Restricted Boltzmann Machine for Interference Pattern Learning in Broadband Receivers

Liset Martínez Marrero
Technological University of Havana
Havana, Cuba
liset@tele.cujae.edu.cu

Thomas Hunziker
Lucerne University of Applied Sciences and Arts
Lucerne, Switzerland
thomas.hunziker@hslu.ch

***Abstract*—Interference and noise mitigation is a critical component of many broadband communication systems. However, interference is often nonstationary, heavily dependent on the environment, and statistical a priori knowledge is not generally available. We propose a restricted Boltzmann machine (RBM) for unsupervised learning of time/frequency interference patterns in orthogonal frequency-division multiplexing (OFDM) receivers. Capable of learning patterns without statistical a priori knowledge, an RBM can be combined with a factor graph underlying a turbo or low-density parity-check decoder. We demonstrate the benefits of the proposed approach using the example of turbo encoded OFDM signal frames exposed to different forms of interference.**

***Index Terms*—interference mitigation, machine learning, restricted Boltzmann machine, turbo coding**

I. Introduction

Intense interference affects the reliability of many communications systems. Possible sources of interference are third party communications equipment operating in the same frequency band, but also a wide range of electromagnetic wave emitting devices such as home appliances, ignition systems, and switching processes in the power distribution network [1]. Cognitive radios have been proposed and shown to evade narrowband interference to a certain extent, but the nonstationary and highly random nature of impulsive noise may make the establishment of reliable data transmission a challenge.

In the following we use the term interference indistinctly for all sorts of interference and noise with statistical properties varying over time or frequency, including narrowband interference and impulsive noise. An example of an environment with heavy interference are high-voltage power substations [2], [3]. The harsh conditions there, mainly due to arc discharges, are difficult to overcome by wireless communications systems. Similar conditions are found in power line communications systems (PLC), where interference originates from switching processes of electrical devices [4]. Detailed descriptions of impulsive noise models and parameters for PLC and wireless communications are given in [5] and [6], respectively.

Many standards for broadband communications build on orthogonal frequency-division multiplexing (OFDM) as modulation scheme, such as digital audio broadcasting, digital video broadcasting, 3GPP long term evolution (LTE), broadband over power line, among others. Besides of evading a need for complex equalizers, OFDM has the advantage of confining interference to a subset of time/frequency slots (i.e., OFDM symbols/subcarriers). In combination with a channel interleaver and an efficient forward error correction (e.g. a low-density parity-check (LDPC) or turbo code), OFDM can cope relatively well with moderate interference. However, for dealing with frequently occurring bursts of interference or high interference power, additional mitigation techniques are required [7], [8].

Concerning impulsive noise mitigation, one technique is to do either clipping or blanking (i.e., nulling) of affected signal sections in the time domain [9]–[11]. The popularity of this simple approach is due to a favorable trade-off between computational complexity and achievable performance. By clipping or blanking signal sections with amplitudes above a predefined threshold, known as blanking threshold, impulsive noise peaks are eliminated. However, at the same time the information-bearing OFDM signal is affected, which is a significant drawback of these techniques [9]. Also, due to the inherently high peak-to-average power ratio, distinguishing between OFDM signal peaks and noisy impulses represents a challenging task [10].

Other approaches pursue to estimate the interference given its sparsity, either in the time domain or in frequency, to subsequently cancel it from the received signal [12], [13]. Sparse Bayesian learning, among other compressed sensing (CS) techniques, has been used for this purpose [14]–[16]. CS takes advantage of information carried on null and pilot tones to estimate impulsive noise. However, this implies a reduced bandwidth efficiency and throughput. In addition, CS methods involve a matrix inversion which may become rather complex for OFDM symbols with a large number of subcarriers.

Combining interference mitigation with LDPC and turbo decoding has been proposed in a few works. In [17] the factor graph used for decoding is extended to incorporate channel gain and noise power estimation, whereas in [18] a particle filter is proposed to account for varying noise power.

In this work we likewise build on factor graphs used for signal decoding, however, we employ a restricted Boltzmann machine (RBM) for assessing the characteristics of the interference. As a method from machine learning, there is no requirement for statistical a priori knowledge about the interference, as opposed to Bayesian approaches to interference mitigation. Whether the interference exhibits narrowband characteristics, from nearby signal transmitters, or intermittent forms, possibly originating from electric devices, the RBM tackles the challenge in the same unbiased manner. This is a significant advantage for the many wireless communication systems encountering varying sorts of interference from environment to environment. To our knowledge, employing unsupervised learning for interference pattern learning and mitigation in OFDM receivers has not yet been studied.

The structure of this work is as follow. Sect. II gives a brief introduction to RBMs. In Sect. III the system model is introduced and the factor graph, extended by the RBM, described. An evaluation of our proposal by means of Monte-Carlo simulations follows in Section IV. Finally, conclusions are drawn in Section V.

II. Restricted Boltzmann Machine

A RBM is a two-layer artificial neural network (ANN), capable of learning patterns in an unsupervised training procedure. A layer of visible units is connected to a layer composed of hidden units. In our setup, the states of the visible units represent the variances $\sigma_1^2, \ldots, \sigma_K^2$ of the noise terms in the signal observation as explained in Sect. III, and they are real positive, i.e., $\sigma_k^2 \in (0, \infty)$. The M hidden states $h_1, \ldots, h_M$, on the other hand, are binary: $h_j \in \{0, 1\}$.

Every combination of visible states $\mathbf{v} = (\sigma_1^2, \ldots, \sigma_K^2)^{\mathrm{T}}$ and hidden states $\mathbf{h} = (h_1, \ldots, h_M)^{\mathrm{T}}$ is assigned an energy through [20]

$$E(\mathbf{v}, \mathbf{h}) = -\left(\mathbf{a}^{\mathrm{T}}\mathbf{v} + \mathbf{b}^{\mathrm{T}}\mathbf{h} + \mathbf{v}^{\mathrm{T}}\mathbf{W}\mathbf{h}\right). \tag{1}$$

The $(K \times M)$-matrix $\mathbf{W}$ in (1) contains the weights, $\mathbf{a}$ and $\mathbf{b}$ are bias (column) vectors of size K and M, respectively, and T in the superscript denotes transposition. In the course of the network training, $\mathbf{W}$, $\mathbf{a}$ and $\mathbf{b}$ are adjusted such that the energy, also sometimes referred to as the cost, decreases.

The energy function is used to define the joint probability

$$f(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp\left(-E(\mathbf{v}, \mathbf{h})\right), \tag{2}$$

where Z is a normalizing constant ensuring that the probabilities of all possible combinations of visible and hidden states sum up to 1.

As a central feature of RBMs, the hidden states $h_1, \ldots, h_M$ are conditionally independent given the visible states. As follows from (2), the conditional probability of $\{h_j = 1\}$ given $\mathbf{v}$ is

$$\Pr(\{h_j = 1\} \mid \mathbf{v}) = \left(1 + \exp\left(-b_j - \sum_{k=1}^{K} \sigma_k^2 w_{kj}\right)\right)^{-1}, \tag{3}$$

where b_j denotes the jth element of $\mathbf{b}$ and w_{kj} the jth element in the kth row of $\mathbf{W}$. Similarly, the visible states $\sigma_1^2, \ldots, \sigma_K^2$ are conditionally independent given the hidden states. As follows from (2), the conditional probability density of σ_k^2 given $\mathbf{h}$ has the form of the density of an exponential distribution with rate parameter

$$\lambda = -a_k - \sum_{j=1}^{M} h_j w_{kj} \tag{4}$$

and mean λ^{-1}.

The mutual conditional independence facilitates the training procedure. As shown in [19], the learning rule can be derived from the method of steepest descent. Given observed visible states $\mathbf{v}_{\mathrm{data}}$ with a sampled vector of hidden states $\mathbf{h}_{\mathrm{data}}$, along with a "typical" combination $(\mathbf{v}_{\mathrm{model}}, \mathbf{h}_{\mathrm{model}})$ of visible and hidden states, the weight matrix is updated according to

$$\Delta\mathbf{W} = \epsilon \cdot \left(\mathbf{v}_{\mathrm{data}}\mathbf{h}_{\mathrm{data}}^{\mathrm{T}} - \mathbf{v}_{\mathrm{model}}\mathbf{h}_{\mathrm{model}}^{\mathrm{T}}\right) \tag{5}$$

with ϵ representing the learning rate. The bias vectors are updated in a similar fashion, that is, $\Delta\mathbf{a} = \epsilon \cdot (\mathbf{v}_{\mathrm{data}} - \mathbf{v}_{\mathrm{model}})$ and $\Delta\mathbf{b} = \epsilon \cdot (\mathbf{h}_{\mathrm{data}} - \mathbf{h}_{\mathrm{model}})$.

The combination $(\mathbf{v}_{\mathrm{model}}, \mathbf{h}_{\mathrm{model}})$, representing the current model, is obtained from $(\mathbf{v}_{\mathrm{data}}, \mathbf{h}_{\mathrm{data}})$ through Gibbs sampling. This involves a number of alternating rounds of sampling the hidden states on the basis of the visible states and vice versa, according to the probability distributions given by (3) and (4), respectively. For more insight into RBMs we refer to [20].

III. System Model

A transmitted signal frame is assumed to convey N information bits, encoded and modulated onto a complex baseband signal vector $\mathbf{s} = (s_1 \ldots, s_K)$. For forward error correction we assume a turbo code, incorporating two constituent encoders and an interleaver. The well-established iterative turbo decoding procedure building on belief propagation – and the sum-product algorithm – can be derived from a factor graph [21], composed of so-called variable nodes and factor nodes (see Fig. 1). The graph can be deduced from the factorized joint probability of the transmitted bits $[\mathbf{i}\ \mathbf{p}\ \mathbf{q}]$, the noise power $\mathbf{v}$, the signal observarion $\mathbf{r}$, and a hidden state vector $\mathbf{h}$:

$$\begin{aligned} f([\mathbf{i}\ \mathbf{p}\ \mathbf{q}], \mathbf{r}, \mathbf{v}, \mathbf{h}) = {} & f(\mathbf{i}) \cdot f(\mathbf{p} \mid \mathbf{i}) \cdot f(\mathbf{q} \mid \mathbf{i}) \\ & \times f(\mathbf{v}, \mathbf{h}) \\ & \times f(\mathbf{r} \mid [\mathbf{i}\ \mathbf{p}\ \mathbf{q}], \mathbf{v}). \end{aligned} \tag{6}$$

- The vector $\mathbf{i}$ contains the N information bits.
- The vectors $\mathbf{p}$ and $\mathbf{q}$ contain the parity bits introduced by the two constituent encoders of the turbo code.
- The composite vector $[\mathbf{i}\ \mathbf{p}\ \mathbf{q}]$ comprises the transmitted bits.
- The vector $\mathbf{r} = (r_1, \ldots, r_K)$ contains the K signals in the frame as observed at the receiver end.
- Each observed signal r_k is subject to a random noise term with variance σ_k^2.

The above factors have the following forms:

- The N bits are assumed independent and $f(\mathbf{i}) = 2^{-N}$.
- The parity bits are deterministically dependent on $\mathbf{i}$ and thus $f(\mathbf{p} \mid \mathbf{i}) \in \{0, 1\}$ and $f(\mathbf{q} \mid \mathbf{i}) \in \{0, 1\}$.

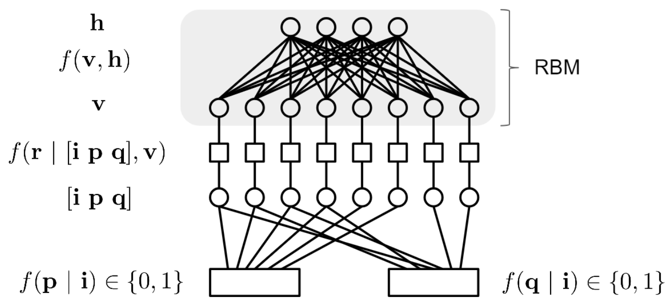

Fig. 1. Factor graph augmented by a RBM: the variable nodes are shown as circles and the factor nodes in rectangular form.

- The varying noise power is dependent on the hidden states through $f(\mathbf{v},\mathbf{h})$ as described in Sect. II.
- The additive noise terms contained in the observed signals are independent and complex Gaussian distributed. Hence,

$$f(\mathbf{r} \mid [\mathbf{i}\ \mathbf{p}\ \mathbf{q}],\mathbf{v}) = \prod_{k=1}^{K} \frac{1}{\pi\sigma_k^2} \exp\left(-\frac{|r_k - s_k|^2}{\sigma_k^2}\right), \quad (7)$$

  where the mapping $[\mathbf{i}\ \mathbf{p}\ \mathbf{q}] \mapsto (s_1 \ldots, s_K)$ depends on the applied modulation method.

The decoding proceeds by passing messages among the variable and factor nodes of the factor graph. The messages have the form of likelihood ratios conveying extrinsic information from the various decoding rounds.

As for the RBM augmenting the factor graph, the bit likelihood ratios need to be translated into noise power values for the visible units. For each signal we compute the probabilities $\Pr\{s_k = \mu_i\}$ of the I possible noise-free signals $(\mu_i)_{i=1,\ldots,I}$, and use these probabilities for estimating the noise power according to

$$\widehat{\sigma}_k^2 = \sum_{i=1}^{I} \Pr\{s_k = \mu_i\} \cdot |r_k - \mu_i|^2. \quad (8)$$

The step is simple for binary modulated signals (where $I = 2$) and more complex in the case of higher-order modulation. The more accurate the bit likelihood values become, the more accurate the noise power estimates.

In downward direction, the noise power estimates are used to update the likelihood ratios of the elements in $[\mathbf{i}\ \mathbf{p}\ \mathbf{q}]$ on the basis of (7).

The order in which messages are passed along the factor graph including the RBM can be chosen arbitrarily. However, letting messages propagate from the two bottom factor nodes upwards and through the RBM, then back downwards, and repeating this procedure a certain number of times seems a natural choice. Convergence towards the optimal solution, i.e., maximizing the a posteriori bit probabilities, is not guaranteed, but often near optimal results are obtained by the sum-product algorithm.

IV. Numerical Results

The performance of the proposed interference power estimation has been investigated using Monte-Carlo simulations. As an example of a state-of-the-art coded OFDM scheme, turbo encoded binary modulated signals are mapped onto 320 subcarriers over 26 consecutive OFDM symbols. The 26×320 bits are obtained from a turbo encoder similar to the rate $\frac{1}{2}$ encoder specified in the IEEE 1901 standard for PLC [22]. Employing two parallel systematic convolutional encoders of rate $\frac{2}{3}$ and a turbo interleaver, 4160 information bits are encoded into codewords of length 8320. Tail biting is used for termination.

As we are interested in the impact of interference, a channel with a flat magnitude response is assumed. After discrete Fourier transform, the receiver employs an iterative turbo decoding procedure on the basis of the factor graph in Fig. 1. Each iteration involves an update of the bit likelihood values as per the first constituent encoder, an update of the bit likelihood values as per the second encoder, and an update of the noise power estimates. For the latter step, the time/frequency plane is subdivided into partitions of size 13×20, i.e., 20 adjacent subcarriers over 13 OFDM symbols. In each partition, the observed noise values are squared and taken to define the visible states in the vector $\mathbf{v}$. The $M = 50$ hidden states in $\mathbf{h}$ are then generated according to (3), and from these the noise power estimates are computed according to (4).

The RBM can be trained during burst receptions by carrying out additional updates of the hidden states and visible units (i.e., Gibbs sampling), or while listening to the idle channel as we did. We carried out a large number of RBM training steps, each involving 10 updates of the hidden states and visible units before adjusting the weights. The results presented in the following are obtained after training of the RBM.

In the first scenario, artificially generated noise patterns are used as shown in Fig. 2 (above). In each partition, one half of the 260 time/frequency slots are subject to Gaussian distributed noise with variance σ_{H}^2 and the other half to variance σ_{L}^2, where $\sigma_{\mathrm{H}}^2/\sigma_{\mathrm{L}}^2 = 10$. One of the 10 noise pattern is randomly chosen for each partition. The average bit error rate (BER) after different numbers of iterations is shown in Fig. 3 versus the signal-to-noise power ratio (SNR), defined as the ratio of the bit energy ε_{b} and the noise variance *averaged* over the time/frequency slots.

As expected, the BERs decay with the number of iterations of the turbo decoder. Including noise power estimation (NPE) through the described RBM clearly improves receiver performance. As the first NPE runs just after the first iteration, the benefits become visible after the second iteration, ending up with a gain of approx. 1.5 dB after 10 iterations compared to a decoder relying on a single average noise power estimate.

In the second scenario we assume an intermittent narrowband interferer, partially overlapping the received signal. Some of the random interference patterns are shown in Fig. 2 (below). The interferer has a power spectral density of 20 dB above the additive white Gaussian noise power density. With-

Fig. 2. Noise power patterns of size 13×20. White colored time/frequency slots are subject to intense noise power and black colored slots are subject to weak noise. Above: artificially generated patterns with 10 dB difference in noise power. Below: patterns resulting from narrowband interference partially overlapping in time.

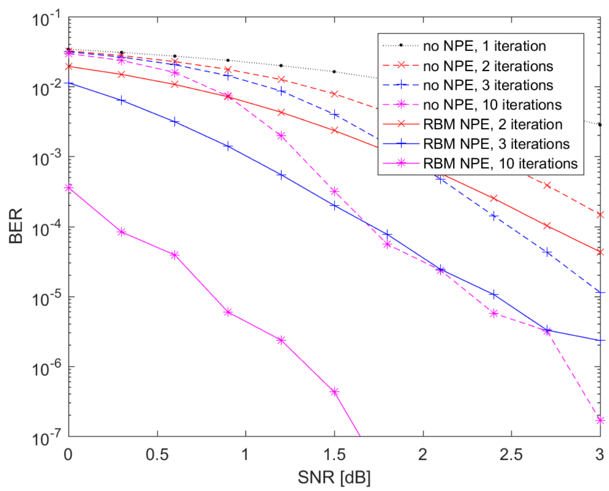

Fig. 3. Randomly generated noise power patterns: decoder performance including trained RBM for NPE, and without NPE (assuming uniform noise power).

out NPE, the large impulsive interference terms in a small subset of input values throw the bit likelihood computations by sum-product algorithm off the track. The RBM, after proper training, recognizes the interference patterns and facilitates an improvement from iteration to iteration, as seen in Fig. 4. Through NPE the decoder assesses the information content in the input values, achieving BERs in the order of 10^{-6} after five iterations at an average interference-plus-noise power level just below the signal power.

V. Discussion

In many environments, broadband wireless communication systems are exposed to nonstationary interference, sometimes

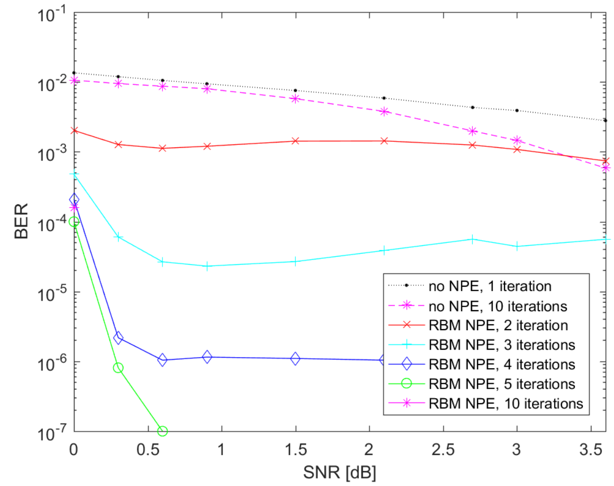

Fig. 4. Intermittent narrowband interferer: decoder performance including trained RBM for NPE, and without NPE (assuming uniform noise power).

from analog electrical devices, and in other situations from third party communications equipment. The characteristics of the interference may change completely from one environment to another environment. Neural networks in the form of RBMs are capable of building up a knowledge of the interference patterns encountered in a certain environment. After training, a receiver can use the RBM to identify a particular interference pattern and in the same time properly weight the observed signals. As a method from machine learning, RBMs do not require a priori information, making them superior to many other noise estimation approaches building on Bayesian statistics.

We have demonstrated the benefits of RBMs for dealing with certain forms of interference, improving receiver performance by several decibels. There are in fact many ways in which ANN can be applied for interference pattern recognition in OFDM receivers. Further research studying the benefits of other forms of ANN seems worthwhile, such as convolutional neural networks, or also some forms of recurrent neural networks.

References

[1] I. Landa, A. Blázquez, M. Vélez and A. Arrinda, "Impulsive radio noise levels interfering wireless systems up to 1 Ghz," Antennas, Propagation & RF Technology for Transport and Autonomous Platforms 2017, Birmingham, pp. 1-5, 2017.

[2] B.L. Agba, F. Sacuto, M. Au, F. Labeano and F. Gagnon, "Wireless Communications for Power Substations: RF Characterization and Modeling," Cham: Springer International Publishing, 2019.

[3] X. Wang, L. Mei and F. Labeau, "Application of WFRFT in Impulsive Noise Channels of Substation Communications," 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), pp. 1-5, 2017.

[4] L. Lampe, "Power Line Communications: Principles, Standards and Applications from multimedia to smart grid," John Wiley&Sons, 2016.

[5] T. Shongwe, A.J. Han Vinck and H.C. Ferreira, "A study on impulse noise and its models," 2015.

[6] J. Lago-Fernandez and J. Salter, "Modeling impulsive interference in DVB-T: statistical analysis, test waveforms and receiver performance," BBC R. and D. Whitepaper (WHP 80), 2004.

[7] C. Soltanpur, K.M. Rabie, B. Adebisi and A. Wells, "Masreliez-equalized VOFDM in non-Gaussian channels: Power line communication systems," IEEE Systems Journal, 2017.
[8] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," IEEE Trans. Commun., vol. 4, pp. 145-147, 1996.
[9] U. Epple, D. Shutin and M. Schnell, "Mitigation of Impulsive Frequency-Selective Interference in OFDM Based Systems," IEEE Wireless Communications Letters, vol. 1, no. 5, pp. 484-487, 2012.
[10] U. Epple and M. Schnell, "Advanced Blanking Nonlinearity for Mitigating Impulsive Interference in OFDM Systems," IEEE Transactions on Vehicular Technology, 2016.
[11] F. H. Juwono, R. Reine, L. Liu, J. Lino.and D.D. Huang, "Performance of impulsive noise blanking in precoded OFDM-based PLC systems," IEEE International Conference on Communication Systems (ICCS), pp. 1-6, 2016.
[12] S. Liu, F. Yang, W. Ding and J. Song, "Double Kill: Compressive Sensing Based Narrowband Interference and Impulsive Noise Mitigation for Vehicular Communications," IEEE Transactions on Vehicular Technology, vol. PP, no. 99, 2015.
[13] P. D. Mariyam, F. H. Juwono, P. D. Pamungkasari and D. Gunawan, "How to deal with impulsive noise in OFDM-based PLC: A survey," International Conference on Electrical Engineering and Informatics (ICELTICs), pp. 163-168, 2017.
[14] W. Ding, Y. Lu, F. Yang, W. Dai, P. Li, S. Lino.and J. Song, "Spectrally Efficient CSI Acquisition for Power Line Communications: A Bayesian Compressive Sensing Perspective," IEEE Journal on Selected Areas in Communications, vol. 34, no. 7, pp. 2022-2032, 2016.
[15] M. Korki, J. Zhangy, C. Zhang and H. Zayyani, "An iterative bayesian algorithm for block-sparse signal reconstruction," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2174-2178, 2015.
[16] L. Bai, M. Tucci and M. Raugi, "Impulsive Noise Mitigation with Interleaving Based on MUSIC in Power Line Communication," IEEE Transactions on Smart Grid, 2018.
[17] M. Nassar, P. Schniter and B.L. Evans, "A factor graph approach to joint OFDM channel estimation and decoding in impulsive noise environments," IEEE Transactions on Signal Processing 62, no. 6, pp. 1576-1589, 2014.
[18] S. Wang, L. Cui, S. Cheng, Y. Zhai, M. Yeary and Q. Wu, "Noise adaptive LDPC decoding using particle filtering," IEEE Transactions on Communications 59, no. 4, pp. 913-916, 2011.
[19] G.E. Hinton, "A practical guide to training restricted Boltzmann machines," In Neural networks: Tricks of the trade, pp. 599-619. Springer, Berlin, Heidelberg, 2012.
[20] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the national academy of sciences 79, no. 8, pp. 2554-2558, 1982.
[21] F.R. Kschischang, B.J. Frey and H.A. Loeliger, "Factor graphs and the sum-product algorithm," IEEE Transactions on Information Theory, vol. 47, pp. 498-519, 2001.
[22] "IEEE Standard for Broadband over Power Line Networks: Medium Access Control and Physical Layer Specifications," IEEE Std 1901-2010, 2010.