



Copyright Notice

©2012 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This document was downloaded from Chalmers Publication Library (<http://publications.lib.chalmers.se/>), where it is available in accordance with the IEEE PSPB Operations Manual, amended 19 Nov. 2010, Sec. 8.1.9 (<http://www.ieee.org/documents/opsmanual.pdf>)

(Article begins on next page)

Uniformly Reweighted Belief Propagation for Estimation and Detection in Wireless Networks

Henk Wymeersch, *Member, IEEE*, Federico Penna, *Student Member, IEEE*,
Vladimir Savic, *Student Member, IEEE*

Abstract—In this paper, we propose a new inference algorithm, suitable for distributed processing over wireless networks. The algorithm, called uniformly reweighted belief propagation (URW-BP), combines the local nature of belief propagation with the improved performance of tree-reweighted belief propagation (TRW-BP) in graphs with cycles. It reduces the degree of freedom in the latter algorithm to a single scalar variable, the uniform edge appearance probability ρ . We provide a variational interpretation of URW-BP, give insights into good choices of ρ , develop an extension to higher-order potentials, and complement our work with numerical performance results on three inference problems in wireless communication systems: spectrum sensing in cognitive radio, cooperative positioning, and decoding of a low-density parity-check (LDPC) code.

Index Terms—Distributed inference, approximate inference, belief propagation, message passing, factor graphs, tree-reweighted belief propagation, variational methods.

I. INTRODUCTION

IN wireless communication, belief propagation (BP) [1] has found applications in many aspects of receiver design, including equalization, demapping, multi-user detection, multi-antenna detection, and decoding [2]–[6]. Most notably in the latter application, the use of BP message passing over a suitable graphical model has led to practical decoding algorithms for powerful error-correcting codes, such as LDPC codes and turbo codes [7]. Recently, there has been an interest from the wireless communication community to extend BP to distributed problems, involving cooperation among multiple spatially separated wireless devices. BP inherently lends itself well to distributed implementation, and is thus a powerful, yet practical algorithm to perform cooperative estimation and detection. In fact, BP-based cooperative message passing algorithms have been applied to a wide variety of problems, including cooperative positioning [8]–[10], artificial intelligence [11], computer vision [12], cognitive radio [13], [14], link loss

H. Wymeersch is with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden (Email: henkw@chalmers.se). F. Penna is with the Politecnico di Torino, Turin, Italy, and now is with the Fraunhofer Heinrich Hertz Institute, Berlin, Germany (E-mail: federico.penna@hhi.fraunhofer.de). V. Savic is with the Universidad Politécnica de Madrid, Madrid, Spain (Email: vladimir@gaps.ssr.upm.es).

This work was supported, in part, by the Swedish Research Council (VR), under grant no. 2010-5889; the European Research Council, under grant COOPNET No. 258418; the Spanish Ministry of Science and Innovation under the grants TEC2009-14219-C03-01 and TEC2010-21217-C02-02-CR4HFDVL; program CONSOLIDER-INGENIO 2010 under the grant CSD2008-00010 COMONSENS; the European Commission under the grant FP7-ICT-2009-4-248894-WHERE-2. V. Savic is supported by the FPU fellowship from Spanish Ministry of Science and Innovation. Simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE.

monitoring [15], network control [16], cooperative beamforming [17], and sensor networks [18]. Despite of its widespread use, BP is faced by a lack of convergence guarantees, as well as over-confidence of beliefs [19]. These problems have been partly addressed in [20], where a novel message passing algorithm was proposed, called tree-reweighted BP (TRW-BP). This algorithm has stronger convergence guarantees [21], and often gives better performance than BP. However, TRW-BP involves the selection of, and optimization over, so-called *edge appearance probabilities*. This makes TRW-BP difficult to implement in a network setting, as the choice of valid edge appearance probabilities involves a problem that is hard to solve in a distributed manner.

In this paper, we present a novel algorithm combining the distributed nature of BP and the improved performance of TRW-BP in graphs with cycles. The new algorithm is called uniformly reweighted BP (URW-BP), as it collapses the edge appearance probabilities into a single scalar variable. As special cases, URW-BP reverts to BP under suitable choice of the variable, and corresponds to the optimal choice of edge appearance probabilities in TRW-BP for certain types of graphs. While URW-BP admittedly is a straightforward modification of TRW-BP, this paper for the first time indicates that URW-BP is a powerful inference algorithm in its own right, and deserves deeper investigation. Our contributions are as follows:

- We propose URW-BP as a practical (and in some case equivalent) alternative to TRW-BP for distributed inference problems;
- We interpret URW-BP as a multi-objective optimization problem, trading off single-variable entropy and correlation among variables, thus enabling URW-BP to mitigate overconfidence of beliefs in BP; and
- Through simulations, we show that URW-BP consistently outperforms BP, for centralized and distributed inference, for discrete and continuous variables, and for pairwise as well as higher-order interactions.

II. INFERENCE THROUGH MESSAGE PASSING

A. Problem Formulation

We consider an a posteriori distribution of the form

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \quad (1)$$

where \mathbf{y} is a (fixed) observation and $\mathbf{x} = [x_1, x_2, \dots, x_N]$ is the N -dimensional unobserved variable of interest. For nota-

tional convenience, we consider \mathbf{x} to be a discrete random variable with finite domain. Our goal is to determine the marginal a posteriori distributions, $p(x_n|\mathbf{y})$, for every component x_n . Brute-force computation of $p(x_n|\mathbf{y})$ is often intractable, due to the high-dimensional nature of \mathbf{x} . Practical algorithms can be developed by harnessing the factorization of both $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$. The factorization depends on conditional independence of the variables and is problem-specific. For concreteness, we will focus on factorizations of the form

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^N \phi_n(x_n) \prod_{(l_1, l_2) \in \mathcal{E}} \psi_{l_1, l_2}(x_{l_1}, x_{l_2}), \quad (2)$$

where \propto denotes proportionality up to a multiplicative constant (including, but not limited to $1/p(\mathbf{y})$) and \mathcal{E} is the set of variable indices for which there is an interaction. Such factorizations can be conveniently expressed by a graphical model, such as a *Markov random field*, where vertices correspond to variables and their single-variable factors (ϕ_n) and edges correspond to factors (ψ_{l_1, l_2}). Note that the set of edges in the Markov random field is exactly \mathcal{E} . Alternatively, the factorization can be represented by a *factor graph*, where vertices correspond to both variables and factors, and edges connect a variable vertex with a factor vertex when the corresponding variable appears in the corresponding factor. We will focus on factor graphs, as they will enable an easier extension to higher-order interactions later on. As an illustration of both graphical models, see Fig. 1. In wireless networks, N may represent the number of nodes in a network, x_n a variable of interest, and ψ_{l_1, l_2} an interaction between pairs of nodes. Detailed examples will be discussed in Section IV.

B. Message Passing Algorithms

Given a factor graph of $p(\mathbf{x}|\mathbf{y})$, we can run a *message passing algorithm* and determine approximations of the marginals $p(x_n|\mathbf{y})$, called beliefs, denoted by $b_n(x_n)$. Various message passing algorithms have been proposed, arguably the most popular of which is belief propagation (BP). Alternatives include naive mean field (MF) [22], TRW-BP [20], and generalized BP (GBP) [23]. When the factor graph is cycle-free, BP and TRW-BP are equivalent and provide the exact marginals, i.e., $b_n(x_n) = p(x_n|\mathbf{y})$, $\forall n, x_n$. When the factor graph contains cycles, BP and TRW-BP are not guaranteed to yield the true marginal posteriors, or even to converge at all.

1) *Belief Propagation*: Assuming the variable X_n appears in the factor $\psi_{m, n}$, the BP message¹ from $\psi_{m, n}$ to X_n is given by

$$\mu_{\psi_{m, n} \rightarrow X_n}(x_n) = \sum_{x_m} \phi_m(x_m) \psi_{m, n}(x_m, x_n) \prod_{k \in \mathcal{N}_m \setminus \{n\}} \mu_{\psi_{m, k} \rightarrow X_m}(x_m), \quad (3)$$

where $k \in \mathcal{N}_m$ indicates that there exists a factor $\psi_{m, k}$. The beliefs are updated according to

$$b_n(x_n) \propto \phi_n(x_n) \prod_{m \in \mathcal{N}_n} \mu_{\psi_{m, n} \rightarrow X_n}(x_n). \quad (4)$$

¹For notational convenience, we will express beliefs and messages in terms of messages from factor vertices to variable vertices. Messages from variable vertices to factor vertices will not be considered, except in Section III-C.

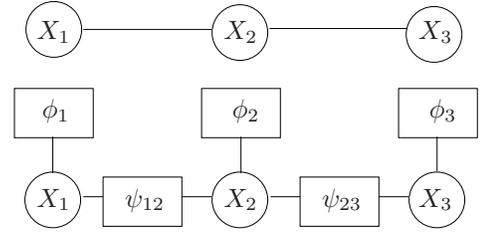


Figure 1. Markov random field (top) and factor graph (bottom) of the factorization $p(\mathbf{x}|\mathbf{y}) \propto \phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\psi_{12}(x_1, x_2)\psi_{23}(x_2, x_3)$.

Equations (3)–(4) are iterated until the beliefs converge.

2) *Tree-Reweighted Belief Propagation*: In TRW-BP, the message from $\psi_{m, n}$ to X_n is given by

$$\mu_{\psi_{m, n} \rightarrow X_n}(x_n) = \quad (5)$$

$$\sum_{x_m} \phi_m(x_m) \psi_{m, n}^{1/\rho_{m, n}}(x_m, x_n) \mu_{\psi_{m, n} \rightarrow X_m}^{\rho_{m, n}-1}(x_m) \quad (6)$$

$$\prod_{k \in \mathcal{N}_m \setminus \{n\}} \mu_{\psi_{m, k} \rightarrow X_m}^{\rho_{m, k}}(x_m),$$

while the belief is given by

$$b_n(x_n) \propto \phi_n(x_n) \prod_{m \in \mathcal{N}_n} \mu_{\psi_{m, n} \rightarrow X_n}^{\rho_{m, n}}(x_n). \quad (7)$$

As in BP, (5)–(7) are iterated until the beliefs converge. The values $\{\rho_{m, n}\}_{(m, n) \in \mathcal{E}}$ are the so-called edge appearance probabilities² (EAPs) of the factor vertices $\{\psi_{m, n}\}_{(m, n) \in \mathcal{E}}$.

The performance of TRW-BP depends on the choice of EAPs. Optimization over the EAPs (referred to here as *optimized TRW-BP*) is possible, but corresponds to an outer iterative loop, involving a high-dimensional optimization problem over $\{\rho_{m, n}\}_{(m, n) \in \mathcal{E}}$. Moreover, the set of valid EAPs is non-trivial: given a graph G and the set $\mathfrak{T}(G)$ of all possible spanning trees, we can introduce a distribution over the spanning trees: $0 \leq p(T) \leq 1$, for $T \in \mathfrak{T}(G)$. For a given distribution, the EAP of factor $\psi_{m, n}$ is then given by

$$\rho_{m, n} = \sum_{T \in \mathfrak{T}(G)} p(T) \times \mathbb{I}\{\text{vertex } \psi_{m, n} \in T\}, \quad (8)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. Note that when G is a tree, every factor vertex $\psi_{m, n}$ appears in every spanning tree, so that $\rho_{m, n} = 1$, $\forall (m, n) \in \mathcal{E}$. When the graph G contains cycles, there must be at least one $\rho_{m, n} < 1$.

BP can now be interpreted as a modification of TRW-BP, where $\rho_{m, n} = 1$, $\forall (m, n) \in \mathcal{E}$, irrespective of the structure of G . Note that this choice of EAPs is not valid for a graph with cycles.

III. UNIFORMLY REWEIGHTED BELIEF PROPAGATION

A. Description

In order to combine the simplicity and distributed nature of BP with the improved performance of TRW-BP in graphs with cycles, we propose a novel approximate inference algorithm, called uniformly reweighted belief propagation (URW-BP). In

²The terminology of *edge appearance* comes from the Markov random field representation of $p(\mathbf{x}|\mathbf{y})$, where the factors $\psi_{m, n}$ are represented by edges.

URW-BP, the EAPs in (5)–(7) are all equal (and denoted by ρ), so their optimization is a scalar optimization problem. In the following sections, we will connect BP, TRW-BP, and URW-BP in a variational interpretation, comment on the conditions under which URW-BP and optimized TRW-BP coincide, and cast URW-BP as a solution to a multi-objection optimization problem.

B. Variational Interpretation

1) *General Formulation*: Given any distribution $b(\mathbf{x})$, the Kullback Leibler divergence [24] between $b(\mathbf{x})$ and $p(\mathbf{x}|\mathbf{y})$ is given by

$$\text{KL}(b||p) = \sum_{\mathbf{x}} b(\mathbf{x}) \log \frac{b(\mathbf{x})}{p(\mathbf{x}|\mathbf{y})} \geq 0. \quad (9)$$

If we insert (2), and perform some straightforward manipulations, we can express this inequality as

$$\log p(\mathbf{y}) \geq \mathcal{H}(b) + \chi(b) \quad (10)$$

where $\mathcal{H}(b)$ denotes the entropy of the distribution $b(\mathbf{x})$, i.e., $\mathcal{H}(b) = -\sum_{\mathbf{x}} b(\mathbf{x}) \log b(\mathbf{x})$ and

$$\begin{aligned} \chi(b) = & \sum_{n=1}^N \sum_{x_n} b_n(x_n) \log \phi_n(x_n) \\ & + \sum_{(m,n) \in \mathcal{E}} \sum_{x_m, x_n} b_{m,n}(x_m, x_n) \log \psi_{m,n}(x_m, x_n). \end{aligned} \quad (11)$$

$$(12)$$

Note that (10) is valid with equality if and only if $b(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$, $\forall \mathbf{x}$. Hence, we can write

$$\log p(\mathbf{y}) = \max_{b \in \mathbb{M}(G)} \{ \mathcal{H}(b) + \chi(b) \}, \quad (13)$$

where $\mathbb{M}(G)$ is the so-called *marginal polytope* corresponding to the factor graph G , which is the set of marginal distributions $\{b_n(x_n)\}_{n=1}^N$ and $\{b_{m,n}(x_m, x_n)\}_{(m,n) \in \mathcal{E}}$ that can be related to a global distribution $b(\mathbf{x})$ that factorizes according to the same factor graph G . The formulation (13) implies that if we can solve the optimization problem, then the solution is $b(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$, with corresponding maximum equal to $\log p(\mathbf{y})$. The optimization problem turns out to be concave,³ but intractable: unless G is a tree, no explicit expression for $\mathcal{H}(b)$ is available and the number of constraints to describe $\mathbb{M}(G)$ is exponential in the size of G [22]. BP, TRW-BP, and URW-BP can be interpreted as approximate methods to solve (13).

2) *Belief Propagation*: In BP, the set $\mathbb{M}(G)$ is outer-bounded by the convex set $\mathbb{L}(G)$, which is the set of marginal beliefs $\{b_n(x_n)\}_{n=1}^N$ and $\{b_{m,n}(x_m, x_n)\}_{(m,n) \in \mathcal{E}}$, that are mutually consistent, but need not be consistent with any global belief $b(\mathbf{x})$. It can be shown that $\mathbb{L}(G)$ is a convex set,⁴ and that when G is a tree, $\mathbb{M}(G) = \mathbb{L}(G)$ [22]. Secondly, the entropy in (13) is replaced by the so-called Bethe entropy

$$\mathcal{H}_{\text{Bethe}}(b) = \sum_{n=1}^N \mathcal{H}(b_n) - \sum_{(m,n) \in \mathcal{E}} \mathcal{I}(b_{m,n}), \quad (14)$$

³Since $\chi(b)$ is linear in b , the entropy $\mathcal{H}(b)$ is concave in b , and the set $\mathbb{M}(G)$ is a convex set.

⁴Since any convex combination of marginal beliefs is again a valid marginal belief.

where $\mathcal{I}(b_{m,n})$ is the mutual information between x_m and x_n :

$$\mathcal{I}(b_{m,n}) = \sum_{x_m, x_n} b_{m,n}(x_m, x_n) \log \frac{b_{m,n}(x_m, x_n)}{b_m(x_m)b_n(x_n)}. \quad (15)$$

It is shown in [23] that BP (3)–(4) is an iterative method to find of a stationary point of the Lagrangian corresponding to the problem (13), with $\mathbb{M}(G)$ replaced by $\mathbb{L}(G)$ and $\mathcal{H}(b)$ replaced by $\mathcal{H}_{\text{Bethe}}(b)$.

3) *Tree-Reweighted Belief Propagation*: In TRW-BP, the set $\mathbb{M}(G)$ is also outer-bounded by the convex set $\mathbb{L}(G)$. For a fixed set of valid EAPs, the entropy $\mathcal{H}(b)$ is replaced by a convex *upper bound*

$$\mathcal{H}(b|\rho) = \sum_{n=1}^N \mathcal{H}(b_n) - \sum_{(m,n) \in \mathcal{E}} \rho_{m,n} \mathcal{I}(b_{m,n}), \quad (16)$$

where we have introduced the vector $\rho \doteq [\rho_{m,n}]_{(m,n) \in \mathcal{E}}$. We note again that BP corresponds to setting $\rho = \mathbf{1}$ in TRW-BP. For a fixed ρ , the stationary points of the Lagrangian lead to (5)–(7), following a similar line of reasoning as in [23].

4) *Uniformly Reweighted Belief Propagation*: In URW-BP, similar to BP and TRW-BP, the set $\mathbb{M}(G)$ is outer-bounded by the set $\mathbb{L}(G)$. For a fixed value of $\rho \in \mathbb{R}$, the entropy $\mathcal{H}(b)$ is approximated by

$$\mathcal{H}(b|\rho) = \sum_{n=1}^N \mathcal{H}(b_n) - \sum_{(m,n) \in \mathcal{E}} \rho \mathcal{I}(b_{m,n}). \quad (17)$$

Again, BP is found by setting $\rho = 1$. Stationary points of the Lagrangian lead to the TRW-BP equations given in (5)–(7), with $\rho_{m,n} = \rho$, $\forall (m,n) \in \mathcal{E}$.

C. Extension of TRW-BP and URW-BP Beyond Pairwise Interactions

In the original formulation of TRW-BP, only pairwise interactions were considered, with a corresponding Markov random field where variables are vertices and edges are pairwise potentials. It was suggested in [20] that the extension to higher-order interactions requires a graphical model that is a hypergraph. A joint a posteriori distribution with higher-order interaction is of the form

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^N \phi_n(x_n) \prod_{l=1}^L \psi_l(x_{C_l}), \quad (18)$$

where $x_{C_l} \subseteq \mathbf{x}$. We now present two distinct TRW-BP message passing algorithms.

1) *Using Factor Appearance Probabilities*: Based on factor graphs, we have derived TRW-BP equations for arbitrary interactions [25]. The message from factor vertex ψ_l to variable vertex X_n is given by

$$\begin{aligned} \mu_{\psi_l \rightarrow X_n}(x_n) = & \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \times \\ & \prod_{m \in C_l \setminus \{n\}} \phi_m(x_m) \mu_{\psi_l \rightarrow X_m}^{\rho_l - 1}(x_m) \prod_{k \in \mathcal{N}_m \setminus \{l\}} \mu_{\psi_k \rightarrow X_m}^{\rho_k}(x_m), \end{aligned} \quad (19)$$

where $\sum_{\sim x_n}$ denotes summation over all variables except x_n . The beliefs are given by

$$b_n(x_n) \propto \phi_n(x_n) \prod_{l \in \mathcal{N}_n} \mu_{\psi_l \rightarrow X_n}^{\rho_l}(x_n). \quad (20)$$

Here, the variables ρ_l refer to the appearance probabilities of the factor vertices ψ_l in the collection of trees. We have made a slight abuse of notation, as here $l \in \mathcal{N}_n$ means that there exists a factor ψ_l that has as variable X_n . For URW-BP, we simply set all $\rho_l = \rho$.

For the sake of completeness, we also provide message passing rules when messages from variable vertices to factor vertices are computed. In that case, the message from variable vertex X_n to factor vertex ψ_l , $n \in C_l$ is given by

$$\begin{aligned} \mathbf{m}_{X_n \rightarrow \psi_l}(x_n) = \\ \phi_n(x_n) \mathbf{m}_{\psi_l \rightarrow X_n}^{\rho_l - 1}(x_n) \prod_{k \in \mathcal{N}_n \setminus \{l\}} \mathbf{m}_{\psi_k \rightarrow X_n}^{\rho_k}(x_n), \end{aligned} \quad (21)$$

while the message from factor vertex ψ_l to variable vertex X_n , $n \in C_l$ is now

$$\begin{aligned} \mathbf{m}_{\psi_l \rightarrow X_n}(x_n) = \\ \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m \in C_l \setminus \{n\}} \mathbf{m}_{X_m \rightarrow \psi_l}(x_m). \end{aligned} \quad (22)$$

Finally, the belief of variable x_n is given by (20), where $\mu_{(\cdot)}(\cdot)$ should be replaced by $\mathbf{m}_{(\cdot)}(\cdot)$. These message passing rules can be seen as extension of the sum-product algorithm from [26].

2) *Using Edge Appearance Probabilities:* An alternative approach is to convert the factor graph to contain only pairwise interactions [27]. Consider a factor $\psi_l(x_1, x_2, x_3)$. We can convert this part of the factor graph as follows (see also [22, page 289]): (i) introduce a new variable \mathbf{z}_l that takes values in the product domain (e.g., if $x_i \in \mathcal{X}_i$, then $\mathbf{z}_l \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$); (ii) introduce 3 factors $f_{l,i}(\mathbf{z}_l, x_i) = \mathbb{I}\{z_{l,i} = x_i\} \psi_l^{\alpha_i}(\mathbf{z}_l)$; (iii) replace in the factor graph the factor vertex $\psi_l(x_1, x_2, x_3)$ with 4 factor vertices $\psi_l(\mathbf{z}_l)^{1 - \sum_i \alpha_i} f_{l,1}(\mathbf{z}_l, x_1) f_{l,2}(\mathbf{z}_l, x_2) f_{l,3}(\mathbf{z}_l, x_3)$, where the parameters $\alpha_i \in \mathbb{R}$ are chosen such that $0 \leq \sum_i \alpha_i \leq 1$. This latter factor graph only contains pairwise interactions, on which we can apply the message passing rules (5). For example, setting $\alpha_i = 0$, these messages can in turn be transformed to messages on the original factor graph, leading to the message from variable vertex X_n to factor vertex ψ_l , $n \in C_l$ being given by

$$\begin{aligned} \mathbf{m}_{X_n \rightarrow \psi_l}(x_n) = \\ \phi_n(x_n) \mathbf{m}_{\psi_l \rightarrow X_n}^{\rho_{l,n} - 1}(x_n) \prod_{k \in \mathcal{N}_n \setminus \{l\}} \mathbf{m}_{\psi_k \rightarrow X_n}^{\rho_{k,n}}(x_n), \end{aligned} \quad (23)$$

where $\rho_{l,n}$ is the edge appearance probability of the edge connecting factor vertices X_n and \mathbf{Z}_l in the transformed factor graph. The message from factor vertex ψ_l to variable vertex is give by

$$\begin{aligned} \mathbf{m}_{\psi_l \rightarrow X_n}(x_n) = \\ \sum_{\sim x_n} \psi_l(x_{C_l}) \mathbf{m}_{X_n \rightarrow \psi_l}^{\rho_{l,n} - 1}(x_n) \prod_{m \in C_l \setminus \{n\}} \mathbf{m}_{X_m \rightarrow \psi_l}^{\rho_{l,m}}(x_m), \end{aligned} \quad (24)$$

and the belief of variable x_n is again given by (20), where $\mu_{(\cdot)}(\cdot)$ should be replaced by $\mathbf{m}_{(\cdot)}(\cdot)$.

D. Discussion of URW-BP

1) URW-BP as a Multi-Objective Optimization Solution:

We can write the URW-BP objective function for general interactions as

$$\chi(b) + \sum_{n=1}^N \mathcal{H}(b_n) - \rho \sum_{l=1}^L \mathcal{I}(b_{C_l}), \quad (25)$$

which we can interpret as a multi-objective problem with trade-off parameter $\rho \geq 0$. The function $\chi(b)$ is linear in $b(\mathbf{x})$, while $\sum_{n=1}^N \mathcal{H}(b_n)$ is concave in $b(\mathbf{x})$.⁵ Suppose we have two candidate solutions: $b(\mathbf{x})$ and $\tilde{b}(\mathbf{x})$, for which $\chi(b) = \chi(\tilde{b})$. When $\rho = 1$, the solution with the maximal Bethe entropy will be chosen. Maximizing Bethe entropy corresponds to maximizing the entropies of $b_n(x_n)$ and at the same time minimizing the dependence among variables. When $\rho = 0$, the solution with the maximal entropy of the individual beliefs will be chosen, irrespective of the dependence (mutual information) among the variables. Hence, *by decreasing ρ , we can force the beliefs to be less concentrated* (i.e., less over-confident).

2) *URW-BP as Optimized TRW-BP:* In general, URW-BP is not guaranteed to outperform TRW-BP. In fact, for some graphs, URW-BP does not even correspond to a valid distribution over spanning trees. However, we can describe sufficient, though not necessary conditions, under which URW-BP and optimized TRW-BP coincide. We will denote set cardinality by $|\cdot|$.

Definition 1 (Symmetric factorization). *Given a factor graph G , corresponding to a factorization*

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^N \phi_n(x_n|\theta_n) \prod_{l=1}^L \psi_l(x_{C_l}|\theta_{C_l}) \quad (26)$$

with L non-trivial⁶ factors, where θ_n and θ_{C_l} represents parameters that fully determine the corresponding functions. We call the factorization symmetric when (i) $|C_l| = |C_k|$, $\forall k, l \in \{1, \dots, L\}$ and $|\mathcal{N}_n| = |\mathcal{N}_m|$, $\forall m, n \in \{1, \dots, N\}$; (ii) $\theta_n = \theta_m$, $\forall m, n \in \{1, \dots, N\}$ and $\theta_{C_l} = \theta_{C_k}$, $\forall k, l \in \{1, \dots, L\}$.

Proposition 1. *The optimal AEPs in TRW-BP for factorizations which are either symmetric, or for which the corresponding factor graph is a tree, are uniform, i.e., $\rho = \rho \mathbf{1}$, for some $\rho \in [0, 1]$.*

The proof follows immediately from symmetry considerations. As a special case, [20, page 2327] points out that for symmetric factorizations with only pairwise interactions, the optimal EAP in TRW-BP should be uniform, and is given by $\rho = (N-1)/L$, where $L = |\mathcal{E}|$. Introducing n_D as the number of pairwise interactions in which each variable is involved,⁷ a simple counting argument yields $n_D \approx 2L/N$, so that

$$\rho \approx \min\left(1, \frac{2}{n_D} \left(1 - \frac{1}{N}\right)\right) \quad (27)$$

$$\approx \min(1, 2/n_D). \quad (28)$$

⁵Since entropy is concave, and summation preserves concavity.

⁶In the sense that $|C_l| \geq 2$.

⁷Note that n_D is also the node degree in the corresponding Markov random field, and that $n_D + 1$ is the variable vertex degree in the corresponding factor graph.

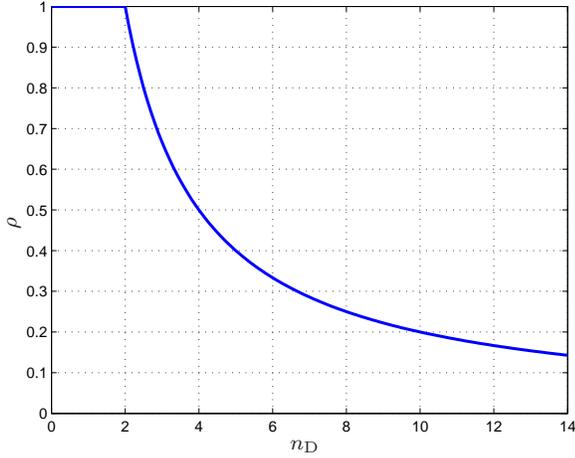


Figure 2. Value of ρ as a function of the average node degree according to (28).

Note that for trees, $L = N - 1$, so that $n_D \approx 2$, with corresponding $\rho = 1$.

While symmetric factorizations are not very interesting from a practical perspective, we can expect that URW-BP should perform close to optimized TRW-BP as long as the factorization exhibits sufficient symmetry or when the factor graph is sufficiently tree-like. The conditions for sufficient symmetry are not explored here, and are a topic for future research. However, as we will see in the following section, the class of factorizations for which URW-BP outperforms BP turns out to be quite large.

In the context of distributed inference, where there are N devices, with a total of L bi-directional communication links among devices, n_D corresponds to the average connectivity, and can be estimated by running a simple consensus algorithm over a network [28]. The relationship (28) is shown in Fig. 2. It is clear that for high network connectivity, we can expect URW-BP to outperform BP.

Finally, we note that choosing ρ according to (27) is not guaranteed to outperform BP. In practice, the best value of ρ will depend on the strength of the interactions between variables. For example, with very strong interactions between variables along a spanning tree, and very weak interactions between variables outside of the spanning tree, the best ρ will be close to 1, irrespective of n_D . Contrary to TRW-BP, an explicit optimization over ρ is not possible. In TRW-BP, the optimization problem for a fixed ρ is

$$\begin{aligned} & \text{maximize} && \mathcal{H}(b|\rho) + \chi(b) \\ & \text{s.t.} && b \in \mathbb{L}(G). \end{aligned} \quad (29)$$

We know that for any $b \in \mathbb{L}(G)$ we have the upper bound

$$\log p(\mathbf{y}) \leq \mathcal{H}(b|\rho) + \chi(b). \quad (30)$$

Hence, the tightest upper bound is found by solving the following problem:

$$\log p(\mathbf{y}) \leq \min_{\rho} \max_{b \in \mathbb{L}(G)} \{ \mathcal{H}(b|\rho) + \chi(b) \}. \quad (31)$$

This outer minimization can be solved numerically to find the best ρ [20]. In contrast, URW-BP and BP generally do not satisfy the two inequalities above, so direct optimization over ρ is not possible. Moreover, similar to BP, URW-BP is not guaranteed to converge. However, following an analysis similar to [29], one can show that the sufficient conditions for convergence of URW-BP are less stringent than for BP.

IV. CASE STUDIES

In this section, we will apply URW-BP to a three practical applications, involving discrete and continuous variables, centralized and distributed processing, and pairwise as well as higher-order interactions. The first application involves a cognitive radio network, where devices determine whether or not the spectrum is being utilized. The second application involves determining positions of wireless devices through cooperation. The goal of every device is to determine its own position based on distance estimates with respect to neighbors and reference nodes. The final application involves decoding of LDPC codes. In all three cases we show that URW-BP outperforms BP. For case studies 1 and 2, we assume perfect medium access control, which is a common assumption for distributed inference problems [30], [31].

A. Case Study 1: Signal Detection for Cognitive Radio

1) *Problem Formulation:* In cognitive radio networks, secondary users (SU) are granted access to the spectrum of primary users (PU), when said spectrum is not used by the PU. This allows spatial reuse of scarce spectrum. Prior to using the channel, the SU must sense the spectrum and decide whether or not any PUs are present. As PUs correspond to signal sources with a precise physical location, it is reasonable to assume that two SU that are close to each other are likely to observe the same PU signal state, whereas observations of SU that are far apart are most likely independent. Based on [32], these properties can be captured through (i) binary variables ($x_n \in \{0, 1\}$, where $x_n = 1$ means that the channel is occupied near the location of the n -th SU), (ii) observations \mathbf{y}_n (reflecting the signals collected by the n -th SU) with corresponding likelihoods $\phi_n(x_n) = p(\mathbf{y}_n|x_n)$, and (iii) spatial correlations⁸ between x_n and x_m , $\psi_{m,n}(x_n, x_m) = \exp(\lambda_{m,n}\mathbb{I}\{x_n = x_m\})$, where $\lambda_{m,n} \geq 0$ generally decreases with the distance between the two SUs. The set of neighbors \mathcal{N}_n in this case is defined implicitly as the set of SU that are within communication range with the n -th SU [13]. The joint a posteriori distribution $p(\mathbf{x}|\mathbf{y})$ is of the form (2):

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^N p(\mathbf{y}_n|x_n) \prod_{(m,n) \in \mathcal{E}} \exp(\lambda_{m,n}\mathbb{I}\{x_n = x_m\}), \quad (32)$$

where the product over couples (m, n) is only over $m < n$, for which $m \in \mathcal{N}_n$. The goal of the n -th SU is to determine $b_n(x_n) \approx p(x_n|\mathbf{y})$, where $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$.

⁸In general, higher-order interactions (e.g., between x_n , x_m , and x_p) are possible as well, though, similar to [32], these interactions are neglected for simplicity.

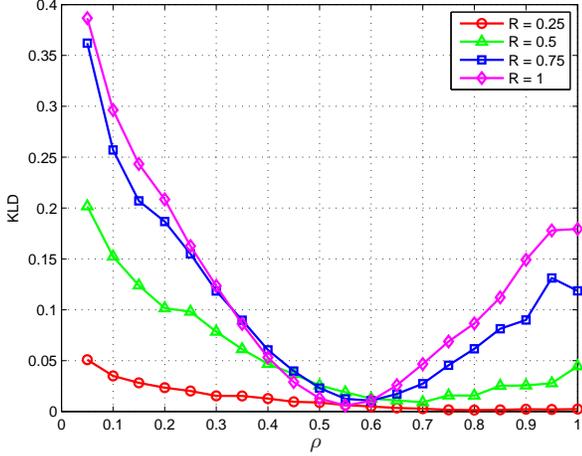


Figure 3. Average KLD vs. ρ (top) and vs. normalized communication range (bottom) for a simple network of 4 nodes, $\lambda_{m,n} \sim \mathcal{U}(0.2, 4)$. The value $\rho = 1$ corresponds to standard BP.

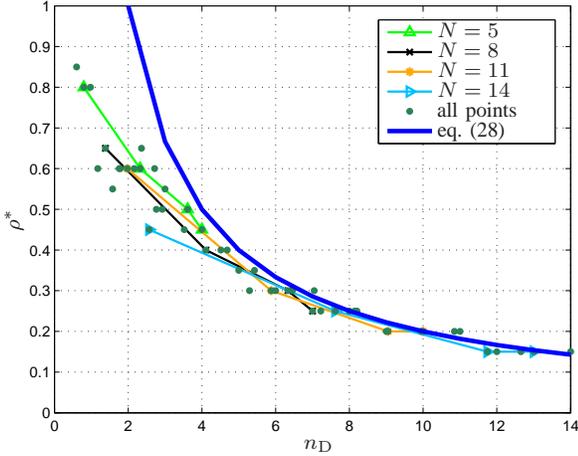


Figure 4. Best EAP (ρ^*) based on average KLD vs. average degree n_D .

The likelihood functions $\phi_n(x_n) = p(\mathbf{y}_n|x_n)$ account for the detection performance of each node n . Assume that each node uses as test statistic for detection a vector of S i.i.d. signal received samples, and that noise and PU signal are both complex Gaussian random variables with variance σ_v^2 and σ_s^2 respectively. Then, it can be readily shown that for $x_n = 0$ (i.e., when no PU is accessing the channel around the n -th SU), $\mathbf{y}_n \sim \mathcal{CN}(0, \sigma_v^2 \mathbf{I}_S)$, while for $x_n = 1$ (i.e., when a PU is active), $\mathbf{y}_n \sim \mathcal{CN}(0, (\sigma_v^2 + \sigma_s^2) \mathbf{I}_S)$, where \mathbf{I}_S is the $S \times S$ identity matrix.

We note that distributed signal detection can be seen as a special case of the more general problem of distributed parameter estimation, or multiple hypothesis testing, studied in the context of wireless sensor networks (see, for example, [33]–[35]). The proposed method can be extended accordingly to these applications.

2) *Implementation Aspects*: As every SU in the network corresponds to a variable x_n , URW-BP can be implemented distributedly by mapping the factor graph to the communi-

cation graph. Thus, the message $\mu_{\psi_{m,n} \rightarrow X_m}(x_m)$ from (5) in the factor graph is transmitted from the n -th SU to the m -th SU. For binary variables, it is convenient to represent messages with log-likelihood ratios (LLRs). Introducing $\gamma_n = \log(\phi_n(0)/\phi_n(1))$, and accounting for the fact that $\psi_{m,n}^{1/\rho}(0, 1) = \psi_{m,n}^{1/\rho}(1, 0) = 1$ and $\psi_{m,n}^{1/\rho}(0, 0) = \psi_{m,n}^{1/\rho}(1, 1) = \exp(\lambda_{m,n}/\rho)$, we can express $\mu_{\psi_{m,n} \rightarrow X_m}(x_m)$ as a scalar quantity:

$$\begin{aligned} M_{n \rightarrow m} &\doteq \log \frac{\mu_{\psi_{m,n} \rightarrow X_m}(0)}{\mu_{\psi_{m,n} \rightarrow X_m}(1)} \\ &= \mathbb{U} \left(\frac{\lambda_{m,n}}{\rho}, \gamma_n + (\rho - 1)M_{m \rightarrow n} + \rho \sum_{k \in \mathcal{N}_n \setminus \{m\}} M_{k \rightarrow n} \right), \end{aligned} \quad (33)$$

where $\mathbb{U}(a, b) = \log(1 + e^{a+b})/(e^a + e^b)$, which is a function that can be computed efficiently for any $a, b \in \mathbb{R}$. Similarly, the beliefs in LLR representation are given by

$$B_n \doteq \log \frac{b_n(0)}{b_n(1)} \quad (35)$$

$$= \log \frac{\phi_n(0) \prod_{m \in \mathcal{N}_n} \mu_{\psi_{m,n} \rightarrow X_n}^\rho(0)}{\phi_n(1) \prod_{m \in \mathcal{N}_n} \mu_{\psi_{m,n} \rightarrow X_n}^\rho(1)} \quad (36)$$

$$= \gamma_n + \rho \sum_{m \in \mathcal{N}_n} M_{m \rightarrow n}. \quad (37)$$

3) *Performance Results*: To evaluate the performance of URW-BP, we compute for each node in the network the Kullback-Leibler divergence (KLD) between belief and true posterior, defined as $\text{KLD}_n \doteq \text{KL}(b_n(x_n) || p(x_n|\mathbf{y}))$ and we average it over all nodes. We first consider a small network of 4 nodes, with correlation factors $\lambda_{n,m}$ modeled as uniformly distributed random variables between 0.2 and 4. Nodes are deployed randomly in a circular area of unit diameter. We then define R as the communication range of nodes in the network: if $R = 1$, all nodes can communicate with each other, therefore many loops are present in the graph; on the contrary, low values of R result in fewer loops. Fig. 3 shows the average KLD of beliefs computed through standard BP and URW-BP as a function of ρ for different values of R . We observe that URW-BP outperforms standard BP (corresponding to $\rho = 1$), especially for large values of R . Also, the best value of ρ (denoted as ρ^*) tends to decrease as $R \rightarrow 1$ as an effect of the increasing number of loops in the graph.

Fig. 4 combines the results of simulations performed on networks of different size (N from 4 through 15) and shows the best value of ρ (according to average KLD) as a function of the average degree n_D . Results confirm the dependency of ρ^* on the average degree of the graph, as discussed in Sec. III-D. In particular, the approximate expression (28) proves to be increasingly accurate as $n_D \gg 1$.

B. Case Study 2: Cooperative Positioning

1) *Problem Formulation*: The second application of URW-BP that we consider is cooperative positioning in wireless networks. In this case, the goal of each node (referred to as targets) is to estimate its own position based on a set distance

measurements with nearby targets and few fixed reference nodes (referred to as anchors), assuming a fully distributed architecture. These distance estimates can be obtained using well-known measurement techniques (e.g., time of arrival, received signal strength). Contrary to the previous case, we deal here with *continuous* variables. Also contrary to the previous case where over-confidence of the beliefs is not detrimental as long as the estimates are correct, here the situation is quite different: for safety-critical applications, e.g., tracking of robots or unmanned aerial vehicles (UAVs), it is important not only to have good position estimates, but also to know the uncertainty. When a belief is overconfident, a target may take improper actions (e.g., crash into another UAV).

We consider N targets and minimum 3 anchors scattered randomly in a planar region, and denote the two-dimensional location of the n -th node (target or anchor) by x_n . Target n obtains a noisy measurement y_{mn} of its distance from node m , given by

$$y_{mn} = \|x_m - x_n\| + v_{mn}, \quad (38)$$

where v_{mn} represents the noise drawn from some distribution p_v (e.g., Gaussian or the empirical distribution found from real measurements). As in the previous case study, we assume nodes can only communicate with nearby devices within a predefined radius $R > 0$. Thus, y_{mn} is available to target n only when $\|x_m - x_n\| < R$. Hence, $m \in \mathcal{N}_n$ when $\|x_m - x_n\| < R$. Letting $\phi_n(x_n) = p_n(x_n)$ be the a priori distribution of the position of the n -th target, $\psi_{m,n}(x_m, x_n) = p(y_{mn}|x_n, x_m)$ the likelihood function, and collecting the target positions in a vector \mathbf{x} , and the distance estimates in a vector \mathbf{y} , the a posteriori distribution $p(\mathbf{x}|\mathbf{y})$ is of the form (2):

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^N p_n(x_n) \prod_{(m,n) \in \mathcal{E}} p(y_{mn}|x_n, x_m), \quad (39)$$

where the product over couples (m, n) is only over nodes m within communication range of target n , i.e., $m \in \mathcal{N}_n$. The goal of target n is to determine its belief $b_n(x_n) \approx p(x_n|\mathbf{y})$. From the belief, the target can easily find a point estimate (e.g., the mean of $b_n(x_n)$) and associated uncertainty (e.g., the covariance matrix, or any percentile).

2) *Implementation Aspects*: Since the variables are now continuous, all sums in message update equations in Section III have to be replaced with integrals. As the localization problem is nonlinear and potentially non-Gaussian, exact message representation and updating is not tractable. Therefore, we resort to a non-parametric version of URW-BP (URW-NBP) where the beliefs and message update equations are performed using particle-based approximations (see [8]–[10] for more details on NBP). Contrary to the first case study, in cooperative localization, BP and URW-BP can harness the broadcast nature of the wireless channel, as detailed in [8].

3) *Performance Results*: We considered a network with 4 anchors and $N = 25$ targets, in a square 20 m by 20 m deployment area. For simplicity, we assume the noise v_{mn} has a zero-mean Gaussian distribution with standard deviation 30 cm. We performed ten iterations of message passing, representing the messages with 500 particles. Performance is

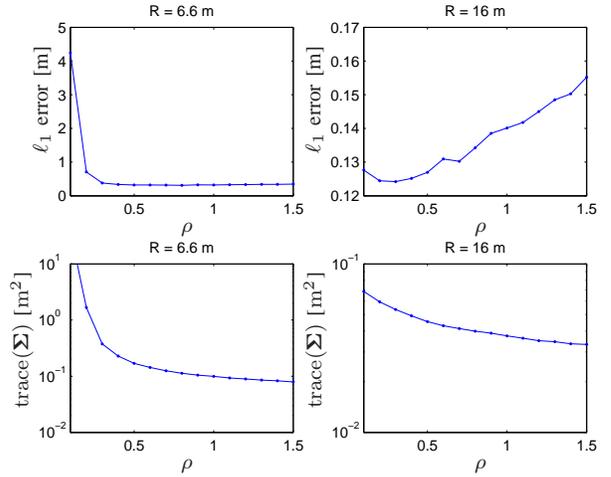


Figure 5. Positioning performance as a function of ρ for different communication radii. The first row shows the average ℓ_1 error, and the second row the trace of the covariance matrix (denoted by Σ) of the belief.

measured in terms of the average ℓ_1 positioning error and the average trace of the covariance of the belief. The ℓ_1 error for target n is given by $\|x_n - \hat{x}_n\|$, where \hat{x}_n is the mean of $b_n(x_n)$. The results are averaged over 100 Monte Carlo runs.

Fig. 5 shows the performance of URW-BP as a function of $\rho \in [0.1, 1.5]$ for $R = 6.6$ m and $R = 16$ m after the tenth iteration. We observe that for low connectivity, the ℓ_1 error is relatively insensitive to ρ for any $\rho > 0.4$. However, the beliefs become more concentrated as ρ increases. Hence, reducing ρ results in a more robust algorithm. When the connectivity is increased ($R = 16$), the best value of $\rho \approx 0.3$, while $\rho = 1$ induces around 20% additional error. Again, we observe more concentrated beliefs with increasing ρ . Overall, the differences between BP and URW-BP are small.

C. Case Study 3: LDPC Decoding

1) *Problem Formulation*: Consider an LDPC code with an $L \times N$ sparse parity check matrix \mathbf{H} . Let \mathbf{x} denote the transmitted codeword and \mathbf{y} the observation over a memoryless channel, with known $p(y_n|x_n)$. The a posteriori distribution of interest is now

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \quad (40)$$

$$= \prod_{n=1}^N p(y_n|x_n) \mathbb{I}\{\mathbf{H}\mathbf{x} = \mathbf{0}\} \quad (41)$$

$$= \prod_{n=1}^N p(y_n|x_n) \prod_{l=1}^L \mathbb{I}\left\{\sum_{m \in C_l} x_m = 0\right\}, \quad (42)$$

where the summation is in the binary field, and C_l is the index set corresponding to the non-zero elements of the l -th row in \mathbf{H} . Clearly, we can make the association with (18) through $\phi_n(x_n) \leftrightarrow p(y_n|x_n)$ and $\psi_l(x_{C_l}) \leftrightarrow \mathbb{I}\{\sum_{m \in C_l} x_m = 0\}$.

2) *Implementation Aspects*: Messages can be represented efficiently in the log-domain, similar to the cognitive radio problem. It turns out that the message from variable node X_n

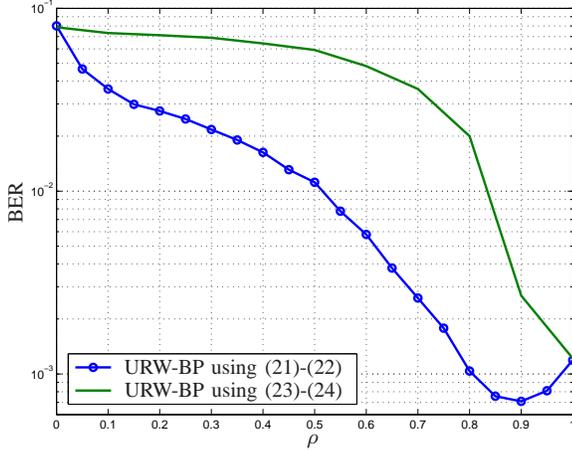


Figure 6. Performance of URW-BP after 20 decoding iterations as a function of the scalar parameter ρ for an LDPC code, at a fixed SNR ($E_b/N_0 = 3\text{dB}$). BP corresponds to $\rho = 1$.

to check node ψ_l , expressed as a log-likelihood ratio (LLR), is given by (see (21) and (23))

$$\lambda_{X_n \rightarrow \psi_l} = \lambda_{\text{ch},n} + \rho \sum_{k \in \mathcal{N}(X_n)} \lambda_{\psi_k \rightarrow X_n} - \lambda_{\psi_l \rightarrow X_n}, \quad (43)$$

where $\lambda_{\text{ch},n} = \log p(y_n|x_n = 1)/p(y_n|x_n = 0)$. The message from variable node to check node depends whether we use (22) or (24). Due to the nature of $\psi_l(x_{C_l})$, $\psi_l^{1/\rho}(x_{C_l}) = \psi_l(x_{C_l})$, irrespective of ρ . Hence, using (22) leads to the standard message passing rule, which we abbreviate as (see also [3, page 160])

$$\lambda_{\psi_l \rightarrow X_n} = f_{\boxplus} \left(\{ \lambda_{X_m \rightarrow \psi_l} \}_{m \neq n} \right). \quad (44)$$

On the other hand, using (24), the message from check ψ_l to X_n becomes

$$\lambda_{\psi_l \rightarrow X_n} = f_{\boxplus} \left(\{ \rho \lambda_{X_m \rightarrow \psi_l} \}_{m \neq n} \right) + (\rho - 1) \lambda_{X_n \rightarrow \psi_l}. \quad (45)$$

In either case, the belief regarding the n -th bit is given by

$$\lambda_{b,n} = \lambda_{\text{ch},n} + \rho \sum_{k \in \mathcal{N}(X_n)} \lambda_{\psi_k \rightarrow X_n}. \quad (46)$$

For additional details see [25]. We perform a maximum of 20 decoding iterations and stop decoding once a codeword has been found.

3) *Performance Results:* In Fig. 6, we plot the bit error rate (BER) performance of a rate 1/2 LDPC code with $N = 256$ and $L = 128$ at a fixed SNR as a function of ρ , over an AWGN channel. We observe that using (23)-(24) yields a performance worse than BP, for all values of ρ . On the other hand, when using (21)-(22), we observe that $\rho = 1$ does not yield the best performance and that the global minimum in the BER is achieved by $\rho \approx 0.9$. The difference in BER between BP and URW-BP is not large, but still significant considering that we have used a real LDPC code, designed such that loops are long and have limited impact on BP decoding. The performance gap can be much greater in case of non-optimized graph configurations, i.e., with many short loops.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a new inference algorithm for estimation and detection in wireless networks. This algorithm, called uniformly reweighted belief propagation (URW-BP), combines the local nature of BP with the improved performance of TRW-BP in graphs with cycles. URW-BP also leads to less concentrated beliefs and is thus more robust. URW-BP includes one scalar tuning parameter (ρ), which depends on the properties of the graphical model. We provided a variational interpretation of URW-BP, give insights into good choices of ρ , and develop an extension to higher-order potentials. While we can only prove the optimality of URW-BP for a restricted class of factorizations (i.e., symmetric factorizations and factorizations for which the factor graph is a tree), numerical results indicate the power of URW-BP over BP for a much larger class of factorizations. In particular, we have illustrated the performance of the proposed method for three practical inference problems in wireless communication systems, involving discrete and continuous variables, centralized and distributed processing, and pairwise as well as higher-order interactions. For each problem, we found that URW-BP consistently outperforms BP, especially for dense graphical models.

Our study did not include the effect of medium access control, packet loss, or time-varying network topology. However, we expect similar results to those found in the literature on belief propagation [36]–[38]. Secondly, we did not perform a detailed study regarding the convergence behavior of URW-BP. For the three case studies we considered, we observed that URW-BP generally converges, and does so at least as fast as BP. Possibly, provably convergent variations of URW-BP can be devised, similar to [39], although this may lead to less practical algorithms. Finally, the choice of optimal ρ is an open problem. An online algorithm to determine ρ for arbitrary network topologies and arbitrary variable interactions is an interesting topic of future research.

REFERENCES

- [1] J. Pearl, "Probabilistic Reasoning in Intelligent Systems. Networks of plausible inference," *Morgan Kaufmann*, 1988.
- [2] A.P. Worthen and W.E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Transactions on Information Theory*, vol. 47, pp. 843–849, Feb. 2001.
- [3] H. Wymeersch, *Iterative Receiver Design*. Cambridge University Press, 2007.
- [4] A.W. Eckford, "Channel estimation in block fading channels using the factor graph EM algorithm," in *Proc. 22nd Biennial Symposium on Communications*, 2004.
- [5] H. Niu, M. Shen, J. A. Ritcey, and H. Liu, "A factor graph approach to iterative channel estimation and LDPC decoding over fading channels," *IEEE Trans. on Wireless Communications*, vol. 4, pp. 1345–1350, July 2005.
- [6] Manyuan Shen, Huaning Niu and Hui Liu, "Iterative receiver design in Rayleigh fading using factor graph," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, (Hong-Kong), Apr. 2003.
- [7] N. Wiberg, *Codes and decoding on general graphs*. PhD thesis, Linköping University, Sweden, 1996.
- [8] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. of the IEEE, special issue on Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*, vol. 97, pp. 427–450, Feb. 2009.

- [9] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Non-parametric belief propagation for self-localization of sensor networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 23, pp. 809–819, Apr. 2005.
- [10] V. Savic, A. Poblacion, S. Zazo, and M. Garcia, "Indoor positioning using nonparametric belief propagation based on spanning trees," *EURASIP Journal on Wireless Communications and Networking*, Aug. 2010.
- [11] B. J. Frey and N. Jovic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1392–1416, Sept. 2005.
- [12] Y. Sheikh and M. Shah, "Bayesian object detection in dynamic scenes," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, June 2005.
- [13] F. Penna, R. Garello, and M. Spirito, "Distributed inference of channel occupation probabilities in cognitive networks via message passing," in *IEEE Symposium on New Frontiers in Dynamic Spectrum*, pp. 1–11, 2010.
- [14] J. Lei, R. Yates, P. Spasojevic, and L. Greenstein, "Cooperative sensing of primary users in cognitive radio networks based on message passing," in *43rd Annual Conference on Information Sciences and Systems (CISS)*, pp. 568–573, 2009.
- [15] Y. Mao, F. Kschischang, B. Li, and S. Pasupathy, "A factor graph approach to link loss monitoring in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 820–829, 2005.
- [16] M. Chiang, "Distributed network control through sum product algorithm on graphs," in *IEEE Global Telecommunications Conference*, vol. 3, pp. 2395–2399, 2003.
- [17] B. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.
- [18] M. Paskin, C. Guestrin, and J. McFadden, "A robust architecture for distributed inference in sensor networks," in *Fourth International Symposium on Information Processing in Sensor Networks*, pp. 55–62, 2005.
- [19] Y. Weiss and W. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173–2200, 2001.
- [20] M. Wainwright, T. Jaakkola, and A. Willsky, "A new class of upper bounds on the log partition function," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2313–2335, 2005.
- [21] T. Roosta, M. Wainwright, and S. Sastry, "Convergence analysis of reweighted sum-product algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. II–541–II–544, 2007.
- [22] M. Wainwright and M. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [23] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, 2005.
- [24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, second ed., 2006.
- [25] H. Wymeersch, F. Penna, and V. Savic, "Uniformly reweighted belief propagation: A factor graph approach," in *IEEE International Symposium on Information Theory*, 2011.
- [26] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," vol. 47, pp. 498–519, Feb. 2001.
- [27] M. Wainwright, T. Jaakkola, and A. Willsky, "MAP estimation via agreement on trees: message-passing and linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3697–3717, 2005.
- [28] Y.-W. Hong, A. Scaglione, and P. K. Varshney, "A communication architecture for reaching consensus in decision for a large network," (Paris, France), pp. 1220–1225, July 2005.
- [29] J. Mooij and H. Kappen, "Sufficient conditions for convergence of the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4422–4437, 2007.
- [30] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors I. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.
- [31] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors II. advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.
- [32] H. Li, "Cooperative spectrum sensing via belief propagation in spectrum-heterogeneous cognitive radio systems," in *Wireless Communications and Networking Conference*, pp. 1–6, 2010.
- [33] A. Dogandzic and B. Zhang, "Distributed estimation and detection for sensor networks using hidden markov random field models," *IEEE Transactions on Signal Processing*, vol. 54, Aug. 2006.
- [34] V. Delouille, R. Neelamani, V. Chandrasekaran, and R. G. Baraniuk, "The embedded triangles algorithm for distributed estimation in sensor networks," in *Proc. 2003 IEEE Workshop Stat. Signal Process.*, 2003.
- [35] S. Venkatesh and M. Alanyali, "M-ary hypothesis testing in sensor networks," in *Proc. 38th Annu. Conf. Inform. Sci. Syst.*, Mar. 2004.
- [36] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4118–4132, Nov. 2006.
- [37] O. Kreidl, *Graphical models and message-passing algorithms for network-constrained decision problems*. PhD thesis, Massachusetts Institute of Technology, 2008.
- [38] C. Crick and A. Pfeffer, "Loopy belief propagation as a basis for communication in sensor networks," in *Uncertainty in Artificial Intelligence*, vol. 18, 2003.
- [39] T. Meltzer, A. Globerson, and Y. Weiss, "Convergent message passing algorithms: a unifying view," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 393–401, AUAI Press, 2009.



Henk Wymeersch (S'99, M'05) received the Ph.D. degree in Electrical Engineering/Applied sciences in 2005 from Ghent University, Belgium. He is currently an Assistant Professor with the Department of Signals and Systems at Chalmers University of Technology, Sweden. He is also affiliated with the FORCE research center on fiber-optic communication. Prior to joining Chalmers, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT). He is a member of the IEEE, Associate Editor for IEEE Communication Letters and the European Transactions on Telecommunication (ETT). He served as Guest Editor for EURASIP Journal on Wireless Communications and Networking (special issue on Localization in Mobile Wireless and Sensor Networks) and is the author of *Iterative Receiver Design* (Cambridge University Press, August 2007). His research interests include algorithm design for wireless transmission, statistical inference and iterative processing.



Federico Penna received a M.S. degree (2008, with honors) in Communications Engineering from Politecnico di Torino, Italy. In 2008 he was a research engineer with Istituto Superiore Mario Boella, Torino, Italy. From 2009 to 2011 he was a Ph.D. student at Politecnico di Torino. During his Ph.D. he spent three months as a visiting researcher at Chalmers University, Gothenburg (Sweden), and six months at UCLA, Los Angeles, CA (USA). At the end of 2011 he joined the Fraunhofer Heinrich Hertz Institute in Berlin, Germany, as a post-doctoral research fellow. His research interests include distributed signal processing, multi-sensor signal detection, and cooperative positioning.



Vladimir Savic received Dipl. Ing. Degree in Electrical Engineering from the University of Belgrade (Serbia), in 2006. From October 2006 to May 2008, he worked as Digital IC Design Engineer in Elsys Eastern Europe, Belgrade. In June 2008, he joined Universidad Politecnica de Madrid (Spain), where he is currently research assistant and PhD student. During Sept-Nov. 2010, he was visiting researcher at the Stony Brook University (NY, USA), and during Sept. 2011 - Jan. 2012, he was visiting researcher at the Chalmers University of Technology (Gothenburg, Sweden). His research interests include cooperative localization and tracking in wireless networks, message passing methods, Monte Carlo methods, particle filtering, and distributed signal processing.