Spectrum Prediction via Long Short Term Memory

Ling Yu, Jin Chen, Guoru Ding
Institute of Communication Engineering
Army Engineering University of PLA
Nanjing, China
e-mail: yl2811284434@126.com, chenjin99@263.net, dr.guoru.ding@ieee.org

Abstract—**Spectrum prediction algorithm has always been a research hot spot of communication because it can reduce considerable time and energy consumption when the receiver senses channel states. Users only sense those channels whose prediction result of channel state in the next time slot will be idle, thus reducing the number of perceptual processes. Many spectrum prediction algorithms have achieved good performance, and with the rise of deep learning it will be a good innovation research in the application of spectrum prediction. A prediction model composed of Long Short Time Memory (LSTM) layers is constructed in this paper and then is trained through supervised learning before prediction. The output of LSTM network is transformed into 0 or 1 which indicates idle state or occupied compared to the threshold. Therefore, the maximum accuracy can be achieved under certain threshold settings. Influence of different LSTM network depth and width on the prediction accuracy is also studied in this paper together with Back Propagation (BP) neural network performance. Results show that LSTM network has better performance than BP network under the condition of same number of hidden layers and neurons. When there are something wrong with spectrum sensing, channel states with error and correct channel states both can be used as learning labels. The simulation results indicate that prediction accuracy for the latter one is about 4% better than the previous one.**

Keywords-spectrum prediction; LSTM network; BP network

I. INTRODUCTION

With the increase of communication traffic and mobile devices, the contradiction between the urgent spectrum demand and the shortage of bandwidth is increasingly fierce [1]. In general, communication users get information about channel states through spectrum sensing while spectrum prediction technology achieves the state information of channels in ahead by mining internal relevance of spectral historical data [2]. This technology can save considerable time and energy consumed by spectrum sensing to improve spectrum utilization. In addition, errors of receiving signals will be reduced if combining spectrum prediction with sensing. Frequency hopping communication is often used in anti-interference technology and spectrum prediction can also play an important role here. Once interference signal is predicted, the frequency of communication should be adjusted in order to operate signal shielding. Therefore, spectrum prediction is of great importance, and more and more researchers explore effective prediction algorithms [3].

With the development of computing power and the explosive growth of data, deep learning as a branch of machine learning has attracted wide attention in recent years. Deep learning has made great progress in image recognition, target detection and natural language processing. How to take advantage of deep learning to efficiently and effectively predict spectrum is an emerging research point. State information of multiple channels is a multi-dimensional time series, and recurrent neural network (RNN) is often used to solve time series problems. Long short time memory (LSTM) network is an improved version of RNN, which can tackle the vanishing gradient issue by introducing threshold mechanism so that long term memory and short term memory both exist. Reference [4] constructs a network with one LSTM layer and three Dense layers to determine the availability of multiple channels.

The main contributions of this paper are summarized as follows:

- Complex network with many LSTM layers is constructed to predict frequency hopping sequence. In addition threshold is set to determine the output of networks to be idle, or occupied, so the optimal threshold value can be found as well as the maximum prediction accuracy when prediction is made.
- The influence of different networks varied in depth and width on prediction accuracy is studied on the same data set. Meanwhile, the performance of conventional back propagation (BP) network with the same size is also studied as the baseline. The simulation shows that LSTM network has better predictive performance than BP network.
- Additionally this paper investigates sensing errors at the receiver end. Two situations which are the correct channel states and channel states with error as learning labels are studied. The simulation shows that choosing the correct channel states as labels has some performance improvements.

The rest of this paper is organized as follows. Section II introduces the structure and principle of BP network and LSTM network. Section III briefly introduces the sliding window method for constructing data sets, which is followed by simulation results and discussions in Section IV. Finally, we conclude the paper in Section V.

978-1-5090-6352-9/17/$31.00 ©2017 IEEE

II. SYSTEM MODEL

High frequency communication is an important technology in communication field. It is easy to exert interference by the adversary because of its locality, therefore frequency hopping technology is used to resist interference. If the interference frequency of the adversary can be predicted correctly by prediction algorithm, we can effectively avoid it. A simple frequency hopping pattern of channels whose number is N is shown in Fig. 1. The slot with red color means occupied while blue represents idle. The main work is to compare the predictive performance of BP and LSTM network, so the following is to illustrate these two networks.

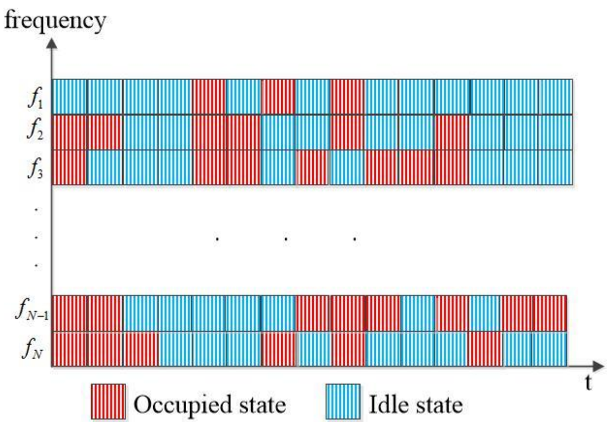

Figure 1. Hopping frequency pattern.

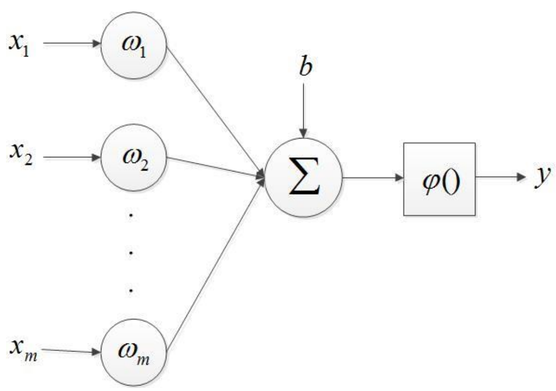

Figure 2. Structure of a single BP network neuron.

A. Conventional BP Network

BP neural networks are composed of neurons that can perform nonlinear operations. Fig. 2 shows a mathematical model of one neuron, and weights $\omega_1, \omega_2 \ldots \omega_m$ and bias b are parameters of neurons. $\varphi()$ represents activation function wh- ich often employs *tanh* or *sigmoid*. When input data is x_1, x_2 ... x_m, the output of single neuron is the nonlinear transformation of the input [5]:

$$y = \varphi(\sum_{i=1}^{m} \omega_i x_i + \mathrm{b}) \tag{1}$$

The structure of the classical neural network has the characteristics of full connection and no ring. Fig. 3 shows the BP network which is used to predict the pattern in Fig. 1. When the time window of channel states is set to T, the number of neurons in the input layer is $N*T$. The input data is denoted as $x \in R^{N*T}$ while the output result is denoted as $y \in R^N$. The output of the ith hidden layer is [5]:

$$\begin{gathered} h^l = \varphi^l(\sum_{i=1}^{n_{l-1}} h_i^{l-1} \omega_i^l + b^l) \qquad l = 1,2,\ldots,L \\ h^0 = x \\ h^L = y \end{gathered} \tag{2}$$

The neural network is trained by historical spectrum data, and its parameters are adjusted automatically using back propagation algorithm to achieve optimal target.

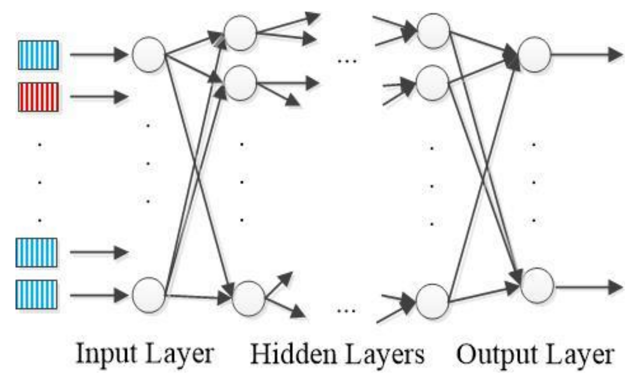

Figure 3. Structure of conventional BP network.

B. Emerging LSTM Network

LSTM cell introduces a variable called cell state and implements more complex nonlinear operations. An LSTM memory cell consists of three gates which are input gate, forget gate and output gate, as shown in Fig. 4 [6]. The result of input gate is calculated as:.

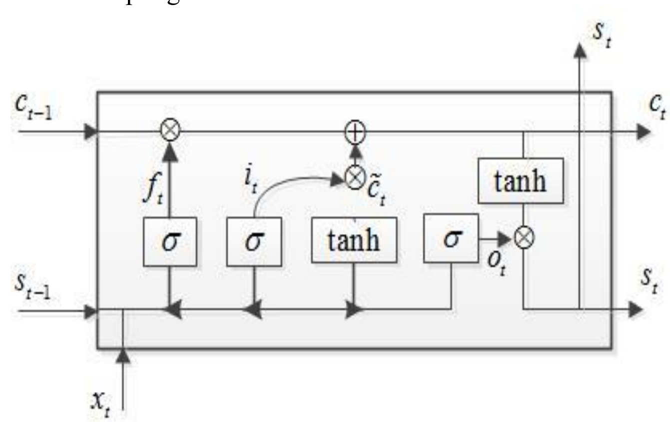

Figure 4. Structure of an LSTM cell memory.

$$i_t = \sigma(\mathrm{W}_x^t \bullet x_t + \mathrm{W}_s^t \bullet s_{t-1} + b_t) \tag{3}$$

The result of forget gate is calculated as:

$$f_t = \sigma(\mathrm{W}_x^f \bullet x_t + \mathrm{W}_s^f \bullet s_{t-1} + b_f) \quad (4)$$

The result of output gate is calculated as:

$$o_t = \sigma(W_x^o \bullet x_t + W_s^o \bullet s_{t-1} + b_o) \quad (5)$$

The input and output states of the cell are:

$$\tilde{c}_t = \tanh(W_x^c \bullet x_t + W_s^c \bullet s_{t-1} + b_c) \quad (6)$$

$$c_t = i_t * \tilde{c}_t + f_t * c_{t-1} \quad (7)$$

Finally, the hidden layer output is calculated as:

$$s_t = o_t * \tanh(c_t) \quad (8)$$

Here, i, f, o and c denotes the input gate, forget gate, output gate, and cell state, respectively. W and b are still the parameters of cell. Fig. 5 illustrates the LSTM network structure used to predict frequency hopping pattern in Fig. 1. There are N neurons in the input layer and historical data of one channel with T slots are entered into the corresponding neuron. LSTM network will output a vector in each slot, but what we are concerned about is predictive state in $T+1$ slot. Therefore, only the final vector of the output layer is chosen as the predictive state of the channels.

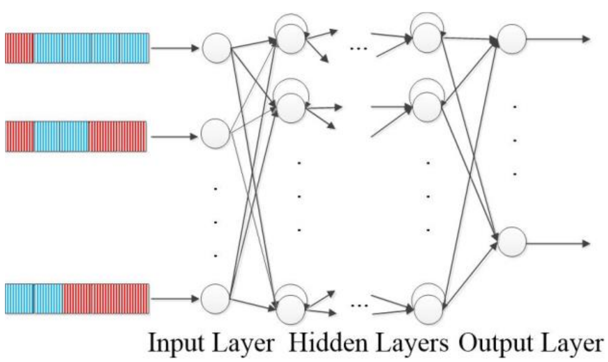

Figure 5. Structure of emerging LSTM network.

C. Main Differences between BP and LSTM Network

Some differences between BP and LSTM network have been mentioned above, and some supplements are made here:

Intuitively, the historical information of multiple channels as the input of BP network must be flat out into a column vector, but there is no need to do so for LSTM network. Therefore, the neurons in the input layer of BP network are more than LSTM network.

For BP network, neurons in the hidden layers have no feedback and their output is only related to the value of the current moment. The neurons of LSTM network have feedback, and their output is also influenced by the previous time slots [7]. The activation function of both network are usually different, and LSTM uses *Relu* instead of *tanh* and *sigmoid*.

In addition, there are some techniques in deep learning that can effectively avoid overfitting such as Dropout [8]. It randomly forces some neurons not to work during the training process so that the rest of neurons can learn better.

III. SLIDING WINDOW

Historical state information of any channel can be achieved by spectrum sensing in actual communication system. The commonly used spectral sensing methods include energy detection [9], matching filtering [10] and cyclic stable feature detection [11] etc. Spectrum prediction is to deal with time series and supervised learning is employed. The original data is a long sequence and the data set is constructed through a sliding window as shown in Fig. 6. Suppose that length of sliding window is 6, the output of network is predictive states of multiple channels in slot *t7* while the known vector perceived by the receiver in slot *t7* is selected as the label of this sample. With this kind of push, the sliding window slides a time slot forward to form a sample, which makes up the data set of the neural network. The data set is split into a training set, a validation set, and a test set according to the proportion. Neural network parameters are trained based on optimization algorithm, and the prediction accuracy is calculated on the test set.

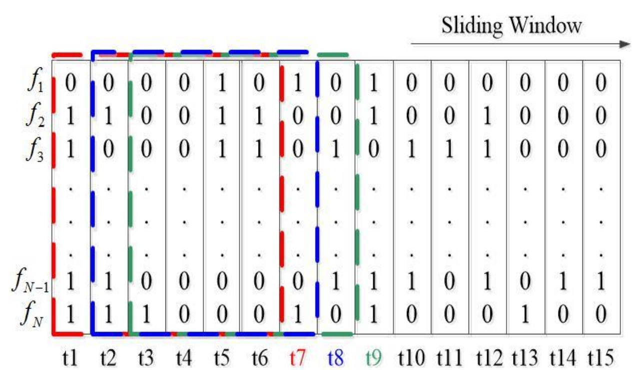

Figure 6. Sliding window approach.

IV. SIMULATION RESULTS

A. Model Structure

The structure of LSTM network constructed in this paper has L hidden layers, which are all LSTM layers. Each layer determines *Relu* as its activation function, and the input of the high-level hidden layer is the output vector of lower hidden layer. The output of the last hidden layer goes into a Dense layer whose activation function is *linear*. The output of one neuron in the output layer corresponds to the predictive state sequence of a channel. The value above the threshold γ_0 is converted to 1, whereas the value is converted to 0. The training target is *Mean Square Error* and the optimizer is *Adam*.

In this paper, the number of neurons in hidden layers of BP network is the same as that of the LSTM network except the input layer. The hidden layer neurons of BP network are all connected with neurons in the last layer, and the optimal method of training is to select the most basic *gradient descent* algorithm.

B. Simulation Analysis

Simulation data are used in these experiments, and a frequency hopping pattern with length of 160 time slots for ten channels is constructed. Fig. 7 shows the predictive result of LSTM network which has two hidden layers with 40 neurons in each layer. As shown in Fig. 7, the prediction accuracy can be reach 90% when the threshold γ_0 is set between 0.35 and 0.45. The accuracy goes up and then down with the threshold changes, and it is unscientific for threshold setting to be too high or too low.

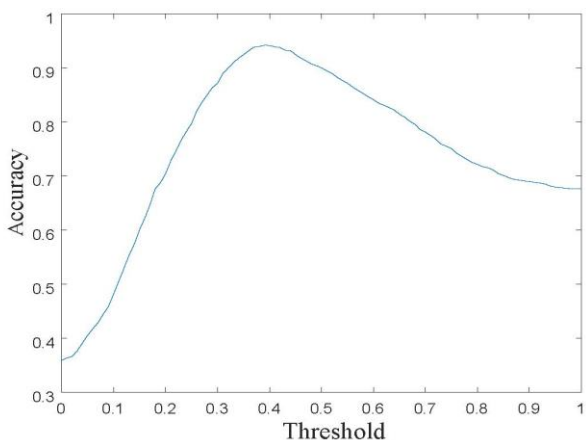

Figure 7. Prediction accuracy of LSTM network with different threshold settings.

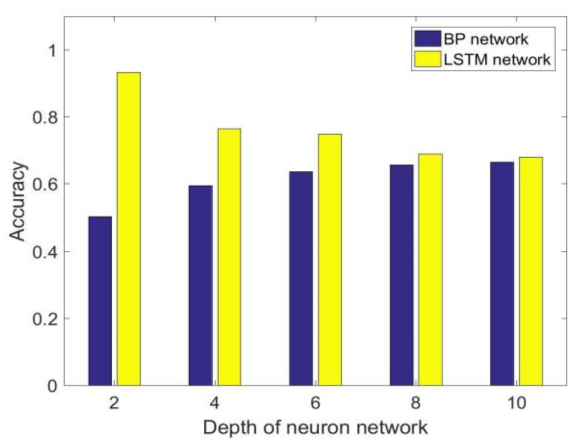

Figure 8. Prediction accuracy of LSTM and BP network with different depth.

Fig. 8 shows the variation of the prediction performance of BP and LSTM network with the same width as network depth changes. Accordingly, Fig. 9 illustrates the performance of BP and LSTM networks with two hidden layers as width varies. For this constructed data set, LSTM network apparently outperforms BP network supposing that both of them have the same number of hidden layers and neurons. In particular, the performance gap between these two networks is more obvious when layers are not so much.

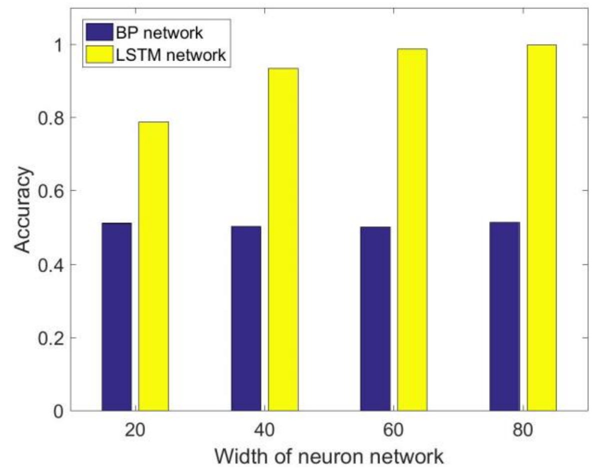

Figure 9. Prediction accuracy of LSTM and BP network with different width.

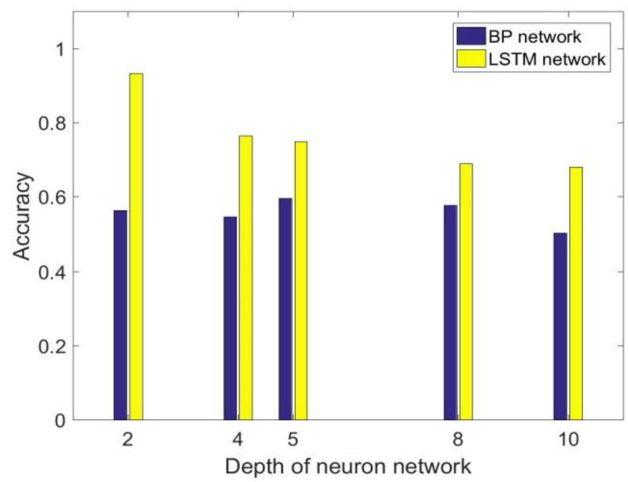

Figure 10. Prediction accuracy of LSTM and BP network with constant total neurons.

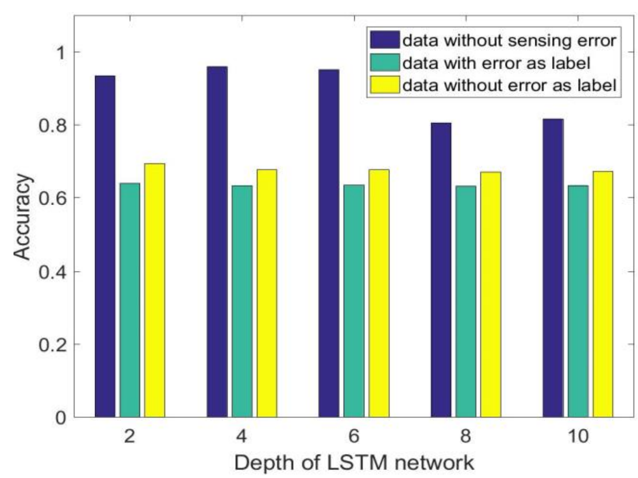

Figure 11. Prediction accuracy of LSTM network varied in depth for wrong sensing states.

Under the constraints of total number of neurons, the changes of accuracy with different network depth are shown in Fig. 10. We can conclude that shallow and wide LSTM network captures the characteristics of the data set better than the deep and narrow network. And the prediction performance of BP network has not changed significantly.

Miss detection and false alarm will inevitably occur during spectrum sensing of receiver. Supposing that miss detection probability and false alarm probability of the ith channel are denoted as p_m^i and p_f^i, channel states detected in the receiver are different from those of transmitter at certain error probability of corresponding channel. Therefore, the label of one sample for this supervised learning has two situations which are the correct information in the transmitter and the error information at the receiving end. In view of the above two situations, prediction accuracy of LSTM network in different depth is shown in Fig. 11. The situation where channel states without sensing error are chosen as the sample input is also drawn in this figure. Fig. 12 shows the performance of LSTM network with different widths in the same three cases. The prediction accuracy for correct channel states as labels is about 4% higher than that for wrong channel states.

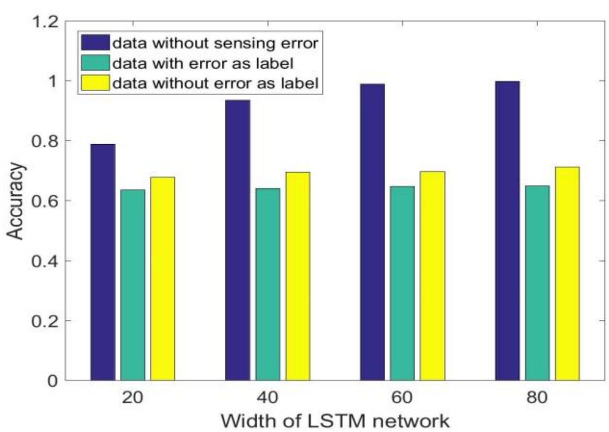

Figure 12. Prediction accuracy of LSTM network varied in width for wrong sensing states.

V. Conclusion

In this paper, we predict the frequency-hopping sequence through the emerging LSTM network and the conventional BP network. The result has demonstrated that LSTM network has some advantages in time series problems and better prediction performance. The effects of different network depth and width on prediction accuracy are also studied, and the results show that the width of network has a greater impact on prediction accuracy. Considering the actual situation, the spectrum sensing may take place phenomenon of false alarm and miss detection. The influence of different labels on the prediction accuracy is also studied in our work. The results show that correct channel states as labels have about 4 percent improvement in prediction performance than wrong channel states. In future studies, the optimization of models and the prediction of actually measured channel data will be considered.

References

[1] G. Ding, J. Wang, Q. Wu, Y. D. Yao, R. Li, H. Zhang, and Y. Zou. "On the limits of predictability in real-world radio spectrum state dynamics: From entropy theory to 5G spectrum sharing," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 178–183, Jul. 2015.

[2] X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng, "Spectrum prediction in cognitive radio networks," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 90–96, Apr. 2013.

[3] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y. D. Yao, and L. Hanzo, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Communications Surveys & Tutorials*, no. 99, Sep. 2017.

[4] L. Yu, Q. Wang, Y. Guo, and P. Li. "Spectrum availability prediction in cognitive aerospace communications: A deep learning perspective," in *Proceedings of Cognitive Communications for Aerospace Applications Workshop*, pp. 1-4, Aug. 2017.

[5] D. C. Plaut, S. J. Nowlan, and G. E. Hinton. "Experiments on learning by back propagation," *Tech. Rep. CMU-CS-86-126*, *Carnegie Mellon Univ.*, Jun. 1986.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural *Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May. 2015.

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.

[9] G. Ding, Q.Wu, Y. D. Yao, J. Wang, and Y. Chen, "Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 126–136, Jul. 2013.

[10] V. Jamali, A. Ahmadzadeh, and R. Schober. "On the design of matched filters for molecule counting receivers," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1711–1714, Aug. 2017.

[11] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.