Informed Scheduling by Stochastic Residual Belief Propagation in Distributed Wireless Networks

Panagiotis-Agis Oikonomou-Filandras, Kai-Kit Wong, and Yangyang Zhang

***Abstract*—This letter devises a novel algorithm for cooperative spectrum sensing based on belief propagation (BP) for distributed wireless networks. The algorithm, called stochastic residual belief propagation (SR-BP), extends the use of residual belief propagation (R-BP) to distributed networks, improving the accuracy, convergence rate, and communication cost for cooperative spectrum sensing. We demonstrate that SR-BP converges to a unique fixed point under conditions similar to those ensuring convergence of asynchronous BP. Then, we develop a way to derive a probability distribution from the residual of each message. Finally, we provide numerical results to showcase the improvements in convergence speed, message overhead and detection accuracy of SR-BP.**

***Index Terms*—Cooperative spectrum sensing, distributed inference, message passing, residual belief propagation.**

I. Introduction

Cognitive radio promises a more efficacious use of spectrum as licensed bands can be used by opportunistic users, namely secondary users (SUs), when the primary users (PUs) are idle. There has been a large body of work addressing the problem of spectrum sensing and one prominent technique is cooperation between SUs to improve channel detection.

In particular, probabilistic graphical models (PGMs) [1], a highly successful tool for reasoning coherently from limited and noisy observations, have recently gained great attention for cooperative spectrum sensing. For example, factor graphs were utilized in [2] to model the cooperative spectrum sensing problem. In addition, belief propagation (BP) was adopted for cooperative spectrum sensing in heterogeneous cognitive radio networks [3]. Earlier in [4], a hidden Markov field model was recommended. Then both central using a fusion center [5] and distributed detection algorithms [6] have been proposed.

Despite the widespread use of BP, several non-trivial problems remain. Problematically, in PGMs with loops, BP often diverges and beliefs can become overpowered. For spectrum sensing where both the capacity for overhead communication and the time the SUs have to decide on transmission are limited and expensive, improving the above can greatly enhance the opportunistic communication. A lot of research has been carried out to address these issues and provide some insight into the dynamics and convergence properties of BP.

In particular, Yedidia *et al.* in [7] demonstrated that BP can be interpreted as performing a constrained minimization of the so-called Bethe free energy. Convergence conditions were proposed in [8]–[10]. Moreover, algorithms that ameliorate the effects of cycles by weighting messages have been proposed in [11], [12]. Remarkably, the importance of message scheduling in BP has also been recognized, and residual BP (R-BP) has hence been proposed as an algorithm implementing a greedy informed schedule for message passing [13]. This gave rise to a number of variants of BP in LDPC decoding that provide more elaborate informed schedules for R-BP, e.g., [14].

Distributed wireless networks, such as in the application of *cooperative spectrum sensing*, however, pose new challenges for BP as there is practically no central entity to manage global information, hence making algorithms like tree reweighing BP (TR-BP) [11] and R-BP *unusable* in the distributed case. In [15], [16], Wymeersch *et al.* devised a distributed variant of TR-BP, called uniformly reweighed BP (URW-BP), but did not study the use of informed scheduling for distributed networks.

This letter proposes a novel R-BP algorithm for cooperative spectrum sensing that employs a stochastic message scheduler based on the residuals at each node. This will result in faster convergence, less overhead, and improved results, but also can be integrated with enhancing algorithms such as URW-BP for further performance gain. Our contributions are as follows:

- We devise for the first time to our knowledge, a general informed scheduling framework for distributed sensors.
- Based on the framework, we then present a stochastic R-BP (SR-BP) technique as a practical alternative to perform R-BP for distributed inference.
- We prove that when BP converges, given similar conditions to R-BP, SR-BP will also converge.
- A probability density function (pdf) parametrized by the residuals is proposed for an efficient message schedule.
- Simulation results illustrate that for cooperative spectrum sensing SR-BP greatly outperforms BP and can be used in conjunction with other BP variants like URW-BP.[1]

II. Cooperative Spectrum Sensing as Distributed Hypothesis Testing

We consider a cognitive radio network where N SUs are distributed within a certain region. A PU transmits in the same area with a probability $\Pr(X)$ in which $X \in \{0, 1\}$ is a binary variable representing the state of the channel, i.e., $X = 0$ is idle and $X = 1$ is busy, respectively. We assume a frequency-flat fading channel between the PU and the SUs. Let $1/T_s$

Manuscript received July 30, 2014; accepted November 14, 2014. Date of publication December 2, 2014; date of current version February 18, 2015. This work is supported under EPSRC grant EP/H011536/1. The associate editor coordinating the review of this paper and approving it for publication was H. Jiang.

P.-A. Oikonomou-Filandras and K.-K. Wong are with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: p.filandras@ee.ucl.ac.uk).

Y. Zhang is with Kuang-Chi Institute of Advanced Technology, Shenzhen 518057, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LWC.2014.2376562

[1] As SR-BP only affects the schedule of the transmitted messages, it can be readily applied to all distributed BP variants designed for cooperative spectrum sensing, as well as more general distributed inference algorithms that use message passing, assuming that a residual can be calculated for the messages.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/

be the sampling rate for all the SUs and N_s be the number of observations (samples) obtained by each SU. Each node s observes the signal vector $y_s = [y_s[1], \dots, y_s[N_s]$, where the received signal $y_s[n] \triangleq y_s(nT_s)$ is given by

$$y_s[n] \sim \begin{cases} \mathcal{CN}\left(0, \rho_s^2 + \sigma^2\right), & \text{if the channel is busy,} \\ \mathcal{CN}(0, \sigma^2), & \text{if the channel is idle,} \end{cases} \tag{1}$$

in which the notation $\mathcal{CN}(a,b)$ denotes the complex Gaussian distribution with mean a and variance b.

Let $y = [y_1, \dots y_N]$ denote[2] the observation of all the SUs, and X_s be the state of the channel inferred by SU s. The aim of SU s is to find the maximum *a posteriori* probability (MAP):

$$\hat{x}_s = \arg\max_{x_s} \Pr(X_s = x_s | Y = y), \text{ for } s = 1, \dots N, \tag{2}$$

where cooperation between the SUs is implied in the condition. For simplicity, we assume that all correlations between the SUs are pairwise.[3] Hence, the joint pdf is written as

$$\Pr(X|Y = y) \propto \prod_{s=1}^{N} \Phi_s(X_s | Y_s = y_s) \prod_{s \neq r} \Phi_{s,r}(X_s, X_r), \tag{3}$$

in which Φ_s is the local potential function, and $\Phi_{s,r}$ is the compatibility potential function [1]. We define the univariate potentials $\Phi_s(X_s) \triangleq \Pr(X_s | Y_s = y_s)$ and the pairwise potentials $\Phi_{s,r}(X_s, X_r) \triangleq \exp(\lambda_{s,r} \mathbb{I}(X_s = X_r))$, with $\lambda_{s,r}$ the correlation factor between nodes s and r which is sampled uniformly between $[0.2, 4]$ as in [16], and $\mathbb{I}(\cdot)$ is the indicator function. Such a factorization as in (3) can be conveniently expressed as a graphical model such that factors are mapped to nodes, referred to as clusters, and edges connect clusters with common variables. To compute the marginals in (2) efficiently avoiding calculation on the joint pdf (3), BP is used.

BP calculates approximations of the marginals $\Pr(X_s|Y)$, called beliefs, $b_s(X_s)$ [11]. Once the graphical model of (3) is created, messages are passed between the clusters through the edges, until convergence or a specific number of iterations has passed.

In sum-product BP, the belief of each cluster (i.e., an approximation to the true marginal) at each iteration t is calculated as

$$b_s^{(t)}(X_s) = \Phi_s(X_s) \prod_{j \in \mathcal{N}_s} \mu_{j \to s}^{(t)}(X_{j,s}), \tag{4}$$

where the messages are calculated by

$$\mu_{r \to s}^{(t)}(X_s) \propto \sum_{X_r} \Psi_r(X_r) \frac{b_l^{(t-1)}(X_l)}{\mu_{s \to r}^{(t-1)}(X_{r,s})}. \tag{5}$$

where $\Psi_r(X_r)$ could be either $\Phi_r(X_r)$ or $\Phi_{r,s}(X_r, X_s)$, depending upon the type of factor in the node. The notation $\mathcal{N}_s$ denotes the set of neighboring clusters for cluster s. In addition, $X_{r,s}$ represents the common variables between r and s, which in this case simplifies to X_s. The practical aspects of implementation can be found in [16] from where they are followed precisely.[4]

[2]Uppercase letters are used to represent random variables, while lowercase letters represent observed values of the respective variables.

[3]This assumption is suitable for this application but does not limit the generality of SR-BP to only pairwise interactions.

[4]Since each message can be represented by only a single digit, we consider that the exchange of such information is free of collisions.

III. BP as an Optimization Problem

To devise SR-BP, we first describe BP as an iterative method. For completeness and clarity, we will present all the necessary definitions as presented in [13]. Specifically, each message can be viewed as residing in some message space $\mathcal{R} \subset (\mathbb{R}^+)^d$, where d denotes the dimension of the messages.[5] Hence, the set of messages, $\mathcal{M}$, in a cluster graph is a subset of $\mathcal{R}^{|\mathcal{M}|}$.[6] Let m and n denote the index of individual messages, v_m and v_n denote the mth and nth message respectively, and $v = [v_m, \dots, v_n] \in \mathcal{R}^{|\mathcal{M}|}$ denote a joint assignment of a subset of the messages. The update (4) can be understood as a mapping function $f_m : \mathcal{R}^{|\mathcal{M}|} \to \mathcal{R}$ that defines the mth message as a function of a subset of the messages. Then we can define an iterative method for each message m:

$$v_m^{(t+1)} = f_m(v^{(t)}). \tag{6}$$

Assuming convergence, we have the fixed point

$$f_m(v^*) = v_m^*. \tag{7}$$

Finally, we consider the *global* update functions that contain the iterative methods for all messages. The messages can be updated *synchronously* (i.e., all together simultaneously at each iteration), or *asynchronously* (i.e., only a subset gets updated at each iteration). The equations for the two schemes are:

$$F^s\left(v_1, \dots, v_{|\mathcal{M}|}\right) = \left(f_1(v), \dots, f_{|\mathcal{M}|}(v)\right), \tag{8}$$
$$F_m^a(v_1, \dots, v_{|\mathcal{M}|}) = \left(v_1, \dots, f_m(v), \dots, v_{|\mathcal{M}|}\right). \tag{9}$$

In the asynchronous case, we assume that there is a set of times $T = \{0, 1, 2, \dots\}$ at which one or more components v_m are updated. Also let $\mathcal{T}_m$ be the set of times v_m is updated. Then for the asynchronous case we adopt [13, Assumption 3.1].

Assumption III.1: For every message m, there is a finite time T_m so that for any time $t \geq 0$, the update $v := f_m^a(v)$ is executed at least once in the time interval $[t, t + T_m]$.

Practically, this means that as long as the algorithm has not converged, *every* message will keep being updated iteratively.

The main tool in convergence analysis is *contraction*. For a finite dimensional vector space $\mathbf{V}$ that has a vector norm $\|\cdot\|$, define a mapping $F : \mathbf{V} \to \mathbf{V}$ to be a $\|\cdot\|$-contraction if

$$\|F(v) - F(w)\| \leq a\|v - w\|, \text{ for some } 0 \leq a < 1, v, w \in \mathbf{V}. \tag{10}$$

If $F(\cdot)$ is a $\|\cdot\|$-contraction, then a unique fixed point v^* is guaranteed to exist, and applying $F(\cdot)$ iteratively,

$$v^{(t+1)} = F\left(v^{(t)}\right) \tag{11}$$

is guaranteed to converge to v^*, for all possible initial vectors $v^{(0)} \in \mathbf{V}$. In the message space $\mathcal{R}$ we define a message norm $\|v_m - w_m\|$ that measures the distances between individual messages and a global norm that measures distances between points in $\mathcal{R}^{|\mathcal{M}|}$. Following the analysis in [13], we also use the max norm $\|\cdot\|_\infty$ for the global norm defined as

$$\|v - w\|_\infty \triangleq \max_{m \in |\mathcal{M}|} \|v_m - w_m\|. \tag{12}$$

[5]For convenience all messages are assumed to have the same dimension d.

[6]$|\mathcal{M}|$ is twice the number of edges $\mathcal{E}$ in the graph.

Given that convergence is guaranteed for the synchronous BP, i.e., F^s is a $\|\cdot\|_\infty$- contraction, Elidan *et al.* revealed that the asynchronous BP will also converge if there is a *propagation schedule* that guarantees that every message will be updated until convergence, i.e., Assumption III.1 [13]. The above is described in following theorem [13, Theorem 3.2], which we will use to analyze the convergence of SR-BP:

Theorem III.2: If F^s is a max-norm contraction, then any asynchronous propagation schedule that satisfies Assumption III.1 will converge to a unique fixed point [13].

Moreover, [13] suggested that in an asynchronous message passing scheme, the messages that "carry" more information be propagated first as they will help the algorithm converge faster, and they defined the residual of a message as

$$r_m(v) \triangleq \|f_m(v) - v_m\|. \tag{13}$$

This has led to the proposal of R-BP (residual BP) message passing, where at iteration t, all residuals are calculated and the message with the largest residual is propagated. That is,

$$m^{(t+1)} = \arg\max_m r_m\left(v^{(t)}\right). \tag{14}$$

R-BP typically converges more often and with less messages than both synchronous and asynchronous BP schemes, which have led to a variety of R-BP variants in the LDPC decoding application. Unfortunately, R-BP requires a *centralized entity* to compare all the residuals, making it unsuitable for BP in distributed networks. In this letter, we solve this by SR-BP.

IV. SR-BP

At every iteration of BP each node can transmit its location belief eq. (4). Consequently, each node has to decide if its message is "important" enough to transmit. Following the intuition, we propose a stochastic message passing schedule, in which each node transmits its belief at time slot t with the probability

$$\Pr\left(r_m^{(t)}\right) = \frac{1}{1+\exp\left(-r_m^{(t)}\right)}, \tag{15}$$

which is the sigmoid function of the residual.

Theorem IV.1: Assuming that the PGM already satisfies the max-norm contraction condition for the synchronous BP case, SR-BP will converge to a unique fixed point.

Proof: We only need to prove that SR-BP satisfies Assumption III.1. By definition, we have $r_m^{(t)} \geq 0,\ \forall v_m \in \mathcal{M}$ and thus

$$\Pr\left(r_m^{(t)}\right) \geq 0.5. \tag{16}$$

Consequently, there will always be a positive probability that message m will be transmitted. Hence, there will be a T_m, so that for any time $t \geq 0$, message m will be updated and Assumption III.1 is satisfied. This completes the proof. □

It should be noted that by always having a positive transmission probability the silent node issue discovered in [14] is also resolved, as all nodes will get a chance to transmit.

TABLE I
RESULTS FOR THE ISING MODEL

	Conv. %	Avrg. Conv. Iterations	Avrg. Messages	KLD
BP	83%	78.5	8634	0.255
ABP	88%	63.2	6905	0.257
R-BP	100%	21.5	315	0.369
SR-BP	94.5%	28.4	393	0.365

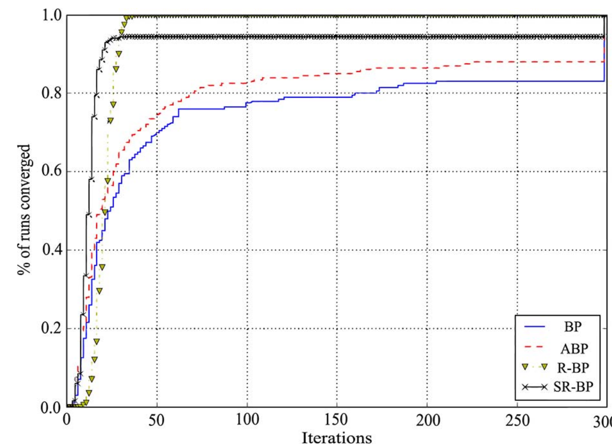

Fig. 1. Cumulative convergence % for the 11×11-node Ising grid.

V. Ising and Cooperative Sensing Results

Monte-Carlo simulations have been carried out to analyze the convergence rate, the message overhead and the quality of the marginals of SR-BP. Comparisons will be made with synchronous BP (BP) and asynchronous BP (ABP). Moreover, we compare SR-BP with the *centralized* R-BP as a benchmark.

A. Ising Model

We consider random grids parameterized by the Ising model [1].[7] A random grid with 11×11 nodes was created with univariate potentials $\Phi_i(X_i)$ sampled from $\mathcal{U}[0,1]$ for each variable, and pairwise potential $\Phi_{i,j}(X_i, X_j) = e^{\lambda C(2\mathbb{I}(X_i,X_j)-1)}$ where λ is sampled uniformly from $[-0.5, 0.5]$ having some nodes to agree and disagree with each other randomly. Finally C is an agreement factor, where higher values impose stronger constraints on the "negotiations" between nodes, making convergence harder. In the simulations, 200 independent realizations were run for the network with 11 nodes and $C = 10$ and the algorithms were allowed to run until convergence or 300 iterations had passed. The results are summarized in Table I.

First, as expected, R-BP achieves convergence every time in the given scenario, requiring 73% less iterations on average and almost 96% less messages. Also, SR-BP performs pretty close to R-BP achieving only slightly worse convergence rate, and requiring 64% less iterations than BP, and around 95% less messages, hence a huge decrease in complexity and overhead. Finally, the KL divergence is almost identical between R-BP and SR-BP. Fig. 1 shows the cumulative percentage of convergence of all the algorithms as a function of iterations passed. Again, please note that even though SR-BP converges less than the *centralized* R-BP, it converges much faster if it does.

[7]The Ising model provides a systematic way to analyze iterative algorithms, e.g., [9], [12], [13].

TABLE II
RESULTS FOR THE COOPERATIVE SPECTRUM SENSING MODEL

	Conv. %	Avrg. Conv. Iterations	Avrg. Messages
BP	9.8%	256.1	2817.26
ABP	13.4%	259.1	2849.6
R-BP	19.5%	269.0	279.0
SR-BP	18.5%	267.9	804.7

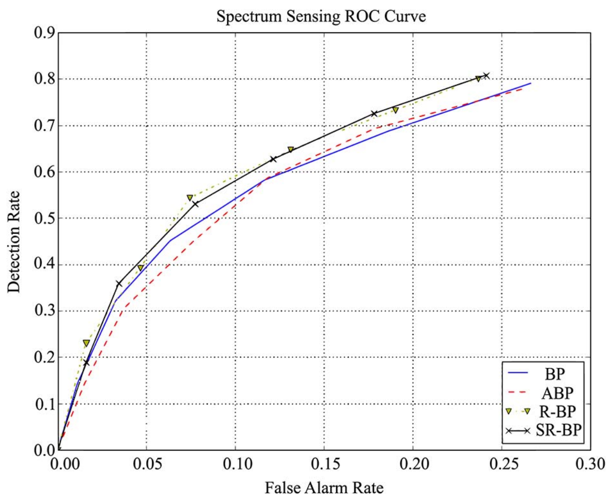

Fig. 2. ROC curves.

B. Cooperative Spectrum Sensing

In this subsection, we assess the performance of SR-BP in the cooperative spectrum sensing scenario.

In the simulations, the SUs were deployed randomly in a circular area with unit diameter, and we define R as the maximum communication range between SUs, where $R = 0.7$. A total of 100 simulations were run, and there were 100 time-slots for each run, with maximum 300 iterations. Simulation results are shown in Table II. In the setup, convergence is much more difficult to achieve, due to the large number of interconnections possible by their high communication range, and the relatively large correlation factors $\{\lambda_{i,j}\}$. Still, the proposed SR-BP manages to double the convergence percentage, using only approximately 29% of the message propagations required by BP. Therefore, in a real life application, SR-BP will achieve better convergence, with a possible 71% decrease in the required transmitted messages (i.e, both computational complexity and overhead).

Note that it is hard to compute the exact marginals for the spectrum sensing case. Instead we compute the ROC curves for the algorithms, as provided in Fig. 2 [17]. As can be seen, SR-BP has a better ROC curve than BP and ABP, achieving a curve that almost matches the one by the centralized R-BP.

VI. CONCLUSION AND DISCUSSION

In this letter, we have presented a novel distributed message scheduling algorithm for running inference algorithms using BP in wireless networks and more specifically it was used for cooperative spectrum sensing in a cognitive radio network. We have proven that SR-BP message schedule will converge to a fixed point if synchronous schedule converges, and showcased the superiority of the proposed algorithm even in the more general non-convergent cases, in which it consistently manages to achieve higher convergence rates, better accuracy, as well as lower overhead and computational cost (the important metrics in cooperative spectrum sensing). It should be noted that this work is far more general and can be used in a large number of applications where distributed iterative algorithms are used. Future work will involve experimentation with more complicated discrete and continuous pdfs. Analysis of the contraction rate of SR-BP and possible alternative distributions that could be used to instigate message propagation. In conclusion, the analysis of message schedules for distributed algorithms has been quite overlooked by the research community despite all the advantages a good message schedule clearly provides. We hope that this work will trigger an increase in interest for this interesting field, with a wide range of applications.

REFERENCES

[1] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques.*. Cambridge, MA, USA: MIT Press, 2009.
[2] S. Zarrin and T. J. Lim, "Belief propagation on factor graphs for cooperative spectrum sensing in cognitive radio," in *Proc. 3rd IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2008, pp. 1–9.
[3] H. Li, "Cooperative spectrum sensing via belief propagation in spectrum-heterogeneous cognitive radio systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2010, pp. 1–6.
[4] A. Dogandzic and B. Zhang, "Distributed estimation and detection for sensor networks using hidden Markov random field models," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 3200–3215, Aug. 2006.
[5] P. Sakulkar and A. Banerjee, "Improved cooperative spectrum sensing using belief propagation algorithm and energy detection," in *Proc. 8th Int. Symp. Wireless Commun. Syst.*, 2011, pp. 472–476.
[6] F. Penna, R. Garello, and M. A. Spirito, "Distributed inference of channel occupation probabilities in cognitive networks via message passing," in *Proc. IEEE Symp. New Frontiers Dyn. Spectr.*, 2010, pp. 1–11.
[7] J. S. Yedidia *et al.*, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
[8] A. T. Ihler, J. W. Fischer, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Lear. Res.*, vol. 6, pp. 905–936, Dec. 2005.
[9] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4422–4437, Dec. 2007.
[10] T. G. Roosta, M. J. Wainwright, and S. S. Sastry, "Convergence analysis of reweighted sum-product algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4293–4305, Sep. 2008.
[11] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Tree-based reparameterization framework for analysis of sum-product and related algorithms," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1120–1146, May 2003.
[12] J. K. Johnson *et al.*, "Walk-sum interpretation and analysis of Gaussian belief propagation," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2006, pp. 579–586.
[13] G. Elidan *et al.*, "Residual belief propagation: Informed scheduling for asynchronous message passing," *arXiv*, p. 6837, Jun. 2012. [Online]. Available: http://arxiv.org/abs/1206.6837
[14] H.-C. Lee, Y.-L. Ueng, S.-M. Yeh, and W.-Y. Weng, "Two informed dynamic scheduling strategies for iterative LDPC decoders," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 886–896, Mar. 2013.
[15] H. Wymeersch, F. Penna, and V. Savic, "Uniformly reweighted belief propagation: A factor graph approach," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 2000–2004.
[16] H. Wymeersch, F. Penna, and V. Savic, "Uniformly reweighted belief propagation for estimation and detection in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1587–1595, Apr. 2012.
[17] J. V. Candy, *Bayesian Signal Processing: Classical, Modern and Particle Filtering Methods*. Hoboken, NJ, USA: Wiley, Apr. 2009.