Joint Channel Estimation and Decoding for Trellis-Coded MIMO Two-Way Relay Networks

Frederic Lehmann

***Abstract*—We present a method for joint decode-and-forward physical layer network coding in two-way relay networks. The two source nodes send their packets simultaneously over time-varying channels to a relay node, then the relay broadcasts the received superimposed packets to the source nodes using network coding. The nodes use trellis coding for the sake of error correction and multi-antenna equipments to combat multipath fading. A challenging multiple access problem occurs at the relay node, which performs joint channel estimation and decoding for the individual source packets. We design message passing algorithms based on factor-graphs to solve this problem. The relay has two separate modules that perform channel estimation and decoding for the packets received from each source node. The interference generated by the other source node is taken into account by exchanging messages between the two modules.**

***Index Terms*—Two-way relay network, physical layer network coding, time-varying MIMO channel, trellis coding, joint channel estimation and decoding, message passing.**

I. Introduction

THE use of network coding in wireless networks has attracted considerable interest during the last decade [1]. The number of timeslots required for packet exchange is reduced by exploiting the broadcast nature of wireless channels. The basic idea is that the achievable throughput in a network can be increased, if intermediate nodes are allowed to perform operations on the incoming data, such as linear combination and coding. Compared with conventional routing protocols, the transmission rate, delay and reliability is improved.

We consider the two-way relay channel [2] with half-duplex constraint (i.e. a node cannot transmit and receive at the same time). Two source nodes nodes n_1 and n_2 exchange their packets through a relay node r. In multiple access broadcast protocols, during the first phase (or multiple access phase) n_1 and n_2 send their packets simultaneously to r and during the second phase (or broadcast phase) node r sends a function of the received packets to n_1 and n_2. Such protocols are particularly efficient, since they require only two time slots to convey the packets, instead of four for the routing protocol. We will focus on the recently proposed Physical-layer Network Coding (PNC) scheme [3]-[8], where the relay r directly processes the superposed baseband signals received from n_1 and n_2. The superposition of the signals at the relay node is due to the wireless channel additivity.

Several modes of operation have emerged for the relay node. In the amplify-and-forward PNC scheme [7], also known as analog network coding (ANC), the relay amplifies and broadcasts the received signal. The relay processing is simple to implement and only coarse synchronization at the packet level is needed. However, the relay also amplifies the noise and the cascaded source-relay-source channels must be known to n_1 and n_2 during the broadcast phase, which leads to non-trivial estimation problems. The relay node in the denoise-and-forward PNC scheme [3]-[6] decodes the received signal to a suitable function of the packets sent by n_1 and n_2, without recovering the packets sent by n_1 and n_2 individually. This method avoids the noise amplification of ANC, provided that synchronization among the source nodes can be achieved. However, these methods are sensitive to the channel conditions [9]. In the joint decode-and-forward PNC scheme [8], the relay decodes both packets from n_1 and n_2 and then re-encodes their modulo-2 sum for broadcasting. This method removes the adverse effect of fading and noise in the source-relay channels, provided that synchronization among the source nodes can be achieved. However, these advantages come at the cost of higher computational complexity at the relay node, which must jointly estimate the source-relay channels and jointly decode the superposed packets.

Several issues regarding the practical design of PNC-based communications still need to be addressed. Firstly, the aforementioned PNC schemes have been evaluated under the simplifying assumptions of quasi-static channels and of perfect channel knowledge at the source and/or relay nodes. However, in practical applications, the channels are unknown to the receivers and must be estimated. Moreover, mobile source and/or relay nodes are subject to time-selective fading channels [10]. Recent efforts have extended maximum-likelihood and least-square channel estimation from traditional point-to-point to two-way relay networks. Training-based channel estimation techniques suitable for PNC have appeared in [11]-[12] for static channels, in [13] for time-varying channels and in [14] for frequency-selective channels in OFDM-based transmissions. Secondly, channel coding must be considered in order to protect the transmitted packets against errors due to fading and packet superposition. We will consider trellis codes, which have desirable properties for practical wireless networks, such as low encoding/decoding complexity at each node and low end-to-end decoding latency. This is particularly important if power consumption at the relay node is a concern. Coding techniques suitable for pseudo amplify-and-forward, denoise-and-forward and joint decode-and-forward

Manuscript received July 27, 2011; revised January 19, 2012.

The author is with Institut Telecom, Telecom SudParis, Departement CITI, UMR-CNRS 5157, 91011 Evry Cedex, France (e-mail: frederic.lehmann@it-sudparis.eu).

Part of this work has been presented at IEEE WCNC 2012, Paris, France [25].

Digital Object Identifier 10.1109/JSAC.2013.1308.10

PNC exist [15]-[16], but all of them assume perfect channel knowledge. Thirdly, it has been recognized that multiple-input multiple-output (MIMO) transmissions increase the channel capacity and provide spatial diversity, resulting in robustness against multipath fading and noise [17]. In point-to-point communications, trellis codes for MIMO systems exist in two forms. A space-time trellis code (STTC) [18]-[19] generates a complex symbol for each transmit antenna, by choosing a trellis code which maximizes the diversity and/or the coding gain. MIMO bit interleaved coded modulation (MIMO-BICM) [20] uses a binary convolutional code with maximal free distance, where the coded bits are interleaved and demultiplexed over the transmit antennas. In the recent literature, the use of MIMO technology for two-way relay networks has also been advocated [9],[12].

In this paper, we consider a PNC scheme with multi-antenna equipments, where n_1 uses MIMO-BICM and n_2 either MIMO-BICM or a STTC. A joint decode-and-forward strategy is used at the relay node r to recover the individual packets sent by n_1 and n_2 during the multiple access phase. During the broadcast phase, the relay combines the received packets through Galois field [1], complex field [21] or soft [22] network coding. Packet decoding at node n_1 and n_2 is out of the scope of the present paper, since the broadcast phase is analog to a traditional point-to-point transmission problem between the relay and the source nodes. We will therefore focus on joint channel estimation and decoding at the relay node. We consider time-varying source-relay MIMO channels, modeled as auto-regressive (AR) processes. Using a factor graph approach [23]-[24], we design soft-output message passing algorithms performing joint MIMO channel estimation and decoding of the packets sent by the source nodes during the multiple access phase. The proposed algorithm performs iteratively joint channel estimation and decoding for the packet sent by n_1 and the packet sent by n_2, by taking into account the interference from the other node.

The main technical contributions of this paper are

- A state-space model of superimposed signals suitable for trellis coded PNC on a time-varying MIMO two-way relay channel.
- A joint channel estimation and decoding method for the superimposed signals, based on belief propagation in a factor graph representation of the proposed state-space model.
- Exact belief propagation involves messages in the form of Gaussian mixtures, which leads to an exponential complexity increase as a function of the time index and iterations. As a remedy, a Gaussian approximation is introduced to collapse Gaussian mixture messages to a single Gaussian. This makes belief propagation not only tractable, by preserving a constant complexity per time recursion and per iteration, but also preserves near-optimal performances for trellis coding with sufficient time interleaving.

Throughout the paper, bold letters indicate vectors and matrices, while $\mathbf{I}_m$ denotes the $m \times m$ identity matrix and $\mathbf{0}_m$ the $m \times m$ all-zero matrix. $\mathcal{N}_{\mathcal{C}}(\mathbf{x} : \mathbf{m}, \mathbf{P})$ denotes a complex Gaussian distribution of the variable $\mathbf{x}$, with mean $\mathbf{m}$ and covariance matrix $\mathbf{P}$. The operator $\det(.)$ will denote the determinant of a matrix. Let $\mathbf{c}_j \in \mathbb{C}^m$, for $j = 1, \ldots, n$ denote the columns of $\mathbf{C} \in \mathbb{C}^{m \times n}$, so that $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_n]$. Then $\mathrm{vec}(\mathbf{C})$ denotes the mn-dimensional vector formed by stacking the columns of $\mathbf{C}$ on top of one another, i.e. $\mathrm{vec}(\mathbf{C}) = [\mathbf{c}_1^T, \ldots, \mathbf{c}_n^T]^T$. The symbol $\otimes$ denotes the Kronecker product.

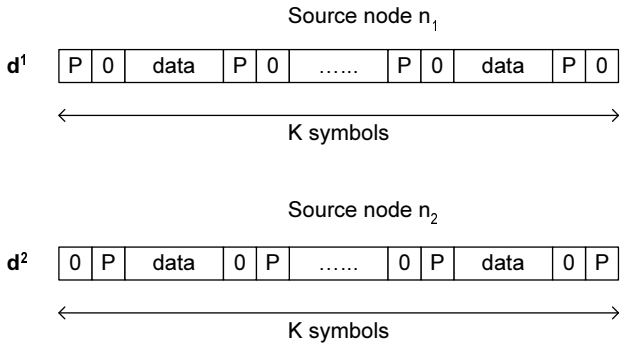

Fig. 1. Data format at the source nodes: pilot symbols (P), zero values (0) and coded data symbols (data) are assigned to the K time slots.

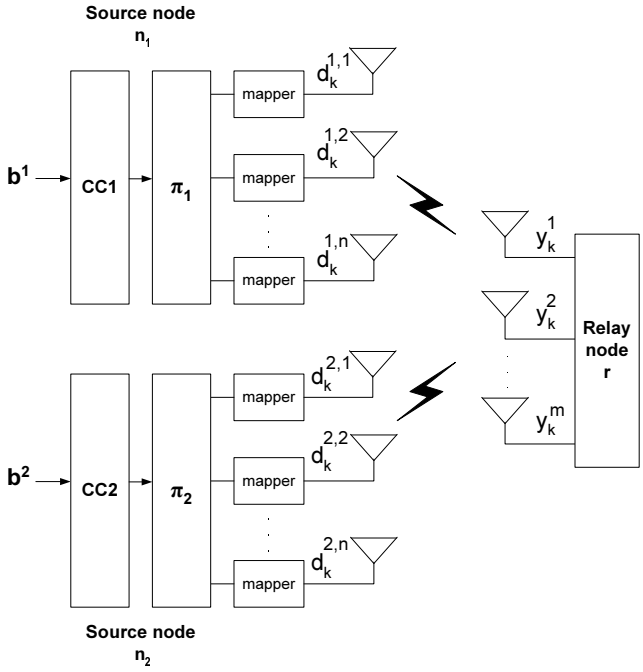

Fig. 2. Multiple access phase of a PNC-based two-way relay system: source node n_1 and n_2 use BICM.

This paper is organized as follows. First, Section II describes the system model adopted for the multiple access phase of PNC on a time-varying MIMO channel. In Section III and IV, we introduce message passing algorithms for joint decode-and-forward at the relay node. Finally, in Section V, the performances of the proposed algorithms are assessed through numerical simulations and compared with existing methods.

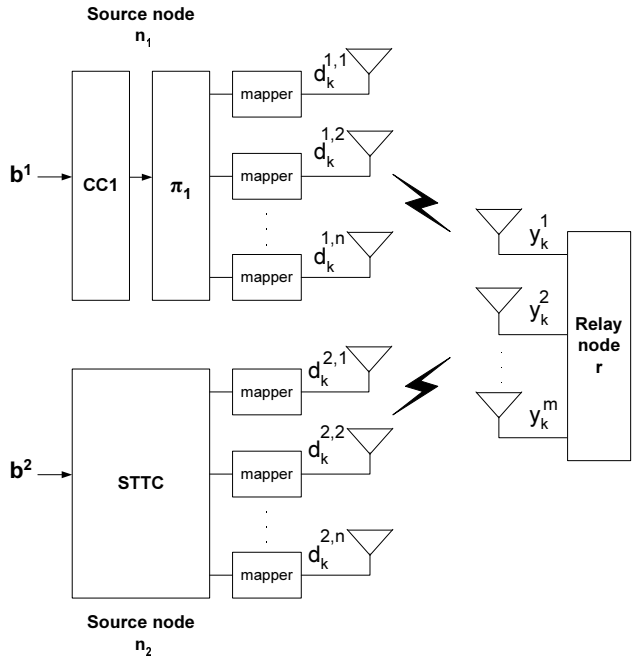

Fig. 3. Multiple access phase of a PNC-based two-way relay system: source node n_1 uses BICM while n_2 uses a STTC.

II. System Model

A. Communication System

We consider the multiple access phase of a PNC-based two-way relay system, with two source nodes n_1 and n_2 having n transmit antennas and a relay node r having m receive antennas. The sequence of uniformly, independently and identically distributed (u.i.i.d.) information bits sent by n_1 (resp. n_2) is denoted by $\mathbf{b}^1 = [b_1^1, b_2^1, \ldots, b_B^1]^T$ (resp. $\mathbf{b}^2 = [b_1^2, b_2^2, \ldots, b_B^2]^T$). Source node n_1 (resp. n_2) encodes its information bits and delivers at instant $k = 1, \ldots, K$ the complex modulated vector $\mathbf{d}_k^1 = [d_k^{1,1}, d_k^{1,2}, \ldots, d_k^{1,n}]^T$ (resp. $\mathbf{d}_k^2 = [d_k^{2,1}, d_k^{2,2}, \ldots, d_k^{2,n}]^T$) sent over the n transmit antennas. The symbol energy is normalized to 1, i.e. $E[|d_k^{i,j}|^2] = 1$, $\forall i, j, k$ and the symbol duration is denoted by T. The sequence of modulated vectors delivered by n_1 (resp. n_2) is denoted by $\mathbf{d}^1 = \{\mathbf{d}_k^1\}_{k=1}^K$ (resp. $\mathbf{d}^2 = \{\mathbf{d}_k^2\}_{k=1}^K$). The data format at the source nodes is depicted in Fig. 1. Pilot symbols are periodically inserted for the purpose of channel estimation at the relay node. In order to avoid pilot interference at the relay node [14], when a node assigns a pilot symbol to a given time slot, the other node assigns a zero value to that time slot.

Without loss of generality, we consider the two scenarios depicted in Fig. 2 and Fig. 3. Fig. 2 corresponds to the homogeneous scenario, where both source nodes encode their information bits using BICM. Source node n_1 (resp. n_2) encodes its information bits with the convolutional code [26] CC1 (resp. CC2), followed by a bit interleaving function π_1 (resp. π_2) and symbol mapping. The encoding function at n_1 (resp. n_2), which maps any sequence of information bits $\mathbf{b}^1$ (resp. $\mathbf{b}^2$) to the corresponding valid sequence of modulated vectors $\mathbf{d}^1$ (resp. $\mathbf{d}^2$) is called $m_1(.)$ (resp. $m_2(.)$). Fig. 3 corresponds to the heterogeneous scenario, where the source node n_1 (resp. n_2) encodes its information bits using BICM (resp. a STTC). Therefore n_1 generates $\mathbf{d}^1 = m_1(\mathbf{b}^1)$ using the same encoding function as in Fig. 2. The STTC used by n_2 is characterized by the encoder state at instant k, s_k^2, which forms a finite-state Markov process [18]-[19], i.e. s_k^2 depends only on s_{k-1}^2 and on the new information bit b_k^2. Moreover, we define the states such that s_k^2 determines completely the value of the complex modulated vector $\mathbf{d}_k^2$ at instant k.

Remark 2.1: The case where n_1 uses a STTC and n_2 uses BICM will not be considered, since it is identical to the configuration of Fig. 3 by exchanging the role of n_1 and n_2. The case where both n_1 and n_2 use a STTC leads to a factor graph representation with many short cycles, so that the message passing detectors introduced latter, which handle the decoding of the packets coming from n_1 and n_2 separately, suffers from an error floor. Thus, this configuration cannot be treated using our algorithm, but needs to be handled using the computationally expensive joint trellis approach of [16], extended to the MIMO case.

Let $\mathbf{y}_k = [y_k^1, y_k^2, \ldots, y_k^m]^T$ be the complex baseband observations received at the m antennas of the relay node r, during the multiple access phase at instant k. Due to the packet superposition on the source-relay MIMO channels and assuming perfect symbol synchronization at the relay node, $\mathbf{y}_k$ has the form

$$\mathbf{y}_k = \sqrt{E_s^1}\mathbf{X}_k^1\mathbf{d}_k^1 + \sqrt{E_s^2}\mathbf{X}_k^2\mathbf{d}_k^2 + \mathbf{n}_k, \tag{1}$$

where E_s^1 (resp. E_s^2) is the average energy per transmit antenna for source node n_1 (resp. n_2). The relative proximity between the source nodes and the relay is measured by the relative path-loss gain defined as

$$G = 10\log_{10}\left(\frac{E_s^1}{E_s^2}\right), \tag{2}$$

which means that for $G \geq 0$, n_1 is closer to r than n_2. $\mathbf{X}_k^1$ (resp. $\mathbf{X}_k^2$) is the $m \times n$ matrix of time-varying complex path gains for the MIMO channel between n_1 (resp. n_2) and r. $\mathbf{n}_k = [n_k^1, n_k^2, \ldots, n_k^m]^T$ is a vector of independent zero-mean additive white Gaussian noise (AWGN) samples, with covariance matrix equal to $\mathbf{R} = N_0\mathbf{I}_m$. Defining the mn-dimensional stacked vectors of channel gains

$$\begin{cases} \mathbf{x}_k^1 = \mathrm{vec}({\mathbf{X}_k^1}^T) \\ \mathbf{x}_k^2 = \mathrm{vec}({\mathbf{X}_k^2}^T), \end{cases}$$

Eq. (1) can be rewritten as

$$\mathbf{y}_k = \mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1 + \mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{x}_k^2 + \mathbf{n}_k, \tag{3}$$

where the $m \times nm$ observation matrices are given by

$$\begin{cases} \mathbf{H}_k^1(\mathbf{d}_k^1) = \sqrt{E_s^1}\mathbf{I}_m \otimes {\mathbf{d}_k^1}^T \\ \mathbf{H}_k^2(\mathbf{d}_k^2) = \sqrt{E_s^2}\mathbf{I}_m \otimes {\mathbf{d}_k^2}^T. \end{cases} \tag{4}$$

B. MIMO Channel Model

Let $f_m^1 T$ (resp. $f_m^2 T$) be the normalized fading rate of the MIMO channel between n_1 (resp. n_2) and r. Define the corresponding coefficients $\zeta^i = 2 - \cos(2\pi f_m^i T) - \sqrt{(2 - \cos(2\pi f_m^i T))^2 - 1}$, for $i \in \{1, 2\}$. We consider the approximate model for a mobile Rayleigh fading channel using an autoregressive model of order one (AR(1)) introduced

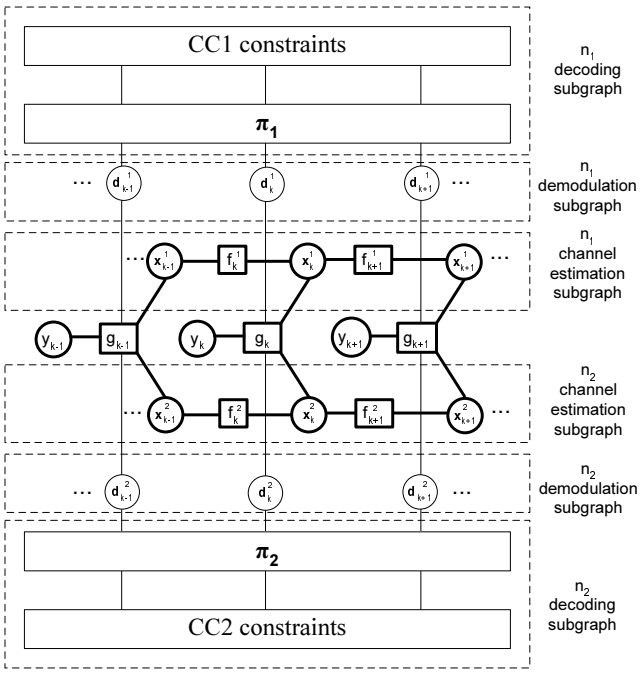

Fig. 4. Factor graph for the relay processing: source node n_1 and n_2 use BICM.

in [27] (pp. 74-75). A state-space representation of the MIMO channel gains is obtained as

$$\begin{cases} \mathbf{x}_k^1 = \mathbf{F}^1 \mathbf{x}_{k-1}^1 + \mathbf{u}_k^1 \\ \mathbf{x}_k^2 = \mathbf{F}^2 \mathbf{x}_{k-1}^2 + \mathbf{u}_k^2, \end{cases} \tag{5}$$

where the state transition matrices are given by

$$\begin{cases} \mathbf{F}^1 = \zeta^1 \mathbf{I}_{nm} \\ \mathbf{F}^2 = \zeta^2 \mathbf{I}_{nm}, \end{cases}$$

and the process noise vector $\mathbf{u}_k^1$ (resp. $\mathbf{u}_k^2$) is zero-mean complex Gaussian distributed with covariance matrix equal to $\mathbf{Q}^1 = [1-(\zeta^1)^2]\mathbf{I}_{nm}$ (resp. $\mathbf{Q}^2 = [1-(\zeta^2)^2]\mathbf{I}_{nm}$). Assuming that n_1 and n_2 are sufficiently far apart, the corresponding random channel gain vectors $\mathbf{x}_k^1$ and $\mathbf{x}_k^2$ are independent. Moreover, assuming that

$$\begin{cases} p(\mathbf{x}_0^1) = \mathcal{N}_{\mathcal{C}}(\mathbf{x}_0^1 : \mathbf{0}_{mn}, \mathbf{I}_{mn}) \\ p(\mathbf{x}_0^2) = \mathcal{N}_{\mathcal{C}}(\mathbf{x}_0^2 : \mathbf{0}_{mn}, \mathbf{I}_{mn}), \end{cases}$$

then at any time instant k, all the channel gains are zero-mean Gaussian distributed random variables, normalized to unit variance.

C. Factor Graph Representation

We first apply the factor graph framework of [24] to the homogeneous scenario of Fig. 2, where both source nodes employ BICM. We obtain the following factorization

$$\begin{aligned} & p\left(\mathbf{b}^1, \mathbf{d}^1, \{\mathbf{x}_k^1\}_{k=0}^K, \mathbf{b}^2, \mathbf{d}^2, \{\mathbf{x}_k^2\}_{k=0}^K | \{\mathbf{y}_k\}_{k=1}^K\right) \\ & \propto p\left(\{\mathbf{y}_k\}_{k=1}^K | \mathbf{d}^1, \{\mathbf{x}_k^1\}_{k=0}^K, \mathbf{d}^2, \{\mathbf{x}_k^2\}_{k=0}^K\right) \\ & \quad \times p(\{\mathbf{x}_k^1\}_{k=0}^K) p(\{\mathbf{x}_k^2\}_{k=0}^K) \\ & \quad \times p(\mathbf{d}^1|\mathbf{b}^1) p(\mathbf{d}^2|\mathbf{b}^2) p(\mathbf{b}^1) p(\mathbf{b}^2) \\ & \propto \prod_{k=1}^K p\left(\mathbf{y}_k | \mathbf{d}_k^1, \mathbf{x}_k^1, \mathbf{d}_k^2, \mathbf{x}_k^2\right) \\ & \quad \times p(\mathbf{x}_0^1) \prod_{k=1}^K p(\mathbf{x}_k^1|\mathbf{x}_{k-1}^1) \times I(\mathbf{d}^1 = m_1(\mathbf{b}^1)) \\ & \quad \times p(\mathbf{x}_0^2) \prod_{k=1}^K p(\mathbf{x}_k^2|\mathbf{x}_{k-1}^2) \times I(\mathbf{d}^2 = m_2(\mathbf{b}^2)), \end{aligned} \tag{6}$$

where $I(\mathbf{d}^1 = m_1(\mathbf{b}^1))$ (resp. $I(\mathbf{d}^2 = m_2(\mathbf{b}^2))$) is the code indicator function, equal to one if $\mathbf{d}^1$ (resp. $\mathbf{d}^2$) is the valid complex modulated coded sequence corresponding to $\mathbf{b}^1$ (resp. $\mathbf{b}^2$) and equal to zero otherwise. To obtain the last expression in (6), we have used the first order Markov assumption for the channel gains in (5), the fact that the AWGN in (3) is memoryless, and that that the information bits are u.i.i.d. The factor graph corresponding to the above factorization is depicted in Fig. 4. Variable nodes are represented as circles and the local functions appearing in the factorization, denoted by

$$\begin{cases} f_k^1 = p(\mathbf{x}_k^1|\mathbf{x}_{k-1}^1) \\ f_k^2 = p(\mathbf{x}_k^2|\mathbf{x}_{k-1}^2) \\ g_k = p\left(\mathbf{y}_k|\mathbf{d}_k^1, \mathbf{x}_k^1, \mathbf{d}_k^2, \mathbf{x}_k^2\right). \end{cases} \tag{7}$$

are represented as squares. The portion of the graph in thick (resp. thin) line corresponds to the channel estimation task (resp. demodulation and decoding tasks) at the relay node. Also the graph is symmetric, so that the upper (resp. lower) half of the graph corresponds to channel estimation, demodulation and decoding of the packet sent by n_1 (resp. n_2). This was expected, since both source nodes use the same type of coding and modulation over the same kind of MIMO channel.

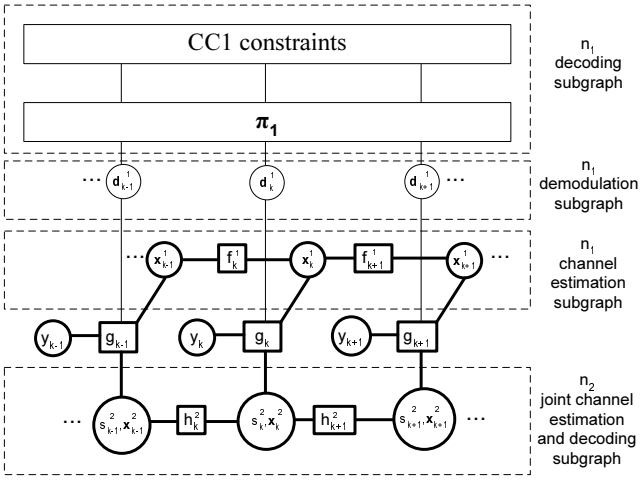

Fig. 5. Factor graph for the relay processing: source node n_1 uses BICM while n_2 uses a STTC.

Now, applying the factor graph framework to the heterogeneous scenario of Fig. 3, where n_1 employs BICM while n_2 employs a STTC, we obtain the following factorization

$$\begin{aligned} & p\left(\mathbf{b}^1, \mathbf{d}^1, \{\mathbf{x}_k^1\}_{k=0}^K, \mathbf{b}^2, \{s_k^2\}_{k=0}^K, \{\mathbf{x}_k^2\}_{k=0}^K | \{\mathbf{y}_k\}_{k=1}^K\right) \\ & \propto p\left(\{\mathbf{y}_k\}_{k=1}^K | \mathbf{d}^1, \{\mathbf{x}_k^1\}_{k=0}^K, \{s_k^2\}_{k=0}^K, \{\mathbf{x}_k^2\}_{k=0}^K\right) \\ & \quad \times p(\{\mathbf{x}_k^1\}_{k=0}^K) p(\{\mathbf{x}_k^2\}_{k=0}^K) p(\mathbf{d}^1|\mathbf{b}^1) p(\{s_k^2\}_{k=0}^K|\mathbf{b}^2) \\ & \quad \times p(\mathbf{b}^1) p(\mathbf{b}^2) \\ & \propto \prod_{k=1}^K p\left(\mathbf{y}_k | \mathbf{d}_k^1, \mathbf{x}_k^1, s_k^2, \mathbf{x}_k^2\right) \\ & \quad \times p(\mathbf{x}_0^1) \prod_{k=1}^K p(\mathbf{x}_k^1|\mathbf{x}_{k-1}^1) \times I(\mathbf{d}^1 = m_1(\mathbf{b}^1)) \\ & \quad \times p(\mathbf{x}_0^2) \prod_{k=1}^K p(\mathbf{x}_k^2|\mathbf{x}_{k-1}^2) \times p(s_0^2) \prod_{k=1}^K p(s_k^2|s_{k-1}^2), \end{aligned} \tag{8}$$

To obtain the last expression in (8), we have used the fact that the STTC states form a first order Markov process and that the value of the STTC state s_k^2 determines completely the value of the complex modulated vector $\mathbf{d}_k^2$ at instant k. The factor graph corresponding to the above factorization is depicted in Fig. 5. The local functions appearing in the factorization are defined as follows

$$\begin{cases} f_k^1 = p(\mathbf{x}_k^1|\mathbf{x}_{k-1}^1) \\ h_k^2 = p(\mathbf{x}_k^2|\mathbf{x}_{k-1}^2) p(s_k^2|s_{k-1}^2) \\ g_k = p\left(\mathbf{y}_k|\mathbf{d}_k^1, \mathbf{x}_k^1, s_k^2, \mathbf{x}_k^2\right). \end{cases} \tag{9}$$

The portion of the graph in thick line corresponds to the channel estimation task for the packet sent by n_1 and to joint channel estimation and decoding for the packet sent by n_2, at the relay node. The portion of the graph in thin line corresponds to the demodulation and decoding tasks for the packet sent by n_1. As expected, this graph is not symmetric, since the source nodes do not use the same type of coding and modulation.

Remark 2.2: The expression of the local function node h_k^2 indicates that another factor graph representation could have been obtained by separating the channel estimation from the STTC decoding task for the packet sent by n_2. However, this decomposition would lead to a subgraph containing many short cycles, instead of the proposed tree-like subgraph for joint channel estimation and decoding. It is well-known that the performance of message passing degrades in the presence of too many short cycles [23].

III. Message Passing Algorithm for Joint Decode-and-Forward in the Homogeneous Scenario

In this section, we derive a Bayesian inference algorithm to obtain a joint decode-and-forward scheme at the relay node in the homogeneous scenario of Fig. 2. We apply the sum-product algorithm (SPA) [23], which implements belief propagation [28], to the factor graph of Fig. 4. This graph has cycles, therefore belief propagation will not implement exact Bayesian inference [28]. However, excellent SPA performances have been reported also on graphs with cycles, provided that the graph is sufficiently sparse. This property has been used successfully in the past, notably for decoding turbo, low-density parity-check (LDPC) and repeat-accumulate codes (RA) [23].

As we shall see, exact belief propagation involves messages in the form of Gaussian mixtures. Thus, an exponential complexity increase as a function of the time index and iterations results. However, in practice we need a detection algorithm with a constant complexity per time recursion and per iteration. To achieve this goal, we first introduce a Gaussian approximation for Gaussian mixture messages in Section III-A. We then develop a belief propagation algorithm based on this approximation in the subsequent sections. Our simulations will show that the resulting iterative detection algorithm achieves an excellent performance/complexity tradeoff.

Let $\mu_{u\to v}(.)$ be the message sent by node u to node v in the factor graph. Due to the symmetry in the factor graph, we only derive the messages corresponding to the upper half of the graph. The message schedule proceeds as follows. Without loss of generality, assuming that $E_s^1 \geq E_s^2$, first the upper half of the graph is treated, i.e. the packet from the source node with the strongest received signal. First, channel estimation is performed (see Section III-B), followed by the demodulation step (see Section III-C) and channel decoding (see Section III-D). Then, the same steps are peformed for the lower half of the graph. This procedure is iterated a number of times, until convergence is reached. For the sake of completeness, the message initialization is detailed in Section III-E.

A. Gaussian Approximation using the Moment-Matching Method

Assume that a message sent by node u to node v in the factor graph is a Gaussian mixture of the form

$$\mu_{u\to v}(.) \propto \sum_i \omega_i \mathcal{N}_{\mathcal{C}}(., \mathbf{a}_i, \mathbf{\Sigma}_i), \tag{10}$$

with $\sum_i \omega_i = 1$. This message can be approximated by a single Gaussian with the same expectation and covariance as the original message, namely

$$\mu_{u\to v}(.) \propto \mathcal{N}_{\mathcal{C}}(., \hat{\mathbf{a}}, \hat{\mathbf{\Sigma}}), \tag{11}$$

where

$$\begin{aligned} \hat{\mathbf{a}} &= \sum_i \omega_i \mathbf{a}_i \\ \hat{\mathbf{\Sigma}} &= \sum_i \omega_i \left[\mathbf{\Sigma}_i + (\mathbf{a}_i - \hat{\mathbf{a}})(\mathbf{a}_i - \hat{\mathbf{a}})^H\right]. \end{aligned}$$

The demonstration is readily available from [29] (p. 107).

B. Channel Estimation

Assume that $\mu_{\mathbf{x}_k^2 \to g_k}(\mathbf{x}_k^2)$ has the form

$$\mu_{\mathbf{x}_k^2 \to g_k}(\mathbf{x}_k^2) \propto \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2 : \hat{\mathbf{x}}_{k\backslash k}^2, \mathbf{P}_{k\backslash k}^2). \tag{12}$$

Let us first apply the sum-product rule at the function node g_k. We obtain

$$\begin{aligned} \mu_{g_k \to \mathbf{x}_k^1}(\mathbf{x}_k^1) \propto & \sum_{\mathbf{d}_k^1} \sum_{\mathbf{d}_k^2} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1) \mu_{\mathbf{d}_k^2 \to g_k}(\mathbf{d}_k^2) \\ & \times \int p\left(\mathbf{y}_k | \mathbf{d}_k^1, \mathbf{x}_k^1, \mathbf{d}_k^2, \mathbf{x}_k^2\right) \mu_{\mathbf{x}_k^2 \to g_k}(\mathbf{x}_k^2) d\mathbf{x}_k^2. \end{aligned}$$

According to (3) and (12), this expression becomes

$$\begin{aligned}
&\mu_{g_k\to\mathbf{x}_k^1}(\mathbf{x}_k^1)\\
&\propto\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\quad\times\int\mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k:\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{x}_k^2,\mathbf{R}\right)\\
&\qquad\times\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2:\hat{\mathbf{x}}_{k\backslash k}^2,\mathbf{P}_{k\backslash k}^2)d\mathbf{x}_k^2 \qquad (13)\\
&\propto\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\quad\times\mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k:\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2,\\
&\qquad\qquad\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H+\mathbf{R}\Big).
\end{aligned}$$

where the closed form expression of the integral in the last expression has been demonstrated in [29] (p. 38). In order to obtain a tractable expression, we collapse the obtained Gaussian mixture to a single Gaussian using the moment-matching method of Section III-A. According to Appendix B, we obtain

$$\mu_{g_k\to\mathbf{x}_k^1}(\mathbf{x}_k^1)\propto\mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k:\hat{\mathbf{H}}_k^1\mathbf{x}_k^1+\mathbf{b}_k^2,\mathbf{S}_k^2\right), \qquad (14)$$

where

$$\begin{cases}\hat{\mathbf{H}}_k^1=\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mathbf{H}_k^1(\mathbf{d}_k^1)\\ \hat{\mathbf{H}}_k^2=\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\mathbf{H}_k^2(\mathbf{d}_k^2)\\ \mathbf{b}_k^2=\hat{\mathbf{H}}_k^2\hat{\mathbf{x}}_{k\backslash k}^2.\end{cases}$$

and

$$\begin{aligned}
\mathbf{S}_k^2=&\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)^H\\
&+\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\Bigg\{\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H+\\
&\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)\hat{\mathbf{x}}_{k\backslash k}^2\hat{\mathbf{x}}_{k\backslash k}^{2\,H}\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)^H\Bigg\}\\
&+\mathbf{R}.
\end{aligned}$$

Remark 3.1: The n_1 to relay channel estimation needs an approximation of the likelihood $p(\mathbf{y}_k|\mathbf{x}_k^1)$, by averaging out the variables $\mathbf{d}_k^1$, $\mathbf{d}_k^2$ and $\mathbf{x}_k^2$. We obtain the Gaussian distribution (14) with mean $\hat{\mathbf{H}}_k^1\mathbf{x}_k^1+\mathbf{b}_k^2$, where $\hat{\mathbf{H}}_k^1$ is the observation matrix averaged over the modulated vectors $\mathbf{d}_k^1$ and $\mathbf{b}_k^2$ is a bias which accounts for the superimposed n_2 to relay transmission. The covariance matrix $\mathbf{S}_k^2$ accounts for the presence of AWGN, residual interference from the superimposed n_2 to relay transmission and channel estimation uncertainty.

Now, the n_1 channel estimation subgraph in Fig. 4 is a tree, therefore if it were disconnected from the rest of the factor graph, the SPA would achieve exact Bayesian inference on this subgraph. Moreover, according to the expression of the function node f_k^1 in (9) and of the message $\mu_{g_k\to\mathbf{x}_k^1}(\mathbf{x}_k^1)$ in (14), the channel estimation subgraph corresponds to a linear Gaussian system. As shown in [23], the forward (resp. backward) pass of the SPA computes Gaussian messages of the form

$$\begin{aligned}
\mu_{f_k^1\to\mathbf{x}_k^1}(\mathbf{x}_k^1)\propto&\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\hat{\mathbf{x}}_{k|k-1}^1,\mathbf{P}_{k|k-1}^1)\\
\mu_{f_{k+1}^1\to\mathbf{x}_k^1}(\mathbf{x}_k^1)\propto&\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\hat{\mathbf{x}}_{k|k+1:K}^1,\mathbf{P}_{k|k+1:K}^1),
\end{aligned} \qquad (15)$$

whose mean and covariance time update rule is given by the two-filter Kalman smoother [30].

Let us finally compute the message returned by the channel estimation subgraph to the function node g_k. The sum-product rule, applied to variable node $\mathbf{x}_k^1$, yields

$$\begin{aligned}
\mu_{\mathbf{x}_k^1\to g_k}(\mathbf{x}_k^1)\propto&\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\hat{\mathbf{x}}_{k|k-1}^1,\mathbf{P}_{k|k-1}^1)\\
&\times\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\hat{\mathbf{x}}_{k|k+1:K}^1,\mathbf{P}_{k|k+1:K}^1).
\end{aligned}$$

Using the expression for a product of Gaussian densities in Appendix A, we obtain the following simplification

$$\mu_{\mathbf{x}_k^1\to g_k}(\mathbf{x}_k^1)\propto\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\hat{\mathbf{x}}_{k\backslash k}^1,\mathbf{P}_{k\backslash k}^1), \qquad (16)$$

where

$$\begin{cases}\mathbf{P}_{k\backslash k}^1=\mathbf{P}_{k|k+1:K}^1\left[\mathbf{P}_{k|k-1}^1+\mathbf{P}_{k|k+1:K}^1\right]^{-1}\mathbf{P}_{k|k-1}^1\\ \hat{\mathbf{x}}_{k\backslash k}^1=\mathbf{P}_{k\backslash k}^1\left[{\mathbf{P}_{k|k-1}^1}^{-1}\hat{\mathbf{x}}_{k|k-1}^1+{\mathbf{P}_{k|k+1:K}^1}^{-1}\hat{\mathbf{x}}_{k|k+1:K}^1\right].\end{cases}$$

We have just shown that under a Gaussian approximation on $\mu_{g_k\to\mathbf{x}_k^1}(.)$, if $\mu_{\mathbf{x}_k^2\to g_k}(.)$ is Gaussian for all k, then all the messages exchanged in the channel estimation subgraph in the upper half of Fig. 4 are Gaussian. Note that the channel estimation subgraph in the upper and lower halves of Fig. 4 are identical. It follows that all the messages exchanged in the channel estimation subgraph in the lower half of Fig. 4 are also Gaussian. Therefore, by induction, if $\mu_{\mathbf{x}_k^2\to g_k}(.)$ is Gaussian for all k at the initial iteration, then all the messages exchanged in the channel estimation subgraphs will be Gaussian for all subsequent iterations as well.

C. Demodulation

The demodulation consists of performing soft symbol vector detection, by calculating $\mu_{g_k\to\mathbf{d}_k^1}(\mathbf{d}_k^1)$. Applying the sum-product rule at the function node g_k,

$$\begin{aligned}
\mu_{g_k\to\mathbf{d}_k^1}(\mathbf{d}_k^1)=&\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\times\int\int p\left(\mathbf{y}_k|\mathbf{d}_k^1,\mathbf{x}_k^1,\mathbf{d}_k^2,\mathbf{x}_k^2\right)\\
&\qquad\times\mu_{\mathbf{x}_k^1\to g_k}(\mathbf{x}_k^1)\mu_{\mathbf{x}_k^2\to g_k}(\mathbf{x}_k^2)d\mathbf{x}_k^1d\mathbf{x}_k^2\\
=&\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\times\mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k:\mathbf{H}_k^1(\mathbf{d}_k^1)\hat{\mathbf{x}}_{k\backslash k}^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2,\\
&\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{P}_{k\backslash k}^1\mathbf{H}_k^1(\mathbf{d}_k^1)^H+\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H+\mathbf{R}\Big).
\end{aligned} \qquad (17)$$

The demonstration is postponed to Appendix C. The message from $\mathbf{d}_k^1$ to the bit interleaver π_1, corresponds to the calculation of *a posteriori* bit probabilities from *a posteriori* symbol probabilities. This is a standard procedure in BICM, which can be found in [31], for instance. Similarly, the message from the bit interleaver π_1 to $\mathbf{d}_k^1$, corresponds to the calculation of *extrinsic* symbol probabilities from *extrinsic* bit probabilities at the decoder output [31].

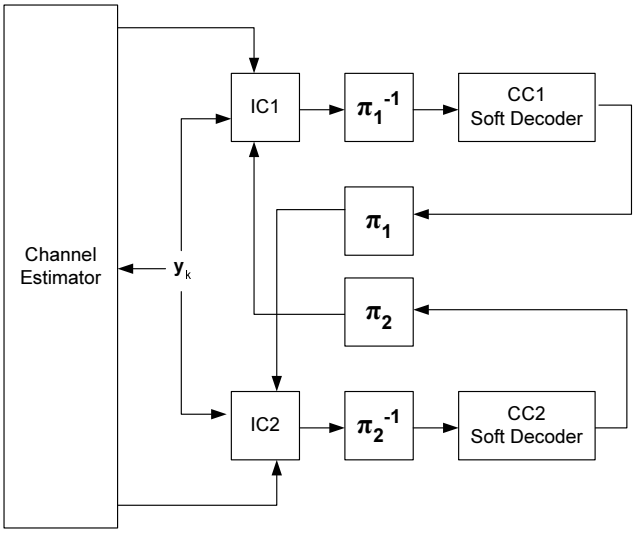

Fig. 6. Iterative serial interference cancellation (SIC) receiver for the PNC system of Fig. 2: iterative channel estimation, interference cancellation (IC) and decoding.

D. Decoding

The SPA applied to the upper box in Fig. 4, corresponding to the convolutional code (CC1) constraints, consists of the well-known BCJR algorithm [32]. Therefore we omit the details here.

E. Initialization

During the first iteration, the messages $\mu_{\mathbf{x}_k^2 \to g_k}(\mathbf{x}_k^2)$ (corresponding to the n_2 to relay MIMO channel distribution), $\mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)$ (corresponding to the n_1 symbol vector distribution) and $\mu_{\mathbf{d}_k^2 \to g_k}(\mathbf{d}_k^2)$ (corresponding to the n_2 symbol vector distribution) must be initialized. We use the prior probability mass function (pmf) for $\mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)$ and $\mu_{\mathbf{d}_k^2 \to g_k}(\mathbf{d}_k^2)$ which is the uniform pmf, since the information packets at the source nodes have been assumed u.i.i.d. Similarly, we use the prior probability distribution function (pdf) for $\mu_{\mathbf{x}_k^2 \to g_k}(\mathbf{x}_k^2)$, which is the Gaussian

$$p(\mathbf{x}_k^2) = \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2 : \mathbf{0}_{mn}, \mathbf{I}_{mn}), \forall k$$

according to the MIMO channel model in (5).

F. Comparison with Existing Message-Passing Methods

For the homogeneous scenario of Fig. 2, two message-passing algorithms have been published in the literature.

Firstly, the single-user bound, which corresponds to the performance of message passing when only the source node of interest communicates with the relay node and the other source node is discarded, has been studied in [33].

Secondly, iterative serial interference cancellation (SIC) [**?**], originally employed for code division multiple access (CDMA) systems, can be easily adapted to our MIMO multiple access relay channel problem. Therefore we use iterative SIC as a benchmark to assess the performances of the proposed algorithm. A complete description of the iterative SIC receiver is given by Fig. 6. At each iteration, the channel is (re-)estimated, using the code-aided technique suitable for time-varying MIMO interference channels, introduced in [35] (see Section V). The n_1 to relay and n_2 to relay channels are jointly estimated using this method, with a number of Gaussian components fixed to one, in order to keep the computational complexity of the benchmark method close to the complexity of the proposed scheme. This procedure is iterated until convergence is reached.

IV. Message Passing Algorithm for Joint Decode-and-Forward in the Heterogeneous Scenario

In this section, we derive a Bayesian inference algorithm to obtain a joint decode-and-forward scheme at the relay node in the heterogeneous scenario of Fig. 3.

The message schedule proceeds as follows on the factor graph of Fig. 5. First, channel estimation, followed by demodulation and channel decoding are performed for the packet sent by n_1 to the relay node. These steps are obtained by applying the SPA to the upper half of the graph (see Section IV-B), with only minor modifications with respect to the message update rules presented in Section III. Then, the SPA is applied to the lower half of the graph (see Section IV-A), which corresponds to joint channel estimation and decoding of the packet sent by n_2 to the relay node. This procedure is iterated a number of times, until convergence is reached. For the sake of completeness, the message initialization is detailed in Section IV-C.

A. n_2 Joint Channel Estimation and Decoding

We first calculate the message $\mu_{g_k \to s_k^2, \mathbf{x}_k^2}(s_k^2, \mathbf{x}_k^2)$, which is instrumental for the derivation of the forward and backward pass of the SPA for the tree-like subgraph at the bottom of Fig. 5. Let us apply the sum product rule at the function node g_k, defined in (9)

$$\mu_{g_k \to s_k^2, \mathbf{x}_k^2}(s_k^2, \mathbf{x}_k^2) = \sum_{\mathbf{d}_k^1} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1) \int p\left(\mathbf{y}_k | \mathbf{d}_k^1, \mathbf{x}_k^1, s_k^2, \mathbf{x}_k^2\right) \mu_{\mathbf{x}_k^1 \to g_k}(\mathbf{x}_k^1) d\mathbf{x}_k^1.$$

Since the n_1 channel estimation subgraphs in Fig. 4 and Fig. 5 are identical, Eq. (16) is still valid. According to (3), we obtain

$$\begin{aligned}\mu_{g_k \to s_k^2, \mathbf{x}_k^2}(s_k^2, \mathbf{x}_k^2) &\propto \sum_{\mathbf{d}_k^1} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1) \int \mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k : \mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1 + \mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2, \mathbf{R}\right) \\ &\qquad \times \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1 : \hat{\mathbf{x}}_{k\backslash k}^1, \mathbf{P}_{k\backslash k}^1) d\mathbf{x}_k^1 \\ &\propto \sum_{\mathbf{d}_k^1} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1) \mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k : \mathbf{H}_k^1(\mathbf{d}_k^1)\hat{\mathbf{x}}_{k\backslash k}^1 + \mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2, \\ &\qquad \mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{P}_{k\backslash k}^1 \mathbf{H}_k^1(\mathbf{d}_k^1)^H + \mathbf{R}\Big),\end{aligned} \tag{18}$$

where the closed form expression of the integral in the last expression has been demonstrated in [29] (p. 38). In order to obtain a tractable expression, we collapse the obtained

Gaussian mixture to a single Gaussian using the moment-matching of Section III-A. According to Appendix D, we obtain

$$\mu_{g_k\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\propto\mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k:\mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2+\mathbf{b}_k^1,\mathbf{S}_k^1\right). \quad (19)$$

where

$$\begin{cases}\hat{\mathbf{H}}_k^1=\sum\limits_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mathbf{H}_k^1(\mathbf{d}_k^1)\\ \mathbf{b}_k^1=\hat{\mathbf{H}}_k^1\hat{\mathbf{x}}_{k\backslash k}^1.\end{cases}$$

and

$$\mathbf{S}_k^1=\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\Bigg\{\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{P}_{k\backslash k}^1\mathbf{H}_k^1(\mathbf{d}_k^1)^H+\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)\hat{\mathbf{x}}_{k\backslash k}^1\hat{\mathbf{x}}_{k\backslash k}^{1\,H}\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)^H\Bigg\}+\mathbf{R}.$$

Remark 4.1: The n_2 to relay joint channel estimation and decoding needs an approximation of the likelihood $p(\mathbf{y}_k|s_k^2,\mathbf{x}_k^2)$, by averaging out the variables $\mathbf{d}_k^1$, and $\mathbf{x}_k^1$. We obtain the Gaussian distribution (19) with mean $\mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2+\mathbf{b}_k^1$, where $\mathbf{b}_k^1$ is a bias which accounts for the superimposed n_1 to relay transmission. The covariance matrix $\mathbf{S}_k^1$ accounts for the presence of AWGN, residual interference from the superimposed n_1 to relay transmission and channel estimation uncertainty.

1) Forward Pass: We seek a time update rule for the message $\mu_{s_{k-1}^2,\mathbf{x}_{k-1}^2\to h_k^2}(s_{k-1}^2,\mathbf{x}_{k-1}^2)$ of the form

$$\mu_{s_{k-1}^2,\mathbf{x}_{k-1}^2\to h_k^2}(s_{k-1}^2,\mathbf{x}_{k-1}^2)\propto\alpha_{k-1|k-1}(s_{k-1}^2)\times\mathcal{N}_{\mathcal{C}}(\mathbf{x}_{k-1}^2:\hat{\mathbf{x}}_{k-1|k-1}^2(s_{k-1}^2),\mathbf{P}_{k-1|k-1}^2(s_{k-1}^2)), \quad (20)$$

where $\alpha_{k-1|k-1}(s_{k-1}^2)$ is the belief of the STTC state s_{k-1}^2 and $\mathcal{N}_{\mathcal{C}}(\mathbf{x}_{k-1}^2:\hat{\mathbf{x}}_{k-1|k-1}^2(s_{k-1}^2),\mathbf{P}_{k-1|k-1}^2(s_{k-1}^2))$ is the belief of the MIMO channel vector $\mathbf{x}_{k-1}^2$, conditional on the value of s_{k-1}^2.

Applying the sum-product rule to the local function h_k^2, defined in (9), we obtain

$$\mu_{h_k^2\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\propto\sum_{s_{k-1}^2}p(s_k^2|s_{k-1}^2)\times\int p(\mathbf{x}_k^2|\mathbf{x}_{k-1}^2)\mu_{s_{k-1}^2,\mathbf{x}_{k-1}^2\to h_k^2}(s_{k-1}^2,\mathbf{x}_{k-1}^2)d\mathbf{x}_{k-1}^2$$

Using (5) and (20) this expression becomes

$$\mu_{h_k^2\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\propto\sum_{s_{k-1}^2}p(s_k^2|s_{k-1}^2)\alpha_{k-1|k-1}(s_{k-1}^2)\times\int\mathcal{N}_{\mathcal{C}}\left(\mathbf{x}_k^2:\mathbf{F}^2\mathbf{x}_{k-1}^2,\mathbf{Q}^2\right)\times\mathcal{N}_{\mathcal{C}}\left(\mathbf{x}_{k-1}^2:\hat{\mathbf{x}}_{k-1|k-1}^2(s_{k-1}^2),\mathbf{P}_{k-1|k-1}^2(s_{k-1}^2)\right)d\mathbf{x}_{k-1}^2.$$

The integral appearing in the previous equation is the well-known prediction step of Kalman filtering [36], therefore

$$\mu_{h_k^2\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\propto\sum_{s_{k-1}^2}p(s_k^2|s_{k-1}^2)\alpha_{k-1|k-1}(s_{k-1}^2)\times\mathcal{N}_{\mathcal{C}}\left(\mathbf{x}_k^2:\hat{\mathbf{x}}_{k|k-1}^2(s_{k-1}^2),\mathbf{P}_{k|k-1}^2(s_{k-1}^2)\right), \quad (21)$$

where

$$\begin{cases}\hat{\mathbf{x}}_{k|k-1}^2(s_{k-1}^2)=\mathbf{F}^2\hat{\mathbf{x}}_{k-1|k-1}^2(s_{k-1}^2)\\ \mathbf{P}_{k|k-1}^2(s_{k-1}^2)=\mathbf{F}^2\mathbf{P}_{k-1|k-1}^2(s_{k-1}^2)\mathbf{F}^{2\,H}+\mathbf{Q}^2.\end{cases}$$

In order to avoid an exponential complexity increase with time, we collapse the Gaussian mixture (21) to a single Gaussian of the form (see Section III-A)

$$\mu_{h_k^2\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\propto\alpha_{k|k-1}(s_k^2)\times\mathcal{N}_{\mathcal{C}}\left(\mathbf{x}_k^2:\hat{\mathbf{x}}_{k|k-1}^2(s_k^2),\mathbf{P}_{k|k-1}^2(s_k^2)\right), \quad (22)$$

where the predicted belief of s_k^2 is given by $\alpha_{k|k-1}(s_k^2)=\sum_{s_{k-1}^2}p(s_k^2|s_{k-1}^2)\alpha_{k-1|k-1}(s_{k-1}^2)$ and

$$\begin{cases}\hat{\mathbf{x}}_{k|k-1}^2(s_k^2)\\ =\sum\limits_{s_{k-1}^2}\dfrac{p(s_k^2|s_{k-1}^2)\alpha_{k-1|k-1}(s_{k-1}^2)}{\alpha_{k|k-1}(s_k^2)}\hat{\mathbf{x}}_{k|k-1}^2(s_{k-1}^2)\\ \mathbf{P}_{k|k-1}^2(s_k^2)\\ =\sum\limits_{s_{k-1}^2}\dfrac{p(s_k^2|s_{k-1}^2)\alpha_{k-1|k-1}(s_{k-1}^2)}{\alpha_{k|k-1}(s_k^2)}\Bigg[\mathbf{P}_{k|k-1}^2(s_{k-1}^2)\\ \qquad+\left(\hat{\mathbf{x}}_{k|k-1}^2(s_{k-1}^2)-\hat{\mathbf{x}}_{k|k-1}^2(s_k^2)\right)\\ \qquad\times\left(\hat{\mathbf{x}}_{k|k-1}^2(s_{k-1}^2)-\hat{\mathbf{x}}_{k|k-1}^2(s_k^2)\right)^H\Bigg].\end{cases}$$

Now, applying the sum-product rule at the variable node $(s_k^2,\mathbf{x}_k^2)$, we get

$$\mu_{s_k^2,\mathbf{x}_k^2\to h_{k+1}^2}(s_k^2,\mathbf{x}_k^2)\propto\mu_{h_k^2\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\mu_{g_k\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2).$$

Injecting (19) and (22) into this expression, we obtain a product of Gaussians corresponding to the correction step of the well-known Kalman filter [36], which can be rewritten as

$$\mu_{s_k^2,\mathbf{x}_k^2\to h_{k+1}^2}(s_k^2,\mathbf{x}_k^2)\propto\alpha_{k|k}(s_k^2)\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2:\hat{\mathbf{x}}_{k|k}^2(s_k^2),\mathbf{P}_{k|k}^2(s_k^2)), \quad (23)$$

where

$$\begin{cases}\mathbf{K}_k^2(s_k^2)=\mathbf{P}_{k|k-1}^2(s_k^2)\mathbf{H}_k^2(s_k^2)^H\\ \qquad\times\left(\mathbf{H}_k^2(s_k^2)\mathbf{P}_{k|k-1}^2(s_k^2)\mathbf{H}_k^2(s_k^2)^H+\mathbf{S}_k^1\right)^{-1}\\ \hat{\mathbf{x}}_{k|k}^2(s_k^2)=\hat{\mathbf{x}}_{k|k-1}^2(s_k^2)\\ \qquad+\mathbf{K}_k^2(s_k^2)\left(\mathbf{y}_k-\mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k|k-1}^2(s_k^2)-\mathbf{b}_k^1\right)\\ \mathbf{P}_{k|k}^2(s_k^2)=\mathbf{P}_{k|k-1}^2(s_k^2)-\mathbf{K}_k^2(s_k^2)\mathbf{H}_k^2(s_k^2)\mathbf{P}_{k|k-1}^2(s_k^2)\\ \alpha_{k|k}(s_k^2)=\alpha_{k|k-1}(s_k^2)\\ \qquad\times\mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k:\mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k|k-1}^2(s_k^2)+\mathbf{b}_k^1,\\ \qquad\qquad\mathbf{H}_k^2(s_k^2)\mathbf{P}_{k|k-1}^2(s_k^2)\mathbf{H}_k^2(s_k^2)^H+\mathbf{S}_k^1\Big).\end{cases}$$

2) Backward Pass: In the same way, the update rule for the backward messages parameterized by

$$\mu_{h_{k+1}^2\to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\propto\beta_{k|k+1:K}(s_k^2)\mathcal{N}_{\mathcal{C}}\left(\mathbf{x}_k^2:\hat{\mathbf{x}}_{k|k+1:K}^2(s_k^2),\mathbf{P}_{k|k+1:K}^2(s_k^2)\right), \quad (24)$$

and

$$\mu_{s_k^2,\mathbf{x}_k^2\to h_k^2}(s_k^2,\mathbf{x}_k^2)\propto\beta_{k|k:K}(s_k^2)\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2:\hat{\mathbf{x}}_{k|k:K}^2(s_k^2),\mathbf{P}_{k|k:K}^2(s_k^2)), \quad (25)$$

are obtained using a time-reversed application of the operations of the forward pass. The details are therefore omitted.

3) Message Sent Back to g_k: Applying the sum-product rule to the variable node $(s_k^2, \mathbf{x}_k^2)$,

$$\mu_{s_k^2,\mathbf{x}_k^2 \to g_k}(s_k^2,\mathbf{x}_k^2) \propto \mu_{h_k^2 \to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)\mu_{h_{k+1}^2 \to s_k^2,\mathbf{x}_k^2}(s_k^2,\mathbf{x}_k^2)$$

Injecting (22) and (24) into the previous expression, we obtain

$$\mu_{s_k^2,\mathbf{x}_k^2 \to g_k}(s_k^2,\mathbf{x}_k^2) \propto \sigma_{k\backslash k}(s_k^2)\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2 : \hat{\mathbf{x}}_{k\backslash k}^2(s_k^2), \mathbf{P}_{k\backslash k}^2(s_k^2)), \tag{26}$$

where

$$\begin{cases} \mathbf{P}_{k\backslash k}^2(s_k^2) = \mathbf{P}_{k|k+1:K}^2(s_k^2) \\ \qquad \times \left[\mathbf{P}_{k|k-1}^2(s_k^2) + \mathbf{P}_{k|k+1:K}^2(s_k^2)\right]^{-1} \mathbf{P}_{k|k-1}^2(s_k^2) \\ \hat{\mathbf{x}}_{k\backslash k}^2(s_k^2) = \mathbf{P}_{k\backslash k}^2(s_k^2)\Big[\mathbf{P}_{k|k-1}^2(s_k^2)^{-1}\hat{\mathbf{x}}_{k|k-1}^2(s_k^2) \\ \qquad + \mathbf{P}_{k|k+1:K}^2(s_k^2)^{-1}\hat{\mathbf{x}}_{k|k+1:K}^2(s_k^2)\Big] \\ \sigma_{k\backslash k}(s_k^2) = \alpha_{k|k-1}(s_k^2)\beta_{k|k+1:K}(s_k^2) \\ \qquad \times \dfrac{1}{\pi^{nm}\det\left(\mathbf{P}_{k|k-1}^2(s_k^2) + \mathbf{P}_{k|k+1:K}^2(s_k^2)\right)} \\ \qquad \times \exp\Big[-(\hat{\mathbf{x}}_{k|k-1}^2(s_k^2) - \hat{\mathbf{x}}_{k|k+1:K}^2(s_k^2))^H \\ \qquad\qquad \times \left(\mathbf{P}_{k|k-1}^2(s_k^2) + \mathbf{P}_{k|k+1:K}^2(s_k^2)\right)^{-1} \\ \qquad\qquad \times (\hat{\mathbf{x}}_{k|k-1}^2(s_k^2) - \hat{\mathbf{x}}_{k|k+1:K}^2(s_k^2))\Big]. \end{cases}$$

The demonstration proceeds from a straightforward application of the formula in Appendix A.

B. n_1 Channel Estimation and Demodulation and Decoding

The n_1 channel estimation subgraphs in Fig. 4 and Fig. 5 are identical. Therefore, the n_1 channel estimation in the heterogeneous scenario is the procedure already described in Section III-B, except that $\mu_{g_k \to \mathbf{x}_k^1}(\mathbf{x}_k^1)$ must be recalculated. Let us first apply the sum-product rule at the function node g_k. We obtain

$$\mu_{g_k \to \mathbf{x}_k^1}(\mathbf{x}_k^1) \propto \sum_{\mathbf{d}_k^1}\sum_{s_k^2} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1) \times \int p\left(\mathbf{y}_k|\mathbf{d}_k^1, \mathbf{x}_k^1, s_k^2, \mathbf{x}_k^2\right) \mu_{s_k^2,\mathbf{x}_k^2 \to g_k}(s_k^2,\mathbf{x}_k^2) d\mathbf{x}_k^2.$$

According to (3) and (26), this expression becomes

$$\begin{aligned} \mu_{g_k \to \mathbf{x}_k^1}(\mathbf{x}_k^1) &\propto \sum_{\mathbf{d}_k^1}\sum_{s_k^2} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)\sigma_{k\backslash k}(s_k^2) \\ &\quad \times \int \mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k : \mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1 + \mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2, \mathbf{R}\right) \\ &\qquad \times \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2 : \hat{\mathbf{x}}_{k\backslash k}^2(s_k^2), \mathbf{P}_{k\backslash k}^2(s_k^2)) d\mathbf{x}_k^2 \\ &\propto \sum_{\mathbf{d}_k^1}\sum_{s_k^2} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)\sigma_{k\backslash k}(s_k^2) \\ &\quad \times \mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k : \mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1 + \mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k\backslash k}^2(s_k^2), \\ &\qquad \mathbf{H}_k^2(s_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(s_k^2)^H + \mathbf{R}\Big). \end{aligned} \tag{27}$$

where the closed form expression of the integral in the last expression has been demonstrated in [29] (p. 38). In order to obtain a tractable expression, we collapse the obtained Gaussian mixture to a single Gaussian using the moment-matching method of Section III-A. We obtain

$$\mu_{g_k \to \mathbf{x}_k^1}(\mathbf{x}_k^1) \propto \mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k : \hat{\mathbf{H}}_k^1\mathbf{x}_k^1 + \mathbf{b}_k^2, \mathbf{S}_k^2\right). \tag{28}$$

where

$$\begin{cases} \hat{\mathbf{H}}_k^1 = \displaystyle\sum_{\mathbf{d}_k^1} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)\mathbf{H}_k^1(\mathbf{d}_k^1) \\ \mathbf{b}_k^2 = \displaystyle\sum_{s_k^2} \sigma_{k\backslash k}(s_k^2)\mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k\backslash k}^2(s_k^2). \end{cases}$$

and

$$\begin{aligned} \mathbf{S}_k^2 = &\sum_{\mathbf{d}_k^1} \mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)\left(\mathbf{H}_k^1(\mathbf{d}_k^1) - \hat{\mathbf{H}}_k^1\right)\left(\mathbf{H}_k^1(\mathbf{d}_k^1) - \hat{\mathbf{H}}_k^1\right)^H \\ &+ \sum_{s_k^2} \sigma_{k\backslash k}(s_k^2)\Big\{\mathbf{H}_k^2(s_k^2)\mathbf{P}_{k\backslash k}^2(s_k^2)\mathbf{H}_k^2(s_k^2)^H + \\ &\left(\mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k\backslash k}^2(s_k^2) - \mathbf{b}_k^2\right)\left(\mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k\backslash k}^2(s_k^2) - \mathbf{b}_k^2\right)^H\Big\} \\ &+ \mathbf{R}. \end{aligned}$$

The proof is omitted, since it follows exactly the same steps as in Appendix B.

Remark 4.2: The n_1 to relay channel estimation needs an approximation of the likelihood $p(\mathbf{y}_k|\mathbf{x}_k^1)$, by averaging out the variables $\mathbf{d}_k^1$, s_k^2 and $\mathbf{x}_k^2$. We obtain the Gaussian distribution (28) with mean $\hat{\mathbf{H}}_k^1\mathbf{x}_k^1 + \mathbf{b}_k^2$, where $\hat{\mathbf{H}}_k^1$ is the observation matrix averaged over the modulated vectors $\mathbf{d}_k^1$ and $\mathbf{b}_k^2$ is a bias which accounts for the superimposed n_2 to relay transmission. The covariance matrix $\mathbf{S}_k^2$ accounts for the presence of AWGN, residual interference from the superimposed n_2 to relay transmission and channel estimation uncertainty.

The demodulation of the symbol vectors sent by n_1 also needs to be modified. Applying the sum-product rule at the function node g_k,

$$\begin{aligned} &\mu_{g_k \to \mathbf{d}_k^1}(\mathbf{d}_k^1) \\ &= \sum_{s_k^2} \int\int p\left(\mathbf{y}_k|\mathbf{d}_k^1, \mathbf{x}_k^1, s_k^2, \mathbf{x}_k^2\right) \\ &\qquad \times \mu_{\mathbf{x}_k^1 \to g_k}(\mathbf{x}_k^1)\mu_{s_k^2,\mathbf{x}_k^2 \to g_k}(s_k^2,\mathbf{x}_k^2) d\mathbf{x}_k^1 d\mathbf{x}_k^2 \\ &= \sum_{s_k^2} \sigma_{k\backslash k}(s_k^2) \\ &\quad \times \mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k : \mathbf{H}_k^1(\mathbf{d}_k^1)\hat{\mathbf{x}}_{k\backslash k}^1 + \mathbf{H}_k^2(s_k^2)\hat{\mathbf{x}}_{k\backslash k}^2(s_k^2), \\ &\quad \mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{P}_{k\backslash k}^1\mathbf{H}_k^1(\mathbf{d}_k^1)^H + \mathbf{H}_k^2(s_k^2)\mathbf{P}_{k\backslash k}^2(s_k^2)\mathbf{H}_k^2(s_k^2)^H + \mathbf{R}\Big). \end{aligned} \tag{29}$$

The demonstration is similar to Appendix C.

Finally, the channel decoding method is unchanged with respect to that of Section III-D.

C. Initialization

During the first iteration, the messages $\mu_{s_k^2,\mathbf{x}_k^2 \to g_k}(s_k^2,\mathbf{x}_k^2)$ (corresponding to the joint STTC state and MIMO channel distribution for n_2) and $\mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)$ (corresponding to the n_1 symbol vector distribution) must be initialized. We use the prior pmf for $\mu_{\mathbf{d}_k^1 \to g_k}(\mathbf{d}_k^1)$ which is the uniform pmf,

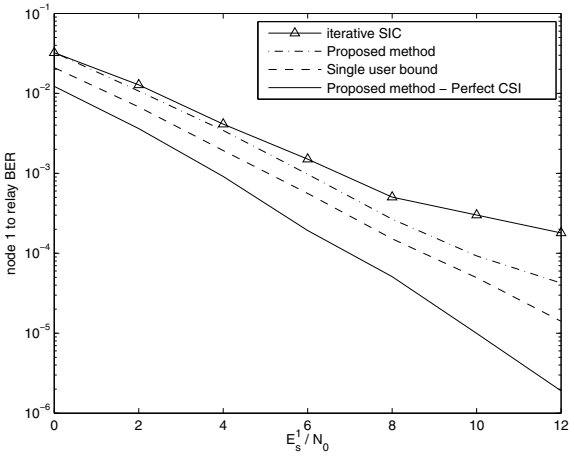

Fig. 7. BER for the packets sent by n_1 after 3 iterations: $G = 0$ dB, $f_m^1 T = f_m^2 T = 10^{-3}$ and $K = 100$.

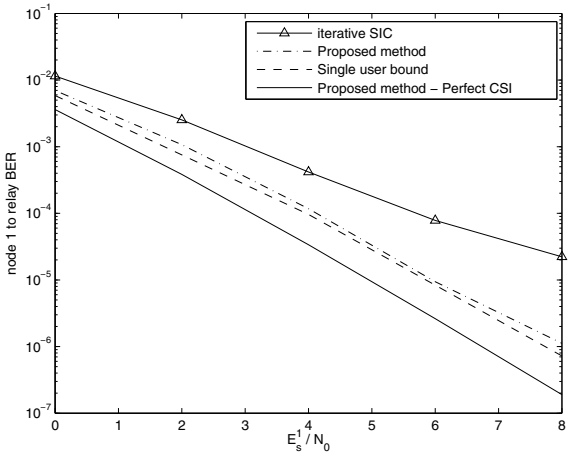

Fig. 8. BER for the packets sent by n_1 after 3 iterations: $G = 0$ dB, $f_m^1 T = f_m^2 T = 10^{-3}$ and $K = 1000$.

since the information packets at the source node n_1 have been assumed u.i.i.d. Similarly for $\mu_{s_k^2, \mathbf{x}_k^2 \to g_k}(s_k^2, \mathbf{x}_k^2)$, we use the prior mixed discrete-continuous distribution $p(s_k^2)p(\mathbf{x}_k^2)$, where $p(s_k^2)$ is the uniform pmf over the STTC states and

$$p(\mathbf{x}_k^2) = \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^2 : \mathbf{0}_{mn}, \mathbf{I}_{mn}), \forall k$$

according to the MIMO channel model in (5).

V. Simulation Results

The following simulation parameters are used for all the scenarios (homogeneous and heterogenerous) and all the relay node processing algorithms (proposed method, iterative SIC and single-user bound). The source nodes are equipped with $n = 2$ transmit antennas and the relay node with $m = 2$ receive antennas. The source nodes employ the data format of Fig. 1 and binary phase shift keying (BPSK) modulation. The pilot insertion rate (PIR) is fixed to $1 : 10$, so due to the insertion of an additional zero symbol after each pilot symbol, 20 percent of the data rate is lost as a consequence of pilot insertion.

Unless otherwise specified, we assume perfect timing synchronization and perfect power control, i.e. $G = 0$ dB, at the relay node. For simplicity, we also assume that the n_1-to-relay and n_2-to-relay MIMO channels have identical normalized fading rates, i.e. $f_m^1 T = f_m^2 T$. This situation corresponds for instance to n_1 and n_2 being fixed and r moving at a given constant maximal velocity. Moreover, the normalized fading rates are known to the relay node. The time-varying Rayleigh fading MIMO channels are simulated with Jakes Doppler spectrum using the method described in [37]. At the relay side, the MIMO channels are modeled as simple AR(1) processes, introduced in Section II-B.

A. Homogeneous Scenario

In the homogeneous scenario, n_1 and n_2 use BICM. The convolutional codes CC1 and CC2 are identical and chosen as the rate-$1/2$ recursive systematic convolutional code (RSC) with polynomials $(1, 5/7)$ in octal representation. However

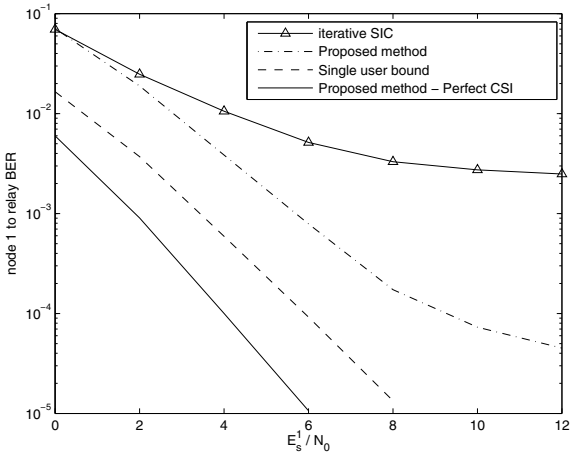

Fig. 9. BER for the packets sent by n_1 after 3 iterations: $G = 0$ dB, $f_m^1 T = f_m^2 T = 0.005$ and $K = 100$.

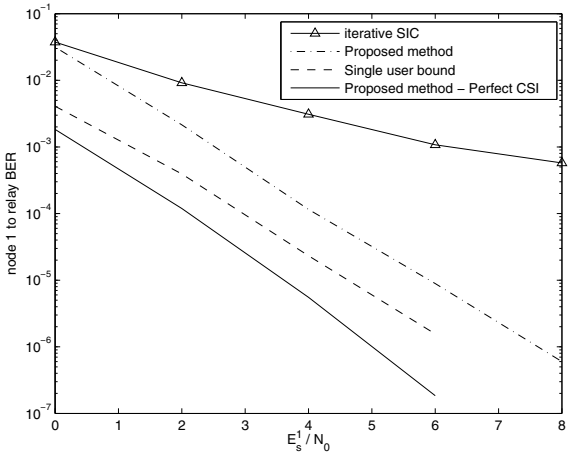

Fig. 10. BER for the packets sent by n_1 after 3 iterations: $G = 0$ dB, $f_m^1 T = f_m^2 T = 0.005$ and $K = 1000$.

π_1 and π_2 are different interleavers picked at random. The

proposed message passing receiver (see Section III), iterative SIC (see Section III-F) and the single user bound (see Section III-F) perform three iterations. No performance improvements were observed by further increasing the number of iterations. We obtain the proposed message passing receiver, with perfect channel state information (CSI) by removing the variable nodes corresponding to MIMO channel gains in the factor graph of Fig. 4, and applying message passing, assuming that all channel gains are perfectly known to the relay.

Fig. 7 (resp. Fig. 8) show the bit-error rate (BER) performances for the packets sent by n_1, for a block length of $K = 100$ (resp. $K = 1000$) at a normalized fading rate of 10^{-3}. These curves allow us to study the influence of the interleaver length, which is proportional to the block length K.

First, we observe that for a given signal-to-noise ratio (SNR), the BER is much lower for $K = 1000$ than for $K = 100$, which can be explained by additional temporal diversity gains obtained for long interleavers. Moreover, the proposed method reaches performances close to the single-user bound, regardless of the SNR, when $K = 1000$. However, the proposed method is between 1 and 2 dB less power efficient than the single-user bound, when $K = 100$. This reveals that the proposed receiver solves the multiple access problem in a very efficient way, provided that the interleaver is long enough. This phenomenon can be interpreted by the fact that the SPA works well only if short cycle events are avoided in the factor graph of Fig. 4 [23], i.e. when long interleavers are used.

At high SNR, iterative SIC, due to the presence of residual interference from user node n_2, performs worse than the proposed method. Similar BER curves (although not shown) were obtained for the packets sent by n_2.

In order to study the robustness of the channel estimation, we choose a two-way relay channel with faster time-variations, while keeping the same amount of pilot overhead. Fig. 9 (resp. Fig. 10) show the BER performances for the packets sent by n_1, for a block length of $K = 100$ (resp. $K = 1000$) at a normalized fading rate of 0.005. The proposed method reaches an error floor for $K = 100$, which disappears for $K = 1000$. However, the error floor of iterative SIC does not disappear for large block lengths, due to residual multiple access interference. Similar BER curves were also obtained for the packets sent by n_2.

B. Heterogeneous Scenario

In the heterogeneous scenario, user node n_1 uses the same BICM scheme as in the homogeneous scenario. User node n_2 uses the algebraic full-diversity STTC for BPSK modulation with encoding polynomials $(1, 5/7)$ in octal representation [19]. We apply the proposed message passing receiver described in Section IV and a genie-aided receiver with perfect CSI (obtained by removing the variable nodes corresponding to MIMO channel gains in the factor graph of Fig. 5, and applying message passing, assuming that all channel gains are perfectly known to the relay). The BER results for the packets sent by n_1 (resp. n_2) are shown in Fig. 11 (resp. Fig. 12), for a block length of $K = 1000$, a normalized fading rate of 10^{-3} and a relative path-loss gain of $G = -3$ dB. Satisfactory results are obtained after 3 iterations and no further improvements could be observed by augmenting the number of iterations.

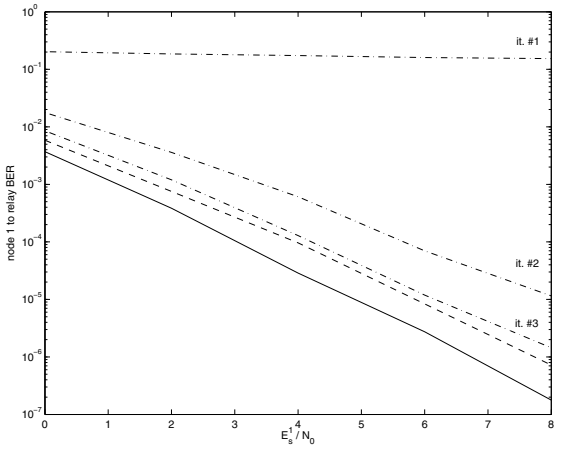

Fig. 11. BER for the packets sent by n_1: $G = -3$ dB, $f_m^1 T = f_m^2 T = 10^{-3}$ and $K = 1000$. Solid curve: proposed method with perfect CSI after 3 iterations - Dashed curve: single-user bound with unknown two-way MIMO channel after 3 iterations - Dashed dotted curves: proposed method with unknown two-way MIMO channel for the 3 first iterations.

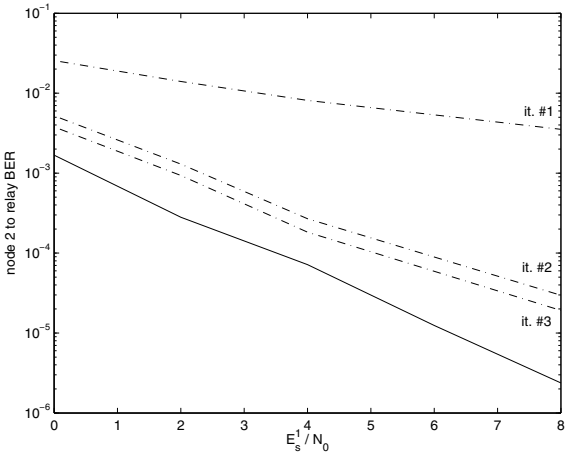

Fig. 12. BER for the packets sent by n_2: $G = -3$ dB, $f_m^1 T = f_m^2 T = 10^{-3}$ and $K = 1000$. Solid curve: proposed method with perfect CSI after 3 iterations - Dashed dotted curves: proposed method with unknown two-way MIMO channel for the 3 first iterations.

VI. Conclusion

In this paper, physical layer network coding using the joint decode-and-forward scheme, was considered. Recovering the messages sent by both source nodes during the first phase is a challenging multiple access problem, especially in the presence of an unknown time-varying MIMO two-way relay channel. Based on a factor graph representation of the problem at hand, a soft-output message passing solution to the problem of joint channel estimation and decoding at the relay node was introduced. The key feature of the proposed

approach is the approximation of Gaussian mixture messages by single-Gaussian messages, using moment-matching. This keeps the computational complexity at an acceptable level, while preserving excellent performances.

It is shown through numerical simulations that the proposed receiver significantly outperforms the conventional iterative serial interference canceller, with a reasonable amount of pilot symbols. Moreover, as long as the interleaver is long enough, the performances of the proposed algorithm stay close to the single user bound.

Future extensions of this work include the consideration of additional channel impairments, such as phase noise and frequency offsets, and the introduction of more powerful coding schemes, such as graph-based codes.

APPENDIX A

The following equality holds for a vector $\mathbf{x}$ of size l

$$\begin{aligned}
&\mathcal{N}_{\mathcal{C}}(\mathbf{x}:\mathbf{x}_1,\mathbf{P}_1)\times\mathcal{N}_{\mathcal{C}}(\mathbf{x}:\mathbf{x}_2,\mathbf{P}_2)\\
=&\mathcal{N}_{\mathcal{C}}\Big(\mathbf{x}:\mathbf{P}_2\left[\mathbf{P}_1+\mathbf{P}_2\right]^{-1}\mathbf{P}_1\left(\mathbf{P}_1^{-1}\mathbf{x}_1+\mathbf{P}_2^{-1}\mathbf{x}_2\right),\\
&\qquad\mathbf{P}_2\left[\mathbf{P}_1+\mathbf{P}_2\right]^{-1}\mathbf{P}_1\Big)\\
&\times\frac{1}{\pi^l\det(\mathbf{P}_1+\mathbf{P}_2)}\\
&\times\exp\left\{-(\mathbf{x}_1-\mathbf{x}_2)^H\left[\mathbf{P}_1+\mathbf{P}_2\right]^{-1}(\mathbf{x}_1-\mathbf{x}_2)\right\}.
\end{aligned}$$

APPENDIX B
GAUSSIAN MIXTURE REDUCTION FORMULA (14)

We first derive the parameters of the Gaussian density $\mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k:\mathbf{m}_k(\mathbf{x}_k^1),\mathbf{S}_k(\mathbf{x}_k^1)\right)$, having the same mean and covariance as the Gaussian mixture (13).

According to the moment-matching method of Section III-A, we have

$$\begin{aligned}
&\mathbf{m}_k(\mathbf{x}_k^1)\\
=&\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\qquad\times\left[\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2\right]\\
&\mathbf{S}_k(\mathbf{x}_k^1)\\
=&\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\quad\times\Big[\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H+\mathbf{R}\\
&\qquad+\left(\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2-\mathbf{m}_k(\mathbf{x}_k^1)\right)\\
&\qquad\times\left(\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2-\mathbf{m}_k(\mathbf{x}_k^1)\right)^H\Big].
\end{aligned}$$

Using the distributivity of multiplication over addition, $\mathbf{m}_k(\mathbf{x}_k^1)$ can also be written as

$$\begin{aligned}
&\mathbf{m}_k(\mathbf{x}_k^1)\\
=&\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1\\
&+\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2\\
=&\left(\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1\right)\left(\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\right)\\
&+\left(\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\backslash k}^2\right)\left(\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\right).
\end{aligned}$$

If the messages are properly normalized, i.e. $\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)=1$ and $\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)=1$, the desired result follows, that is

$$\mathbf{m}_k(\mathbf{x}_k^1)=\hat{\mathbf{H}}_k^1\mathbf{x}_k^1+\hat{\mathbf{H}}_k^2\hat{\mathbf{x}}_{k\backslash k}^2.$$

Injecting this result in the expression of the covariance matrix $\mathbf{S}_k(\mathbf{x}_k^1)$, and using the distributivity of multiplication over addition we obtain

$$\begin{aligned}
&\mathbf{S}_k(\mathbf{x}_k^1)\\
=&\sum_{\mathbf{d}_k^1}\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\quad\times\left[\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)\mathbf{x}_k^1+\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)\hat{\mathbf{x}}_{k\backslash k}^2\right]\\
&\qquad\times\left[\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)\mathbf{x}_k^1+\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)\hat{\mathbf{x}}_{k\backslash k}^2\right]^H\\
&+\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H\\
&+\mathbf{R}
\end{aligned}$$

The cross terms of the double summation are equal to zero, due to the independence of $\mathbf{d}_k^1$ and $\mathbf{d}_k^2$. Therefore, we obtain the following simplification

$$\begin{aligned}
&\mathbf{S}_k(\mathbf{x}_k^1)\\
=&\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)\mathbf{x}_k^1\mathbf{x}_k^{1\,H}\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)^H\\
&+\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\Big\{\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\backslash k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H\\
&\qquad+\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)\hat{\mathbf{x}}_{k\backslash k}^2\hat{\mathbf{x}}_{k\backslash k}^{2\,H}\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)^H\Big\}\\
&+\mathbf{R}.
\end{aligned}$$

We average out $\mathbf{x}_k^1$, whose prior density is

$$p(\mathbf{x}_k^1)=\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\mathbf{0}_{mn},\mathbf{I}_{mn}),\forall k$$

according to the MIMO channel model in (5). So the final

expression of the covariance matrix becomes

$$\begin{aligned}
&\mathbf{S}_k^2\\
=&E\left[\mathbf{S}_k(\mathbf{x}_k^1)\right]\\
=&\int \mathbf{S}_k(\mathbf{x}_k^1)p(\mathbf{x}_k^1)d\mathbf{x}_k^1\\
=&\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)\\
&\qquad\times\left(\int \mathbf{x}_k^1{\mathbf{x}_k^1}^H p(\mathbf{x}_k^1)d\mathbf{x}_k^1\right)\left(\mathbf{H}_k^1(\mathbf{d}_k^1)-\hat{\mathbf{H}}_k^1\right)^H\\
&+\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\Big\{\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\setminus k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H\\
&\qquad+\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)\hat{\mathbf{x}}_{k\setminus k}^2{\hat{\mathbf{x}}_{k\setminus k}^2}^H\left(\mathbf{H}_k^2(\mathbf{d}_k^2)-\hat{\mathbf{H}}_k^2\right)^H\Big\}\\
&+\mathbf{R}.
\end{aligned}$$

which is the desired result since

$$\int \mathbf{x}_k^1{\mathbf{x}_k^1}^H\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\mathbf{0}_{mn},\mathbf{I}_{mn})d\mathbf{x}_k^1=\mathbf{I}_{mn}.$$

APPENDIX C
DEMODULATION FORMULA (17)

Eq. (17) can be rewritten as

$$\begin{aligned}
&\mu_{g_k\to\mathbf{d}_k^1}(\mathbf{d}_k^1)\\
=&\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\qquad\times\int\left[\int p\left(\mathbf{y}_k|\mathbf{d}_k^1,\mathbf{x}_k^1,\mathbf{d}_k^2,\mathbf{x}_k^2\right)\mu_{\mathbf{x}_k^2\to g_k}(\mathbf{x}_k^2)d\mathbf{x}_k^2\right]\\
&\qquad\qquad\times\mu_{\mathbf{x}_k^1\to g_k}(\mathbf{x}_k^1)d\mathbf{x}_k^1.
\end{aligned}$$

The inner integral already appeared in (13), so we have

$$\begin{aligned}
&\mu_{g_k\to\mathbf{d}_k^1}(\mathbf{d}_k^1)\\
=&\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\qquad\times\int\mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k:\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\setminus k}^2,\\
&\qquad\qquad\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\setminus k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H+\mathbf{R}\Big)\\
&\qquad\qquad\times\mu_{\mathbf{x}_k^1\to g_k}(\mathbf{x}_k^1)d\mathbf{x}_k^1.
\end{aligned}$$

Now injecting (16) in the previous result, we obtain

$$\begin{aligned}
&\mu_{g_k\to\mathbf{d}_k^1}(\mathbf{d}_k^1)\\
=&\sum_{\mathbf{d}_k^2}\mu_{\mathbf{d}_k^2\to g_k}(\mathbf{d}_k^2)\\
&\qquad\times\int\mathcal{N}_{\mathcal{C}}\Big(\mathbf{y}_k:\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{x}_k^1+\mathbf{H}_k^2(\mathbf{d}_k^2)\hat{\mathbf{x}}_{k\setminus k}^2,\\
&\qquad\qquad\mathbf{H}_k^2(\mathbf{d}_k^2)\mathbf{P}_{k\setminus k}^2\mathbf{H}_k^2(\mathbf{d}_k^2)^H+\mathbf{R}\Big)\\
&\qquad\qquad\times\mathcal{N}_{\mathcal{C}}(\mathbf{x}_k^1:\hat{\mathbf{x}}_{k\setminus k}^1,\mathbf{P}_{k\setminus k}^1)d\mathbf{x}_k^1.
\end{aligned}$$

The integral in the previous expression can be expressed in closed form (see [29] - p. 38). We obtain the desired final expression in (17).

APPENDIX D
GAUSSIAN MIXTURE REDUCTION FORMULA (19)

We derive the parameters of the Gaussian density $\mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_k:\mathbf{m}_k(s_k^2,\mathbf{x}_k^2),\mathbf{S}_k^1\right)$, having the same mean and covariance as the Gaussian mixture (18).

According to the moment-matching method of Section III-A, we have

$$\mathbf{m}_k(s_k^2,\mathbf{x}_k^2)=\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\left[\mathbf{H}_k^1(\mathbf{d}_k^1)\hat{\mathbf{x}}_{k\setminus k}^1+\mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2\right]$$

$$\begin{aligned}
&\mathbf{S}_k^1\\
&=\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)\\
&\qquad\times\Big[\mathbf{H}_k^1(\mathbf{d}_k^1)\mathbf{P}_{k\setminus k}^1\mathbf{H}_k^1(\mathbf{d}_k^1)^H+\mathbf{R}\\
&\qquad+\left(\mathbf{H}_k^1(\mathbf{d}_k^1)\hat{\mathbf{x}}_{k\setminus k}^1+\mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2-\mathbf{m}_k(s_k^2,\mathbf{x}_k^2)\right)\\
&\qquad\qquad\times\left(\mathbf{H}_k^1(\mathbf{d}_k^1)\hat{\mathbf{x}}_{k\setminus k}^1+\mathbf{H}_k^2(s_k^2)\mathbf{x}_k^2-\mathbf{m}_k(s_k^2,\mathbf{x}_k^2)\right)^H\Big].
\end{aligned}$$

If the messages are properly normalized, i.e. $\sum_{\mathbf{d}_k^1}\mu_{\mathbf{d}_k^1\to g_k}(\mathbf{d}_k^1)=1$, the desired result is obtained.

REFERENCES

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[3] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. ACM MOBICOM*, Sep. 2006, pp. 358–365.

[4] S. Zhang and S. C. Liew, "Physical-layer network coding schemes over finite and infinite fields," in *Proc. IEEE Globecom*, Nov.–Dec 2008, pp. 1–6.

[5] P. Popovski and H. Yomo, "The anti-packets can increase the achievable throughput of a wireless multi-hop network," in *Proc. IEEE ICC*, June 2006, pp. 3885–3890.

[6] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 773–787, June 2009.

[7] P. Popovski and H. Yomo, "Bi-directional amplification of throughput in a wireless multihop network," in *Proc. IEEE VTC*, May 2006, pp. 588–593.

[8] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels," in *Proc. IEEE ICC*, June 2007, pp. 707–712.

[9] F. Rossetto and M. Zorzi, "Mixing network coding and cooperation for reliable wireless communications," *IEEE Trans. Wireless Commun.*, pp. 15–21, Feb. 2011.

[10] B. Talha and M. Pätzold, "Channel models for mobile-to-mobile cooperative communication systems: A state of the art review," *IEEE Veh. Technol. Mag.*, vol. 6, no. 2, pp. 33–43, June 2011.

[11] F. Gao, R. Zhang, and Y. C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.

[12] T.-H. Pham, Y.-C. Liang, A. Nallanathan, and H. K. Garg, "Optimal training sequences for channel estimation in bi-directional relay networks with multiple antennas," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 474–479, Feb. 2010.

[13] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation and training design for two-way relay networks in time-selective fading environments," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2681–2691, Aug. 2011.

[14] F. Gao, R. Zhang, and Y. C. Liang, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.

[15] T. Koike-Akino, P. Popovski, and V. Tarokh, "Denoising strategy for convolutionally encoded bidirectional relaying," in *Proc. IEEE ICC*, June 2009, pp. 1–5.

[16] D. To and J. Choi, "Convolutional codes in two-way relay networks with physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2724–2729, Sep. 2010.

[17] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.
[18] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[19] H. El Gamal and A. R. Hammons, "On the design and performance of algebraic space-time codes for BPSK and QPSK modulation," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 907–913, June 2002.
[20] E. Biglieri, G. Taricco, and E. Viterbo, "Bit-interleaved time-space codes for fading channels," in *Proc. CISS*, Mar. 2000.
[21] T. Wang and G. B. Giannakis, "Complex field network coding for multiuser cooperative communications," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 561–571, Apr. 2008.
[22] S. Yang and R. Koetter, "Network coding over a noisy relay: A belief propagation approach," in *Proc. ISIT*, June 2007, pp. 801–804.
[23] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graph and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.
[24] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
[25] F. Lehmann, "A joint decode-and-forward strategy for physical network coding based on factor graphs," in *Proc. IEEE WCNC*, Apr. 2012.
[26] S. Lin and D. J. Costello, *Error Control Coding*. Englewood Cliffs, NJ: Prentice Hall, 1983.
[27] G. L. Stüber, *Principles of Mobile Communications*. Norwell, MA: Kluwer Academic Publishers, 1999.
[28] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kaufmann, 1988.
[29] H. Tanizaki, *Nonlinear Filters: Estimation and Applications*. Berlin: Springer, 1996.
[30] D. C. Fraser and J. E. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Trans. Autom. Control*, vol. 14, pp. 387–390, Aug. 1969.
[31] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8-PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.
[32] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, pp. 284–287, Mar. 1974.
[33] F. Simoens and M. Moeneclaey, "Code-aided estimation and detection on time-varying correlated MIMO channels: A factor-graph approach," *EURASIP J. Applied Signal Process.*, pp. 1–11, 2006.
[34] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1450–1460, Aug. 2001.
[35] Y. Zhu, D. Guo, and M. L. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6008–6018, Dec. 2009.
[36] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.
[37] Y. Li and X. Huang, "The simulation of independent Rayleigh faders," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1503–1514, Sep. 2002.

Frederic Lehmann received the E.E. and the M.S.E.E. degrees from ENSERG, France, in 1998. In 2002, he received the Ph.D. in electrical engineering from the National Polytechnical Institute, Grenoble (INPG), France. He worked as a Research Engineer with STMicroelectronics from 1999 to 2002. From 2003 to 2004, he was a Post-Doctoral Researcher at LAAS (Laboratory for Analysis and Architecture of Systems), CNRS, Toulouse, France. Currently, he is an Assistant Professor at Institut TELECOM, Telecom SudParis, Evry, France. His main research interests are in the areas of communication theory, non-linear signal processing, and statistical image processing.