m = 40$ kHz, $f_z = 10$ kHz, namely $m_p = 0.8$;

10) PM signal with $k_p = 0.8$ rad/V, $f_m = 30$ kHz, $f_z = 10$ kHz, namely $m_p = 0.8$.

Normalized deviation and RMSE of blindly estimated AM index with respect to SNR is shown in **Figure 3**. The mean and the variance of estimated error for the code chip

(a)

(b)
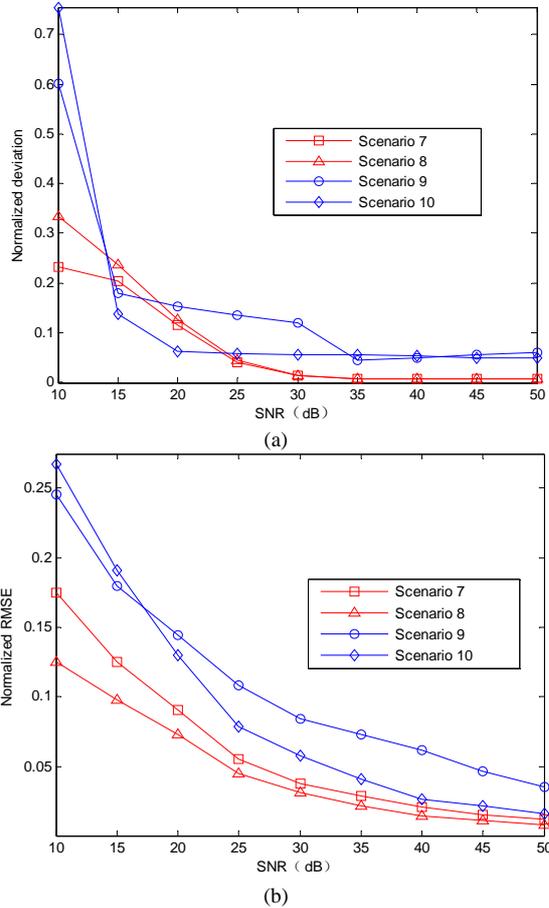
Figure 3. Blindly estimated result of AM index with BPSK baseband. (a) Normalized deviation; (b) Normalized RMSE.

(a)

(b)

Figure 4. Blindly estimated result of AM index with 2FSK baseband. (a) Normalized deviation; (b) Normalized RMSE.

are 4.9×10^{-5} and 3.1×10^{-9} respectively [11]. The data length used by the algorithm is 1M. The noise is additive white Gaussian noise. Simulation results are averaged from 100 Monte Carlo runs.

The AM signal with 2FSK baseband is generated. The sampling frequency is 200 MHz, the carrier frequency is 70 MHz, and $V_m = 1$ V. The following four simulation scenarios are conducted:

11) FM signal with $k_f = 50$ kHz/V, $f_m = 40$ kHz, $f_z = 10$ kHz, frequency interval 20 kHz, namely $m_f = 1.25$;

12) FM signal with $k_f = 50$ kHz/V, $f_m = 30$ kHz, $f_z = 10$ kHz, frequency interval 20 kHz, namely $m_f = 5/3$;

13) PM signal with $k_p = 0.8$ rad/V, $f_m = 40$kHz, $f_z = 10$ kHz, frequency interval 20 kHz, namely $m_p = 0.8$;

14) PM signal with $k_p = 0.8$ rad/V, $f_m = 30$ kHz, $f_z = 10$ kHz, frequency interval 20 kHz, namely $m_p = 0.8$.

Normalized deviation and RMSE of blindly estimated AM index with respect to SNR is shown in **Figure 4.** The data length used by the algorithm is 1M. The noise is additive white Gaussian noise. Simulation results are averaged from 100 Monte Carlo runs.

4.3. Analysis of Simulation Results

According to the above experimental results, conclusions can be drawn as:

1) When SNR≥20 dB and the AM signal is modulated by baseband modulations such as monotone, PSK, FSK, the normalized deviation and the RMSE of blindly estimated AM index are smaller than 0.2 and 0.15 respectively. Hence, the algorithm can be used in engineering practice.

2) The normalized deviations and RMSEs of blindly estimated AM index for FM and PM signals are almost the same.

3) Referring to the modulation index range shown in **Table 1**, high correct identification probabilities can be expected in the recognition of analogue FM and PM signals with the assistance of the AM index estimations when SNR ≥ 15 dB.

5. Conclusions

Based on analyzing of the AM signals in the TF domain,

this paper establishes a unified mathematical expression for the signal with baseband modulations of monotone, PSK, and FSK. Blind AM index estimation algorithm is further deduced. Moreover, measures to alleviate noise influence are also presented. Simulation results shows that the algorithm is effective, and it can be used in the cases with low SNR. The algorithm can be used in the blind demodulation and the recognition of AM signal for the communication situations without cooperation.

6. Acknowledgements

7. References

[1]   R. W. Jeremy and D. B. Gary, "A Method for Differentiating between Frequency and Phase Modulated Signals," *Proceedings of* 1999 *Information, Decision and Control Conference*, Adelaide, February 1999, pp. 489-494.

[2]   T. Shintaro and M. Eiichiro, "Automatic Classification of Analogue Modulation Signals by Statistical Parameters," *Proceedings of 1999 Military Communications Conference*, New Jersey, November 1999, pp. 202-207.

[3]   X. W. Zhong, H. Chen and K.-C. Yi, "Automatic Modulation Recognition of TTC Signals of Satellite," *Chinese Space Science and Technology*, Vol. 23, No. 3, 2003, pp. 57-64.

[4]   J. He and Q. Guo, "Automatic Recognition of Modulation Scheme in Satellite TT&C Channel," *Proceedings of the 7th International Conference on Electronic Measurement & Instrument*, Beijing, August 2005.

[5]   A. Engin, "Performance Comparison of Wavelet Families for Analog Modulation Classification Using Expert Discrete Wavelet Neural Network System," *Expert Systems with Applications*, Vol. 33, No. 4, 2007, pp. 23-25.

[6]   A. Engin and A. Derya, "Using Combination of Support Vector Machines for Automatic Analog Modulation Recognition," *Expert Systems with Applications*, Vol. 36, No.1, 2009, pp. 3956-3964.

[7]   A. E. EI-Mahdy, "Automatic Modulation Classification of Composite FM/PM Speech Signals in Sensor Arrays over Flat Fading Channel," *IET Communications*, Vol. 1, No. 2, 2007, pp. 157-164.

[8]   X. D. Zhangx and Z. Bao, "Communication Signal Processing," Publishing House of National Defense Industry, Beijing, 2000.

[9]   Z. T. Huang, Y. Y. Zhou and W. L. Jiang, "The Automatic Analysis of Intra-Pulse Modulation Characteristics Based on the Relatively Non-Ambiguity Phase Restoral," *Journals on Communications*, Vol. 24, No. 4, 2003, pp.153-160.

[10] Y. X. Wu, L. D. Ge and Z. Y. Xu, "A Novel Identification Method for Commonly Used Digital Modulations," *Acta Electronica Sinica*, Vol. 35, No. 4, 2007, pp. 782-785.

[11] Y. F. Zhanx, Z. G. Caox and Z. X. Ma, "Modulation Classification of M-QAM Signals," *Journals on Communications*, Vol. 25, No. 2, 2004, pp. 68-74.
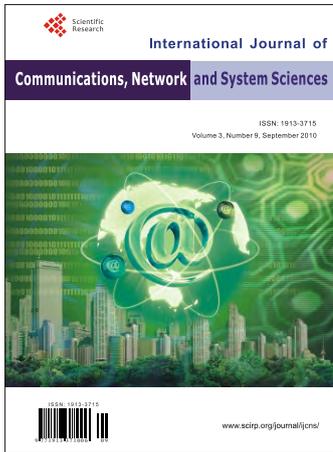
International Journal of

Communications, Network and System Sciences (IJCNS)

IJCNS is an international refereed journal dedicated to the latest advancement of communications and network technologies. The goal of this journal is to keep a record of the state-of-the-art research and promote the research work in these fast moving areas.

Editors-in-Chief

Prof. Huaibei Zhou	Wuhan University, China
Prof. Tom Hou	Virginia Tech, USA

Subject Coverage

This journal invites original research and review papers that address the following issues in wireless communications and networks. Topics of interest include, but are not limited to:

Ad Hoc and mesh networks	Network security
Coding , detection and modulation	Next generation network architectures
Cognitive Radio	Resource management and quality of service
Embedded distributed systems	Signal processing and channel modeling
MIMO and OFDM technologies	Simulation and optimization tools
Network protocol, QoS and congestion control	3G and 4G technologies
Network reliability, security and privacy	UWB technologies
	Wave propagation and antenna design

We are also interested in:

- Short reports—Discussion corner of the journal :
  2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data.
- Book reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

http://www.scirp.org/journal/ijcns   ijcns@scirp.org

TABLE OF CONTENTS

Volume 3   Number 9   **September 2010**