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

Vinayak Nagpal, I-Hsiang Wang, Milos Jorgovanovic, David Tse, Borivoje Nikolić
EECS Dept. University of California Berkeley, USA
Email: {vnagpal,ihsiang,milos,dtse,bora}@eecs.berkeley.edu

Abstract—**Quantize-map-and-forward (QMF) is a relaying scheme that has been shown to achieve the capacity of Gaussian relay networks to within a constant gap. Under QMF the compression indices forwarded by relays are not decoded explicitly. Instead, the message is decoded jointly with the compression indices. In this work we present a practical coding and signaling framework that captures this aspect of QMF. We outline the framework for a simple Gaussian network with a single half-duplex relay. We propose a scheme where binary LDPC codes are used for encoding at the source and at the relay. For suitable choice of codes and quantizer we show that the joint decoding operation for QMF can be reduced to belief propagation over a Tanner graph. Simulation results are given to justify the proposed coding scheme. The proposed framework can be extended for use with high order signal constellations.**

I. INTRODUCTION

A cooperation strategy outlines the procedure used to forward information from source to destination in relay networks. The amplify-and-forward (AF), decode-and-forward (DF) and compress-and-forward (CF) [1][2] strategies have received the most attention in literature. In DF the information forwarded by relays is based on a hard estimate of the encoder's message whereas in AF and CF it is based on a soft estimate. Additionally DF and CF forward coded information whereas AF forwards it in uncoded form. The simplest network configuration has a single relay and in the context of wireless networks the relay is modeled as a half-duplex terminal. For such a network with additive white Gaussian noises it is known that CF outperforms DF and AF [3]. In fact, CF can perform within a constant gap to the information theoretic capacity, regardless of channel parameters [4][5].

More recently there has been a growing interest in co-operative networks with more relays. In such configurations CF no longer performs within a constant gap to capacity. Moreover CF requires the relays to have forward channel knowledge. The overhead for this becomes large for more complex networks. Recent work [5] has shown that an improvement of CF called quantize-map-and-forward (QMF) achieves performance within constant gap of capacity for Gaussian networks with arbitrary number of relays[1]. In the QMF scheme the relay quantizes its received signal at noise level, maps it randomly to a Gaussian codeword and forwards it. QMF does not require forward channel knowledge at relays and has played a key role in several information theoretic

results on cooperative networks [6][4][7][8]. These properties make QMF an interesting candidate for further investigation.

Information theoretic results prove the existence of codes and relay operations that can be used in cooperative networks. For the purpose of system design we need to explicitly construct efficient coding and signaling techniques. Our goal is to develop the QMF scheme from this perspective. In this paper we present a practical coding and signaling framework based on QMF for the simple network configuration with one half-duplex relay.

The fundamental difference between CF and QMF schemes is in their respective decoding operations. CF uses *successive decoding* i.e. side information from the relay (compression index) is decoded first before decoding the message. The CF relay explicitly encodes side information such that it can be reliably decoded at the destination and therefore requires forward channel knowledge. On the other hand quantize-map-and-forward uses *joint decoding* of the message and side information at the destination. Joint decoding usually requires higher complexity and this presents a unique challenge in the design of practical codes for QMF. The key contribution of this paper is to present techniques that significantly reduce the complexity of this step.

Recently a coding technique for QMF [9] was proposed based on lattice strategies. The proposed scheme reduced the complexity of mapping at the relay to polynomial-time while the joint decoding complexity remained exponential-time. In this paper we use an alternative approach wherein we treat coding and modulation separately, by employing bit-interleaved coded modulation (BICM) [10]. The technique allows us to combine channel codes for binary alphabet and higher order signal constellations with only minor loss in performance. Therefore we focus our discussion to the problem of code design and specifically joint decoding for the *binary input* Gaussian relay channel (Sec II).

The binary-input relay channel with QMF has two distinct code components, first is the encoder at the source and second is the mapping at the relay. We choose both these components to be LDPC codes. This is a natural choice given the fact that they have the best known performance (for point-to-point channels). LDPC codes are well suited for practical implementation and have a well developed set of analysis and design tools. LDPC codes for the relay channel with DF relaying have been developed in [11][12][13]. The DF coding ideas initially developed for the binary-input case have

---

[1]Lim *et.al.* describe a generalized version of this scheme in [6] and name it *noisy network coding*.

been extended to high SNR regime using bit-interleaved coded modulation (BICM) in [14]. Code designs for CF relaying using binary inputs have been discussed in [15].

We focus on QMF relaying and represent the joint decoding problem as a sum-product algorithm over a factor graph (Sec.III). The resulting factor graph is not well suited for efficient implementation and good performance. We then present three techniques that allow us get a factor graph that is well conditioned for practical implementation (Sec IV). We show that for a suitable choice of component codes the factor graph contains the Tanner graphs of both component codes as subgraphs. The two subgraphs are connected via *new* nodes that represent the quantization operation at the relay. We call them *quantize* (QU) nodes and compute their sum product update rules (Sec V). For a simple one-bit scalar quantizer at the relay we show that the QU nodes can be simplified into LDPC check nodes connected to a dummy variable node sending a constant message which is the bit error probability in the source-to-relay link. In Sec. VI we present results from simulating the resulting joint decoder using off-the-shelf LDPC codes. Finally in Sec VII we discuss the extension of the ideas in this paper to spectrally efficient (high-SNR) regime using BICM and present preliminary simulation results for the same.

II. PROBLEM SETUP

The half-duplex Gaussian relay channel has three half-duplex terminals: source (S), relay (R) and destination (D) with Gaussian channels between them.

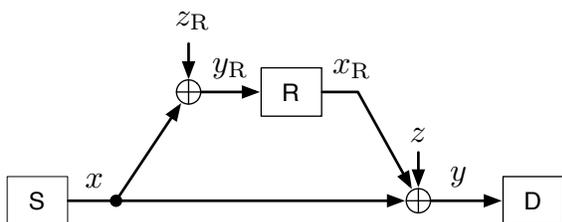### A. Half-duplex Single-Relay Network

Fig. 1.   Half-Duplex Gaussian Relay Channel

The half-duplex single-relay network is shown in Figure 1. R listens for fraction $f \in [0, 1]$ of the total communication time and transmits for fraction $(1 - f)$. The block lengths for the transmitted signals at the source S and the relay R are N_S and N_R respectively. They satisfy the half-duplex constraint $N_R = (1-f)N_S$. The transmit signals $\mathbf{x}_S \in \mathcal{A}^{N_S}, \mathbf{x}_R \in \mathcal{A}^{N_R}$, where $\mathcal{A}$ denotes the alphabet of the signal. Here we use the bold-face lower case alphabet to denote a sequence of symbols. In the rest of our paper we shall focus on the binary alphabet $\{\pm 1\}$.

The received signal sequences are $\mathbf{y} \in \mathbb{R}^{N_S}$, $\mathbf{y}_R \in \mathbb{R}^{fN_S}$, where

$$y = h_1 x_S + h_2 x_R + z \tag{1}$$
$$y_R = h_R x_S + z_R, \tag{2}$$

and h_1, h_2, h_R denote the corresponding channel gains. $h_2 = 0$ when R is listening. $\mathbf{z}$ and $\mathbf{z_R}$ are independent zero-mean Gaussian noise vectors with identity covariance matrices.

B. Quantize-Map-and-Forward Relaying Scheme

For the general continuous alphabet with average power constraints, the quantize-map-and-forward scheme [5] is described as follows.

Source S has a sequence of messages $w_b \in \{1, \ldots, 2^{N_S R}\}$, $b = 1, 2, \ldots$ to be transmitted. At both source S and relay R we create random Gaussian codebooks $\mathcal{C}_S$ and $\mathcal{C}_R$ respectively. S randomly maps each message to one of its Gaussian codewords and transmits it using N_S symbol times giving an overall transmission rate of R. Due to the half-duplex nature of the relay, it must operate using listen-transmit cycles. Relay listens to the first fN_S time symbols of each block. It quantizes its observation and then randomly maps it into a Gaussian codeword using a random mapping function. It transmits this codeword during the next $(1 - f)N_S$ symbol times. Given the knowledge of all the encoding functions and signals received, destination D attempts to decode the message sent by S.

We adapt the QMF scheme described above to the binary-input Gaussian relay network. The only difference is that, the codebooks $\mathcal{C}_S$ and $\mathcal{C}_R$ are now binary codes. Relay R quantizes its received signal into $\mathbf{y}_R$ into K_R bits and maps the quantization index to a codeword $\mathbf{x}_R$ in its codebook $\mathcal{C}_R$. For the majority of this paper we will focus on this simplified channel model. In Sec VII we use a parallel combination of these binary-input channels using BICM to approximate the original channel model described in Sec II-A.

III. JOINT DECODING ALGORITHM AND FACTOR GRAPH

The main result leading to a factor graph for the joint decoding algorithm is summarized in the following theorem.

Theorem 1: The maximum *a posterior* (MAP) joint decoding rule for the channel described in Sec. II-A can be expressed as a sum-product algorithm over a factor graph that has the Tanner graphs of $\mathcal{C}_S$ and $\mathcal{C}_R$ as its subgraphs.

Proof: The decoder searches for the codeword $\mathbf{x}_S \in \mathcal{C}_S$ which maximizes the *a posterior* probability $p(\mathbf{x}_S|\mathbf{y})$. An efficient way to do this search is to consider the bitwise maximum *a posterior* (MAP) decoder, where the aim is to compute $p(x_{S,i}|\mathbf{y}) = \sum_{\sim x_{S,i}} p(\mathbf{x}_S|\mathbf{y})$ for all $i = 1, 2, \ldots N_S$. Here the subscript i denotes the symbol time.

$$p(\mathbf{x}_S|\mathbf{y}) = \sum_{\mathbf{x}_R} \frac{f(\mathbf{y}|\mathbf{x}_S, \mathbf{x}_R) \, p(\mathbf{x}_S, \mathbf{x}_R)}{f(\mathbf{y})}$$
$$\propto \sum_{\mathbf{x}_R} f(\mathbf{y}|\mathbf{x}_S, \mathbf{x}_R) \, p(\mathbf{x}_S, \mathbf{x}_R). \tag{3}$$

The first term $f(\mathbf{y}|\mathbf{x}_S, \mathbf{x}_R)$ can be factorized as

$$f(\mathbf{y}|\mathbf{x}_S, \mathbf{x}_R)$$
$$= \prod_{i=1}^{fN_S} f(y_i|x_{S,i}) \prod_{j=1}^{N_R} f(y_{fN_S+j}|x_{R,j}, x_{S,(fN_S+j)}) \tag{4}$$

since the additive noises are independent over time. Now we focus on the use of LDPC codes for both $\mathcal{C}_S$ and $\mathcal{C}_R$. The codes have characteristic functions $1(\mathbf{x}_S \in \mathcal{C}_S)$ and $1(\mathbf{x}_R \in \mathcal{C}_R)$ respectively. Hence,

$$
\begin{aligned}
p(\mathbf{x}_S, \mathbf{x}_R) &= p(\mathbf{x}_S)\, p(\mathbf{x}_R|\mathbf{x}_S) \\
&\propto 1(\mathbf{x}_S \in \mathcal{C}_S)\, p(\mathbf{x}_R|x_S) \quad (5) \\
&\overset{(a)}{=} 1(\mathbf{x}_S \in \mathcal{C}_S)\, 1(\mathbf{x}_R \in \mathcal{C}_R)\, p(\mathbf{x}_R|\mathbf{x}_S). \quad (6)
\end{aligned}
$$

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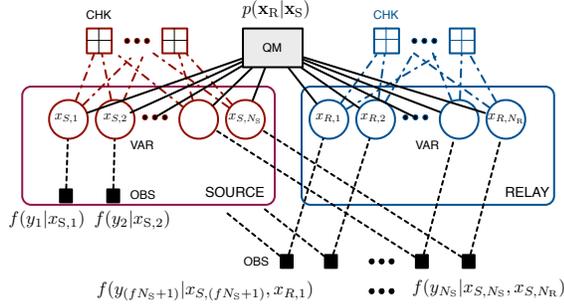

Fig. 2.   Factor Graph

To this end we have shown that the MAP joint decoding procedure can be expressed as a sum-product algorithm over a factor graph containing the Tanner graphs of $\mathcal{C}_S$ and $\mathcal{C}_R$ as its subgraphs hence the proof of Theorem 1 is complete. The factor graph is shown in Fig. 2. $\blacksquare$

The update rules at the function node $p(\mathbf{x}_R|\mathbf{x}_S)$ may get very complex since its arguments are of lengths N_S and N_R. Moreover the node $p(\mathbf{x}_R|\mathbf{x}_S)$ has a degree $N_S + N_R$ which can introduce very short cycles in the graph. Running the sum-product algorithm over such a graph would result in decoding performance which will not approach that of the MAP decoder. In order to get reasonable performance and low decoding complexity we need to simplify this factor graph. In the next section we discuss three specific techniques that allow this.

IV. SIMPLIFIED FACTOR GRAPH

A. Systematic Mapping

If we restrict the encoding at the relay to be in *systematic* form, the first K_R bits $x_R[1 : K_R]$ are the quantization index of $y_R[1 : fN_S]$. This reduces the degree of the $p(\mathbf{x}_R|\mathbf{x}_S)$ function node to $fN_S + K_R$. The complexity is also reduced because the node now represents only a quantization operation.

$$
p(\mathbf{x}_R|\mathbf{x}_S) = p(\mathbf{x}_R|\mathbf{x}_S[1 : fN_S]) \quad (7)
$$

B. Scalar Quantizer

It has been shown [5] that QMF is within constant gap of capacity even with a scalar quantizer. Therefore if we use a scalar quantizer at the relay i.e. each each symbol $y_{R,i}$ is quantized into $x_R[\alpha_i]$, for $i = 1, 2, \dots fN_S$ then the

$p(\mathbf{x}_R|\mathbf{x}_S)$ function node factorizes into fN_S separate nodes each representing a scalar quantization operation.

$$
p(\mathbf{x}_R|\mathbf{x}_S) = \prod_{i=1}^{fN_S} p(x_R[\alpha_i]|x_{S,i}) \quad (8)
$$

$$
\bigcup_{i=1}^{fN_S} \alpha_i = \{1, 2, \dots, K_R\}
$$

$$
\alpha_i \cap \alpha_j = \phi \; \forall i \neq j
$$

where α_i denotes the subset of time indices in $\mathbf{x}_R$ that symbol $y_{R,i}$ is quantized into. Now, the variable nodes of the two
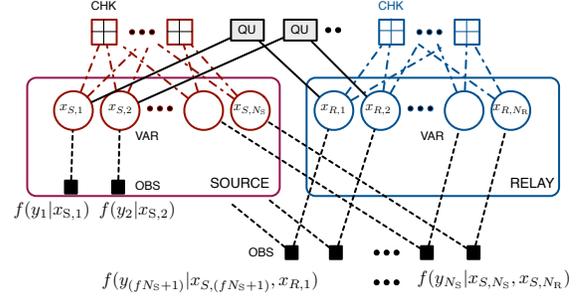

Fig. 3.   Factor Graph: Systematic code at relay and scalar quantizer.

Tanner graphs are connected by function nodes representing the stochastic relations $p(x_R[\alpha_i]|x_{S,i})$ among them. We call such nodes *quantize* (QU) nodes, as they are induced by the quantization procedure at the relay. In summary there are four kinds of nodes in the resulting factor graph: (1) Observation (OBS) nodes, (2) Variable (VAR) nodes, (3) Check (CHK) nodes, and (4) Quantize (QU) nodes. An illustration is given in Figure 3 for the case where each symbol observation $y_{R,i}$ is quantized into one bit.

As depicted in Fig. 3, some VAR nodes in the Tanner graph for the source's code share OBS nodes with some VAR nodes in the Tanner graph for the relay. This may not be desirable in a practical implementation due to the complexity of updates at these OBS nodes and the possibility of introducing short cycles and thereby degrading the performance of iterative message passing. In the next sub-section we propose a DBLAST-like layered system architecture to resolve this issue. This architecture allows us to use a simplified channel model where the destination observes two orthogonal AWGN links - one from the source and the other from the relay. Hence the OBS nodes connecting the source VAR nodes and the relay VAR nodes factorize and the resulting factor graph is simplified.

C. DBLAST Architecture

DBLAST (diagonal Bell-labs space time) was originally introduced for MIMO channels to allow coding across antennas for slow fading scenarios. The relay channel can be thought of as a distributed MIMO channel with coding across the transmit antennas (source and relay). This is a natural fit for the relay channel and we show that it provides several system design advantages. The system architecture is depicted
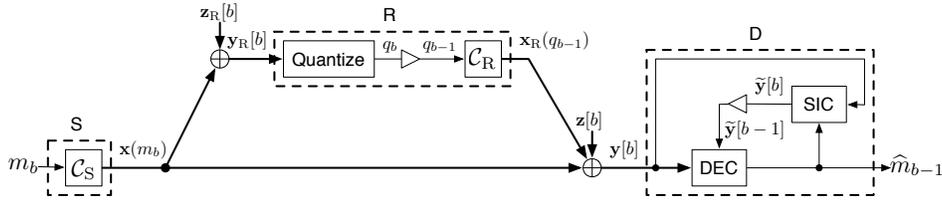
Fig. 4.   System Architecture at Block b

in Fig. 4 and its equivalent channel model is shown in Fig. 5. The architecture relies on introducing a delay of one block at the relay. At the b-th block the destination receives the superposition of the following:

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

Across blocks the messages sent from the source are independent and the destination can perform successive interference cancellation (SIC). At the b-th block, it jointly decodes block $b - 1$ (message m_{b-1} and side information $\mathbf{x}_R(q_{b-1})$) by treating $\mathbf{x}_S(m_b)$ as Gaussian noise. Then it subtracts relay's codeword $\mathbf{x}_R$ from its received signal $\mathbf{y}[b]$ and keeps the residual $\widetilde{\mathbf{y}}[b]$ for decoding the next block. This architecture allows us to use a simplified equivalent channel model.

Simplified Equivalent Channel Model:

Consider the end-to-end system in Figure 4. For decoding the block $b - 1$ message m_{b-1}, the decoder takes two inputs $\mathbf{y}[b]$ and $\widetilde{\mathbf{y}}[b - 1]$. Note that the noises are independent over time and the messages sent at all blocks are independent. We can think of $\mathbf{y}[b]$ and $\widetilde{\mathbf{y}}[b - 1]$ as two orthogonal links with independent additive Gaussian noises. Therefore, for the purpose of code design we can alternatively investigate a simpler model depicted in Figure 5. In this model,

$$y_{SR} = h_{SR}x_S + z_{SR} \qquad (9)$$

$$y_{SD} = h_{SD}x_S + z_{SD} \qquad (10)$$

$$y_{RD} = h_{RD}x_R + z_{RD} \qquad (11)$$

where $h_{SR} = h_R$, $h_{SD} = h_1$, $h_{RD} = h_2$, and z_{SR}, z_{SD}, and z_{RD} are independent Gaussian noises with variances

$$\mathrm{Var}[z_{SR}] = \mathrm{Var}[z_{SD}] = 1 \qquad (12)$$

$$\mathrm{Var}[z_{RD}] = 1 + |h_1|^2. \qquad (13)$$

The equivalent noise z_{RD} has higher variance since the signal from the source in block b is treated as Gaussian noise. We denote the signal-to-noise ratios in the three links by SNR_{SR}, SNR_{SD}, and SNR_{RD} respectively.

Remark 1: Consider the original channel and the equivalent channel under DBLAST *without* binary input constraint i.e. average power constraint is put at the transmitters. It is not difficult to show that the capacities of these two channels are within a constant gap. Note that the cut-set upper bounds for
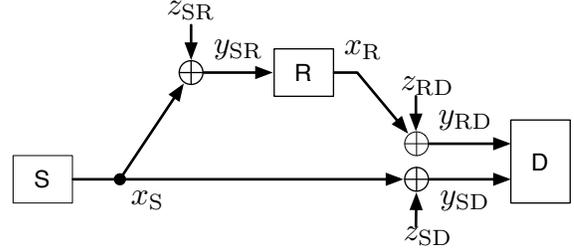

Fig. 5.   Equivalent Channel Model

both channels are within 1 bit (for any listening time $f \in [0, 1]$). Compared with the original channel in Fig. 1, the cut at the source remains unchanged, while the cut at the destination has a $[(1 - f) \times 1]$-bit power-gain loss since SIC achieves the sum capacity of multiple-access channels, and this cut in the original channel allows full cooperation of S and R. Since QMF achieves the cut-set bound to within a constant gap for arbitrary half-duplex Gaussian relay networks, the capacities are within a constant gap. Therefore for our original purpose of implementing QMF, such DBLAST architecture does not incur much loss.

This greatly simplifies the factor graph shown in Fig. 3. The OBS nodes of degree 2 (representing multiple access) can now be factorized. The destination observes two orthogonal set of observations. The factorization is shown in equation (14) where $\underline{\mathbf{y}} := [\mathbf{y}_{SD} \ \mathbf{y}_{RD}]$.

An example of the simplified factor graph is depicted in Figure 6. We see that in such factor graphs, VAR nodes in the two Tanner graphs are connected *only* through QU nodes. Since $y_{RD,i} = 0$ for $i = 1, \ldots, fN_S$, we rename $y_{RD,(fN_S+j)} \equiv y_{RD,j}$, for $j = 1, \ldots, N_R$.
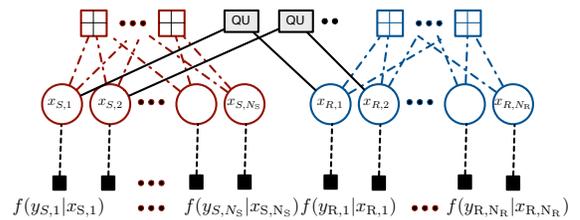

Fig. 6.   Simplified Factor Graph with One-bit Scalar Quantizer, that is, $\alpha_i = \{i\}$ for $i = 1, 2, \ldots, fN_S$.

After using these three techniques the resulting factor graph is very similar to a Tanner graph for an LDPC code. The only difference in the presence of QU nodes which have small

$$f(\underline{\mathbf{y}}|\mathbf{x}_S,\mathbf{x}_R) = \prod_{i=1}^{fN_S} f(y_{SD,i}|x_{S,i}) \prod_{j=1}^{N_R} f(y_{SD,(fN_S+j)}, y_{RD,(fN_S+j)}|x_{R,j}, x_{S,(fN_S+j)})$$

$$= \prod_{i=1}^{fN_S} f(y_{SD,i}|x_{S,i}) \prod_{j=1}^{N_R} f(y_{SD,(fN_S+j)}|x_{S,(fN_S+j)}) f(y_{RD,(fN_S+j)}|x_{R,j})$$

$$= \prod_{i=1}^{N_S} f(y_{SD,i}|x_{S,i}) \prod_{j=1}^{N_R} f(y_{RD,(fN_S+j)}|x_{R,j}), \tag{14}$$

degree (in our example 2). In the next section we derive the message passing updates for this graph following the general principle outlined in [16] and show that for a simple one-bit quantizer each QU node can be further factorized into an LDPC check constraint and a dummy variable node. This reduces the joint decoding algorithm to belief propagation over a Tanner graph.

V. MESSAGE PASSING ALGORITHM

For the point-to-point system, *belief propagation* is an iterative algorithm that computes the *a posterior* probability to decode message bits. The algorithm computes this exactly if the factor graph has no cycles. Otherwise it computes the approximate *a posterior* probability for each bit. For the joint decoding framework presented in previous sections the messages being passed on the edges of the factor graph and the update rules at the variable/check nodes remain the same. The only new ingredient in the mix are the QU nodes introduced by our framework. In this section we derive the update rules for QU nodes.

Let the subscripts V, C, and Q denote VAR nodes, CHK nodes, and QU nodes respectively. For $F \in \{C, Q\}$, let $\omega_{VF}^{(l)}$ denote the message sent from variable node V to function node F in the lth iteration. Every edge in the graph is connected to exactly one variable node and a message on the edge represents the *a posterior* probability for the respective variable. Since all the variables in our graph are binary, the messages can be represented as likelihood or log-likelihood ratios. For the sake of simplicity, in this discussion we represent the messages as a two-dimensional vector

$$\omega_{VF} = [p_1 \ p_{-1}]$$

where $\omega_{VF}(1) = p_1 \in [0, 1]$ represents the probability that the bit is 0 and $\omega_{VF}(2) = p_{-1} \in [0, 1]$ represents the probability that the bit is 1. ($p_1 + p_{-1} = 1$)

At a VAR node V, the incoming messages from CHK nodes and QU nodes are the likelihoods of the bit represented by V. The message sent from V to node $F \in \{C, Q\}$ is the normalized product of all incoming messages into node V except the message from F. The normalization ensures that $p_1 + p_{-1} = 1$ for the outgoing message.

At a CHK node C the incoming message from node V is the likelihoods the bits represented by V. Hence the message sent from C to any VAR node V' is the indicator function

that the check is satisfied, marginalized on the bit represented by V'. Note that the update rules at CHK node represent a deterministic constraint on the variables.

At a QU node Q, the message passed from Q to a VAR node V, is the marginalization of the function $p(x_R[\alpha_i]|x_{S,i})$ on the symbol represented by V. x_R is computed from a noisy observation of $x_{S,i}$ and therefore the QU node imposes a probabilistic constraint on the variables. Since the quantization is scalar, we have: $\forall \mathbf{u} \in \{\pm 1\}^{|\alpha_i|}$ and $v \in \{\pm 1\}$,

$$g(\mathbf{u}, v) := p(x_R[\alpha_i] = \mathbf{u}|x_{S,i} = v). \tag{15}$$

This function can be fully represented by a lookup table with $2^{|\alpha_i|+1}$ values, and hence the update rule can be derived.

As an example we consider a one-bit scalar quantizer at the relay and derive the update rules. We have the following theorem.
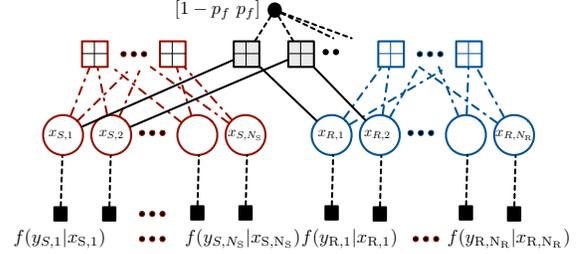

Fig. 7. Euivalent Factor Graph of that in Fig. 6

Theorem 2: If the relay uses a one-bit scalar quantizer and a systematic encoding i.e. $\alpha_i = \{i\}$ for $i = 1, 2, \ldots, fN_S$ and $K_R = fN_S$, then each QU node can be decomposed into to a CHK node connected with a dummy VAR node that sends a constant message:

$$\left[1 - Q\left(\sqrt{\text{SNR}_{SR}}\right) \quad Q\left(\sqrt{\text{SNR}_{SR}}\right)\right]. \tag{16}$$

Proof: Note that

$$\alpha_i = \{i\}, \ i = 1, 2, \ldots, fN_S, \tag{17}$$

and $K_R = fN_S$. The factor graph is depicted in Figure 6. We have: $\forall u \in \{\pm 1\}$ and $v \in \{\pm 1\}$,

$$g(u, v) := p(x_{R,i} = u|x_{S,i} = v)$$
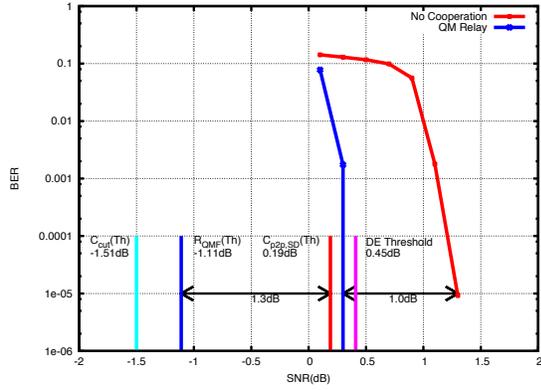$$= (1 - p_f)1\{u = v\} + p_f 1\{u \neq v\}, \tag{18}$$

Fig. 8. Simulation Results for $q_{SR} = 2$, $q_{RD} = 1$, $N_S = 20400$, $N_R = 13600$, $f = \frac{1}{3}$, $r_S = r_R = 0.5$

where $p_f := Q(\sqrt{SNR_{SR}})$ denotes the probability of error in scalar quantization. Since the function g is symmetric in u and v, we can assume that the VAR node is of the source, and the marginalization is on v. Let the other VAR node be V'. Hence,

$$\omega_{QV}(1) = (1 - p_f)\omega_{V'Q}(1) + p_f\omega_{V'Q}(2) \quad (19)$$
$$\omega_{QV}(2) = (1 - p_f)\omega_{V'Q}(2) + p_f\omega_{V'Q}(1), \quad (20)$$

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

Theorem 2 implies that the implementation of the iterative decoder follows the same structure as that in the conventional LDPC decoder. In the next section we present some simulation results based such implementation.

VI. SIMULATION RESULTS

In this section we present simulation results for sum product decoding on the factor graph shown in Fig. 6 i.e. using a one-bit quantizer at the relay. We use LDPC codes that are optimized for point-to-point communication on a AWGN channel from [17]. The signal to noise ratios on the channels are modeled as described in Sec. IV-C using linear scaling factors q_{SR} and q_{RD}.

$$SNR_{SR} = q_{SR}SNR_{SD}$$
$$SNR_{RD} = q_{RD}SNR_{SD}$$

Fig. 8 shows the case with $q_{SR} = 2$ and $q_{RD} = 1$ using rate 0.5 codes of block length 20400 and 13600 for the source and relay respectively. The codes are generated using profiles from Table 1 in [17] using maximum variable node degree of 8.

We compare our simulation results with the following information theoretic quantities

1) Point-to-point Channel:
- Shannon capacity:

$C_{p2p} = I(X;Y)$ with $X \sim \text{Unif}\{\pm 1\}$, $Y = X + Z$, Z is zero-mean Gaussian with variance $\frac{1}{SNR}$, and Z is independent of X.
- Density evolution threshold of the chosen code [17]

2) Half-Duplex Relay Channel:
- Cut-set upper bound of the original channel (Fig. 1): $C_{cut} = \min\{C_S, C_D\}$, where

$$C_S = fC_{MISO} + (1 - f)C_{p2p,SD} \quad (21)$$
$$C_D = fC_{p2p,SD} + (1 - f)C_{SIMO}. \quad (22)$$

Here $C_{p2p,SD}$ denotes the channel capacity for the S-D link, C_{MISO} denotes the channel capacity of the MISO channel from (S, R) to D, and C_{SIMO} denotes the channel capacity of the SIMO channel from S to (R, D). Note the all calculations are based on the channel model in Fig. 1. Note that the input alphabet $\mathcal{A} = \{\pm 1\}$.

- Cut-set upper bound of the equivalent channel under DBLAST (Fig. 5):
The expression of this bound, $C_{cut,DBLAST}$, is the same as above, except that the calculations are based on the channel model in Fig. 5. Since the S-D and R-D links are orthogonal, C_{MISO} becomes the sum of point-to-point capacities of S-D and R-D links: $C_{MISO} = C_{p2p,SD} + C_{p2p,RD}$.

- QMF achievable rate using the scalar one-bit quantizer:
This is the QMF achievable rate based on the scalar one-bit quantizer and the DBLAST transmission architecture. The calculation is based on the general expression in [6]. We omit the details of derivation and just give the result: for a given listening time f, based on the channel model in Fig. 5, $R_{QMF} = \min\{R_S, R_D\}$, where

$$R_S = C_{p2p,SD} + f(H(\widehat{y}_{SR}|y_{SD}) - H_b(p_f)) \quad (23)$$
$$R_D = (1 - f)C_{p2p,RD} + C_{p2p,SD} - fH_b(p_f), \quad (24)$$

where $H_b(\cdot)$ denotes the binary entropy function and $p_f = Q(\sqrt{SNR_{SR}})$.

Table I lists the information theoretic quantities in terms of SNR in dB at rate $= \frac{1}{2}$ and listening time $f = \frac{1}{3}$.

TABLE I
INFORMATION THEORETIC QUANTITIES IN TERMS OF SNR (DECIBELS) AT RATE $\frac{1}{2}$

Point-to-Point		Relay: $q_{RD} = 1, q_{SR} = 2, f = \frac{1}{3}$		
$C_{p2p,SD}$	DE Threshold	C_{cut}	$C_{cut,DBLAST}$	R_{QMF}
0.19	0.45	-1.51	-1.51	-1.11

Discussion on Simulation Results

Since we have not designed specialized LDPC code profiles for the joint factor graph, we use the BER performance of the point-to-point (no cooperation) case as the baseline for comparison. In Fig. 8 it can be seen that the BER performance of the relay channel is ≈ 1dB better than the point-to-point case (*cooperation gain*). The plot also shows the Shannon

threshold for BIAWGN and threshold corresponding to R_{QMF} for for chosen parameters f, q_{SR}, q_{RD}. The gap between these two thresholds can be seen to be ≈ 1.3dB. This can be considered as the theoretical maximum *cooperation gain* using a 1-bit quantizer and QMF relaying. Moreover it can be seen that the threshold corresponding to R_{QMF} is only 0.4dB away from the cut-set bound.

The simulation results above indicate that the joint decoding algorithm presented here converges and performs reasonably well using off-the-shelf LDPC codes designed for the binary input AWGN channel. Using tools described in [18][17] we can design LDPC code profiles which are specialized for use with the joint decoder to get the performance closer to fundamental limits.

VII. CONCLUDING REMARKS

A. Summary

In this paper we propose a framework for designing efficient QMF relaying schemes over a half-duplex relay channel. The framework is based on LDPC codes for encoding at the source and mapping at the relay. We derive the factor graph for joint decoding at the destination, and propose three techniques that make the graph suitable for efficient implementation:

1) systematic codes at the relay
2) scalar quantizers at the relay,
3) a DBLAST-like transmission architecture that results in an effective parallel channel at the destination.

The proposed framework is not constrained to a Gaussian channel and can be extended to any binary-input symmetric-output memoryless half-duplex relay channel.

B. Extension to High-Order Constellation

The capacity advantage of cooperative relaying and the effectiveness of QMF are most significant in the high SNR regime. In this regime coding and modulation both play an important role in system performance. In practice most systems use BICM [10] to combine channel codes designed for binary alphabet with higher order signal constellations. BICM has been proposed for various cooperative channel scenarios [19][20][21][14]. To extend our work to high-order constellations we follow the BICM approach. This is work-in-progress and the results will be reported in a forthcoming paper. Following is a brief preview of the same.

Under classical BICM [10] the effective channel from a bit B to the channel output Y is equivalent to a state-dependent binary-input channel $p_{Y|B,\Theta}(y|b,\theta)$ with state $\Theta \in \{1, 2, ..., m\}$ known to both the transmitter and the receiver (2^m is the cardinality of the chosen constellation). Ideally we would like to use our coding framework for the binary-input channel at each BICM "sub-channel" (under a given state θ). However the sub-channel corresponding to a state θ is no longer guaranteed to be output-symmetric.

For binary-input channels that are not output-symmetric, classical analytical tools for LDPC codes such as density evolution become very complicated since the all-zero codeword can no longer represent the performance of all codewords.
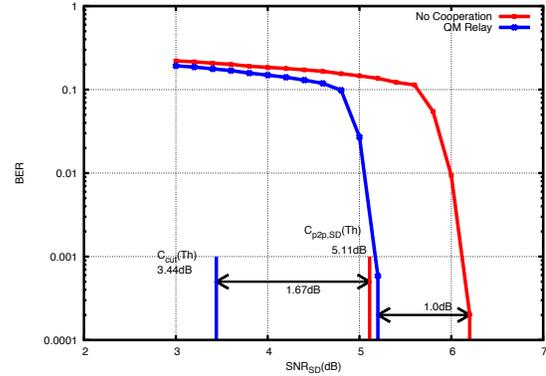

Fig. 9. 16QAM Simulation Results for $q_{SR} = 2$, $q_{RD} = 1$, $N_S = 20400$, $N_R = 13600$, $f = \frac{1}{3}$, $r_S = r_R = 0.5$

This is a well-known issue and can be resolved by adding random dithers into the system [22][23]. Dithers are i.i.d. Bernoulli $\left(\frac{1}{2}\right)$ random variables known to both the transmitter and receiver. Dithering guarantees that the sub-channels under BICM are output-symmetric.

Secondly we need to define the quantize-and-map operation under BICM. The natural approach is to have the relay R quantize its received signal encode it using a *binary* LDPC code and then forward it using a second independent BICM operation. However this choice does not yield a practical decoder because we can no longer "decompose" the original channel into sub-channels as in the point-to-point scenario. This defeats the purpose of BICM which is to allow design of coding and modulation subsystems separately. We propose an alternative approach for the quantize-and-map operation to retain this separation. Under this approach the relay quantizes *each effective sub-channel* (instead of its received signal). The relays operation is described below:

1) compute the LLR (log-likelihood-ratio) of each bit based on knowledge of dither
2) quantize each bit on the effective sub-channel that it went through
3) encode the quantization output using a LDPC code
4) pass the resultant binary codeword to another BICM module

Results of a simulation experiment based on the above approach using 16QAM modulation is given in Fig. 9. The LDPC codes used in this experiment were the same as those used for the binary-input case. The overall design rate is 2 bits/channel use.

For comparison Table II lists the information theoretic quantities for rate 2 and 16QAM constrained inputs.

VIII. ACKNOWLEDGEMENTS

TABLE II
INFORMATION THEORETIC QUANTITIES IN TERMS OF SNR (DECIBELS) AT
RATE 2 FOR 16QAM SCENARIO

Point-to-Point		Relay: $q_{RD} = 1, q_{SR} = 2, f = \frac{1}{3}$	
CM Capacity	BICM Capacity	$C_{\text{cut,DBLAST}}$	R_{QMF}
5.11	5.28	3.44	3.95

REFERENCES

[1] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *Information Theory, IEEE Transactions on*, vol. 25, no. 5, pp. 572 –584, sep 1979.

[2] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3062 – 3080, dec. 2004.

[3] M. Yuksel and E. Erkip, "Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective," *Information Theory, IEEE Transactions on*, vol. 53, no. 10, pp. 3371 –3393, oct. 2007.

[4] S. Pawar, A. Avestimehr, and D. Tse, "Diversity-multiplexing tradeoff of the half-duplex relay channel," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, 23-26 2008, pp. 27 –33.

[5] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *ArXiv e-prints*, Jun. 2009.

[6] S. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *ArXiv e-prints*, Feb. 2010.

[7] V. Nagpal, S. Pawar, D. Tse, and B. Nikolic, "Cooperative multiplexing in the multiple antenna half duplex relay channel," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, june 2009, pp. 1438 –1442.

[8] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited receiver cooperation," *Submitted to IEEE Transactions on Information Theory*, November 2009, http://arxiv.org/abs/0911.2053.

[9] A. Ozgur and S. Diggavi, "Approximately achieving gaussian relay network capacity with lattice codes," *ArXiv e-prints*, May 2010.

[10] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 927 –946, may 1998.

[11] A. Chakrabarti, A. D. Baynast, A. Sabharwal, and B. Aazhang, "Low density parity check codes for the relay channel," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 2, pp. 280 –291, february 2007.

[12] P. Razaghi and W. Yu, "Bilayer low-density parity-check codes for decode-and-forward in relay channels," *Information Theory, IEEE Transactions on*, vol. 53, no. 10, pp. 3723 –3739, oct. 2007.

[13] M. Khojastepour, N. Ahmed, and B. Aazhang, "Code design for the relay channel and factor graph decoding," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, vol. 2, 7-10 2004, pp. 2000 – 2004 Vol.2.

[14] P. Razaghi, M. Aleksic, and W. Yu, "Bit-interleaved coded modulation for the relay channel using bilayer ldpc codes," in *Information Theory, 2007. CWIT '07. 10th Canadian Workshop on*, 6-8 2007, pp. 101 –104.

[15] M. Uppal, Z. Liu, V. Stankovic, and Z. Xiong, "Compress-forward coding with bpsk modulation for the half-duplex gaussian relay channel," *Trans. Sig. Proc.*, vol. 57, no. 11, pp. 4467–4481, 2009.

[16] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 498 –519, feb 2001.

[17] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 619 –637, feb 2001.

[18] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 599 –618, feb 2001.

[19] G. Kramer, "Distributed and layered codes for relaying," in *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, october 2005, pp. 1752 – 1756.

[20] G. Kraidy, N. Gresset, and J. Boutros, "Coding for the non-orthogonal amplify-and-forward cooperative channel," in *Information Theory Workshop, 2007. ITW '07. IEEE*, 2-6 2007, pp. 626 –631.

[21] M. Benjillali and L. Szczecinski, "A simple detect-and-forward scheme in fading channels," *Communications Letters, IEEE*, vol. 13, no. 5, pp. 309 –311, may 2009.

[22] 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," *Information Theory, IEEE Transactions on*, vol. 49, no. 9, pp. 2141–2155, 2003.

[23] A. Ingber and M. Feder, "Parallel bit interleaved coded modulation," *Proceedings of Annual Allerton Conference on Communications, Control, and Computing*, September 2010.