Coding and System Design for Quantize-Map-and-Forward Relaying

Vinayak Nagpal, *Member, IEEE,* I-Hsiang Wang, *Member, IEEE,* Milos Jorgovanovic, *Student Member, IEEE,* David Tse, *Fellow, IEEE,* and Borivoje Nikolić *Senior Member, IEEE*

***Abstract*—In this paper we develop a low-complexity coding scheme and system design framework for the half duplex relay channel based on the Quantize-Map-and-Forward (QMF) relaying scheme. The proposed framework allows linear complexity operations at all network terminals. We propose the use of binary LDPC codes for encoding at the source and LDGM codes for mapping at the relay. We express *joint decoding* at the destination as a belief propagation algorithm over a factor graph. This graph has the LDPC and LDGM codes as subgraphs connected via probabilistic constraints that model the QMF relay operations. We show that this coding framework extends naturally to the high SNR regime using bit interleaved coded modulation (BICM). We develop density evolution analysis tools for this factor graph and demonstrate the design of practical codes for the half-duplex relay channel that perform within 1dB of information theoretic QMF threshold.**

***Index Terms*—Relay channels, low density parity check (LDPC) codes, low density generator matrix (LDGM) codes, iterative decoding, modulation, interleaving, MIMO.**

I. Introduction

COOPERATIVE relaying has been proposed as a promising technique to resolve the increasing demand for data throughput in wireless networks. Recently a lot of progress has been made in establishing the theoretical foundations of cooperative communication. To apply these principles towards the design of practical wireless systems, various system design tradeoffs must be taken into consideration. This paper presents progress towards this goal. We propose a system design and coding framework for quantize-map-and-forward (QMF) [1] relaying that has low complexity and performs close to information theoretic bounds.

A. Cooperative Systems

A cooperative wireless link typically consists of an information source, a destination and one or more cooperating *half duplex* relays. The relays are usually assumed to operate *in-band*. i.e. no additional channel resources are allocated for cooperation. Without loss of generality, it is assumed that relays use time-division-duplexing i.e. they listen to transmission from the source for some fraction of total time, then forward a description of their observation in the remaining fraction.

Manuscript received 1 August 2011; revised 1 May 2012. The material in this paper was presented in part at the Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, USA, September 2010.

The authors are with the Department of EECS, University of California at Berkeley, Berkeley, California, 94720, USA (e-mail: vinayak.nagpal@nokia.com; i-hsiang.wang@epfl.ch; {milos,dtse,bora}@eecs.berkeley.edu).

Digital Object Identifier 10.1109/JSAC.2013.130807

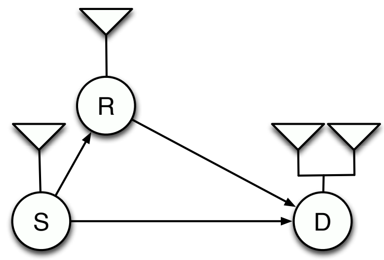


Fig. 1. Example relay network. With multiple antennas at destination, source-relay cooperation provides additional degrees of freedom for communication.

There are several aspects involved in the design of a cooperative relaying system. Listening fractions and forwarding schemes must be determined for each relay. Suitable modulation and channel coding schemes must be designed for various terminals. Rate adaptation mechanisms must be considered to account for changes in availability of relays and channel strengths. Practical constraints must be considered e.g. minimizing the overall system complexity, reuse of building blocks from traditional (non-cooperative) systems as much as possible, compatibility with protocols at higher layers and handling of system imperfections like synchronization, channel estimation errors etc. In this paper, we focus on the coding and signaling aspects of cooperative relaying. Other components are discussed in brief towards the end of the paper.

B. Relaying Schemes

Most wireless systems operate at moderate to high SNR i.e. in a regime where transmit power is not the major limiting factor for link capacity. At high SNR, the link capacity is limited by spatial degrees of freedom. Relay cooperation is of special interest for practical systems because it has the potential to provide additional spatial degrees of freedom. For illustration, consider a relay channel with single-antenna source, a half-duplex single antenna relay and a destination with two antennas shown in Fig. 1. Since the destination has multiple antennas, the source can spatially multiplex traffic to the destination using relay cooperation. If the source-to-relay channel is strong, this network can approach the high-SNR performance of the 2×2 MIMO channel [2].

Various strategies for relay cooperation are proposed in literature. Among these amplify-and-forward (AF), decode-and-forward (DF) and compress-and-forward (CF) [3], [4] have received the most attention. Under DF, the relay decodes the source's message and forwards a hard estimate of it, whereas under AF and CF it forwards a soft estimate without explicitly decoding it. In DF and CF, the relay maps its estimate to a random codeword before forwarding, whereas in AF it forwards an uncoded signal. The QMF scheme [1] also uses soft estimate forwarding with random coding similar to CF. For the example network in Fig. 1, the CF and QMF schemes are close to optimal at high SNR. In fact they achieve within one bit/sec/Hz to the information-theoretic capacity [1], [5]. An intuitive explanation for why QMF/CF performs better than both AF and DF is given in the context of the example in Fig. 1 below.

In the example network of Fig. 1, the destination receives continuously from the source. The relay receives a stronger version, when it is listening. Since the destination has two antennas, it can resolve simultaneous transmissions from the source and relay. In order to achieve spatial multiplexing, the relay should extract the less significant bits (from its observation), which the destination cannot resolve, and forward them. Under AF, the relay only forwards the more significant bits, which the destination can already resolve. Therefore AF cooperation provides limited benefit. Under DF, the relay decodes the entire message before it forwards anything. Since the listening time is limited, this approach is inefficient. The QMF/CF schemes implicitly extract the less significant bits from the relay's observation by using quantization/compression and random mapping. Therefore these schemes provide the most cooperation gain.

Despite having similar performance for the single relay network, the CF and QMF schemes have significant differences. In the conventional CF scheme, the relay compresses its observed signal and performs a random code mapping before forwarding. The compression rate is chosen in order for the destination to perform *two-step decoding* i.e. first decode compressed signals from relay and then use it as side information to decode the message from source. For configurations that involve multiple relays, two-step decoding is sub-optimal and conventional CF is not within bounded gap from information-theoretic capacity [1]. Even for the single-relay configuration, conventional CF requires that the relay have full knowledge of the quality of its forward channel. This introduces a large estimation and feedback overhead for fading channels and increases the complexity of rate adaptation schemes.

Under QMF, the relay quantizes its received signal at noise level, randomly maps it to a codeword and forwards it. Unlike CF, the quantization and mapping is performed without regard to the quality of forward channel at the relay. This reduces the channel estimation and feedback overhead for the link. It also simplifies rate adaptation protocols. Additionally, QMF uses *joint decoding* (as opposed to successive decoding) and performs within bounded gap from capacity for networks having an arbitrary number of relays [1]. QMF has played a key role in several recent information theoretic results on cooperative networks [2], [5]–[7]. Due to these favorable properties, the QMF scheme is superior to CF from the perspective of practical cooperative systems.

Since mapping at a QMF relay is performed without any knowledge of forward channel strength, side information from relays cannot be decoded at the destination independently. QMF requires *joint decoding* of the message (from source) and side information (from relays) [1]. This presents a unique challenge because joint decoding typically requires higher complexity and makes it harder to design a practical cooperative coding scheme. The key contribution of this paper is to develop a low-complexity cooperative coding framework for QMF that significantly reduces the complexity of joint decoding and yet performs close to information theoretic bounds.

C. Related Work

Majority of previous work on code design for cooperative relaying is focused on the DF scheme. DF relays fully decode the source's message. Therefore, DF coding schemes involve partitioning a large codebook into two parts. The source transmits one part of the codeword and the relay transmits the remaining part [8]–[10]. Turbo code designs which perform $\approx$ 1dB away from the DF information theoretic threshold are demonstrated in [11], [12]. LDPC profiles are developed for DF in [13]. A bilayer LDPC structure [14] and the protograph method [15] has been used to get LDPC designs ≤ 0.5dB from the DF threshold. The bilayer structure is extended for use at high SNR using bit-interleaved coded modulation (BICM) [16]. As for CF relaying, a coding scheme using a combination of LDPC and irregular repeat accumulate (IRA) codes is presented in [17]. Rateless coding schemes are developed in [18]. As for QMF relaying, a coding scheme is proposed in independent work [19] based on lattice strategies. The scheme in [19] reduces the complexity of mapping at the relay to polynomial-time while the joint decoding complexity remains exponential-time.

D. Summary of Results

In this paper, a coding scheme for QMF relaying with linear complexity encoding at the source, mapping at the relays and joint decoding at the destination is developed. For a network with one relay, the proposed scheme performs within $(0.5-1)$dB gap from the information-theoretic QMF threshold. For the code design example considered in Section V, the QMF threshold is ≈ 1.5dB better than DF. The key techniques used in this paper are summarized as follows:

1) *BICM:* Design of *binary* channel codes with standard higher order signal constellations is considered based on the widely used BICM technique [20].
2) *LDPC-LDGM:* The scheme uses low density parity check (LDPC) codes at the source for channel coding and low density generator matrix (LDGM) codes at the relays for mapping.
3) *Joint Factor Graph:* The joint decoding procedure at the destination is formulated as a belief propagation algorithm over a factor graph. This graph contains the original channel code (LDPC) and relay mapping functions (LDGM) as subgraphs connected via probabilistic constraints that model the QMF relay operations.

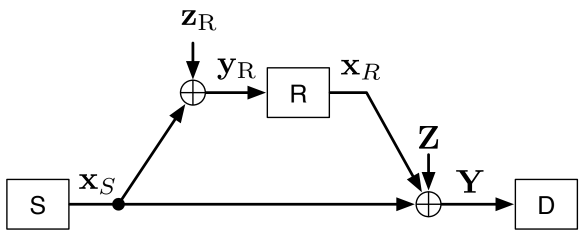


Fig. 2. Half-duplex binary input Gaussian relay channel.

4) *Practical Decoding Algorithm:* Using a DBLAST space-time architecture, scalar quantization procedure at relays and specific choice of component codes, the resulting factor graph is greatly simplified, making it suitable for practical decoder implementation.
5) *Code Design:* Density evolution analysis tools [21], [22] are developed for the systematic design of joint LDPC-LDGM factor graphs.

E. Organization

In Section II, the coding framework for QMF and corresponding joint decoding algorithm is developed. The treatment focusses on a canonical system model with one relay and binary inputs. In Section III, density evolution and code design tools are developed. In Section IV, the framework is extended to the high SNR regime i.e. for high order modulation inputs using BICM. In Section V, the design of codes for an example cooperative link is demonstrated. Finally in Section VI a sketch is provided for extending the proposed framework to scenarios with multiple relays.

II. Coding Framework

A. System Model

Initially, this paper focuses on the design of codes for a *binary memoryless symmetric* (BMS) relay channel as described below. In SectionIV, this model is extended to high order modulation inputs for use at high SNR.

The BMS Gaussian relay channel has three half-duplex terminals: source (S), relay (R) and destination (D) with binary input additive white Gaussian (BIAWGN) channels between them, as shown in Fig. 2. R listens for a fraction $f \in [0,1]$ of the total communication time and transmits for the fraction $(1-f)$. The block lengths for the transmitted codewords at S and R are N_S and N_R respectively. They satisfy the half-duplex constraint $N_R = (1-f)N_S$. The codeword messages sent from S and R are $\mathbf{b}_S \in \{0,1\}^{N_S}$ and $\mathbf{b}_R \in \{0,1\}^{N_R}$ respectively. The corresponding transmitted signals are $\mathbf{x}_S \in \{\pm\sqrt{P_S}\}^{N_S}$ and $\mathbf{x}_R \in \{\pm\sqrt{P_R}\}^{N_R}$ where P_S and P_R are *per-node symbol constraints* on average power i.e. $E|x_{S,i}^2| \leq P_S$ and $E|x_{R,i}^2| \leq P_R$. Bold-face lower case letters are used to denote a sequence of symbols.

Multiple (M) receive antennas are assumed at the destination. This permits consideration of network scenarios where *cooperative spatial multiplexing* is possible [23][2]. An example scenario with $M = 2$ is discussed in the Appendix. The received signals at D and R are denoted as $\mathbf{y}_i \in \mathbb{C}^M$ and $y_{R,j}$ for each symbol time $i \in \{1,2,\ldots,N_S\}$ and $j \in \{1,2,\ldots,fN_S\}$ respectively. They are modeled as follows:

$$\mathbf{y}_i = \mathbf{h}_1 x_{S,i} + \mathbf{h}_2 x'_{R,i} + \mathbf{z}_i,\ y_{R,j} = h_R x_{S,j} + z_{R,j}$$

Here $\mathbf{h}_1, \mathbf{h}_2, h_R$ denote the corresponding channel gains. $x'_{R,i} = 0$ and $\mathbf{h}_2 = \mathbf{0}$ for $i \in \{1,2,\ldots,fN_S\}$ when R is listening. For the remaining time $x'_{R,i} = x_{R,i-fN_S}$, $i \in \{fN_S+1,\ldots,N_S\}$. $\mathbf{z_i}$ and $z_{R,j}$ are i.i.d. zero-mean Gaussian noise vectors with identity covariance matrices. All the channel observations at D are denoted by $\mathbf{Y} \in \mathbb{C}^{M\times N_S}$ i.e. $\mathbf{Y} = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_{N_S}]$. Observations at R are denoted as $\mathbf{y}_R \in \mathbb{C}^{1\times fN_S}$ i.e. $\mathbf{y}_R = [y_{R,1}\ \ldots\ y_{R,fN_S}]$. The channel is characterized by the following parameters: $\mathsf{SNR}_{SR} = P_S|h_R|^2$, $\mathsf{SNR}_{SD} = P_S||\mathbf{h}_1||^2$ and $\mathsf{SNR}_{RD} = P_R||\mathbf{h}_2||^2$.

B. Quantize-Map-Forward Scheme

The quantize-map-and-forward scheme [1] is summarized as follows. S has a sequence of messages $m_k \in \{1,\ldots,2^{N_S\mathcal{R}}\}$, $k = 1,2,\ldots$ to be transmitted. At both S and R, codebooks $\mathcal{C}_S$ and $\mathcal{C}_R$ are created respectively. S maps each message to one of its codewords and transmits it using N_S symbols resulting in an overall transmission rate of $\mathcal{R}$. Relay listens to the first fN_S time symbols of each block. It quantizes its observation at noise level i.e. the quantization distortion is equal to the noise power at the relay. Relay maps the quantized bits to a codeword in $\mathcal{C}_R$. It transmits this codeword using $(1-f)N_S$ symbols. The destination D attempts to decode the message sent by S from received signals ($\mathbf{Y}$). In order to decode, D must know all channel parameters $\mathsf{SNR}_{SD}, \mathsf{SNR}_{RD}$ and SNR_{SR}, the relay listening fraction f and both codebooks $\mathcal{C}_S$ and $\mathcal{C}_R$.

It is assumed that SNR_{SD},SNR_{RD} are measured at D and SNR_{SR} is measured at R using pilot symbols. It is further assumed that SNR_{SR} is forwarded to D by R. The estimation and forwarding overhead of these steps is ignored for the analysis presented in this paper.

C. Factor Graph for Joint Decoding

In the context of the system model and cooperation scheme outlined above, let us focus on binary *linear* codebooks $\mathcal{C}_S^b$ and $\mathcal{C}_R^b$. These can be represented as bipartite Tanner graphs using respective parity check matrices. In such a representation bit (variable) nodes represent the codeword and check (function) nodes represent parity constraints that must be satisfied in order for the codeword to be valid. Let us consider the maximum *a posteriori* (MAP) rule for joint decoding at D. In this subsection, joint decoding is expressed as a sum-product algorithm over a factor graph that contains the Tanner graphs of component codes ($\mathcal{C}_S^b, \mathcal{C}_R^b$) as sub-graphs connected via probabilistic constraints that represent the QMF relaying operation [24][25][26].

Joint decoding involves searching for the codeword $\mathbf{b}_S \in \mathcal{C}_S^b$ that maximizes the *a posteriori* probability $p(\mathbf{b}_S|\mathbf{Y})$. An efficient way to do this search is to consider the bitwise maximum *a posteriori* (MAP) decoder, where the aim is to compute

$p\left(b_{S,i}|\mathbf{Y}\right) \;=\; \sum_{\sim b_{S,i}} p\left(\mathbf{b}_S|\mathbf{Y}\right)$ for all $i = 1, 2, \ldots N_S$.

$$\begin{aligned} p\left(\mathbf{b}_S|\mathbf{Y}\right) &= \sum_{\mathbf{b}_R} \frac{f\left(\mathbf{Y}|\mathbf{b}_S, \mathbf{b}_R\right) p\left(\mathbf{b}_S, \mathbf{b}_R\right)}{f\left(\mathbf{Y}\right)} \\ &\propto \sum_{\mathbf{b}_R} f\left(\mathbf{Y}|\mathbf{b}_S, \mathbf{b}_R\right) p\left(\mathbf{b}_S, \mathbf{b}_R\right). \end{aligned}$$

For the first fN_S bits, R is listening and D observes an interference-free signal from S. During the remaining transmissions, D observes a superposition of signals from S and R. Therefore, the first term $f\left(\mathbf{Y}|\mathbf{b}_S, \mathbf{b}_R\right)$ factorizes as follows:

$$\begin{aligned} &f(\mathbf{Y}|\mathbf{b}_S, \mathbf{b}_R) \\ &= \prod_{i=1}^{fN_S} f(\mathbf{y}_i|b_{S,i}) \prod_{j=1}^{N_R} f(\mathbf{y}_{(fN_S+j)}|b_{S,(fN_S+j)}, b_{R,j}) \end{aligned}$$

The codes $\mathcal{C}_S^b$ and $\mathcal{C}_R^b$ have characteristic functions $\mathbf{1}(\mathbf{b}_S \in \mathcal{C}_S^b)$ and $\mathbf{1}(\mathbf{b}_R \in \mathcal{C}_R^b)$ respectively.

$$\begin{aligned} p\left(\mathbf{b}_S, \mathbf{b}_R\right) &= p\left(\mathbf{b}_S\right) p\left(\mathbf{b}_R|\mathbf{b}_S\right) \\ &\propto \mathbf{1}\left(\mathbf{b}_S \in \mathcal{C}_S^b\right) p\left(\mathbf{b}_R|\mathbf{b}_S\right) \\ &\overset{(a)}{=} \mathbf{1}\left(\mathbf{b}_S \in \mathcal{C}_S^b\right) \mathbf{1}\left(\mathbf{b}_R \in \mathcal{C}_R^b\right) p\left(\mathbf{b}_R|\mathbf{b}_S\right). \end{aligned}$$

(a) is due to the fact that $\mathbf{b}_R$ must be a codeword in $\mathcal{C}_R^b$.

The resulting factor graph in Fig. 3 shows that in addition to nodes representing channel observations i.e. $f(\mathbf{y}_i|b_{S,i}, b_{R,j})$ the subgraphs $\mathbf{1}\left(\mathbf{b}_S \in \mathcal{C}_S^b\right)$ and $\mathbf{1}\left(\mathbf{b}_R \in \mathcal{C}_R^b\right)$ are connected by $p\left(\mathbf{b}_R|\mathbf{b}_S\right)$ that represents the quantization operation at R.

If the component codes $\mathcal{C}_S^b$ and $\mathcal{C}_R^b$ are sparse, the overall factor graph is also sparse. A sum-product algorithm for decoding over such a factor graph has complexity that grows linearly with the length of component codes. However, the sum-product update rules at the function node $p\left(\mathbf{b}_R|\mathbf{b}_S\right)$ is very complex due to its high degree ($N_S + N_R$). Moreover, it introduces very short cycles in the graph, which deteriorates the performance of sum-product decoding. In order to get reasonably close to MAP performance and low decoding complexity, the $p\left(\mathbf{b}_R|\mathbf{b}_S\right)$ node must be factorized further. In the following subsections, choice for component codes $\mathcal{C}_S^b$ and $\mathcal{C}_R^b$ and specific techniques for factorization are discussed.

D. Choice of Component Codes

In the discussion above, general binary linear codes $\mathcal{C}_S^b$ and $\mathcal{C}_R^b$ are considered. A natural choice is to use sparse graph codes (like LDPC) that are known to have good performance and linear complexity decoding/encoding operations.

In a previous communication [24], preliminary results for such factor graphs were presented using off-the-shelf LDPC codes at both S and R. As observed in [24], off-the-shelf (point-to-point) LDPC codes do not allow close-to-optimal performance over cooperative channels. An information-theoretic understanding of this observation is presented in [27]. The authors point out that capacity-achieving codes for the point-to-point channel exhibit higher estimation errors whenever the SNR is below the Shannon limit. Therefore, in cooperative networks where the operating SNR is below the point-to-point Shannon limit, such off-the-shelf codes are no longer suitable to utilize side information from the relay at the destination. As a consequence, specialized codes are required for cooperative channels. For sparse graph codes, specialized code profiles that are optimized for relaying can be designed using standard tools such as density evolution analysis [21], [22]. However, for the LDPC-LDPC combination [24] density evolution does not extend readily to QMF joint factor graphs.

In this paper, the use of LDPC codes at S and LDGM codes at R is proposed. LDPC codes are known to perform very close to information theoretic limits when used for channel coding. Similarly LDGM codes are commonly used for lossy data compression [28] and the LDPC-LDGM combination is a good fit for the QMF relay channel. Moreover, density evolution analysis tools can be extended to LDPC-LDGM joint factor graphs. Such an extension is developed in Section III. This permits explicit construction of code profiles optimized for relaying.

Based on the LDPC-LDGM choice, let us introduce auxiliary variable nodes $\mathbf{b}_Q = \{b_{Q,i}\}_{i=1}^{K_R}$ in the factor graph. $\mathbf{b}_Q$ represents the K_R bits after quantization at R. These are mapped to the codeword $\mathbf{b_R}$ of length N_R obtained after passing through a low density generator matrix having K_R rows, N_R columns and characteristic function $\mathbf{1}(\mathbf{b}_R \in \mathcal{C}_R^b)$. Since $\mathbf{b}_R$ is a deterministic function of $\mathbf{b}_Q$, $p\left(\mathbf{b}_R|\mathbf{b}_S\right)$ can be factorized as follows (Fig 4):

$$\begin{aligned} p(\mathbf{b}_R|\mathbf{b}_S) &= p(\mathbf{b}_R, \mathbf{b}_Q|\mathbf{b}_S) \\ &= p(\mathbf{b}_R|\mathbf{b}_Q, \mathbf{b}_S)p(\mathbf{b}_Q|\mathbf{b}_S) \\ &= \mathbf{1}(\mathbf{b}_R \in \mathcal{C}_R^b)p(\mathbf{b}_Q|\mathbf{b}_S) \end{aligned}$$

The LDGM mapping can either *compress* or *expand* the K_R quantized bits i.e. the LDGM coding rate can be greater than 1. The $\mathbf{b}_R$ nodes always have degree 2 and they simply perform forwarding of messages under the sum-product algorithm.

E. Scalar Quantizer

In general, a vector quantizer can be used at R. However, it is shown [1] that QMF performs within bounded gap of capacity even with a scalar quantizer. Under scalar quantization, the observation for every bit from S is quantized independently. If each $y_{R,i}$ is quantized into $b_Q[A_i]$ for $i = 1, 2, \ldots fN_S$, then the $p\left(\mathbf{b}_Q|\mathbf{b}_S\right)$ function node factorizes into fN_S separate nodes each representing a scalar quantization operation.

$$\begin{aligned} &p\left(\mathbf{b}_Q|\mathbf{b}_S\right) = \prod_{i=1}^{fN_S} p\left(b_Q[A_i]|b_{S,i}\right), \text{ where} \\ &\bigcup_{i=1}^{fN_S} A_i = \{1, 2, \ldots, K_R\},\ A_i \cap A_j = \emptyset\ \forall i \neq j \end{aligned}$$

where A_i denotes the subset of indices in $\mathbf{b}_Q$ that observation $y_{R,i}$ is quantized into. As a result, the variable nodes of the two Tanner graphs are connected by function nodes representing the stochastic relations $p\left(b_Q[A_i]|b_{S,i}\right)$ among them. Henceforth, these are called *quantize* (Q) nodes, as they are induced by the quantization procedure at the relay. An example factor graph showing the LDPC-LDGM construction is illustrated in Fig. 4 where each symbol observation $y_{R,i}$ is quantized into one bit (i.e. $K_R = fN_S$ and $A[i] = \{i\}$). As shown, there are four kinds of nodes in the resulting factor graph: observation

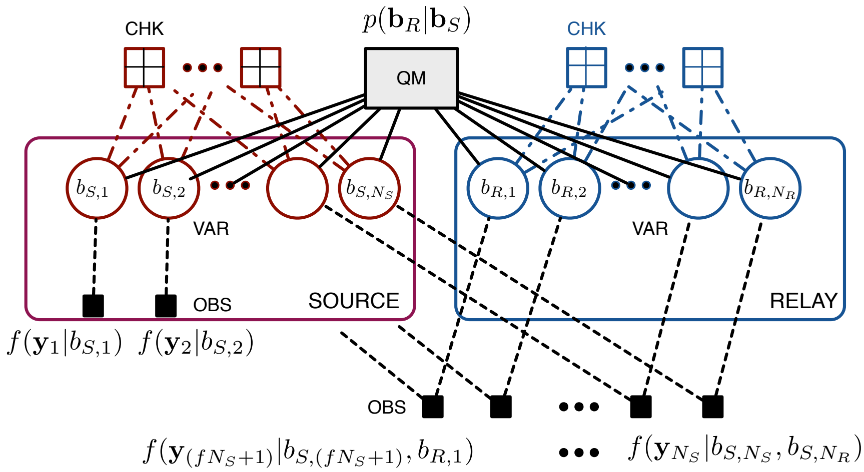


Fig. 3. Factor graph for joint decoding.

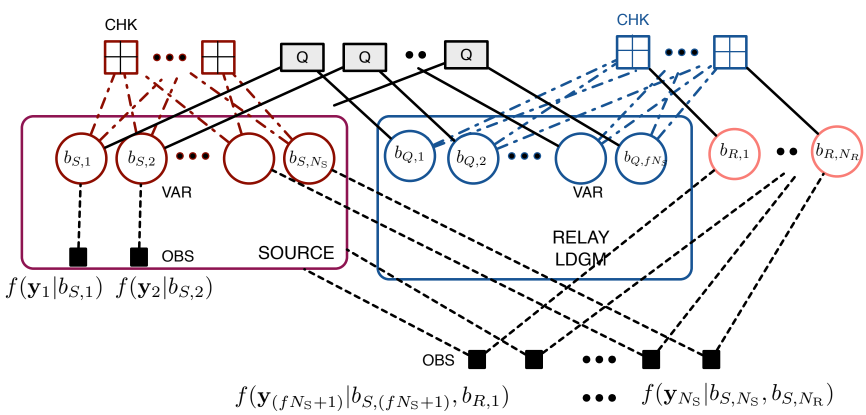


Fig. 4. Factor graph: LDPC code at S, LDGM code at R with 1 bit scalar quantizer.

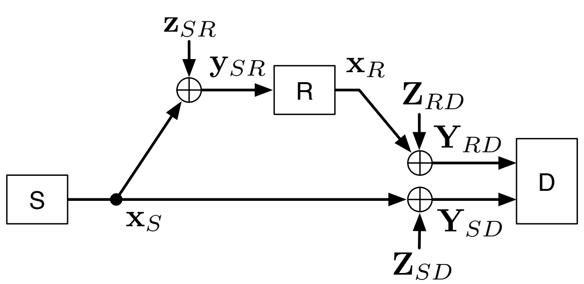


Fig. 5. Channel model using DBLAST.

(OBS) nodes, variable (VAR) nodes, check (CHK) nodes, and quantize (Q) nodes. Some VAR nodes in the $\mathcal{C}_S^b$ subgraph share OBS nodes with VAR nodes in the $\mathcal{C}_R^b$ subgraph. This is because of multiple access at D.

F. DBLAST Scheme

The factor graph shown in Fig. 4 can be simplified further using the Diagonal Bell Labs Space-Time architecture (DBLAST) [29]. As discussed in the Appendix, the degree 2 OBS nodes (representing multiple access) are factorized using DBLAST. Under DBLAST, the destination observes two orthogonal sets of observations (see Fig. 5). The factorization is shown in equation (1) where $\underline{\mathbf{Y}} := [\mathbf{Y}_{SD}\ \mathbf{Y}_{RD}]$.

An example of the simplified factor graph is depicted in Fig. 6. In this graph, VAR nodes in the two Tanner graphs are connected *only* through Q nodes. Since $\mathbf{y}_{RD,i} = \mathbf{0}$ for $i = 1, \ldots, fN_S$, we rename $\mathbf{y}_{RD,(fN_S+j)} \equiv \mathbf{y}_{RD,j}$, for $j = 1, \ldots, N_R$.

The resulting graph has a structure similar to an irregular LDPC code but with special Q constraints. In Section II-G, sum-product updates for this graph are derived following the general principle outlined in [25]. It is shown that for a simple one-bit quantizer each Q node further factorizes into a CHK constraint and a dummy VAR node. This reduces the factor graph to a Tanner graph that does not have any special nodes. Such a property is useful to leverage existing techniques used for the design of low-power, high-throughput LDPC decoders.

G. Decoding Algorithm

For the point-to-point system, *belief-propagation* is an iterative algorithm that computes the *a posteriori* probability to decode message bits. The algorithm computes this exactly if the factor graph has no cycles. Otherwise, it computes the approximate *a posteriori* probability for each bit [25]. For the

$$
\begin{aligned}
f(\underline{\mathbf{Y}}|\mathbf{b}_S,\mathbf{b}_Q) &= \prod_{i=1}^{fN_S} f(\mathbf{y}_{SD,i}|b_{S,i}) \prod_{j=1}^{N_R} f(\mathbf{y}_{SD,(fN_S+j)},\mathbf{y}_{RD,(fN_S+j)}|b_{Q,j},b_{S,(fN_S+j)}) \\
&= \prod_{i=1}^{fN_S} f(\mathbf{y}_{SD,i}|b_{S,i}) \prod_{j=1}^{N_R} f(\mathbf{y}_{SD,(fN_S+j)}|b_{S,(fN_S+j)}) f(\mathbf{y}_{RD,(fN_S+j)}|b_{Q,j}) \\
&= \prod_{i=1}^{N_S} f(\mathbf{y}_{SD,i}|b_{S,i}) \prod_{j=1}^{N_R} f(\mathbf{y}_{RD,(fN_S+j)}|b_{Q,j})
\end{aligned} \tag{1}
$$

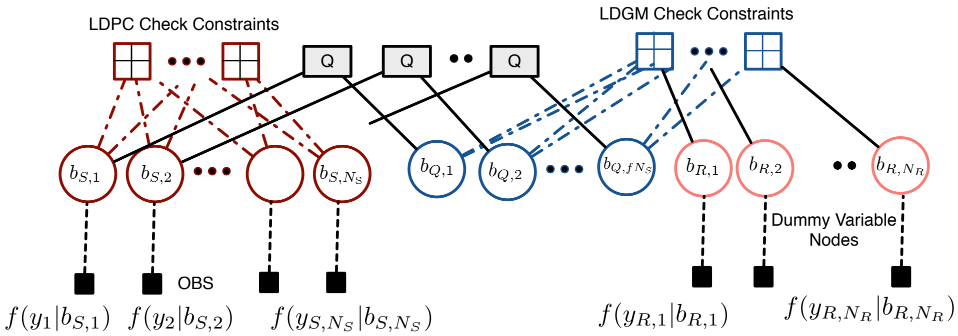


Fig. 6. Simplified factor graph with one-bit scalar quantizer and DBLAST.

factor graph in Fig. 6, messages being passed on the edges of the factor graph and the update rules at the variable/check nodes stay unchanged. The only new ingredient in the mix is the Q nodes introduced by our framework.

Let the subscripts V, C, and Q denote VAR nodes, CHK nodes, and Q nodes respectively. For $\mathsf{F} \in \{\mathsf{C},\mathsf{Q}\}$, let $\omega_{\mathsf{VF}}^{(l)}$ denote the message sent from variable node V to function node F in the l^{th} iteration. Every edge in the graph is connected to exactly one variable node and a message on the edge represents the *a posteriori* probability for the respective variable. The messages can be represented as LLRs, but for the sake of simplicity we represent the messages as a two-dimensional vector in this subsection[1]. $\omega_{\mathsf{VF}} := [p_0\ p_1]$, where $\omega_{\mathsf{VF}}(1) = p_0 \in [0,1]$ represents the probability that the bit is 0 and $\omega_{\mathsf{VF}}(2) = p_1 \in [0,1]$ represents the probability that the bit is 1 $(p_0 + p_1 = 1)$.

The message sent from V to $\mathsf{F} \in \{\mathsf{C},\mathsf{Q}\}$ is the normalized product of all incoming messages into V except for the message from F. The normalization ensures that $p_0 + p_1 = 1$ for the outgoing message. The message sent from C to V′ is the indicator function that the check is satisfied, marginalized on the bit represented by V′. The message sent from Q to V is the marginalization of the function $p\left(b_Q[A_i]|b_{S,i}\right)$ on the symbol represented by V. b_Q is computed from a noisy observation of $b_{S,i}$, the node Q imposes a probabilistic constraint on the variables. Since the quantization is scalar: $\forall \mathbf{u} \in \{0,1\}^{|A_i|}$ and $v \in \{0,1\}$,

$$g(\mathbf{u},v) := p\left(b_Q[A_i] = \mathbf{u}|b_{S,i} = v\right).$$

This function is fully represented by a lookup table with $2^{|A_i|+1}$ values, which is used to derive the update rule for Q.

As an example, let us consider a one-bit scalar quantizer at the relay and derive the update rule. For this case, the Q node can be further factorized into a CHK node and a dummy VAR node that sends a constant message. Note that $A_i = \{i\},\ i = 1,2,\ldots,fN_S$ and $K_R = fN_S$. The factor graph is depicted in Fig. 6. Consider $\forall u \in \{0,1\}$ and $v \in \{0,1\}$,

$$
\begin{aligned}
g(u,v) :=& p\left(b_{Q,i} = u|b_{S,i} = v\right) \\
=& (1-p_f)\mathbf{1}\{u = v\} + p_f\mathbf{1}\{u \neq v\}
\end{aligned}
$$

here $p_f := \frac{1}{2}\operatorname{erfc}\sqrt{\frac{\mathsf{SNR}_{SR}}{2}}$ denotes the probability of bit error for scalar one-bit quantization over a BIAWGN channel. Since the function g is symmetric in u and v, it can be assumed that the VAR node is of the source, and the marginalization is on v. Let the other VAR node be V'. This leads to the following update rule:

$$
\begin{aligned}
\omega_{\mathsf{QV}}(1) &= (1-p_f)\omega_{\mathsf{V'Q}}(1) + p_f\omega_{\mathsf{V'Q}}(2) \\
\omega_{\mathsf{QV}}(2) &= (1-p_f)\omega_{\mathsf{V'Q}}(2) + p_f\omega_{\mathsf{V'Q}}(1),
\end{aligned}
$$

This takes the same form of a CHK node update with incoming messages $\omega_{\mathsf{V'Q}}$ and $[1-p_f\ p_f]$. Therefore, the Q node in this set-up specializes to a CHK node with additional dummy VAR nodes sending constant message $[1-p_f\ p_f]$ that depends on SNR_{SR}. The resulting factor graph is depicted in Fig. 7.

III. Code Design

In this section, design of specific codes for QMF relaying is discussed. Typically sparse graph codes like LDPC and LDGM are drawn randomly from ensembles, which are described using degree profiles. In the point-to-point case, if the

[1]Later we will replace ω by w, the commonly used message $\log\frac{p_0}{p_1}$ (LLR) in belief propagation.

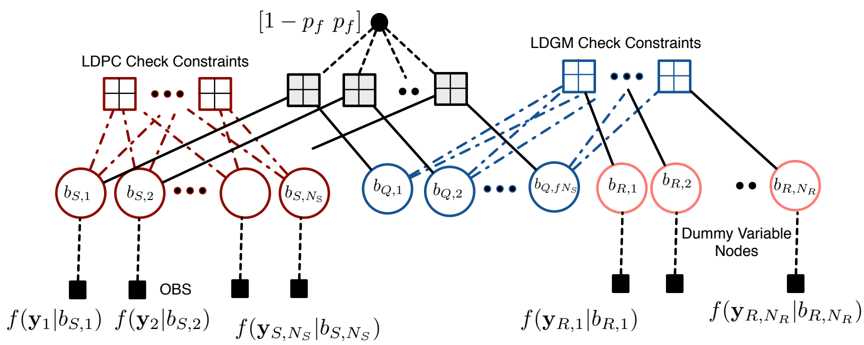


Fig. 7. Equivalent factor graph of that in Fig. 6. The Q nodes are factorized into a CHK node and a dummy variable.

block-length is sufficiently large, the decoding performance of such codes converges to the *ensemble average* [22]. Let us consider degree profiles (λ_S, ρ_S) and (λ_R, ρ_R) for the LDPC and LDGM codes at source and relay respectively. λ_S and ρ_S are polynomials representing variable and check degree distributions for the LDPC code:

$$\lambda_S(x) = \sum_{i=2}^{\infty} \lambda_{S,i} x^{i-1}, \ \rho_S(x) = \sum_{i=2}^{\infty} \rho_{S,i} x^{i-1}$$

Here $\lambda_{S,i}$ and $\rho_{S,i}$ denote the fraction of edges with degree i at a variable node and at a check node respectively. For the LDGM code, we have the similar definition for λ_R and ρ_R except that these are regarding the edges connecting check nodes and variable nodes for $\mathbf{b}_Q$ (not $\mathbf{b}_R$).

These profiles must satisfy the following constraints:

$$\mathcal{R} = 1 - \frac{\int_0^1 \rho_S(x)dx}{\int_0^1 \lambda_S(x)dx}, \ \frac{K_R}{N_R} = \frac{\int_0^1 \rho_R(x)dx}{\int_0^1 \lambda_R(x)dx} = \frac{f}{1-f},$$
$$\lambda_S(1) = \lambda_R(1) = \rho_S(1) = \rho_R(1) = 1$$

In addition to the sub-graphs representing the two component codes, the joint factor graph shown in Fig. 7 also includes edges connecting them (via Q nodes). As discussed previously in Section II-C, let us consider a fixed one bit scalar quantizer. The edges connecting with Q nodes are considered *fixed* in the rest of this section. In contrast, the edges in LDPC and LDGM subgraphs are drawn randomly from the ensemble using construction procedure described in [22].

In point-to-point channels, the typical method to analyze and design sparse graph codes is to compute the ensemble average performance for given degree profiles assuming infinite block length (*convergence to computation trees*). The ensemble average performance (decoding error probability for given SNRs) is calculated using *density evolution* developed in [21], [22]. The two key elements of classical density evolution, namely, *concentration around ensemble average* and *convergence to computation tree channels* for sufficiently large block length hold for the proposed QMF relaying system as well. The proofs can be readily extended from those of point-to-point channels [22] [21].

Without loss of generality it is assumed that the all-zero codeword is transmitted from S. This is a result of the symmetry of the relay channel. However, R does not transmit the all-zero codeword because the source-to-relay channel is noisy. For sufficiently large block lengths and a given value of SNR_{SR} there is a *typical* sequence $\mathbf{b}_Q$ that is mapped to a *typical* $\mathbf{b}_R$ based on the LDGM code. The probability of occurrence for atypical codewords vanishes as the block length becomes large. Therefore, it is ignored for computing the ensemble average performance. A typical $\mathbf{b}_Q$ comprises of $K_R(1-p_f)$ 0's and $K_R p_f$ 1's. p_f is defined in Section II-G. For a given degree profile, (λ_R, ρ_R), each bit of the typical $\mathbf{b}_R$ is i.i.d. Bernoulli(q), where q is the probability of having an odd number of 1's in a column of the generator matrix (drawn randomly from the LDGM ensemble).

$$q = \sum_j \left(\frac{\rho_R(j)/j}{\sum_i \rho_R(i)/i} \right) \frac{1-(1-2p_f)^j}{2}$$

To develop density evolution rules for QMF relaying, we consider the belief-propagation algorithm with log-likelihood ratios as the messages passed among variable nodes and various function nodes. Let $w_{\mathsf{FV}}^{(l)}$ and $w_{\mathsf{VF}}^{(l)}$ denote the message sent from the function node F to the variable node V and vice-versa, at the l-th iteration. $\mathsf{F} \in \{\mathsf{C}_S, \mathsf{C}_R, \mathsf{Q}, \mathsf{O}_S, \mathsf{O}_R\}$ represent the LDPC CHK nodes, LDGM CHK nodes, Q nodes, OBS nodes at S and R respectively. $\mathsf{V} \in \{\mathsf{V}_S, \mathsf{V}_Q, \mathsf{V}_R\}$ represent the VAR nodes corresponding to $\mathbf{b}_S, \mathbf{b}_Q$ and $\mathbf{b}_R$ respectively.

The sum-product update rules in terms of the commonly used LLR's are written as follows:

$$w_{\mathsf{VF}}^{(l)} = \sum_{\mathsf{F}' \in \mathcal{N}(\mathsf{V}) \setminus \{\mathsf{F}\}} w_{\mathsf{F'V}}^{(l)} \tag{2}$$

$$w_{\mathsf{OV}}^{(l)} = w_{\mathsf{V}}, \quad (\mathsf{O}, \mathsf{V}) = \{(\mathsf{O}_S, \mathsf{V}_S), (\mathsf{O}_R, \mathsf{V}_R)\}$$

$$w_{\mathsf{FV}}^{(l+1)} = 2 \tanh^{-1} \left(\prod_{\mathsf{V}' \in \mathcal{N}(\mathsf{F})} \tanh\left(\frac{1}{2} w_{\mathsf{V'F}}^{(l)} \right) \right) \tag{3}$$
$$\text{if } \mathsf{F} = \mathsf{C}_S, \mathsf{C}_R$$

$$w_{\mathsf{FV}}^{(l+1)} = 2 \tanh^{-1} \left((1-2p_f) \prod_{\mathsf{V}' \in \mathcal{N}(\mathsf{F})} \tanh\left(\frac{1}{2} w_{\mathsf{V'F}}^{(l)} \right) \right)$$
$$\text{if } \mathsf{F} = \mathsf{Q}$$

Here $\mathcal{N}(\cdot)$ here denote the set of neighboring nodes and w_{V} represents the LLR from channel observation.

Compared to the point-to-point case where there is only one kind of variable node (V) and one kind of CHK node (C) the update rules can be expressed simply by (2) and (3) where $\mathsf{F} = \mathsf{C}$. Density evolution analysis tracking the density of these messages in each iteration. For the point-to-point case with degree distribution (λ, ρ), there is only one type of edge and the evolution is expressed using a pair of coupled recursive equations as follows:

$$P^{(l+1)}_{\mathsf{CV}} = \Gamma^{-1}\left(\sum_j \rho_j \left(\Gamma\left(P^{(l)}_{\mathsf{VC}}\right)\right)^{\otimes(j-1)}\right)$$

$$P^{(l)}_{\mathsf{VC}} = P_{\mathsf{V}} \otimes \sum_i \lambda_i \left(P^{(l)}_{\mathsf{CV}}\right)^{\otimes(i-1)}$$

Here $\Gamma(\cdot)$ denotes a transformation on the density as defined in [21], $\otimes$ denotes the convolution operator and $P^{(l)}_{\{\cdot\}}$ denotes the density of message $w^{(l)}_{\{\cdot\}}$. P_{V} represents the conditional density of the LLR of the point-to-point channel.

For the QMF relaying case, there are 4 types of edges and densities for messages along all of them must be tracked. The recursive density updates are derived similarly:

Function Nodes to Variable Nodes:

$$P^{(l+1)}_{\mathsf{C}_S\mathsf{V}_S} = \Gamma^{-1}\left(\sum_j \rho_{S,j}\left(\Gamma\left(P^{(l)}_{\mathsf{V}_S\mathsf{C}_S}\right)\right)^{\otimes(j-1)}\right)$$

$$P^{(l+1)}_{\mathsf{C}_R\mathsf{V}_Q} = \Gamma^{-1}\left(\sum_j \rho_{R,j}\left(\Gamma\left(P^{(l)}_{\mathsf{V}_Q\mathsf{C}_R}\right)\right)^{\otimes(j-1)} \otimes \Gamma\left(P^{(l)}_{\mathsf{V}_R\mathsf{C}_R}\right)\right)$$

$$P^{(l+1)}_{\mathsf{QV}_S} = \Gamma^{-1}\left(\Gamma\left(P^{(l)}_{\mathsf{V}_Q\mathsf{Q}}\right) \otimes \Gamma\left(\delta_{\log\frac{1-p_f}{p_f}}\right)\right)$$

$$P^{(l+1)}_{\mathsf{QV}_Q} = \Gamma^{-1}\left(\Gamma\left(P^{(l)}_{\mathsf{V}_S\mathsf{Q}}\right) \otimes \Gamma\left(\delta_{\log\frac{1-p_f}{p_f}}\right)\right)$$

Variable Nodes to Function Nodes:

$$\begin{aligned}
P^{(l)}_{\mathsf{V}_S\mathsf{C}_S} =& P_{\mathsf{V}_S} \otimes \sum_i \Big\{ f\lambda_{S,i}\left(P^{(l)}_{\mathsf{C}_S\mathsf{V}_S}\right)^{\otimes(i-1)} \otimes P^{(l)}_{\mathsf{QV}_S} + \\
& (1-f)\lambda_{S,i}\left(P^{(l)}_{\mathsf{C}_S\mathsf{V}_S}\right)^{\otimes(i-1)}\Big\} \\
P^{(l)}_{\mathsf{V}_Q\mathsf{C}_R} =& \sum_i \lambda_{R,i}\left(P^{(l)}_{\mathsf{C}_R\mathsf{V}_Q}\right)^{\otimes(i-1)} \otimes P^{(l)}_{\mathsf{QV}_Q} \\
P^{(l)}_{\mathsf{V}_R\mathsf{C}_R} =& P_{\mathsf{V}_R} \\
P^{(l)}_{\mathsf{V}_S\mathsf{Q}} =& P_{\mathsf{V}_S} \otimes \sum_i \lambda_{S,i}\left(P^{(l)}_{\mathsf{C}_S\mathsf{V}_S}\right)^{\otimes(i)} \\
P^{(l)}_{\mathsf{V}_Q\mathsf{Q}} =& \sum_i \lambda_{R,i}\left(P^{(l)}_{\mathsf{C}_R\mathsf{V}_Q}\right)^{\otimes(i)}
\end{aligned}$$

Here $\delta_r(\cdot)$ denotes the Dirac delta function at point $r \in \mathbb{R}$. $\delta_{\log\frac{1-p_f}{p_f}}$ shows up in the expressions because the Q node is equivalent to a CHK node connected to a *constant*. The differences in evolution rules between the QMF relaying and point-to-point channel arise due to the probabilistic Q constraints in the joint factor graph.

As in the point-to-point case, P_{V_S} is the conditional density of the LLR of the source to destination channel, given that an all-zero codeword is sent from S. P_{V_R} is the marginal density of the LLR of the relay to destination channel under the marginal law that $\mathbf{b}_R$ is i.i.d. Bernoulli(q).

The density evolution rules derived above are used to compute the probability of error in decoding of $\mathbf{b}_S$. For successive interference cancellation using DBLAST, $\mathbf{b}_R$ must also be reliably decoded. Density evolution rules to compute probability of decoding error for $\mathbf{b}_R$ can be similarly derived.

IV. Bit Interleaved Coded Modulation

So far, the discussion has focussed on the BMS Gaussian relay defined in Section II-A. For the high SNR regime, input alphabet $\mathbf{x}_S \in \mathcal{A}^{N_S}$ and $\mathbf{x}_R \in \mathcal{A}^{N_R}$ where $\mathcal{A}$ represents constellation points in a high-order modulation scheme must be considered. In practice many systems use BICM [20] to combine channel codes designed for binary alphabet with high-order signal constellations. BICM has also been proposed for various cooperative channel scenarios [30][31][32][16]. In this subsection, a procedure is discussed for extending the coding framework from the BMS relay channel to a relay channel with inputs from high-order alphabets.

Under classical BICM [20], a point to point Gaussian channel is decomposed into *parallel independent memoryless* "sub-channels". Every "sub-channel" $p_{Y|B,S}(y|b,s)$ has binary inputs $b \in \{0,1\}$ and depends on state $s \in \{1,2,\ldots,L\}$ which is chosen uniformly and known to both the terminals (2^L is the cardinality of the chosen signal constellation). At the receiver, LLR for a bit that was mapped to state s is calculated from symbol observation $y \in \mathbb{C}$ (in case of MIMO receiver $\mathbf{y} \in \mathbb{C}^M$).

$$LLR(y,s) = \log\frac{P_{B|Y,S}(b=0|y,s)}{P_{B|Y,S}(b=1|y,s)}$$

However, this binary channel is not guaranteed to be output-symmetric i.e. the crossover probability for a bit is not independent of its value. Let $f_\Lambda(\lambda)$ represent the PDF of $LLR(y,s)$. The channel is output symmetric if the following condition holds:

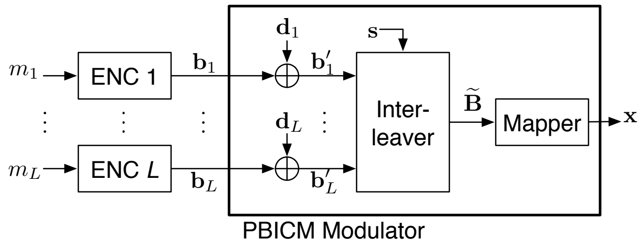
$$f_{\Lambda|B}(\lambda|b=0) = f_{\Lambda|B}(-\lambda|b=1)$$

Conventional methods for designing linear coding schemes such as density evolution etc. cannot be used with asymmetric channels. This issue is resolved by adding random dithers at every bit to make the channel output-symmetric as proposed in [20], [33], [34]. Dithers are i.i.d. $\mathrm{Bernoulli}\left(\frac{1}{2}\right)$ variables known to both the transmitter and receiver. For a dither $d \in \{0,1\}$ the channel $p_{Y|B,S,D}(y|b,s,d)$ is binary, memoryless and symmetric (BMS).

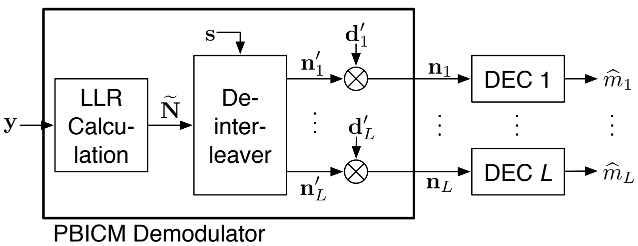
$$LLR(y,s,d) = (-1)^d LLR(y,s)$$

This method is called parallel BICM (PBICM) in [34] and Fig. 8 shows the architecture for a PBICM point to point link having L states i.e. signal constellation of size 2^L. $\{m_i\}_{i=1}^L$ represent messages and $\mathbf{b}_i$ and $\mathbf{b}'_i$ the transmit codewords before and after dithering. The equivalent BMS channel can be characterized by L, the SNR of the underlying AWGN channel and the symbol mapping in modulation. In the rest of this paper we consider that *Gray* mapping is used.

In order to use PBICM with the relay channel, a definition of quantize-and-map operation under PBICM is required. With a PBICM modulator at source S, the observations at relay



(a) PBICM Transmitter



(b) PBICM Receiver

Fig. 8. PBICM architecture. $\{\mathbf{d}_i\}_{i=1}^{L}$ are dithers, and $\mathbf{d}'_i = 1 - 2\mathbf{d}_i$.

R ($\mathbf{y}_R$) represent L interleaved codewords. If R performs quantization at the *symbol* level, then the decomposition into independent binary sub-channels will be lost. As an alternative, it is proposed that R perform quantization at the *bit* level.

S and R both use PBICM modulator blocks with constellation size 2^L having state and dither vectors given by $\mathbf{s}_S, \mathbf{s}_R, \mathbf{D}_S$ and $\mathbf{D}_R$ respectively. The QMF operation at R is described below (depicted in Fig. 9):

1) For observed symbol sequence $\mathbf{y}_{SR} := \{y_{SR,j}\}_{j=1}^{fN_S}$ perform PBICM demodulation. The output is represented as $\{\mathbf{n}_{SR,i}\}_{i=1}^{L}$ where each $\mathbf{n}_{SR,i} := \{n_{SR,i,j}\}_{j=1}^{fN_S}$ represents LLRs for the i^{th} codeword.
2) Quantize every LLR in $\{\mathbf{n}_{SR,i}\}_{i=1}^{L}$. As an example, for a one bit scalar quantizer this simply involves observing the sign of LLRs.
3) Encode the quantizer output $\{m_{R,i}\}_{i=1}^{L}$ using an LDGM code.
4) Transmit the resultant codewords $\{\mathbf{b}_{R,1}\}_{i=1}^{L}$ using a PBICM modulator.

Using this definition of QMF, the Gaussian relay channel is decomposed into parallel BMS relay channels. The BMS relay channel is shown in Fig. 10. It is characterized by constellation size at S and R and the SNR of the underlying AWGN links i.e. $\mathsf{SNR}_{SR}, \mathsf{SNR}_{SD}, \mathsf{SNR}_{RD}$.

V. Link Design Example

In Section III an extension of density evolution tools was developed [21][22] for joint LDPC-LDGM factor graphs based on QMF relaying. In this section, a link design example with construction of explicit codes is shown for a DBLAST-equivalent channel shown in Fig. 5 BMS relay channel. The performance of designed codes is presented using simulations with high order modulation based on PBICM principles described in Section IV.

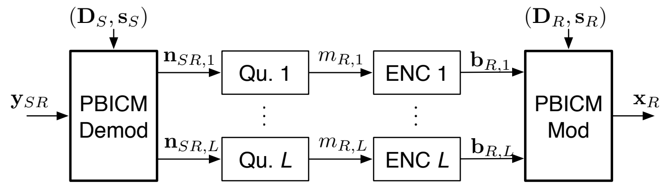


Fig. 9. QMF relaying with PBICM.

A. System Parameters

The capacity advantage of cooperative relaying is most pronounced when the source to relay link is significantly better than the direct link between source and destination. We therefore consider an example scenario where the S to R link is 10 dB stronger than the others.

$$\mathsf{SNR}_{SD} = \mathsf{SNR}_{RD}, \quad \mathsf{SNR}_{SR} = 10 \times \mathsf{SNR}_{SD} \tag{4}$$

1) Modulation Order: As a guideline for system design use the following information-theoretic bound on maximal achievable rate using QMF relaying with continuous Gaussian inputs x_S and x_R and a vector Gaussian quantizer at the noise level.

$$\mathcal{R}_{\text{QMF,G}} = \tag{5}$$
$$\min\left\{ \begin{array}{l} (1-f)\mathfrak{C}_{\text{G}}(\mathsf{SNR}_{SD}) + f\mathfrak{C}_{\text{G}}\left(\frac{\mathsf{SNR}_{SR}}{2} + \mathsf{SNR}_{SD}\right), \\ (1-f)\mathfrak{C}_{\text{G}}(\mathsf{SNR}_{RD}) + \mathfrak{C}_{\text{G}}(\mathsf{SNR}_{SD}) - f \end{array} \right\}$$

Here $\mathfrak{C}_{\text{G}}(x) := \log(1+x)$ is the AWGN point-to-point capacity at signal-to-noise ratio x. If the inputs are constrained to structured constellations such as 16 QAM, 64 QAM, then the achievable rate with 2^{2n}-QAM modulation and BICM is computed as follows:

$$\mathcal{R}_{\text{QMF},n} = \tag{6}$$
$$\min\left\{ \begin{array}{l} (1-f)\mathfrak{C}_{n}(\mathsf{SNR}_{SD}) + f\mathfrak{C}_{n}\left(\frac{\mathsf{SNR}_{SR}}{2} + \mathsf{SNR}_{SD}\right), \\ (1-f)\mathfrak{C}_{n}(\mathsf{SNR}_{RD}) + \mathfrak{C}_{n}(\mathsf{SNR}_{SD}) - f \end{array} \right\}$$

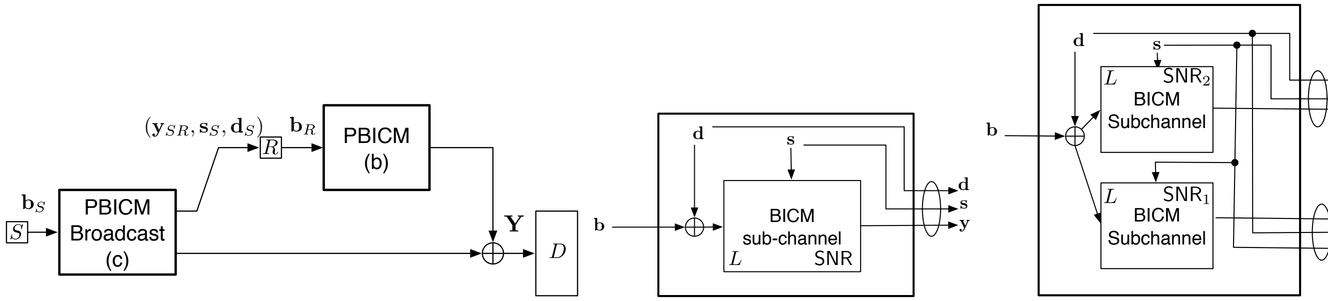
Here too we use a vector Gaussian quantizer at the noise level. Note that $n \in \{2,3,4\}$ and $\mathfrak{C}_n(x)$ denotes the 2^{2n}-QAM constellation-constrained point-to-point capacity at signal-to-noise ratio x under BICM.

2) Listening-Time Fraction: For QMF, the listening-time fraction f at R can be independently optimized to maximize system throughput [2], [5], [35]. The optimal f^* is found by balancing the two terms in the minimization of (5):

$$(1-f^*)\mathfrak{C}_{\text{G}}(\mathsf{SNR}_{SD}) + f^*\mathfrak{C}_{\text{G}}\left(\frac{\mathsf{SNR}_{SR}}{2} + \mathsf{SNR}_{SD}\right)$$
$$= (1-f^*)\mathfrak{C}_{\text{G}}(\mathsf{SNR}_{RD}) + \mathfrak{C}_{\text{G}}(\mathsf{SNR}_{SD}) - f^*$$

Alternatively a sub-optimal listening fraction f can be used based on reduced channel knowledge at relay. It is shown in [2] that this does not have a significant impact on throughput.

For system parameters in Eq (4), $\mathcal{R}_{\text{QMF,G}}$ and $\mathcal{R}_{\text{QMF},n}$ are plotted for $n = 2,3,4$ in Fig. 11 vs. SNR_{SD}. For each point, the optimized listening fraction f^* is used. To design a link with throughput of 5.4 information bits per symbol, both 64 QAM and 256 QAM are potentially good choices for



(a) Equivalent relay channel. (b) Equivalent point-to-point channel. (c) Equivalent broadcast channel.

Fig. 10. Equivalent binary-input system.

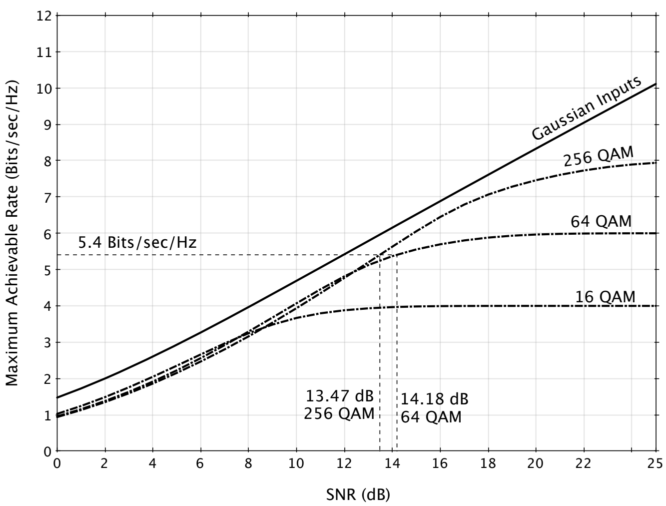


Fig. 11. Maximum achievable rate for QMF relaying with modulation constraints on channel inputs plotted vs SNR_{SD} for SNR relationships in Eq. (4).

modulation having QMF information theoretic thresholds at 14.18 dB and 13.47 dB respectively. Let us choose 64 QAM (6 coded bits per symbol) for the example design, which means that S should use an LDPC code of rate $\mathcal{R} = \frac{5.4}{6} = 0.9$. The optimal listening fraction corresponding to $\mathsf{SNR}_{SD} = 14.18$ dB is $f^* \approx \frac{2}{3}$. This determines the LDGM coding rate

$$\frac{K_R}{N_R} = \frac{f}{1-f} \approx 2$$

B. Code Design

Codes $\mathcal{C}_S^b$ and $\mathcal{C}_R^b$ optimized for the above system parameters can be designed using density evolution tools [21]. This involves finding good degree profiles that have the lowest possible decoding SNR threshold and randomly generating finite block length codes from them.

In order to reduce the computational complexity of density evolution we use the Gaussian approximation to density evolution developed in [36]. Additionally, we use the following heuristics to reduce the search space for profiles.

1) For $\mathcal{C}_S^b$ we consider check degree profiles that are concentrated [36] i.e. all check degrees (from edge perspective) are either k or $k+1$ for some integer $k \geq 2$.
2) For $\mathcal{C}_S^b$ we consider variable degree profiles with maximum degree of 8.
3) For $\mathcal{C}_R^b$ we limit ourselves to regular LDGM profiles.

Using these heuristics we design the following degree profile for the system parameters in this example.

$$\lambda_S(x) = 0.28x + 0.32x^2 + 0.28x^3 + 0.12x^6 + 0.0009x^7$$
$$\rho_S(x) = 0.04x^{28} + 0.96x^{29}$$
$$\lambda_R(x) = x^4, \rho_R(x) = x^9$$

Simulation results for the bit error rate in decoding of $\mathbf{b}_S$ using codes (with block lengths $\approx 10^4$ and $\approx 10^5$) drawn from above profiles are shown in Fig. 12(a) using PBICM with 64QAM modulation, one bit scalar quantizer and an ideal interleaver. As shown the BER performance is $\leq$ 1dB of the QMF threshold. For the single relay scenario, the information-theoretic thresholds for QMF and CF are identical, therefore as a reference for comparison thresholds for DF, AF, and the no-cooperation case are also shown. The DF and the AF thresholds are computed using the following expressions. Derivations follow standard analysis of the schemes and are omitted here.

$$\mathcal{R}_{\mathrm{DF},n} = \max_{f\in[0,1]}$$
$$\min\left\{f\mathfrak{C}_n\left(\mathsf{SNR}_{SR}\right), (1-f)\mathfrak{C}_n\left(\mathsf{SNR}_{RD}\right) + \mathfrak{C}_n\left(\mathsf{SNR}_{SD}\right)\right\}$$
$$\mathcal{R}_{\mathrm{AF},n} = \frac{1}{2}\mathfrak{C}_n\left(\mathsf{SNR}_{SD}\right) + \frac{1}{2}\mathfrak{C}_n\left(\mathsf{SNR}_{\mathrm{eff}}\right)$$
$$\mathsf{SNR}_{\mathrm{eff}} = \mathsf{SNR}_{SD} + \frac{\mathsf{SNR}_{SR}\mathsf{SNR}_{RD}}{1+\mathsf{SNR}_{SR}+\mathsf{SNR}_{RD}}$$

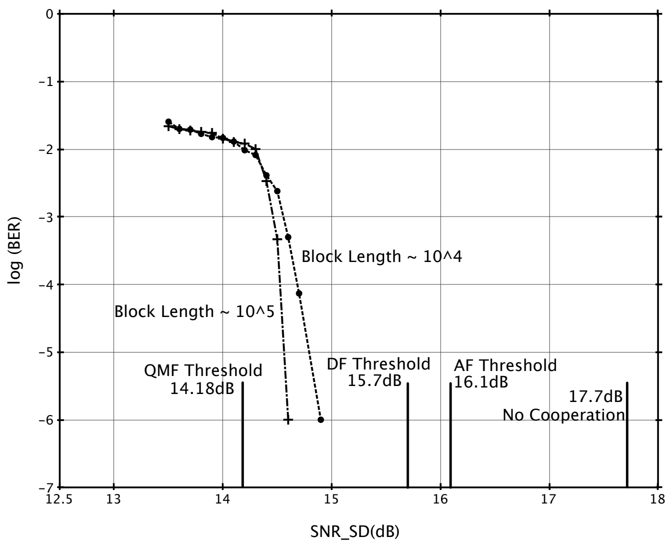
The optimal listening time for DF is determined by the channel parameters, while that for AF is always $1/2$.

For the DBLAST architecture, $\mathbf{b}_R$ must also be reliably decoded at or below the target SNR (for successive interference cancellation to work). Fig. 12(b) shows the BER for $\mathbf{b}_R$ which is also within $\leq$ 1dB of the QMF threshold for both of the block-lengths.

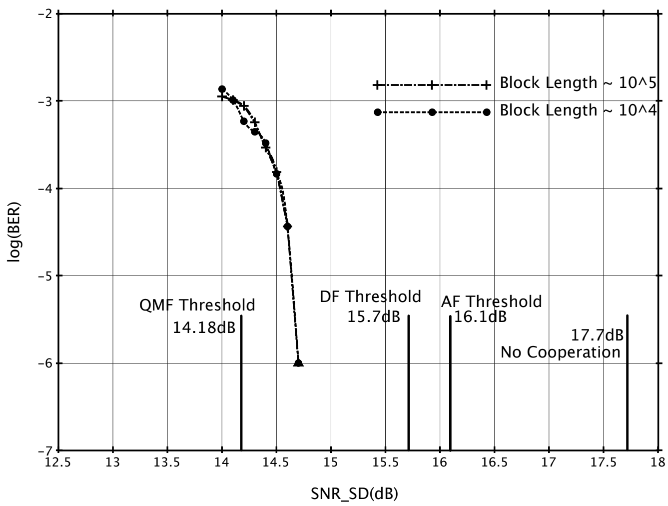
VI. Conclusions

The QMF relaying scheme has the following key advantages over other known relaying schemes such as AF, DF, and CF.

1) For the single relay network, it outperforms AF and DF at high SNR.



(a) BER for $\mathbf{b}_S$ using design rate of 5.4bits/sec/Hz with 64QAM.



(b) BER simulation for $\mathbf{b}_R$ using design rate of 5.4bits/sec/Hz with 64QAM.

Fig. 12. Code design simulation results.

2) For the single relay network, it achieves the same performance as CF but reduces channel feedback overhead. Unlike CF, QMF does not require knowledge of forward channel strength at the relay.
3) For arbitrary relay networks with multiple relays, QMF achieves better high SNR performance than AF, DF and CF.

In this paper, a low-complexity channel coding framework is developed for QMF relaying. For the single relay network, the framework performs within $(0.5-1)$dB of fundamental limits.

The techniques presented here can be extended to complex system scenarios, which are discussed below.

A. Multiple Relays

When there is more than one relay in the system, the proposed factor graph extends in a straightforward manner. Optimal listening schedules can be computed for each of the relays. As proposed, the source would use an LDPC code and each relay would use an LDGM code based on its respective schedule. The joint factor graph would include multiple LDGM sub-graphs.

The DBLAST architecture proposed in this paper extends naturally to networks with one level of multiple non-interfering relays e.g. the diamond network. As discussed previously, DBLAST significantly reduces the complexity of the factor graph. DBLAST requires that all codewords from relays are decoded correctly at destination in order to permit successive interference cancellation. This additional constraint does not lead to a reduction in the QMF information-theoretic achievable rate. In fact, such a requirement is explicitly considered in the probability of error analysis for the QMF scheme in [6].

However, some challenges for multiple relay networks remain to be addressed. When the relays can hear one another or the source can reach the destination via *multiple hops*, it is unclear how the DBLAST architecture can be applied. In such scenarios, an alternate space-time architecture must be considered. Moreover as the number of relays increase, the channel knowledge overhead required to compute optimal listening schedules becomes large. Practical techniques at the physical and MAC layers are required to address this complexity. These are considered as directions for future work.

B. Rate Adaptation and Hybrid ARQ

In the link design example, suitable coding rates, constellation and listening fraction are computed for a given set of operating channel conditions. However, optimizing codes based on instantaneous channel conditions is not feasible in practice. Under commonly used rate adaptation mechanisms, terminals switch between a few candidate codes and a few candidate constellations based on channel conditions. Cooperative links need to consider multiple channel parameters to determine transmission rates i.e. for a single relay three SNR parameters are required as opposed to just one for a point-to-point link. This makes rate adaptation schedules for relay networks more complex. An advantage of QMF relaying is that rate adaptation schedules depend only on the ability of the destination to decode as opposed to DF, where adaptation must consider decoding at relays as well.

Modern adaptation mechanisms like hybrid automatic repeat request (HARQ) can be incorporated into the proposed framework. Additional parity bits for refinement sent from the source after receiving a repeat request from the destination. It can be cooperatively delivered to the destination using QMF relaying. The joint decoding factor graph is expanded to incorporate these refinement parity bits and the decoding algorithm remains unchanged.

ACKNOWLEDGEMENTS

The authors acknowledge Prof. Rüdiger Urbanke for fruitful discussions leading to the choice of LDPC-LDGM structures. We also acknowledge the students, faculty and sponsors of the Berkeley Wireless Research Center and support of the Center for Circuit & System Solutions (C2S2) Focus Center, one of six research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program.

APPENDIX

The QMF relaying scheme introduces correlation between $\mathbf{x}_S$ and $\mathbf{x}_R$, which can be thought of as coding across transmit antennas in a MIMO channel. A natural space-time architecture for such a channel is DBLAST. Using DBLAST for the relay channel has also been proposed in [12][11][29][30]. It relies on introducing a delay of one block at the relay and using successive interference cancellation (SIC) at the destination. At the k-th block the destination receives the superposition of the following:

- signal from the source containing the codeword sent at block k, namely, $\mathbf{x}_S(m_k)$
- signal from the relay containing the side information about the source's codeword at block $k-1$, namely, $\mathbf{x}_R(q_{k-1})$

Messages sent from the source are independent across blocks. At the k-th block, the destination jointly decodes block $k-1$ (message m_{k-1} and side information $\mathbf{x}_R(q_{k-1})$) by treating $\mathbf{x}_S(m_k)$ as Gaussian noise. The receiver subtracts relay's codeword $\mathbf{x}_R(q_{k-1})$ from its received signal $\mathbf{Y}[k]$ and keeps the residual $\widetilde{\mathbf{Y}}[k]$ for decoding the next block. This architecture allows the use of a simplified equivalent channel model. Note that the one-block delay introduced at R has the added benefit of allowing time for QMF processing at R.

1) Simplified Channel Model: The equivalent channel model is shown in Fig. 5. For decoding the block $k-1$ message m_{k-1}, the decoder takes two inputs $\mathbf{Y}[k]$ and $\widetilde{\mathbf{Y}}[k-1]$. We can think of $\mathbf{Y}[k]$ and $\widetilde{\mathbf{Y}}[k-1]$ as two orthogonal links with independent Gaussian noise. Therefore, for the purpose of code design we can alternatively investigate a simpler model depicted in Fig. 5. In this model,

$$\mathbf{Y}_{ij} = \mathbf{h}_{ij}\mathbf{x}_i + \mathbf{Z}_{ij},\ (i,j) = (R,D),(S,D),$$
$$\mathbf{y}_{SR} = h_{SR}\mathbf{x}_S + \mathbf{z}_{SR}$$

As an example, let us consider a scenario where D has two receive antennas ($M=2$). In that case, the DBLAST equivalent channel becomes [37]:

$$\mathbf{y}_{ij} = h_{ij}\mathbf{x}_i + \mathbf{z}_{ij},\ (i,j) = (R,D),(S,D),$$

where,

$$h_{SD} = ||\mathbf{h}_1||,\ h_{RD} = \sqrt{||\mathbf{h}_{2\perp 1}||^2 + \frac{||\mathbf{h}_{2\parallel 1}||^2}{1+P_S||\mathbf{h}_1||^2}}$$

$\mathbf{h}_{2\perp 1}$ and $\mathbf{h}_{2\parallel 1}$ denote the perpendicular and parallel components of $\mathbf{h}_2$ with respect to $\mathbf{h}_1$, respectively. The signal-to-noise ratios of the three links are $\mathsf{SNR}_{SR} = |h_{SR}|^2P_S$, $\mathsf{SNR}_{SD} = |h_{SD}|^2P_S$, and $\mathsf{SNR}_{RD} = |h_{RD}|^2P_R$ respectively.

Remark 1: Consider the original channel and the DBLAST-equivalent channel. Note that the capacities of these two channels are within two bits of each other. This is based on the following observations:

1) *The min-cut upper bound for both channels are within one bit of each other (for any listening fraction $f \in [0,1]$).*
The mutual information across cut $\{S\},\{R,D\}$ remains unchanged between the two channels. Consider the mutual information across the cut $\{S,R\},\{D\}$. It is known that SIC achieves the sum capacity of multiple-access channels. In the original channel (Fig. 2) S and R have unlimited cooperation. As a result, the min-cut bound for DBLAST incurs a power-gain loss of at most $(1-f)$ bits.
2) *QMF relaying scheme achieves the min-cut upper bound to within one bit for the two channels* [1].

REFERENCES

[1] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
[2] V. Nagpal, S. Pawar, D. Tse, and B. Nikolic, "Cooperative multiplexing in the multiple antenna half duplex relay channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2009, pp. 1438–1442.
[3] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
[4] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[5] S. Pawar, A. Avestimehr, and D. Tse, "Diversity-multiplexing tradeoff of the half-duplex relay channel," in *Proc. 46th Annual Allerton Conf. Commun., Control, Computing*, 2008, pp. 27–33.
[6] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
[7] I.-H. Wang and D. Tse, "Interference mitigation through limited receiver cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2913–2940, May 2011.
[8] B. Zhao and M. Valenti, "Distributed turbo coded diversity for relay channel," *Electron. Lett.*, vol. 39, no. 10, pp. 786–787, May 2003.
[9] T. Hunter and A. Nosratinia, "Cooperation diversity through coding," in *Proc. IEEE Int. Symp. Inf. Theory*, 2002, p. 220.
[10] M. Janani, A. Hedayat, T. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: Space-time transmission and iterative decoding," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 362–371, Feb. 2004.
[11] Z. Zhang and T. Duman, "Capacity-approaching turbo coding and iterative decoding for relay channels," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1895–1905, Nov. 2005.
[12] Z. Zhang and T. M. Duman, "Capacity approaching turbo coding for half-duplex relaying," *IEEE Trans. Commun.*, vol. 55, no. 9, p. 1822, Sep. 2007.
[13] A. Chakrabarti, A. D. Baynast, A. Sabharwal, and B. Aazhang, "Low density parity check codes for the relay channel," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 280–291, Feb. 2007.
[14] P. Razaghi and W. Yu, "Bilayer low-density parity-check codes for decode-and-forward in relay channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3723–3739, Oct. 2007.
[15] T. V. Nguyen, A. Nosratinia, and D. Divsalar, "Bilayer protograph codes for half-duplex relay channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2010, pp. 948–952.
[16] P. Razaghi, M. Aleksic, and W. Yu, "Bit-interleaved coded modulation for the relay channel using bilayer LDPC codes," in *Proc. 10th Canadian Workshop Inf. Theory (CWIT)*, 2007, pp. 101–104.
[17] M. Uppal, Z. Liu, V. Stankovic, and Z. Xiong, "Compress-forward coding with BPSK modulation for the half-duplex Gaussian relay channel," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4467–4481, 2009.
[18] M. Uppal, G. Yue, X. Wang, and Z. Xiong, "A rateless coded protocol for half-duplex wireless relay channels," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 209–222, Jan. 2011.
[19] A. Ozgur and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," *ArXiv e-prints*, May 2010.
[20] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[21] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
[22] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
[23] Y. Fan, H. Poor, and J. Thompson, "Cooperative multiplexing in full-duplex multi-antenna relay networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2008, pp. 1–5.

[24] V. Nagpal, I.-H. Wang, M. Jorgovanovic, D. Tse, and B. Nikolić, "Quantize-map-and-forward relaying: Coding and system design," in *Proc. 48th Annual Allerton Conf. Commun., Control, Computing*, Oct. 2010, pp. 443–450.
[25] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[26] S. Aji and R. McEliece, "The generalized distributive law," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 325–343, Mar. 2000.
[27] A. Bennatan, S. Shamai, and A. R. Calderbank, "In praise of bad codes for multi-terminal communications," *CoRR*, vol. abs/1008.1766, 2010.
[28] E. Martinian and J. S. Yedidia, "Iterative quantization using codes on graphs," *CoRR*, vol. cs.IT/0408008, 2004.
[29] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical J.*, vol. 1, no. 2, pp. 41–59, 1996. [Online]. Available: http://dx.doi.org/10.1002/bltj.2015
[30] G. Kramer, "Distributed and layered codes for relaying," in *Proc. 39th Asilomar Conf. Signals, Syst. Computers*, Oct. 2005, pp. 1752–1756.
[31] G. Kraidy, N. Gresset, and J. Boutros, "Coding for the non-orthogonal amplify-and-forward cooperative channel," in *Proc. Inf. Theory Workshop (ITW)*, 2007, pp. 626–631.
[32] M. Benjillali and L. Szczecinski, "A simple detect-and-forward scheme in fading channels," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 309–311, May 2009.
[33] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, 2003.
[34] A. Ingber and M. Feder, "Parallel bit interleaved coded modulation," in *Proc. Annual Allerton Conf. Commun., Control, Computing*, Sep. 2010.
[35] M. Yuksel and E. Erkip, "Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3371–3393, Oct. 2007.
[36] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.
[37] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

Vinayak Nagpal received the B. Eng. degree from the University of Pune, India, in 2003, and the M.S. degree from Chalmers University of Technology, Sweden, in 2006. He received the Ph.D. degree from the University of California at Berkeley, USA, in 2012 under the guidance of Prof. Borivoje Nikolić. Since then, he is affiliated with the Nokia Research Center, Berkeley, USA. Previously, he held positions at Conexant Systems, Pune, India (2003), the National Radio Astronomy Observatory, Charlottesville, VA (2005), and the Harvard Smithsonian Center for Astrophysics, Cambridge, MA (2006). His research interests include wireless networks and real time signal processing.

I-Hsiang Wang received the B.S. degree in electrical engineering from National Taiwan University, Taiwan, in 2006. He received a Ph.D. degree in electrical engineering and computer sciences from the University of California at Berkeley, USA, in 2011. Since 2011, he has been affiliated with the École Polytechnique Fédérale de Lausanne, Switzerland, as a postdoctoral researcher. His research interests include network information theory, wireless networks, coding theory, and network coding. Dr. Wang received a 2-year Vodafone Graduate Fellowship in 2006.

Milos Jorgovanovic received his Dipl. Ing. degree in electrical engineering from the University of Belgrade, Serbia, in 2007, and the M.Sc. degree from the University of California at Berkeley in 2010. He is currently working towards his Ph.D. degree at the University of California at Berkeley under the guidance of Prof. Borivoje Nikolić. He has held internship positions with the Kodak European Research Center in Cambridge, UK (2006), the Technical University of Berlin, Germany (2009), and Samsung Mobile in Richardson, TX (2010). His research interests include MIMO detection algorithms and architectures, wireless communication systems design, signal processing for digital communications, and digital integrated circuit design.

David Tse received the B.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1989, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1991 and 1994, respectively. From 1994 to 1995, he was a Postdoctoral Member of Technical Staff at the Department of AT&T Bell Laboratories. Since 1995, he has been with the Department of Electrical Engineering and Computer Sciences at the University of California at Berkeley, where he is currently a Professor. Dr. Tse received a 1967 NSERC 4-year graduate fellowship from the government of Canada in 1989, a NSF CAREER award in 1998, Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, the IEEE Communications and Information Theory Society Joint Paper Award in 2001, the Information Theory Society Paper Award in 2003, and the 2009 Frederick Emmons Terman Award from the American Society for Engineering Education. He has given plenary talks at international conferences such as ICASSP in 2006, MobiCom in 2007, CISS in 2008, and ISIT in 2009. He was the Technical Program Co-chair of the International Symposium on Information Theory in 2004 and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2003. He is a co-author, with P. Viswanath, of the text *Fundamentals of Wireless Communication*, which has been used in over 60 institutions around the world.

Borivoje Nikolić received the Dipl.Ing. and M.Sc. degrees in electrical engineering from the University of Belgrade, Serbia, in 1992 and 1994, respectively, and the Ph.D. degree from the University of California, Davis in 1999. He lectured electronics courses at the University of Belgrade from 1992 to 1996. He spent 2 years with Silicon Systems, Inc., Texas Instruments Storage Products Group, San Jose, CA, working on disk-drive signal processing electronics. In 1999, he joined the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, where he is now a Professor. His research activities include digital and analog integrated circuit design and VLSI implementation of communications and signal processing algorithms. He is a co-author of the book *Digital Integrated Circuits: A Design Perspective* (2nd ed., Prentice-Hall, 2003). Dr. Nikolić received the NSF CAREER award in 2003, the College of Engineering Best Doctoral Dissertation Prize and Anil K. Jain Prize for the Best Doctoral Dissertation in Electrical and Computer Engineering at University of California, Davis in 1999, as well as the City of Belgrade Award for the Best Diploma Thesis in 1992. For work with his students and colleagues, he has received best paper awards at the IEEE International Solid-State Circuits Conference, Symposium on VLSI Circuits, IEEE International SOI Conference, and the ACM/IEEE International Symposium of Low-Power Electronics.