Journal of Information Hiding and Multimedia Signal Processing ©2015 ISSN 2073-4212
Ubiquitous International Volume **6**, Number **5**, September **2015**

Power Spectrum Compliant QIM Watermarking for Autoregressive Host Signals

Bin Yan, Ya-Fei Wang, Ling-Yun Song

College of Electronics, Communication and Physics
Shandong University of Science and Technology
579 Qian Wan Gang Road, Qingdao, 266590, P. R. China
yanbinhit@hotmail.com

Hong-Mei Yang

College of Information Science and Engineering
Shandong University of Science and Technology
579 Qian Wan Gang Road, Qingdao, 266590, P. R. China
yhm1998@163.com

Received January, 2015; revised May, 2015

Abstract. *The autoregressive (AR) model is widely used in modeling image, speech and EEG signals. Using this model as the model for the host signal, we have devised a watermarking algorithm which is compliant with the power spectrum condition. This is achieved by embedding the quantization watermark in the residual signal of the AR model, both for dither modulation (DM) watermarking and spread-transform dither modulation (STDM) watermarking. This paper also analyzes the decoding performance. An analytic result is obtained, which describes the relationship between the decoding error rate and the signal to noise ratio, model parameters and the length of the vector. This analysis result is verified through numerical experiments. Using this analysis result, a designer of the watermarking system can determine the design parameters based on the specification of the given system performance index.*

Keywords: Performance analysis; Auto-regressive host model; Dither modulation; Spread transform dither modulation; Power spectrum compliant watermarking.

1. **Introduction.** For the design of a robust watermarking system, an appropriate model for the potential attack must be incorporated. If the potential attacker uses an optimum filter (such as Wiener filter) to estimate and then remove the watermark, then the watermark must satisfy the power spectrum density (PSD) condition in order to resist this attack [1]. The PSD condition was first derived in the context of image watermarking and correlation detection. Based on this, Hwang *et. al.* derived the PSD condition for optimum detector [2]. Recently, Panda applied this principle to the design of audio watermarking system [3]. However, the current research on PSD condition focuses on the spread spectrum watermarking and leaves the quantization based watermarking out of the scope. In addition, for image, speech and EEG signals, autoregressive (AR) models are widely used in their processing. This paper aims at filling the gap of current research, i.e., the design and analysis of PSD compliant QIM (Quantization Index Modulation) watermarking for AR signal model.

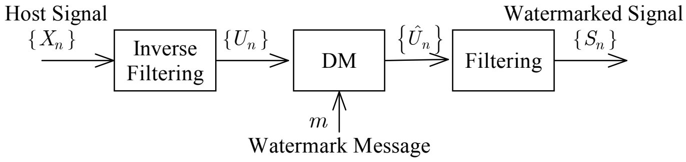

FIGURE 1. Embedder of AR-SSDM watermarking

The contribution of this paper can be summarized as follows. Using an AR model for the host signal, a PSD compliant QIM watermarking algorithm is designed. The theoretical decoding performance is derived. To improve the decoding performance, we also extend the proposed framework to spread transform dither modulation (STDM). The theoretical result can be used to guide the design of practical watermarking systems for speech, image and EEG signals, i.e., the watermark designer can utilize our result to design system parameters based on performance requirement.

We use capital letters with time index to denote a random process. For example, random process $\{X_n\}$ or X_n. The range of the random process is assumed to be $-\infty < n < \infty$ unless a finite range is specified. We name the spectrum shaped watermarking as *spectrum shaped dither modulation (SS-DM)* watermarking.

The proposed embedding framework is presented in section 2. We then prove that this framework leads to PSD compliant watermarking in section 3. Its decoding performance is analyzed in section 4. We then extend the framework to STDM in section 5. In section 6 we test our theoretical analysis through simulation experiments. Finally, we conclude the paper in section 7.

2. **The Proposed Embedding Method.** The whole system consists of three parts: embedding, attacking and decoding. The model of the host signal is assumed to be AR model. So we call this watermarking system AR-SSDM system. The block diagram of the embedding process is shown in Fig.1.

The host signal is characterized by an AR model

$$X_n = -\sum_{\ell=1}^{P} a_\ell X_{n-\ell} + U_n, -\infty < n < \infty, \tag{1}$$

where U_n is white Gaussian noise (WGN), i.e., $U_n \sim \mathcal{N}(0, \sigma_u^2)$. $a_1, \cdots, a_P$ are model parameters. To make the final watermark signal compliant with the PSD condition, we embed the watermark into the residual signal. So the host signal is first inverse filtered as $U_n = X_n + \sum_{\ell=1}^{P} a_\ell X_{n-\ell}$. Each watermark bit is embedded into U_n using QIM, resulting in watermarked residual signal $\hat{U}_n$. Here we consider only DM without distortion compensation. Using $\hat{U}_n$ as driving noise of the original AR model, we get the watermarked host signal $S_n = -\sum_{\ell=1}^{P} a_\ell S_{n-\ell} + \hat{U}_n$.

The attacking channel is modelled as WGN channel. So the received signal is $R_n = S_n + Z_n$, where Z_n is WGN: $Z_n \sim \mathcal{N}(0, \sigma_Z^2)$.

The structure of the decoder is shown in Fig.2. In order to extract the watermark bits, the decoder need to recover the residual signal by inverse filtering, i.e., $V_n = R_n + \sum_{\ell=1}^{P} a_\ell R_{n-\ell}$. Since each bit was embedded independently, so we decode the watermark message bit by bit. But it should be mentioned that, since the filtered channel noise in the residual signal is not white, so bit-wise decoding is not optimal. It is adopted here for its simplicity. The decoder uses the minimum distance decoding, i.e., $b_n = \arg\min_{b\in\{-1,1\}} \|V_n - \mathcal{Q}_b(V_n)\|$, where $\mathcal{Q}_b(\cdot)$ is the quantizer used to embed the bit b.

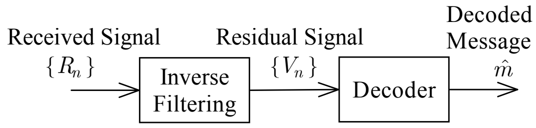

FIGURE 2. Decoder of AR-SSDM watermarking

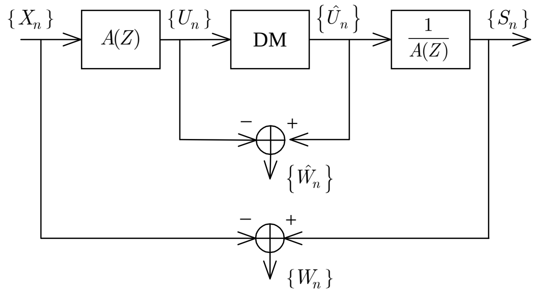

FIGURE 3. Relationship between watermark signal and watermark signal in residual signal

3. **PSD Shaping.** In this section, we prove that the watermark signal generated in last section is PSD compliant, i.e., the PSD of the watermark is proportional to the PSD of the host signal. Let the inverse filter in the $\mathcal{Z}$-transform domain be $A(z) = 1 + \sum_{\ell=1}^{P} a_\ell z^{-\ell}$. So $1/A(z)$ is the corresponding synthesis filter. Since the host signal is modelled as the result of passing WGN through an all-pole filter $1/A(z)$, so the PSD of the host signal should be

$$P_x(f) = \frac{\sigma_u^2}{|A(e^{j2\pi f})|^2} = \frac{\sigma_u^2}{\left|1 + \sum_{\ell=1}^{P} a_\ell e^{-j2\pi f\ell}\right|^2}, \; -\frac{1}{2} < f < \frac{1}{2}. \tag{2}$$

Since the watermark is embedded using QIM, the watermark signal $\hat{W}_n$ can be characterized by uniformly distributed white noise in the interval $[-\Delta, \Delta]$. So the mean function and auto-correlation function (ACF) of the watermark is: $\mathrm{E}(\hat{W}_n) = 0$, and $R_{\hat{w}}(k) = \frac{\Delta^2}{3}\delta_k$, where δ_k is the Kronecker δ function. Taking the inverse Fourier transform of this ACF, we can obtain the PSD of the watermark in the residual signal: $P_{\hat{w}}(f) = \frac{\Delta^2}{3}$.

The watermark signal is defined as the difference signal between the watermarked host signal and the original host signal, i.e., $W_n = S_n - X_n$. Referring to Fig.3, and use $\mathcal{Z}$ transform, the watermark signal W_n and the watermark signal $\hat{W}_n$ in the residual signal is related by:

$$w(z) = s(z) - x(z) = \hat{u}(z)\frac{1}{A(z)} - u(z)\frac{1}{A(z)} = \frac{\hat{w}(z)}{A(z)}.$$

So the PSD of the watermark is

$$P_w(f) = \frac{P_{\hat{w}}(f)}{|A(e^{j2\pi f})|^2} = \frac{\Delta^2/3}{\left|1 + \sum_{\ell=1}^{P} a_\ell e^{-j2\pi f\ell}\right|^2}. \tag{3}$$

Comparing Eq.(2) and Eq.(3), we can conclude that, using the watermark embedding framework in section 2, the PSD of the watermark signal is proportional to the PSD of the host signal.

4. **Decoding Performance.** The performance of this system is characterized by the probability of decoding error subject to given distortions. So we first introduce measures that are used to quantify the distortions involved in embedding and attacks.

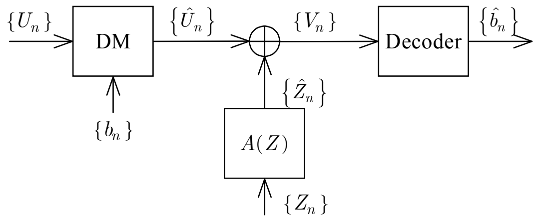

FIGURE 4. Equivalent diagram for watermark embedding and decoding in the residual signal

4.1. **Distortion measures.** The embedding distortion is characterized by the average power of the watermark signal, i.e.,

$$D_w = \lim_{M\to\infty} \mathrm{E}\left(\frac{1}{2M+1}\sum_{k=-M}^{M} W_k^2\right) = \lim_{M\to\infty}\frac{1}{2M+1}\sum_{k=-M}^{M}\sigma_w^2 = \sigma_w^2. \quad (4)$$

The channel distortion is characterized by the average power of the channel noise, i.e., the variance of the WGN: $D_c = \sigma_z^2$. Similarly, we can obtain the average power of the host signal: $D_x = \sigma_x^2$. From Eq.(2) and Eq.(3), we obtain:

$$D_x = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_x(f)\,\mathrm{d}f = \gamma\sigma_u^2,\;\; D_w = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_w(f)\,\mathrm{d}f = \gamma\frac{\Delta^2}{3},$$

where γ is:

$$\gamma = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\left|1+\sum_{\ell=1}^{P} a_\ell e^{-j2\pi f\ell}\right|^2}\mathrm{d}f$$

So the embedding distortion can be described by document to watermark ratio (DWR) as $\mathrm{DWR} = 10\log_{10}\frac{D_x}{D_w} = 10\log_{10}\frac{\sigma_u^2}{\Delta^2/3}$. The distortion induced by channel attack can be described by watermark to noise ratio(WNR) as $\mathrm{WNR} = 10\log_{10}\frac{D_w}{D_c} = 10\log_{10}\gamma\frac{\Delta^2/3}{\sigma_z^2}$.

4.2. **Probability of decoding error.** The equivalent block diagram of the whole system is show in Fig.4. The channel noise is filtered by the inverse filter, so the variance of the equivalent channel noise is $\sigma_{\hat{Z}}^2 = \sum_{\ell=0}^{P} h_\ell^2\sigma_Z^2 = \sigma_Z^2\left(1+\sum_{\ell=1}^{P} a_\ell^2\right)$. Using this equivalent noise, the decoding of watermark is exactly the same as those in white noise. So following the approaches in [4], we obtain the probability of decoding error:

$$P_e = \sum_{k=-\infty}^{\infty}\left[Q\left(\frac{2k\Delta+\frac{\Delta}{2}}{\sigma_Z\sqrt{1+\sum_{\ell=1}^{P} a_\ell^2}}\right) - Q\left(\frac{2k\Delta+\frac{3\Delta}{2}}{\sigma_Z\sqrt{1+\sum_{\ell=1}^{P} a_\ell^2}}\right)\right]. \quad (5)$$

5. **Extension to STDM.** Due to its high sensitivity to channel noise, DM is seldom used alone. It is usually combined with error correction code (ECC) or spread transform. In this section, we extend the above framework to STDM. When embedding the watermark in the residual siganl, we collect L signal components and consider it as a high dimensional vector. This vector is projected onto a random direction. We then use DM to embed one bit of watermark in this projection. Other steps are the same as in section 2. The random projection improves the security of DM watermarking. In addition, the quantization noise is distributed to L samples, thus allowing larger quantization step in DM embedding. Hence the robustness is improved. Two properties of the STDM is useful for the performance analysis in this section. (1) If the quantization step in DM is Δ_P,

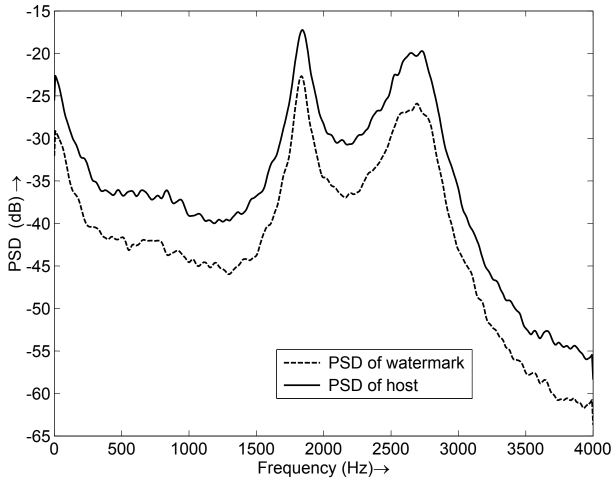

FIGURE 5. Welch estimation of the PSD of the host and watermark signal

then the embedding error introduced to each sample is no larger than $\Delta_p/\sqrt{L}$. (2) If the random direction vector is of unit length, then the projected channel noise has the same variance as before projection.

The performance analysis of our system utilizing STDM is similar to the analysis in section 4, except that the quantization step should be the quantization step Δ_P in projection domain. Similar DWR and WNR expressions can be obtained: $\mathrm{DWR} = 10\log_{10}\frac{3L\sigma_u^2}{\Delta_p^2}$ and $\mathrm{WNR} = 10\log_{10}\frac{\gamma\Delta_p^2}{3L\sigma_z^2}$. The probability of decoding error can be obtained as:

$$P_e = \sum_{k=-\infty}^{\infty}\left[Q\left(\frac{2k\Delta_{\mathrm{P}} + \frac{\Delta_{\mathrm{P}}}{2}}{\sigma_Z\sqrt{1+\sum_{\ell=1}^{P}a_\ell^2}}\right) - Q\left(\frac{2k\Delta_{\mathrm{P}} + \frac{3\Delta_{\mathrm{P}}}{2}}{\sigma_Z\sqrt{1+\sum_{\ell=1}^{P}a_\ell^2}}\right)\right]$$

Comparing this result with that of Eq.(5), we notice that the equivalent quantization step is increased. So an appropriate L can be chosen to combat the channel noise.

6. **Experiments.**

To verify the theoretical analysis of PSD and P_e, we use Monte Carlo simulation to simulate the watermarking systems in this paper.

6.1. **Verification of the PSD of the watemrark.**

The calculated average error rate is compared with the theoretical analysis given in Eq.(5). A set of AR parameters estimated from real speech signal is used in the simulations [5]:

$$\mathbf{a} = [1, -0.6, 1.205, -1.588, 1.153, -1.427, 1.018, -0.536, 0.352, -0.314, -0.055]^T$$

In the simulation, we use a fixed DWR = 6 dB, and WNR changes from 15dB to 30dB. The length of the sequence is chosen to be 2000. The estimated PSD for the host signal and watermark signal are shown in Fig.5 using Welch estimation, and in Fig.6 using autoregressive estimation. The proportionality of the watermark PSD and host PSD can be easily identified from the two figures. This verifies that the PSD of the watermark satisfies the PSD condition for energy efficient watermarking. In Fig.7, theoretical result of P_e is compared with the estimate P_e from Monte Carlo simulation. These two results coincide with each other, thus validates the theoretical analysis.

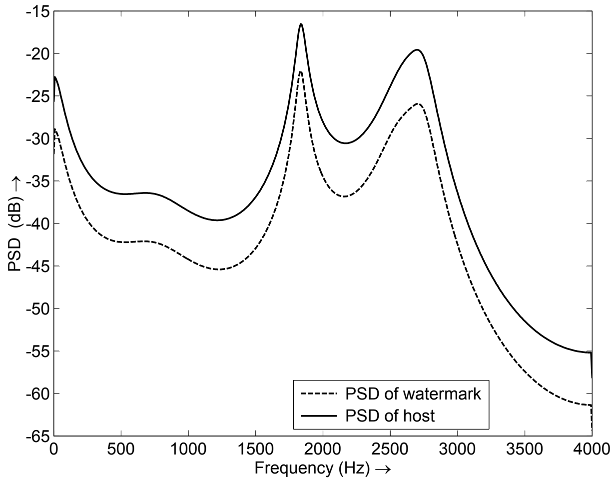

FIGURE 6. AR estimation of the PSD of the host and watermark signal

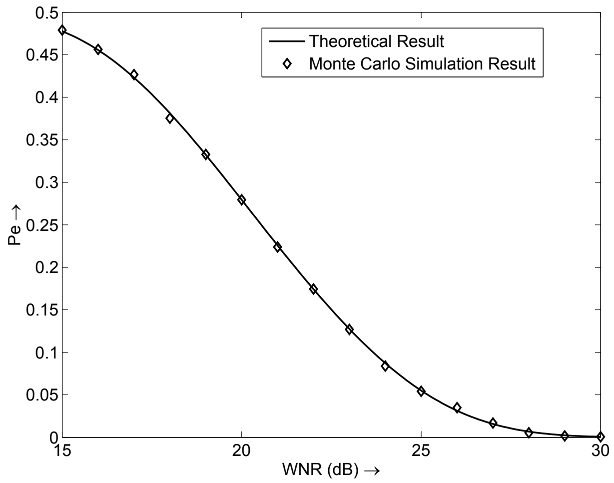

FIGURE 7. Comparison between theoretical result and simulation result of P_e

6.2. **The decoding performance of STDM with unknown AR parameters.** In the above analysis, we assume that the model parameters are available. This is true for stationary random process and informed detection. But most practical signals are non-stationary. For example, speech signal can only be modelled as short-term stationary. So the watermark decoder needs to estimate the model parameters from the received signals. Then these estimated parameters are used in inverse filtering. But the estimated parameters are usually different from the true parameters, especially when facing channel noise [6, 7]. This estimation error may lead to extra noise after inverse filtering, causing the P_e to increase. The decoding performance for unknown AR parameters is shown in Fig.8. We also plot the decoding performance of spread spectrum (SS) based watermarking with known AR parameters. Obviously, even though the estimation error may degrade the decoding performance of our spectrum shaped STDM algorithm, but it is still outperforms the SS watermarking with known AR parameters. Considering that the performance of SS watermarking with unknown AR parameters is worse, so spectrum shaped STDM is still preferred than spectrum shaped SS watermarking for practical watermarking systems.

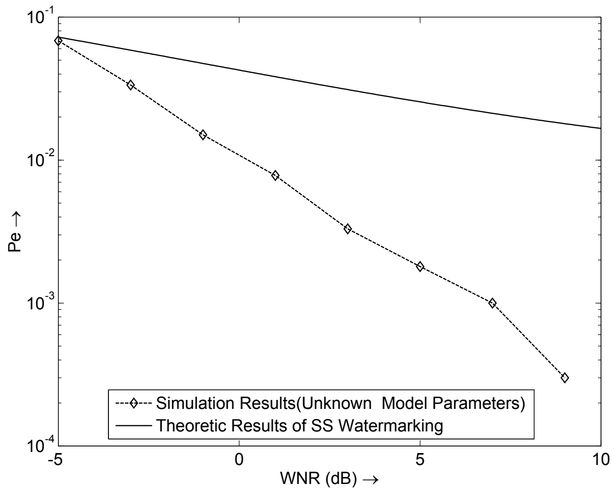

FIGURE 8. Decoding error of power spectrum shaped STDM Vs. theoretical decoding error of power spectrum shaped spread spectrum watermarking

7. **Conclusion.** Using AR model as the model of host signal, a PSD compliant QIM watermarking system is designed. It is proved that the generated watermark signal satisfies the power spectrum condition. The theoretical decoding performance is analyzed and verified by simulation experiments. Our result is useful in designing watermarking system for speech, image and EEG signals facing the optimum filtering attacks. The theoretical analysis result is useful to help the designer to choose design parameters such as P, Δ and L from the given performance requirement specified by the users, such as P_e.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (NSFC) under the project grant number: 61272432, Shandong Provincial Natural Science Foundation (No. ZR2014JL044), Qingdao science and technology development plan (No. 12-1-4-6-(10)-jch). The work of Ya-Fei Wang is also supported in part by "Graduate Innovation Fund of Shandong University of Science and Technology" with grant no: YC140340.

REFERENCES

[1] J. K. Su, B. Girod, Power-spectrum condition for energy-efficient watermarking, IEEE Transactions on Multimedia 4 (4) (2002) 551–560.
[2] Y. Hwang, K. A. Moon, M. J. Kim, Watermark design for the Wiener attack and whitening filtered detectioin, in: Proceedings of the SPIE, Vol. 4675, 2002, pp. 441–449.
[3] J. Panda, M. Kumar, Application of energy efficient watermark on audio signal for authentication, in: International Conference on Computational Intelligence and Communication Networks, 2011, pp. 202–206.
[4] J. Barni, F. Bartolini, Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications, Marcel Dekker Inc., 2004.
[5] W. C. Chu, Speech Coding Algorithms: Foundation and Evolution of Standardized Coders, Wiley-Interscience, New Jersey, 2003.
[6] S. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, 1993.
[7] S. M. Kay, Modern Spectral Estimation: Theory and Application, Prentice-Hall, Englewood Cliffs, NJ, 1988.