IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 3, 2017, accepted May 22, 2017, date of publication May 26, 2017, date of current version June 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2708739

Blind Classification of Line-Coding Schemes Based on Characteristic Features

JANGHOON OH[1], JINWOO JEONG[1], YEONSOO JANG[2], JAEYOON LEE[1], AND DONGWEON YOON[1], (Member, IEEE)
[1]Department of Electronics and Computer Engineering, Hanyang University, Seoul 04763, South Korea
[2]Second Research and Development Institute, Agency for Defense Development, Daejeon 34188, South Korea

Corresponding author: Dongweon Yoon (dwyoon@hanyang.ac.kr)

This work was supported by the research fund of Signal Intelligence Research Center, supervised by the Defense Acquisition Program Administration and Agency for Defense Development of Korea.

ABSTRACT In modern defense technology, signal intelligence plays an important role in military operations as an intelligence source. Among its target signals, telemetry and telecommand signals can use line-coding schemes for their advantageous properties in the presentation of binary data of transmission systems. Therefore, in a non-cooperative context, blind classification of the line-coding scheme is the final crucial step in recovering target information from an unknown received signal. In this paper, we examine the characteristic features of line-coding schemes and then propose a simple blind classification algorithm for the schemes. We also analyze correct classification probabilities of the proposed algorithm in a noiseless and noisy environment through computer simulations. The proposed method can discern line-coding schemes, allowing reconstruction of the original information data.

INDEX TERMS Blindness, classification algorithms, signal reconstruction, line coding, signal intelligence.

I. INTRODUCTION

The information-bearing signal is generally converted to digital data in binary bits and then the bits are encoded into electrical pulses or waveforms in order to transmit information over the channel. The procedure for choosing a particular pair of waveforms and turning the digital data into electrical waveforms is commonly called line coding, which is widely used in various signals for telecommunication, telemetry and telecommand.

In modern defense technology, signal intelligence (SIGINT) plays an important role in military operations as an intelligence source. Among its target signals, telemetry and telecommand signals can use line-coding schemes for their advantageous properties in the presentation of binary data of transmission systems. To be specific with telemetry and telecommand signals in military technical standards such as Inter-Range Instrumentation Group (IRIG) standard, they use conventional line-coding schemes [1]. Therefore, in the viewpoint of SIGINT, it is important to be able to blindly classify line-coding schemes adopted in a received signal. A good introduction to line coding is given in [2]–[4]. Line coding is a popular scheme for binary data transmission due to several desirable properties, including a small transmission bandwidth, power efficiency, zero dc value, simple synchronization in the pulses, and transparency for the transmission and reception of arbitrary symbols or bit patterns. Therefore, it is highly probable that an unknown received signal could include a certain line-coding scheme for the efficient transmission of information data.

In a non-cooperative context, such as signal intelligence or spectrum surveillance applications, a receiver cannot understand received signal directly when it has no information on system parameters of the transmitter. Consequently, the receiver must blindly estimate the parameters to interpret the received signal without help of transmitter to fulfill the mission [5]–[13]. FIGURE 1 shows the relationship among a received signal, information data, and a receiver in a non-cooperative and cooperative context with regard to line-coding schemes.

Classification of the line-coding schemes is the final crucial step in recovering target information from an unknown received signal. Ample research is available on the blind estimation of unknown transmission parameters, such as interleaver parameters [5]–[7], modulation schemes [8]–[10], and channel coding schemes [11]–[13], however, blind classification of line codes, to the best of our knowledge, has not yet been reported.

There are two general algorithm classes for signal identification: likelihood-based (LB) and feature-based (FB) algorithms. The former uses the likelihood function of an

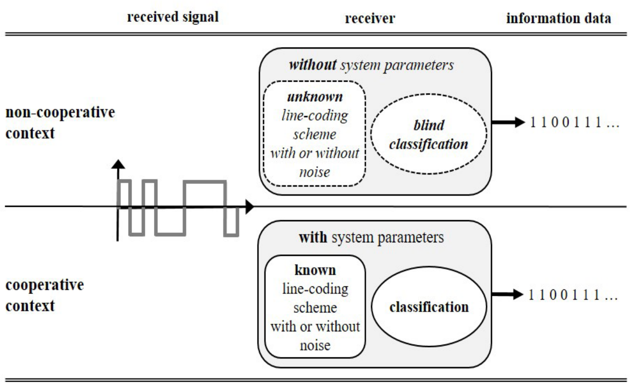

FIGURE 1. Relationship among a received signal, information data, and a receiver in a non-cooperative and cooperative context.

unknown received signal and makes a decision of identification by comparing the likelihood ratio to a threshold value. In theory, though the LB algorithm gives an optimal solution, it requires high computational complexity. In the case of FB algorithm, expediently-selected signal features are used to make decisions on an unknown received signal by comparing values. Even though the algorithm presents sub-optimal performance, it can be easily implemented [14]. In this paper, we propose a simple FB method using characteristic features for the blind classification of line codes from an unknown received signal. It is a sequential classification algorithm to discriminate among multiple line codes blindly. We derive equations for correct classification probabilities of the core function of the proposed method and validate the classification performance in both noiseless and noisy environments by computer simulations.

This paper is organized as follows. Section 2 presents the characteristic features of line-coding schemes. Section 3 explains the proposed blind classification algorithm and section 4 provides numerical and simulation results. Finally, section 5 offers conclusion.

II. CHARACTERISTIC FEATURES OF LINE-CODING SCHEMES

It is well known that there are two representative categories of line codes: return-to-zero (RZ) and non-return-to-zero (NRZ). In RZ-coded signals, the waveforms revert to a zero-voltage level at some point, normally half of a bit interval; this does not take place in NRZ-coded signals. The line-coding schemes are further categorized as unipolar, polar, bipolar, and Manchester codes according to the manner of assignment of voltage levels to the binary data [2]–[4]. These codes are selectively applied to maximize the bit rate for a given channel, to minimize transmission power, to recover synchronization from the received signal, or to reduce dc value in various digital communication systems.

Since every line-coding scheme has its unique waveform format and characteristic features based on its polarity of pulses, dc value, number of consecutive 1's or 0's, spectral occupancy, and average power, we typically consider more than one criterion when selecting a scheme. Note that the given average power feature, which is important in the proposed blind classification algorithm, can be considered as peak-to-average power ratio (PAPR) changing according to the mark ratio in an unknown received signal.

TABLE 1. Classification of line-coding schemes according to characteristic features.

Schemes		Negative polarity	Zero level	Number of consecutive 1's (0's)	PAPR
NRZ	unipolar	N	Y	≥0 (≥0)	0.5
NRZ	polar	Y	N	≥0 (N/A)	1
NRZ	bipolar	Y	Y	1 (≥0)	0.5
RZ	unipolar	N	Y	≥0 (≥0)	0.25
RZ	polar	Y	Y	1 (1)	0.5
RZ	bipolar	Y	Y	1 (≥0)	0.25
Manchester		Y	N	≤2 (N/A)	1

TABLE I shows the classification of line-coding schemes according to four characteristic features: pulse polarity, zero level, number of consecutive 1's or 0's, and PAPR. With regard to pulse polarity, the codes divide into two types: one using negative value to represent binary bits and the other using no negative value. Zero level lets us identify two schemes, polar NRZ and Manchester, which have no zero level in their waveforms. The number of consecutive 1's or 0's in line-coded waveforms can be utilized to identify some schemes. Finally, line codes of interest have three different PAPR values. Since every line-coded waveform has its unique format for the same binary sequence, these characteristic features can be used to perform blind classification of line-coding schemes in a non-cooperative context.

III. PROPOSED ALGORITHM FOR BLIND CLASSIFICATION

We propose an algorithm based on the characteristic features of pulse polarity, zero level, number of consecutive 1's or 0's, and PAPR to classify the line-coding scheme blindly from an unknown received signal and FIGURE 2 shows the major steps of the proposed algorithm.

First, the levels of the unknown received waveforms S_r need to be decided among 1, 0, and -1. It is straightforward in a noiseless environment. However, in a noisy environment, we use a threshold ε to make the decision among levels. After the level decision, it is possible to detect unipolar signaling by examining the presence of negative values to represent polarities of line codes. Then, by PAPR comparison, we can distinguish between NRZ and RZ signaling. The remainder after detecting unipolar signaling should be one of three line-coding schemes: polar, bipolar, and Manchester codes. We distinguish Manchester and polar NRZ by detecting zero

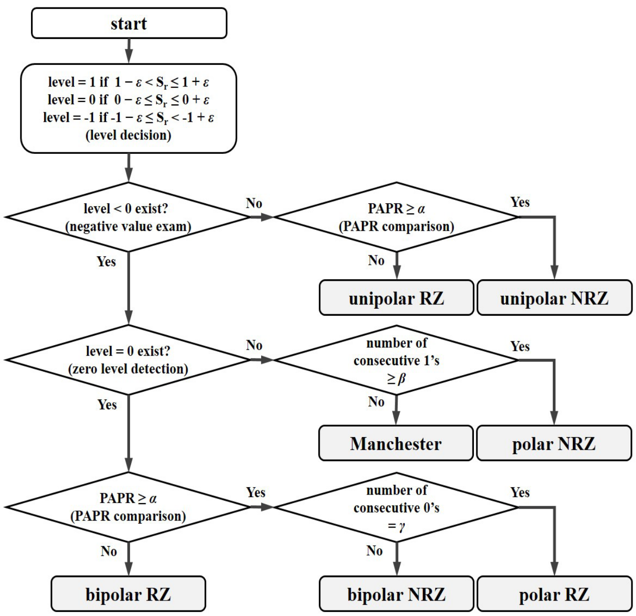

FIGURE 2. Block diagram of algorithm for blind classification of line-coding schemes.

value because they have no zero level in waveforms. Then we can count the number of consecutive 1's in the received waveforms. Since Manchester signaling can have consecutive 1's having the length of 2 at most, we can discriminate it from polar NRZ by setting the threshold β to be 3. Now three choices, bipolar NRZ and RZ, and polar RZ signaling, remain and they can be distinguished by PAPR comparison and the number of consecutive 0's. Since polar RZ signaling can only have consecutive 0's as long as 1 in its waveform, we can separate it from bipolar NRZ by setting the threshold γ to be 1. With regard to PAPR comparison in the block diagram of FIGURE 2, the PAPR threshold α is used to discern between NRZ and RZ for both unipolar and bipolar cases of the unknown received signal. In TABLE I, the rightmost column presents nominal PAPR values of typical line-coding schemes for the same binary sequence. The actual PAPR value, however, is variable according to the number of marks (or spaces). In consequence, the classification is probabilistic in both noiseless and noisy environments.

In this manner, the proposed algorithm allows us to classify line-coding schemes without any prior information about the applied waveforms, which is a crucial advantage especially in applications of signal intelligence or spectrum surveillance.

IV. NUMERICAL AND SIMULATION RESULTS

The simple theoretical formulations of correct classification probability for unipolar NRZ and RZ in a noiseless channel are derived and analyzed as follows. Since the PAPR values of unipolar NRZ and RZ are determined by the probabilities of occurrence of mark and space, we can derive the correct classification probability based on them. The simple theoretical formulation for the probability of occurrence of x marks out of N line-coded waveforms can be obtained from the binomial probability density function (pdf) as in (1).

$$f_x(x) = \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} \delta(x-k), \tag{1}$$

where N is the number of data, x is the number of mark occurrences, p is the probability of occurrence of mark, and $\delta(\cdot)$ is the delta function.

Unipolar NRZ signaling can be successfully classified with our PAPR-based algorithm when mark ratio is greater than predefined threshold value as in (2) and the range of the integer x can be defined as (3).

$$\alpha \le \frac{x}{N} \le 1 \tag{2}$$

and

$$\lceil \alpha N \rceil \le x \le N, \tag{3}$$

where α is the threshold for PAPR comparison and $\lceil\rceil$ is the ceiling function.

By taking advantage of the range of x, the correct classification probability for unipolar NRZ can be formulated as (4). With regard to unipolar RZ scheme, the probability of appearance of x marks out of N line-coded waveforms can also be represented as (1). However, a receiver can be considered to receive $2N$ waveforms to be blindly classified in a non-cooperative context because data bits 0 and 1 are line-coded into 00 and 10 waveforms, respectively. Therefore, the correct classification for unipolar RZ signaling can be obtained when the mark ratio x to $2N$ is smaller than predefined threshold value as (5).

$$\begin{aligned} P_{uni_N RZ}(N) &= \sum_{x=\lceil \alpha N \rceil}^{N} f_x(x) \\ &= \sum_{x=\lceil \alpha N \rceil}^{N} \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} \delta(x-k) \\ &= \sum_{x=\lceil \alpha N \rceil}^{N} \binom{N}{x} p^x (1-p)^{N-x} \end{aligned} \tag{4}$$

$$0 \le \frac{x}{2N} < \alpha \tag{5}$$

In this case, the range of the integer x for the correct classification becomes

$$0 \le x \le \lfloor 2\alpha N \rfloor, \tag{6}$$

where $\lfloor\rfloor$ is the floor function. Therefore, the probability of correct classification for unipolar RZ can be achieved by

$$\begin{aligned} P_{uni_R Z}(N) &= \sum_{x=0}^{\lfloor 2\alpha N \rfloor} f_x(x) \\ &= \sum_{x=0}^{\lfloor 2\alpha N \rfloor} \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} \delta(x-k) \\ &= \sum_{x=0}^{\lfloor 2\alpha N \rfloor} \binom{N}{x} p^x (1-p)^{N-x}. \end{aligned} \tag{7}$$

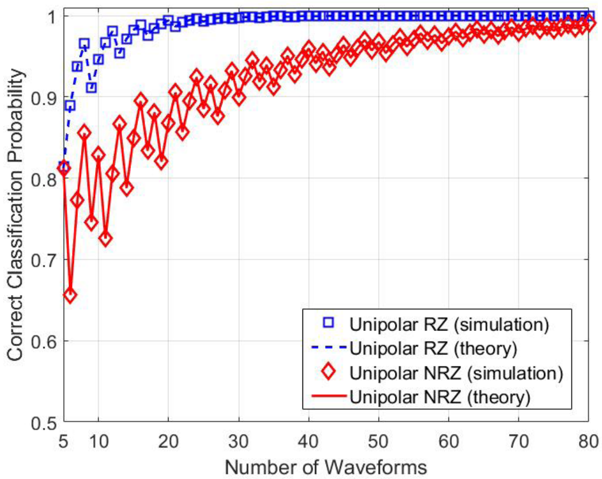

FIGURE 3. Comparison of simulation and theoretical results regarding correct classification probabilities of the PAPR comparison of the proposed algorithm for unipolar cases (results for every 1-waveform increase of the number of input waveforms).

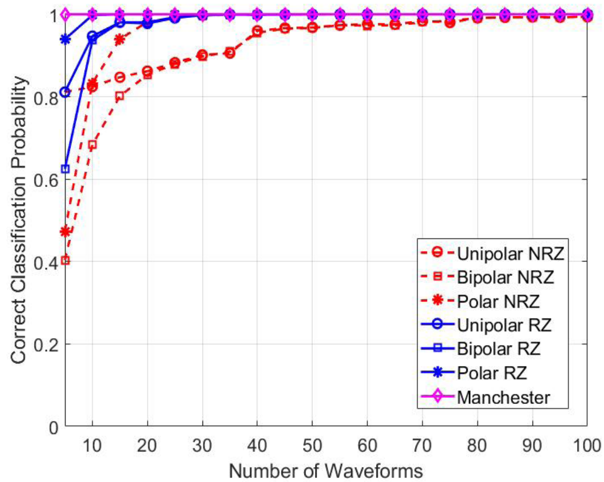

FIGURE 4. Comprehensive simulation results regarding correct classification probabilities of the proposed algorithm in a noiseless environment.

For the bipolar NRZ and RZ signaling in a noiseless channel, the correct classification probabilities can also be obtained by (4) and (7), respectively, since the absolute-valued bipolar signaling becomes identical to unipolar signaling in the waveform and PAPR value. The PAPR comparison of the proposed algorithm can be verified by exploiting (4) and (7) because the equations are based on a binomial distribution which explains properly the probabilistic occurrence of mark and space in the received line-coded signal. When the mark ratio in the signal is higher than the predefined threshold α, the applied line-coding scheme will be classified as unipolar NRZ, and therefore the correct classification probability can be formulated as (4). On the other hand, when the mark ratio is lower than the threshold, the scheme will be classified as unipolar RZ. Hence, the correct classification probability can be given by (7). To verify the PAPR comparison part, the correct classification probabilities for the unipolar signaling are presented versus the number of received waveforms with theoretical and simulation results in a noiseless channel. We compare the results from the derived equations with numerical simulations in FIGURE 3. In the simulations, we assume $p = 0.5$ and $\alpha = 0.375$; for the threshold α, it is median between the PAPR values of unipolar NRZ and RZ. The simulation results are obtained by 100,000 times of iterations for Monte Carlo simulation. FIGURE 3 shows the excellent match between theoretical and simulation results, where we depict results for every 1-waveform increase of the number of waveforms. The correct classification probabilities initially fluctuate markedly for every 1-waveform increase of the number of waveforms up to around 30 waveforms; this is mainly because of the rapid change of the instantaneous PAPR values obtained from the input waveforms. Then, as the number of input waveforms becomes larger enough, the instantaneous PAPRs become stable and the correct classification probabilities converge to 1. More than 30 input waveforms are needed to achieve the correct classification probability higher than 0.9.

To confirm the functionality of the proposed algorithm comprehensively, we extend the simulation over multiple line-coding schemes in a noiseless environment. The simulations are designed to follow the proposed algorithm in FIGURE 2 and performed by using Monte Carlo simulation assuming average power of 1 for the received signal. FIGURE 4 presents the results of simulations to classify the line-coding schemes of TABLE I in a noiseless environment; the resulting correct classification probabilities are given versus the number of input waveforms and plotted for every 5-waveform increase of the number of input waveforms from 0 to 100. We see the improved correct classification probabilities in most of schemes as the length of input waveforms increases, except in Manchester code. Since the probabilistic factors of the proposed algorithm, such as PAPR comparison, become stable as the length of input waveforms increases, the correct classification probabilities approach to 1. Manchester code presents the correct classification probability of 1 for all the values of abscissa since there is no probabilistic factor in the process of blind classification using the proposed algorithm.

We now turn our attention to the blind classification of line-coding schemes in a noisy environment. To investigate the functionality of the proposed algorithm, we perform computer simulations by considering additive white Gaussian noise (AWGN) environment. In this case, the line-coded waveforms are affected by AWGN and the values of levels become probabilistic. Thus, the levels in the received waveforms need to be decided using threshold $\varepsilon = 0.5$. After the level decision, we can handle waveforms in the same manner as the noiseless case. FIGURE 5 and FIGURE 6 show the results of simulations using the proposed algorithm to classify the target line-coding schemes with 50 and 100 input waveforms, respectively, in a noisy environment; the resulting

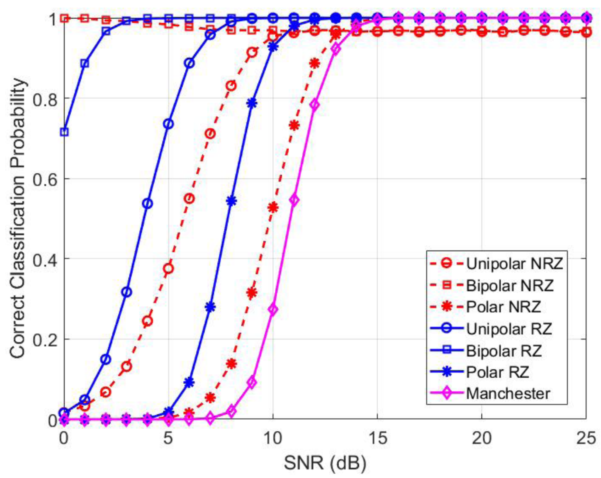

FIGURE 5. Comprehensive simulation results regarding correct classification probabilities of the proposed algorithm with 50 input waveforms in AWGN.

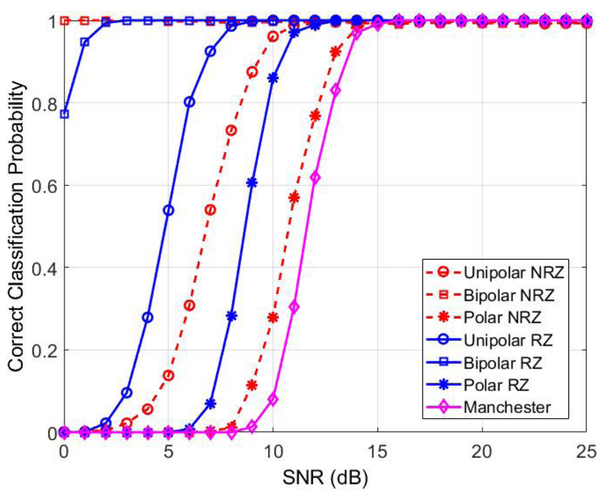

FIGURE 6. Comprehensive simulation results regarding correct classification probabilities of the proposed algorithm with 100 input waveforms in AWGN.

correct classification probabilities are given versus SNRs. As we found in FIGURE 4, the correct classification probabilities are influenced by the length of input waveforms. Therefore, the performance in a noisy environment needs to be analyzed according to the number of input waveforms as well as signal-to-noise ratio (SNR). Regarding FIGURE 5 and FIGURE 6, the correct classification probabilities are influenced by the number of waveforms as presented in FIGURE 4. It is mainly because of the characteristics of function modules of the proposed algorithm. Among the modules, when the number of input waveforms is small, PAPR comparison will be affected by an instantaneous PAPR value and produce the comparison results biased to negative decision. As a result, in the case of 50 input waveforms, the correct classification probabilities for unipolar NRZ and bipolar NRZ schemes converge toward around 0.97 even in the high SNR region. To be specific, the PAPR values of unipolar NRZ and bipolar NRZ schemes vary with the number of 0 levels while the polar RZ scheme has a fixed PAPR value independent of the number of 0 levels. On the other hand, other line-coding schemes show slightly enhanced correct classification probabilities in the case of 50 input waveforms. This seems related to other modules, including negative value exam, zero level detection and consecutive level counters, which can be affected by only one erroneous level change in a noisy environment. Therefore, the correct classification probabilities become enhanced with 50 input waveforms compared to the case of 100 input waveforms in the same SNR region, and we can obtain reasonable performance with the limited number of unknown received waveforms in a noisy environment. Regarding the 50 input waveforms case, to achieve the correct classification probability 0.9 for all the target line-coding schemes, SNR should be higher than around 12.5 dB in a noisy environment. Therefore, acquiring necessary waveforms and securing SNR condition will play important roles in the blind classification of target line-coding schemes using the proposed algorithm for signal intelligence or spectrum surveillance.

V. CONCLUSION

In order to reconstruct target information from an unknown received signal in a non-cooperative context, the receiver must be able to estimate system parameters of the transmitter blindly. An algorithm for the blind classification of line-coding schemes is crucial to SIGINT operations as the last step in recovering target information from an unknown received signal since the line coding is essential in the presentation of binary data for telecommunication, telemetry and telecommand systems.

In this paper, we proposed a sequential blind classification algorithm for typical line-coding schemes, derived numerical expressions for the core function, and validated comprehensive performance of the proposed algorithm in both noiseless and noisy environments by computer simulations. The algorithm is based on a systematic method using characteristic features of line codes and able to classify typical line-coding schemes. We expect that this algorithm will play its role in the blind classification of unknown received signals for the purposes of signal intelligence or spectrum surveillance. It will help to make the techniques simpler and more accurate.

REFERENCES

[1] *Telemetry Standards*, IRIG Standard 106-13 (Part 1), Jun. 2013, ch. 4.
[2] R. E. Ziemer and W. H. Tranter, "Principles of baseband digital data transmission," in *Principles of Communications: Systems, Modulations, and Noise*, 6th ed. Hoboken, NJ, USA: Wiley, 2009.
[3] B. Sklar, "Formatting and baseband modulation," in *Digital Communications: Fundamentals and Applications*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
[4] J. D. Gibson, "Line coding," in *The Communications Handbook*, 2nd ed. Boca Raton, FL, USA: CRC Press, 1999.
[5] G. Sicot, S. Houcke, and J. Barbier, "Blind detection of interleaver parameters," *Signal Process.*, vol. 89, no. 4, pp. 450–462, Apr. 2009.
[6] L. Lu, K. H. Li, and Y. L. Guan, "Blind detection of interleaver parameters for non-binary coded data streams," in *Proc. ICC*, Dresden, Germany, Jun. 2009, pp. 3668–3671.

[7] G. Burel and R. Gautier, "Blind estimation of encoder and interleaver characteristics in a non cooperative context," in *Proc. IASTED*, Scottsdale, AZ, USA, Nov. 2003, pp. 275–280.
[8] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.
[9] H. C. Wu, M. Saquib, and Z. Yun, "Novel automatic modulation classification using cumulant features for communications via multipath channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3098–3105, Aug. 2008.
[10] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," in *Proc. ICCSPA*, Sharjah, United Arab Emirates, Feb. 2013, pp. 1–6.
[11] V. Choqueuse, M. Marazin, L. Collin, K. C. Yao, and G. Burel, "Blind recognition of linear space–time block codes: A likelihood-based approach," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1290–1299, Mar. 2010.
[12] R. Moosavi and E. G. Larsson, "A fast scheme for blind identification of channel codes," in *Proc. IEEE GLOBECOM Conf.*, Houston, TX, USA, Dec. 2011, pp. 1–5.
[13] Y. G. Debessu, H.-C. Wu, and H. Jiang, "Novel blind encoder parameter estimation for turbo codes," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 1917–1920, Dec. 2012.
[14] O. A. Dobre and R. Inkol, "Blind signal identification: Achievements, trends, and challenges," in *Proc. IEEE COMM*, Ottawa, ON, Canada, Jun. 2012, pp. 349–352.

JANGHOON OH received the B.S. degree in electronic communication engineering from Hanyang University, Seoul, South Korea, in 1996 and the M.S. degree in electrical and computer engineering from University of California at Irvine, Irvine, CA, USA, in 2006. He is currently working toward the Ph.D. degree in wireless communications with the Department of Electronics and Computer Engineering, Hanyang University, Seoul. From 1996 to 2004, he was a Technical Manager with the Network Strategy Division, SK Telecom, South Korea, and from 2009 to 2011, he was a Senior Manager with the Corporate Planning Division, NCsoft, South Korea.. His research interests include wireless communication networks and military communications.

JINWOO JEONG received the B.S., M.S., and Ph.D. degrees in electronic communication engineering from Hanyang University, Seoul, South Korea, in 1999, 2001, and 2016, respectively. From 2001 to 2007, he was with the Digital Media Laboratory, LG Electronics Inc., South Korea. Since 2016, he has been a Research Professor with the Department of Electronic Engineering, Hanyang University, Seoul. His research interests include military communications systems, electronic warfare, and satellite and deep space communications.

YEONSOO JANG received the B.S., M.S., and Ph.D. degrees in electronic communications engineering from Hanyang University, Seoul, South Korea, in 2015. He is currently with the Second Research and Development Institute, Agency for Defense Development, Daejeon, South Korea. His research interest includes digital communication systems and radar systems.

JAEYOON LEE received the B.S. and M.S. degrees in information and communications engineering from Daejeon University, Daejeon, South Korea, in 2002 and 2004, respectively, and the Ph.D. degree in electronics and communications engineering from Hanyang University, Seoul, South Korea, in 2008. From 2010 to 2012, he was a Research Professor with the Department of Electronic Engineering, Hanyang University. From 2012 to 2014, he was with the Electronics and Telecommunication Research Institute, Daejeon, as a Senior Researcher. From 2014 to 2016, he was with Korean Intellectual Property Office, Daejeon, as a Patent Examiner. Since 2016, he has been a Research Professor with the Department of Electronic Engineering, Hanyang University. His research interests include new modulation techniques, accurate performance evaluations, and digital communications theory.

DONGWEON YOON received the B.S. (*summa cum laude*), M.S., and Ph.D. degrees in electronic communications engineering from Hanyang University, Seoul, South Korea, in 1989, 1992, and 1995, respectively. From 1995 to 1997, he was an Assistant Professor with the Department of Electronic and Information Engineering, Dongseo University, Pusan, South Korea. From 1997 to 2004, he was an Associate Professor with the Department of Information and Communications Engineering, Daejeon University, Daejeon, South Korea. Since 2004, he has been with the Faculty of Hanyang University, Seoul, South Korea. He has twice been an Invited Researcher with the Electronics and Telecommunications Research Institute, Daejeon, in 1997 and from 2002 to 2005. He was a Visiting Professor with Pennsylvania State University, PA, USA, from 2001 to 2002, and with University of California at Riverside, Riverside, CA, USA, from 2010 to 2011. He was the Department Chair of the academic year from 2011 to 2013 with the Department of Electronic Engineering, Hanyang University. He is the Director of Signal Intelligence Research Center. He is currently a Professor with the Department of Electronic Engineering and the Director of Signal Intelligence Research Center, Hanyang University. His research interests include digital communications theory and system, satellite and space communications, wireless communications, military communications, and signal intelligence.