Low-Complexity Codes for Approximately Achieving Wireless Network Capacity

Ayan Sengupta
ARNI, I&C, EPFL

***Abstract*—Characterizing the fundamental limits of communication in wireless networks has been a holy grail of information theory for decades. Recent progress on this problem for wireless relay networks [1] has led to a Quantize-Map-and-Forward (QMF) relaying strategy which approximately achieves the network capacity. The scheme employs Gaussian codebooks for transmission and random mappings at the relays. For communication system design however, efficient and implementable coding and signaling techniques are required. Taking inspiration from the phenomenal success of iterative codes for point-to-point channels, we aim to develop low-complexity QMF-based iterative schemes for networks that provably retain the universal performance guarantees. In preliminary work, we present a structured QMF scheme for the parallel-relay network, employing LDPC ensembles for node operations. We demonstrate via simulations, that our scheme, with no transmit channel state information, offers a robust performance over fading channels and achieves the full diversity order of the network at moderate SNRs.**

***Index Terms*—Wireless Networks, Quantize-Map-and-Forward Relaying, Cooperative Communication, LDPC Codes**

I. Introduction

Cooperative communication as a means to provide reliable high data-rate wireless transmission has been extensively studied in the literature for the past several decades, both from an information-theoretic viewpoint as well as from the perspective of communication protocols. However, the capacity characterization of general wireless networks has been open.

Proposal submitted to committee: June 21th, 2011; Candidacy exam date: June 28th, 2011; Candidacy exam committee: Bixio Rimoldi, Christina Fragouli, Rüdiger Urbanke.

This research plan has been approved:

Date: ________________

Doctoral candidate: ________________
(name and signature)

Thesis director: ________________
(name and signature)

Thesis co-director: ________________
(if applicable) (name and signature)

Doct. prog. director: ________________
(R. Urbanke) (signature)

Recently, an approximate characterization of the capacity of wireless relay networks has been established in [1]. This is done via a two-stage approach where first an analytically simple *linear deterministic model* is developed that captures the three key features of wireless communication in the interference limited (high SNR) regime–channel strength, broadcast and superposition. A complete capacity characterization of networks under such a model is provided, which is a generalization of the max-flow min-cut theorem for wired networks. Using insights from this model, the authors develop a relaying scheme for Gaussian networks, termed Quantize-Map-and-Forward (QMF), in which the relays quantize their received signal (at noise-level) and (uniformly at random) map it onto their transmit codebooks. QMF is shown to achieve all rates within a constant gap (independent of the channel gains) to the cutset upper bound. Moreover, the relay operations require only local knowledge, and transmit channel state information (CSI) is not required at any point, rendering QMF to be a practical and robust cooperative diversity scheme for (fading) wireless channels.

The design and analysis of QMF relaying in [1] was done without regard to computational complexity constraints. A natural question to ask is whether one can design (computationally) implementable strategies that maintain similar advantages over more traditional cooperative communication schemes. In the realm of point-to-point communication, low-density parity-check (LDPC) codes have emerged in the last decade as a class of codes that can exhibit near Shannon-limit performance while admitting low-complexity message-passing decoding algorithms on sparse bipartite graphs. Moreover, as demonstrated in work of Richardson and Urbanke [2], there exists a well developed set of tools to analyze these codes and provide provable performance guarantees (in terms of decoding thresholds at given rates) for a large class of channels. Given the success of iterative codes in point-to-point channels, it is of interest to examine the performance of such techniques for the design of QMF-based cooperative strategies.

The first efforts in this direction were made in [3], where a QMF-based coding scheme for the half-duplex single-relay network was presented, using LDPC codes for encoding and mapping. However, by invoking a one-block delay at the relay, and performing successive interference cancelation at the destination, [3] simplifies the decoding operation whereby they orthogonalize the interfering streams. In such a simplified scenario, the factor graph for the decoding operation effectively reduces to that of two point-to-point channels connected by nodes representing the quantization operation at the relays. This strategy does not maintain the desired universality property of the original QMF scheme. In particular, it does not naturally extend to larger configurations.

In our preliminary work, we develop a practical QMF-inspired coding scheme for the full-duplex parallel-relay network. This scenario, originally studied in [4], is applicable to next-generation (for example LTE-based) wireless networks, where relays are placed for coverage extension. In our scheme (i) We use LDPC ensembles for the encoding and relay maps, to develop a compound graphical representation of the network operation that takes into account the multiple-access and the broadcast due to the wireless channels as well as the processing at various nodes–namely quantization and modulation. We devise scheduling and message passing algorithms on this graph (ii) We evaluate the performance of our codes and demonstrate that our implementation of QMF retains several advantages over conventional strategies–in particular, it achieves the full diversity order of the network at moderate SNRs without transmit CSI, and is within 3 dB of the optimal outage performance.

While there is promise from the initial results, we believe the performance of such iterative systems can be improved much further. As an instance, we have used off-the-shelf codes for the source and relay operations. Insights into how component codes (and their combinations thereof) influence the iterative decoding performance would be a key to optimizing the structures. With optimized structures, the performance, especially under symmetric channel configurations is expected to be better than that at present. To understand the fundamental bottlenecks (be it code design, tackling multiple-access or improving the quantizer) will be of help in designing better schemes, taking us a step closer to providing provable guarantees on iterative decoding thresholds for general (moderately sized) networks that closely approach optimal.

The rest of this report is organized as follows: Section II summarizes the work of Avestimehr *et. al* [1] on the approximate capacity characterization for wireless networks. In section III, we review the work of Richardson and Urbanke on characterizing the performance limits of LDPC codes. Section IV describes the coding scheme of Nagpal et.al [3] for the single-relay network. We provide an outline and give numerical evaluations for our preliminary work on the parallel-relay network in section V and also mention future goals.

II. Wireless Network Information Flow

In this section, we summarize the work of [1], by first introducing the linear deterministic framework for modeling wireless signal interactions and then demonstrating how insights from the capacity-achieving schemes for this model can be translated to approximately optimal communication strategies for Gaussian networks.

A. Linear Deterministic Framework

To introduce the linear deterministic model, we consider a real scalar Gaussian point-to-point channel.

$$y = hx + z$$

where $\mathbb{E}[|x|^2] \leq 1$, $z \sim \mathcal{N}(0,1)$ and the channel gain is $|h| = \sqrt{\text{SNR}}$. The capacity of this channel is $\frac{1}{2}\log(1+\text{SNR})$ bits/transmission. Assume for simplicity that the transmit signal and noise have a peak power 1. Writing the received signal in terms of the binary expansions of x and z, we have

$$y = 2^{\frac{1}{2}\log(\text{SNR})}\sum_{i=1}^{\infty} x(i)2^{-i} + \sum_{i=1}^{\infty} z(i)2^{-i}$$

Fixing $n = \lceil \frac{1}{2}\log(\text{SNR})\rceil^{+}$,

$$y \approx 2^{n}\sum_{i=1}^{n} x(i)2^{-i} + \sum_{i=1}^{\infty}(x(i+n)+z(i))2^{-i}$$

Ignoring the carry-over from the second summation term, we see that the channel can be approximated by a pipe that only allows the bits of the signal that are above the noise level to pass. This is illustrated in fig. 1. If $\mathbf{x}, \mathbf{y}$ are q-bit vectors, this

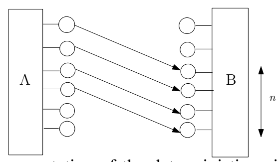


Fig. 1: Pictorial representation of the deterministic point-to-point channel

can be represented by a $q \times q$ shift matrix $\mathbf{S}^{q-n}$ where

$$\mathbf{S}^{q-n} = \begin{pmatrix} \mathbf{0}_{(q-n)\times n} & \mathbf{0}_{(q-n)\times(q-n)} \\ \mathbf{I}_{n\times n} & \mathbf{0}_{n\times(q-n)} \end{pmatrix} \text{ and } \mathbf{y} = \mathbf{S}^{q-n}\mathbf{x}$$

Hence, the capacity of the point-to-point channel in this model is $n = \lceil \frac{1}{2}\log(\text{SNR})\rceil^{+}$ which is within $\frac{1}{2}$-bit of the capacity of the Gaussian channel (1-bit for the complex case).

Broadcast can be modeled in a similar manner. The deterministic model corresponding to a real scalar Gaussian broadcast channel, with two receivers having channel SNR's as SNR_i $(i = 1, 2)$ and $n_i = \lceil \frac{1}{2}\log(\text{SNR}_i)\rceil^{+}$ is shown in fig. 2(a). The stronger receiver B_1 receives the most significant n_1 bits and receiver B_2 gets the most significant n_2 bits $(n_2 < n_1)$. As shown in fig. 2(b), the capacity region of the deterministic broadcast channel closely approximates that of the Gaussian broadcast channel and the difference between the corresponding rate regions is less than 1 bit per user.

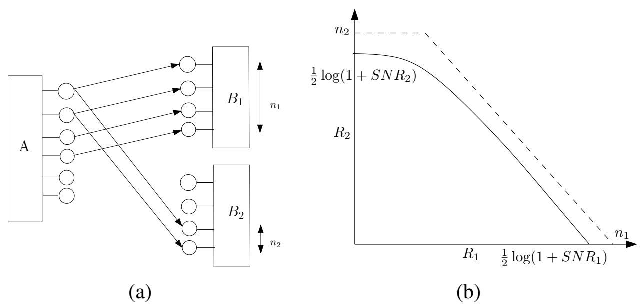
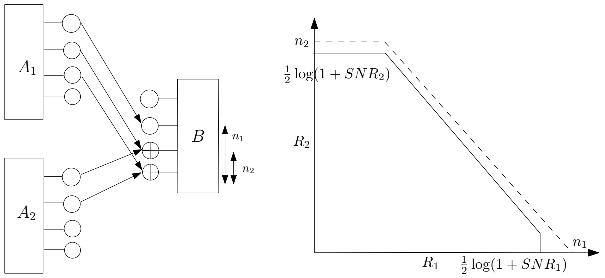


Fig. 2: Pictorial representation of the deterministic broadcast channel

The superposition of two signals is modeled in the deterministic network by XOR-ing the bits from interfering signals that arrive at the same signal level. Consider two transmitters simultaneously trying to send data to a single receiver as shown in fig. 3(a). The SNRs at the transmitters A_i are denoted by SNR_i and n_i is defined as above ($i = 1, 2, n_1 > n_2$). In this case, the receiver gets the most significant $n_1 - n_2$ bits of A_1 without any interference, while the remaining n_2 bits of A_1 and all the bits of A_2 that are above noise level get XOR-ed at the receiver. The remaining bits below the noise level get truncated. This can written algebraically as

$$\mathbf{y} = \mathbf{S}^{q-n_1}\mathbf{x}_1 \oplus \mathbf{S}^{q-n_2}\mathbf{x}_2$$

Again, it can be shown that the capacity region of the deterministic MAC is at most 1 bit per user away from the Gaussian MAC (fig. 3(b)).



(a) (b)

Fig. 3: Pictorial representation of the deterministic multiple-access channel

The above examples show that the linear deterministic model is a good approximation to the Gaussian model in the sense that the capacity regions for one closely approximate those of the other. This is not true in general however, and one can construct networks (e.g Gaussian MIMO channel) where the gaps can become unbounded. Nevertheless, it is possible to gain insights from the deterministic model and construct schemes that are close to optimal for Gaussian networks.

As an example, consider the diamond network shown in fig. 4(a). It can be shown that the capacity of the linear deterministic network in this case is

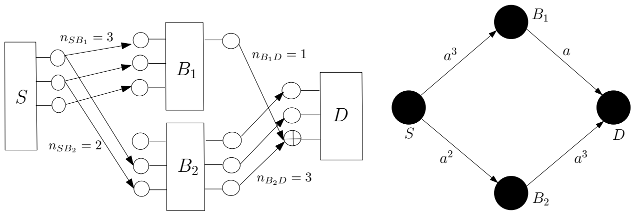
$$\begin{aligned} C^d_{\text{dia}} = \quad & \min\{\max(n_{SB_1}, n_{SB_2}), \max(n_{B_1D}, n_{B_2D}), \\ & n_{SB_1} + n_{B_2D}, n_{SB_2} + n_{B_1D}\} \end{aligned}$$

Interestingly, the relaying scheme in the deterministic model that achieves capacity (routing independent messages through non-interfering paths) leads to an analogous partial decode-and-forward strategy for the Gaussian case, which achieves rates within 1-bit of the cutset bound for all channel gains, whereas schemes like amplify-forward and decode-forward can be arbitrarily far away from it.

B. Main Results

1) Capacity of Linear Deterministic Networks: Let $\mathcal{V}$ be the set of all nodes in the network with $\mathbf{x}_j$ and $\mathbf{y}_j$ denoting the transmitted and received signals at node j, $j \in \mathcal{V}$. The link from node i to node j has a non-negative integer gain n_{ij} associated with it. This number models the channel gain in the corresponding Gaussian setting. The received signal at each node (at each time step) is a deterministic function of the transmitted signals at the other nodes,

$$\mathbf{y}_j = \sum \mathbf{S}^{q-n_{ij}}\mathbf{x}_i \tag{1}$$



(a) (b)

Fig. 4: A deterministic diamond network and its Gaussian counterpart

where $q = \max(n_{ij})$. Denote by Λ_D, the set of all cuts separating S and D. Then, the cut-set upper bound on the capacity of this network is

$$\begin{aligned} \bar{C} &= \max_{p(\mathbf{x}_{\mathcal{V}})} \min_{\Omega \in \Lambda_D} I(\mathbf{y}_{\Omega^c}; \mathbf{x}_\Omega | \mathbf{x}_{\Omega^c}) \\ &\overset{(a)}{=} \max_{p(\mathbf{x}_{\mathcal{V}})} \min_{\Omega \in \Lambda_D} H(\mathbf{y}_{\Omega^c} | \mathbf{x}_{\Omega^c}) \overset{(b)}{=} \min_{\Omega \in \Lambda_D} \text{rank}(G_{\Omega,\Omega^c}) \end{aligned}$$

where G_{Ω,Ω^c} is the transfer matrix associated with the cut, i.e, the matrix relating the vector of all the inputs at the nodes in Ω to the vector of all the outputs in Ω^c induced by (1). Claim (a) follows from the fact that we have deterministic networks and (b) follows since all cut values $H(\mathbf{y}_{\Omega^c} | \mathbf{x}_{\Omega^c})$ are simultaneously optimized by independent and uniform distribution of $\mathbf{x}_{\mathcal{V}}$ and the optimum value of each cut is the logarithm of the size of the range space of the transfer matrix.

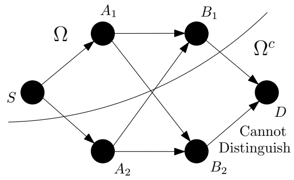
The following theorem says that, in the linear deterministic setting, this upper bound can in fact be reached.

Theorem 1: The capacity C of a linear finite-field deterministic relay network is given by

$$C = \min_{\Omega \in \Lambda_D} \text{rank}(G_{\Omega,\Omega^c})$$

The theorem is initially proved for *layered* networks i.e. all paths from the source to the destination have equal lengths. This results in a simplification, whereby sequence of messages can each be encoded into a block of symbols, with no interaction *across* blocks as they flow through the network.

The encoding strategy is as follows. The source S has a sequence of messages $w_k \in \{1, 2, \cdots 2^{TR}\}, k = 1, 2, \cdots$. Each message is encoded by the source into a signal over T transmission times (symbols), giving an overall transmission rate of R. Every relay j operates over blocks of time T symbols, and uses independent linear mappings from its received vector in the previous block to the transmit vector in the present block, represented by $\mathbf{x}_j = \mathbf{F}_j \mathbf{y}_j$. The matrix $\mathbf{F}_j$ is chosen uniformly at random over all matrices in $\mathbb{F}_2^{qT \times qT}$. Given the encoding functions at the source and relays, the destination attempts to decode the message w_k.



$$G_{\Omega,\Omega^c} = \begin{bmatrix} G_{S,A_2} & 0 & 0 \\ 0 & G_{A_1,B_2} & 0 \\ 0 & 0 & G_{B_1,D} \end{bmatrix}$$

(a) (b)

Fig. 5: An example of a layered relay network, a particular cut and the corresponding transfer matrix.

The proof proceeds by bounding the *probability of indistinguishability*. Since we are dealing with deterministic networks, the only reason why the destination may not be able to decode is if two messages w, w' from the source get mapped to the same received signal at the destination by the intermediate relay nodes. Consider the cut Ω defined by the set of nodes that can distinguish between w, w' and the ones that cannot. Since the events that correspond to the occurrence of the distinguishability sets $\Omega \in \Lambda_D$ are disjoint, we have

$$\mathbb{P}\{w \to w'\} = \sum_{\Omega \in \Lambda_D} \mathcal{P}\begin{pmatrix} \text{Nodes in } \Omega \text{ can distinguish \&} \\ \text{nodes in } \Omega^c \text{ cannot distinguish} \end{pmatrix}.$$

$\mathcal{P}$ can be upper bounded by the products of (conditional) probabilities that each node in Ω^c cannot distinguish between the messages. In the example network of fig. 5(a), for the particular cut Ω shown, we have

$$\mathcal{P} \leq 2^{-T\text{rank}(G_{S,A_2})} 2^{-T\text{rank}(G_{A_1,B_2})} 2^{-T\text{rank}(G_{B_1,D})} = 2^{-T\text{rank}(G_{\Omega,\Omega^c})}$$

where G_{Ω,Ω^c} is the transfer matrix across the cut. Summing over all $\Omega \in \Lambda_D$ and over all the other w' for a fixed w, the probability of error can be upper bounded by

$$\mathbb{P}_e \leq 2^{RT} |\Lambda_D| 2^{-T \min\limits_{\Omega \in \Lambda_D} \text{rank}(G_{\Omega,\Omega^c})}$$

Clearly, $\mathbb{P}_e$ can be made as small as possible as long as $R < \min\limits_{\Omega \in \Lambda_D} \text{rank}(G_{\Omega,\Omega^c})$. The formal proof follows along these lines and is extended to non-layered networks using a time expansion argument.

2) Approximate capacity of Gaussian networks: To mimic the operations in the deterministic model for the Gaussian case, one has to extract the bits above the noise floor at each relay, before proceeding with the random mappings. This is proposed in [1] by means of a quantization operation on the received signals. The quantized bits are then mapped randomly to a transmit Gaussian codeword. The main claim in the case of Gaussian networks is as follows

Theorem 2: The capacity C of the Gaussian relay network satisfies

$$\bar{C} - \kappa \leq C \leq \bar{C}$$

where $\bar{C}$ is the cut-set upper bound on the capacity of the network and κ is $O(|\mathcal{V}|)$, independent of the channel SNRs.

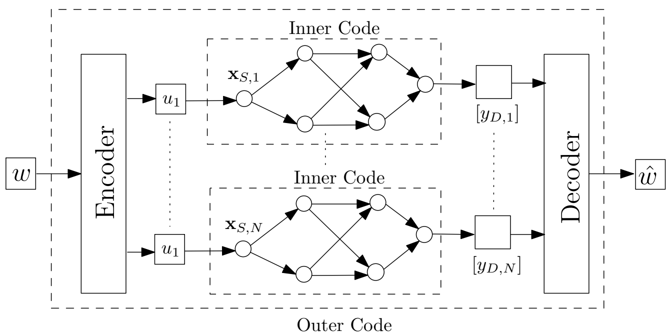


Fig. 6: Setup for Gaussian Networks

The quantization operation $[.] : \mathbb{C} \to \mathbb{Z} \times \mathbb{Z}$ maps a complex number $c = x + iy$ to $[c] = ([x], [y])$ where $[x], [y]$ are the closest integers to x, y. Since the Gaussian noise at all receive antennas has variance 1, this operation is basically scalar quantization at noise-level. The encoding consists of an inner and outer code as shown in Figure 6(a).

- *Inner Code*: Source maps inner code symbol $u \in \{1, \cdots, 2^{R_{in}T}\}$ to $F_S(u) \in \mathcal{T}_{x_S}$. Relay i receives $\mathbf{y}_i$ of length T, quantizes it to $[\mathbf{y}_i]$ and maps it to $F_i([\mathbf{y}_i]) \in \mathcal{T}_{x_i}$. $\mathcal{T}_{x_S}$, $\mathcal{T}_{x_i}$ are length-T random Gaussian codebooks.
- *Outer code*: The message is encoded by the source into N inner code symbols, $u_1, \cdots, u_N$. Each inner code symbol is then sent via the inner code over T transmission times, giving an overall transmission rate of R. The received signal at the destination, corresponding to inner code symbol u_i, is denoted by $\mathbf{y}_{D,i}, i = 1, \cdots, N$. Given the knowledge of all the encoding functions F_i's at the relays and the quantized received signals $[\mathbf{y}_{D,1}], \cdots, [\mathbf{y}_{D,N}]$, the destination node tries to decode the source message.

The proof for layered networks proceeds by lower bounding the mutual information between u and $[\mathbf{y}_D]$ averaged over the choices of random mappings $F_\mathcal{V} = \{F_i, i \in \mathcal{V}\}$. Note that

$$\frac{1}{T} I(u; [\mathbf{y}_D] | F_\mathcal{V}) \geq \frac{1}{T} I(u; [\mathbf{y}_D] | \mathbf{z}_\mathcal{V}, F_\mathcal{V}) - \frac{1}{T} H([\mathbf{y}_D] | u, F_\mathcal{V})$$

where $\mathbf{z}_\mathcal{V}$ is the vector of channel noises at all nodes. The first term on the RHS above is the average end-to-end mutual information conditioned on the noise vector. Once conditioned on a noise vector, the network turns into a deterministic one. A technique similar to the one for linear deterministic networks is used to upper bound the probability that the destination will confuse an inner code symbol with another, which in turn (in conjunction with Fano's inequality) is used to lower bound the end-to-end mutual information. The second term represents roughly the penalty due to noise-forwarding at the relays, and is proportional to the number of relays. Through a series of technical lemmas it is shown that

$$\frac{1}{T} I(u; [\mathbf{y}_D] | \mathbf{z}_\mathcal{V}, F_\mathcal{V}) \geq R_{in} - (\frac{1}{T} + \min\{1, 2^{|\mathcal{V}|} 2^{-T(\bar{C} - 3|\mathcal{V}| - R_{in})}\} R_{in})$$

and

$$H([\mathbf{y}_D] | u, F_\mathcal{V}) \leq 12T|\mathcal{V}|$$

By choosing R_{in} arbitrarily close to $\bar{C} - 3|\mathcal{V}|$ and letting T be arbitrarily large, for any $\delta > 0$, we get

$$\frac{1}{T} I(u; [\mathbf{y}_D] | F_\mathcal{V}) \geq \bar{C} - 15|\mathcal{V}| - \delta$$

Therefore, there exists a choice of mappings at the source and relays that provide an end-to-end mutual information close to $\bar{C} - 15|\mathcal{V}|$. We can now use a good outer code to reliably send a message over N uses of this channel at any rate upto $\bar{C} - 15|\mathcal{V}|$. This constant gap result extends to other scenarios including multicast networks, MIMO networks, frequency-selective networks, half-duplex networks and fading networks as shown in [1].

III. Capacity of Low-Density Parity-Check Codes

In this section, we review the work of Richardson and Urbanke which presents a general method for determining the *capacity* of LDPC codes under iterative message-passing decoding over binary-input memoryless channels. The analysis involves the consideration of an *ensemble* of codes rather than individual entities. The paper develops a set of tools to prove the existence of a decoding threshold (defined as the "worst" channel parameter for which reliable transmission is possible at a given communication rate) under message-passing decoding for a broad class of channels and compute this threshold via a deterministic algorithm for the case when the ensembles of bipartite graphs representing the codes contain no cycles. The authors then show that as the blocklength of the code increases, the behavior of individual instances concentrates around the ensemble behavior, and that the ensemble behavior itself converges to the behavior for the cycle-free case.

A. Preliminaries

We focus on (d_v, d_c)-regular LDPC codes, which contain d_v ones per column and d_c ones per row of its parity-check matrix H. Let n be the length of the (d_v, d_c)-regular LDPC code with codewords given by the solution set of $H\mathbf{x} = 0$. For such a code, we construct a bipartite graph with n *variable* nodes, and $m = \frac{nd_v}{d_c}$ *check* nodes. Each variable node corresponds to one bit of the codeword, and each check node represents one parity-check equation. Edges connecting the variable nodes to the check nodes are in one-to-one correspondence with the 1's in H. The design rate of such a code is $1 - \frac{d_v}{d_c}$.

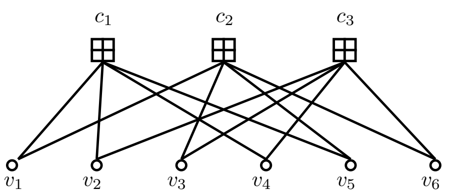


Fig. 7: A (2,4)-regular LDPC code of length 6

1) Ensembles and Neighbourhoods: The ensemble $\mathcal{C}^n(d_v, d_c)$ of (d_v, d_c)-regular LDPC codes of length n is defined as follows. We assign to each node d_v or d_c sockets, according to whether it is a variable or check node. The variable and check sockets, numbering $nd_v(= md_c)$ in total, are labeled separately with the set $[nd_v] = \{1, 2, \ldots, nd_v\}$ in some arbitrary fashion. We pick a permutation π on nd_v letters uniformly at random from the set set of all $(nd_v)!$ permutations. The corresponding labeled bipartite graph is then defined by placing edges between the i^{th} variable node socket and the $\pi(i)^{th}$ check node socket. This induces a uniform distribution on the set $\mathcal{C}^n(d_v, d_c)$ of bipartite graphs.

An important notion associated to the above bipartite graphs is that of *decoding neighbourhoods*. Broadly, a decoding neighbourhood of depth d of a given node or an edge comprises the set of nodes and edges in the graph that are connected to it by paths (sequence of edges) of length less than or equal to d. An useful result associated with the decoding neighbourhoods of (d_v, d_c)-regular LDPC codes is that the number of distinct nodes as well as edges in a neighbourhood of a node or an edge of depth $2l$ is upper bounded by $2(d_v d_c)^l$, which is independent of n. This result, together with the symmetry properties associated with neighbourhoods [2] play an important role in establishing the concentration results.

2) Binary-input memoryless channels: The class of channels considered are memoryless and have an input alphabet $\mathcal{I} = \{\pm 1\}$. The output alphabet $\mathcal{O}$ may be discrete (e.g BSC) or continuous (e.g BIAWGNC).

3) Message-passing Decoders: The class of message-passing algorithms that are used for decoding the codebits behave as follows. At time zero, every variable node v_i has a received value r_i taking values in $\mathcal{O}$. Messages are exchanged between nodes in the graph along the edges in discrete time steps. First, every variable node v_i sends to each neighbouring check node c_j a *message* taking values in $\mathcal{M}$. At time zero, all messages sent by node v_i are equal to r_i. Subsequently, each check node processes the messages it receives and sends back to each of its neighbours a message taking values in $\mathcal{M}$. Each variable node now processes its incoming messages together with its received value r_i to produce new messages to be sent to the check nodes and hence the iterations continue. An important condition followed by these algorithms is that the message passed on an edge e in an iteration does not depend on the previously received message along e. The message maps at variable and check nodes (same for every edge associated with the corresponding node) are represented as $\psi_v^{(l)} : \mathcal{O} \times \mathcal{M}^{d_v-1} \to \mathcal{M}$ and $\psi_c^{(l)} : \mathcal{M}^{d_c-1} \to \mathcal{M}$ respectively. Messages can be interpreted as representing estimates of a particular bit. The sign of the message conveys the hard estimate on the bit $(+1 \text{ or} -1)$, and the magnitude provides the reliability of the message.

A simplification in the analysis of the decoders arises from the fact that under certain symmetry conditions satisfied by the channels and message maps, the error probability conditioned on a transmitted codeword becomes independent of it, allowing analysis assuming the all-one codeword was transmitted.

B. Threshold Determination

1) Density Evolution: In this section, we focus on determining the decoding threshold for the case of tree-like neighbourhoods by analyzing the evolution of message densities as a function of the iteration number. Under the all-one codeword assumption, an incorrect message is one with a nonpositive sign. For discrete message alphabets, the expected fraction of erroneous messages in the l^{th} iteration can be expressed via a system of coupled recursive equations. A closer investigation of these recursions point to the existence of a (computable) threshold. For infinite message alphabets, the situation is much more involved. [2] addresses these questions for the important case of the belief-propagation (BP) decoder.

In BP, the messages sent on an edge e represent the posterior probabilities on the bit associated with the incident variable node, i.e the 2-tuple (p_1, p_{-1}). A representation for the above is in the form of a log-likelihood ratio (llr) $\log \frac{p_1}{p_{-1}}$. Every node acts under the assumption that each message communicated to it in a given iteration is a conditional probability mass function (pmf) on the bit and that the random variables on which the messages are based are independent. On receiving its messages, a node transmits to each neighbouring node the conditional pmf of the bit conditioned on *all* observed information apart from that coming from the particular neighbouring node. In this scenario, at the variable nodes, the likelihood ratio of the outgoing message is the product of the likelihood ratios of the incoming messages. Hence in the llr domain

$$\psi_v(m_0, m_1, \ldots, m_{d_v-1}) = \sum_{i=0}^{d_v-1} m_i \tag{2}$$

For the check-node maps, it is convenient to revert to the equivalent representation where the random variable corresponding to the bit takes values in $\{0, 1\}$, instead of $\{\pm 1\}$ with $0 \leftrightarrow 1$, $1 \leftrightarrow -1$, $p_0 \leftrightarrow p_1$ and $p_1 \leftrightarrow p_{-1}$. The probability of the outgoing message being 0, denoted by $\tilde{p}_0$ is now the probability that the mod-2 sum of the $(d_c - 1)$ independent $(0, 1)$-valued variable is equal to zero. In this setting, one observes that the 2-tuple $(\tilde{p}_0, \tilde{p}_1)$ for the outgoing message is given by the circular convolution of the incoming $(d_c - 1)$ 2-tuples. This operation can be efficiently implemented with a 2-point DFT with the transform coefficients given by

$$\mathcal{F}(p)(0) = p_0 + p_1 = 1$$
$$\mathcal{F}(p)(1) = p_0 - p_1$$

Thus by the point-wise product of DFT coefficients we have

$$\tilde{p}_0 - \tilde{p}_1 = \prod_{k=1}^{d_c-1} (p_0^k - p_1^k) \tag{3}$$

In the llr domain, this translates to the following message-map at the check nodes

$$\psi_c(m_1, m_2, \ldots, m_{d_c-1}) = \log\left(\frac{1 + \prod_{i=1}^{d_c-1} \tanh \frac{m_i}{2}}{1 - \prod_{i=1}^{d_c-1} \tanh \frac{m_i}{2}}\right) \tag{4}$$

To compute the evolution of message densities, we note that (3) points to an efficient update method at the check nodes. In particular, representing the messages by an ordered pair $(\lg \text{sgn}(p_0 - p_1), -\log|p_0 - p_1|) \in \text{GF}(2) \times [0, \infty)$, the check-node message map reduces to an addition over the new space. Also one can obtain the densities of log likelihoods P from the equivalent density $\tilde{P}$ over $\text{GF}(2) \times [0, \infty)$ and vice-versa via a change of measure. The density of the outgoing message $\tilde{Q}$ is then obtained by a convolution of the incoming densities $\tilde{P}$'s. This can be implemented using generalized Fourier transforms which in this case is the Laplace transform in the second coordinate and the 2-point DFT in the first coordinate. Denoting by $\tilde{P}^0(y)$ and $\tilde{P}^1(y)$ the densities $\tilde{P}(y, 0)$ and $\tilde{P}(y, 1)$ $(y \in [0, \infty))$ and the corresponding Laplace transforms by $\hat{\tilde{P}}^0$ and $\hat{\tilde{P}}^1$, the density evolution equations at the check-nodes as a function of the iteration number l are given by

$$\hat{\tilde{Q}}^{(l),0} - \hat{\tilde{Q}}^{(l),1} = \left(\hat{\tilde{P}}^{(l-1),0} - \hat{\tilde{P}}^{(l-1),1}\right)^{d_c-1}$$
$$\hat{\tilde{Q}}^{(l),0} + \hat{\tilde{Q}}^{(l),1} = \left(\hat{\tilde{P}}^{(l-1),0} + \hat{\tilde{P}}^{(l-1),1}\right)^{d_c-1}$$

where analogous notation is used for the outgoing density Q. From (2), the density evolution at the variable nodes represents the convolution of the incoming densities and is given in the Fourier domain by

$$\mathcal{F}\left(P^{(l)}\right) = \mathcal{F}\left(P^{(0)}\right)\left(\mathcal{F}\left(Q^{(l)}\right)\right)^{d_v-1} \tag{5}$$

2) Monotonicity: For most channels of interest, there exists a parameter α which reflects a natural ordering of the channels–the capacity decreases with increasing α. In such cases, it is natural to ask whether convergence for a parameter α automatically implies convergence for every α' such that $\alpha' \leq \alpha$. To address this question, the authors first introduce the notion of *physical degradation* of channels. A channel W' is said to be physically degraded with respect to a channel W if the channel transition probabilities satisfy $p_{W'}(y'|x) = p_Q(y'|y)p_W(y|x)$ for some auxiliary channel Q. They show that in such a setting, the expected fraction of incorrect messages (denoted by p and p' for channels W and W' respectively) during BP decoding in the l^{th} iteration assuming tree-like neighbourhoods satisfies $p \leq p'$. This implies that in a family of single parameter channels that are *ordered* by physical degradation, there exists a threshold, i.e, a maximum value of the parameter for which the error probability goes to zero (and remains zero for all values less than this maximum) in the limit of infinite iterations, separating the decodable and undecodable regions. Such channel families are termed *monotone with respect to the decoder*. An example of such a family is the BSC, where a BSC with crossover probability $\epsilon' \geq \epsilon$ can be obtained from a BSC with parameter ϵ by cascading it with another BSC of appropriate parameter. Decoding thresholds obtained by numerical evaluation are given in [2] for the BSC and BIAWGNC with BP decoding for various (d_v, d_c) ensembles.

C. Concentration and Convergence to Cycle-Free case

The concentration results in [2] are stated as follows:

Theorem 3: Over the probability space of all graphs $\mathcal{C}^n(d_v, d_c)$ and channel realizations, let Z denote the number of incorrect messages among all nd_v variable-to-check messages passed in iteration l. Let p denote the expected number of incorrect messages passed along an edge with a tree-like directed neighbourhood of depth at least $2l$ at the l^{th} iteration. Then there exists positive constants β and γ such that
(i)[Concentration around expected value] For any $\epsilon > 0$

$$\Pr\{|Z - \mathbb{E}[Z]| > nd_v\epsilon/2\} \leq 2e^{-\beta\epsilon^2 n}. \tag{6}$$

(ii)[Convergence to the Cycle-Free Case] For any $\epsilon > 0$ and $n > \frac{2\gamma}{\epsilon}$,

$$|\mathbb{E}[Z] - nd_v p| < nd_v\epsilon/2. \tag{7}$$

The main idea in the proof of (6) is to form a martingale by revealing more and more about the object of interest. First, the nd_v edges are *exposed* one-by-one, followed by the channel observations in the next n steps. At the beginning of the exposure process, we deal with the average over all graphs and channel realizations. At the end of the $n(d_v + 1)$ steps, the exact graph and channel realization get exposed. Using the properties of *local* decoding neighbourhoods, it is shown that the change in the quantity of interest is bounded by a constant (depending on d_v, d_c) independent of n across the revelations. Azuma's inequality is then used to link the average case to specific instances. The proof of (7) follows by conditioning the expectation on whether the decoding neighbourhood is tree-like, and observing that the probability of occurence of a non-tree like neighbourhood of depth $2l$ falls off as $1/n$.

The results of this paper extend to irregular LDPC codes (which are crucial to approaching capacity) and non-binary LDPC codes among others. The analysis tools developed would also be of potential interest in network environments where there are more complex interactions among multiple component codes corresponding to the network nodes, and we wish to provide guarantees on iterative performance.

IV. Coding and System Design for QMF Relaying

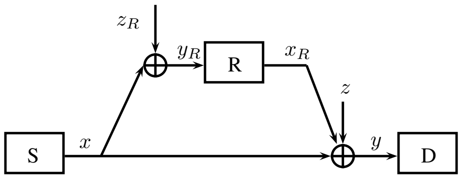


Fig. 8: Half-Duplex Gaussian Relay Channel

In this section, we summarize the QMF-based coding scheme of Nagpal et. al [3] for the half-duplex single relay network shown in fig 8. R listens for fraction f and transmits for fraction $(1-f)$. The blocklengths at the source and relay, N_S and N_R, satisfy the half-duplex constraint $N_R = (1-f)N_S$. Codebooks $\mathcal{C}_S$ and $\mathcal{C}_R$ at the source and relay are binary LDPC codes. The relay quantizes its received vector $\mathbf{y}_R$ into K_R bits and maps it to a codeword $\mathbf{x}_R \in \mathcal{C}_R$ for transmitting. The source-to-relay channel gain is h_R, and source-to-destination and relay-to-destination gains are h_1 and h_2.

By factorizing the *a posteriori* probability $p(\mathbf{x}_S|\mathbf{y})$, the authors derive a factor graph containing the Tanner graphs of $\mathcal{C}_S$ and $\mathcal{C}_R$ as subgraphs, and also consisting of nodes representing the function $p(\mathbf{x}_R|\mathbf{x}_S)$ (corresponding to the source-to-relay channel), and the multiple-access check nodes for the destination. They propose a systematic encoding at the relay where the first K_R bits $x_R[1:K_R]$ are the quantization indices of $y_R[1:fN_S]$. This reduces the function nodes for $p(\mathbf{x}_R|\mathbf{x}_S)$ to represening only a quantization operation, which further reduce to degree 2 nodes with scalar quantizers.

To simplify the multiple-access check nodes, the authors modify the system architecture that allows them to orthogonalize the interference at the destination, and in effect split up the multiple-access nodes. To do this, they introduce a one-block delay at the relay. Now, at the b-th block the destination receives the superposition of the following: (i) signal from the source containing the codeword sent at block b, $\mathbf{x}_S(m_b)$ (ii) signal from the relay containing the side information about the source's codeword at block $b-1$, $\mathbf{x}_R(q_{b-1})$. 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 noise. Then it subtracts $\mathbf{x}_R(q_{b-1})$ from $\mathbf{y}[b]$ and keeps the residual $\tilde{\mathbf{y}}[b]$ for decoding the next block. Subsequently, [3] works on an equivalent simplified channel (shown in fig. 9) with $h_{SR} = h_R$, $h_{SD} = h_1$, $h_{RD} = h_2$, Var[z_{SR}] = Var[z_{SD}] = 1, Var[z_{RD}] = $1 + |h_1|^2$. The higher variance of z_{RD} is due to treating the source transmission in block b as Gaussian noise.

The decoding on this simplified graph (fig. 10) is carried out using the BP algorithm. The messages passed by the variable

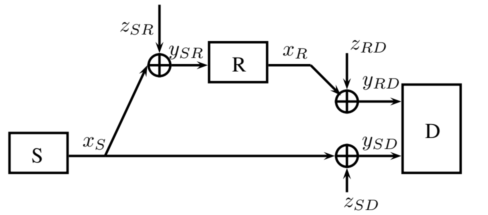


Fig. 9: Equivalent Channel Model

and check nodes in the subgraphs corresponding to $\mathcal{C}_S$ and $\mathcal{C}_R$ are exactly as outlined in [2]. For the quantizer nodes, [3] derives the message-passing rules as a marginalization of the corresponding function. This depends on the structure of the quantizer and the channel statistics. Simulations for a particular channel configuration show that the scheme provides a 1 dB cooperation gain over the strategy of not using the relay at all. Similar results are provided for a 16-QAM alphabet with the bits quantized on each BICM sub-channel separately.

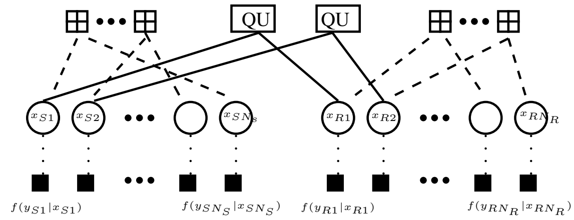


Fig. 10: Simplified factor graph with one-bit scalar quantizers

Due to the architectural modifications proposed, the scheme requires explicit decoding of the relay transmissions. Note that in the original QMF strategy, the relay transmissions do not need to be reconstructed at any point in the network, and this approach is in fact suboptimal for multiple-relay networks. As pointed out in section I, the scheme presented in this paper does not naturally extend to larger configurations.

V. Contributions and Future Directions

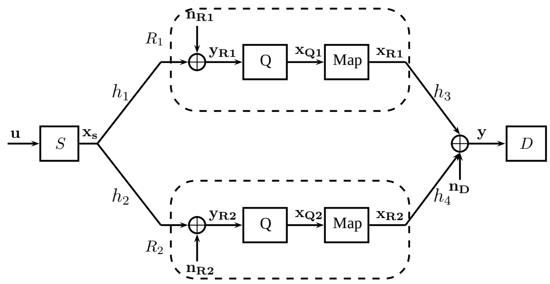


Fig. 11: QMF Strategy for the Parallel Relay Network

Fig. 11 shows the schematic diagram of our QMF strategy for the parallel-relay network, together with the signal vectors at the nodes. The code at the source is an LDPC code of the desired rate. Our mappings can be thought of as encoding $\mathbf{x_Q}$ with a rate $\frac{1}{2}$ LDPC code and letting the parity bits be the transmit sequence $\mathbf{x_R}$. The compound graphical model for decoding is shown in fig. 12. The type α nodes correspond to quantization at the relays, and facilitate soft-information exchange between $\mathcal{G}_0$ and $\{\mathcal{G}_1, \mathcal{G}_2\}$. The multiple-access (type β) nodes, connecting the $\mathbf{x_{R1}}$ and $\mathbf{x_{R2}}$ variable nodes, represent the function $p(\mathbf{y}|\mathbf{x_{R1}}, \mathbf{x_{R2}})$. The information-exchange between $\mathcal{G}_1$ and $\mathcal{G}_2$ that occurs via these nodes is an important ingredient in harnessing the benefits of co-operation from the relays.

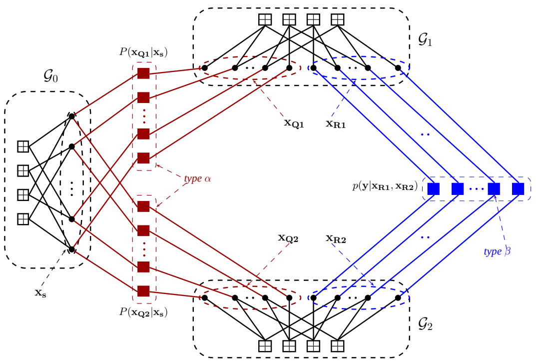


Fig. 12: Decoding Graph for binary signaling with 1-bit scalar quantizers

We derive the message-passing rules for the type α and type β check nodes and specify the following schedule for information exchange among components in the compound graphical model. Intrinsic message-passing rounds in $\mathcal{G}_1$ and $\mathcal{G}_2$ are performed in parallel, once they receive the corresponding incoming messages from the type α and type β nodes. This simultaneously updates the messages $m_{v\to c}^{\mathcal{G}_i\to\alpha}$ and $m_{v\to c}^{\mathcal{G}_i\to\beta}$ $(i = 1, 2)$. In the next step $\mathcal{G}_0$ performs local message-passing rounds with its inputs from the type α nodes, and subsequently updates $m_{v\to c}^{\mathcal{G}_0\to\alpha}$. This in turn updates $m_{c\to v}^{\alpha\to\mathcal{G}_i}$. Also $\mathcal{G}_1$ and $\mathcal{G}_2$ update their information via the type β update equations to have new estimates for the subsequent global iteration.

We incorporate non-binary signaling into our model in a 2-step procedure–namely coding in binary, followed by modulating the bits to a constellation. The symbols are quantized according to the Voronoi regions of the constellations.

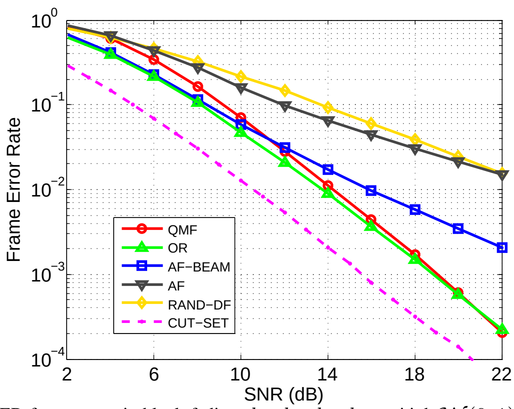


Fig. 13: FER for symmetric block-fading: $h_1, h_2, h_3, h_4 \sim$ i.i.d $\mathcal{CN}(0, 1)$. Rate = 1 bit/sec/Hz. Modulation: 4-QAM.

For the simulations, the binary codebook at the source is chosen from the ensemble designed for the AWGN channel. The mapping codebooks at the relays are samples from a (1,2)-regular LDPC ensemble[1]. We use *gray labeled* QAM constellations for signaling. Fig. 13-14, show that the performance of QMF is within 2–3 dB of the outage performance of the information-theoretic cut-set bound, at moderate SNRs. More

[1]Although a (1,2)-regular code does not by itself exhibit good performance, we found it to work better than most *ad-hoc* choices, including the (3,6)-regular ensemble, repeat-accumulate codes, and also the specially designed ensemble for the MAC channel.

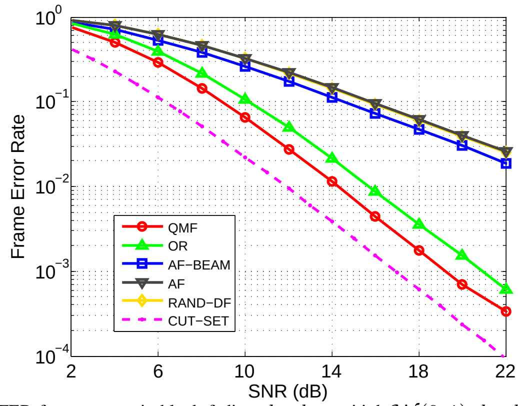


Fig. 14: FER for asymmetric block-fading: $h_1, h_4 \sim$ i.i.d $\mathcal{CN}(0, 1)$; $h_2, h_3 \sim$ i.i.d $\mathcal{CN}(0, \frac{1}{100})$. Rate = 2 bits/sec/Hz. Modulation: 16-QAM.

importantly, it performs as well or better than amplify-forward with beamforming (AF-BEAM) and opportunistic routing (OR) schemes, which require transmit CSI at the relays. This demonstrates that our implementable designs behave similar to the information-theoretic predictions.

At higher spectral efficiencies (3 bits/s/Hz) however, we observe that QMF requires relatively higher SNRs to achieve full diversity in statistically symmetric channels. Also, over static (non-fading) channels, the performance of QMF is better in asymmetric settings (providing a 2 dB gain over decode-forward (DF) relaying), whereas it is worse than DF in the perfectly symmetric (all gains equal) case, where handling multiple-access becomes challenging. Although information theory predicts that rate advantages of QMF over other schemes is more pronounced in asymmetric conditions [1], we believe that optimized code designs will help bridge the gap from the present setting and thus improve the performance at higher spectral efficiencies in symmetric fading.

One way forward would be to examine the performance of *spatially coupled* LDPC ensembles for the network. These are known to approach the MAP thresholds of the underlying graphical models in several settings, which are not as sensitive to the choice of particular ensembles as BP thresholds. This would provide insight as to whether code-design alone is the principle issue, or other aspects like noise-accumulation due to the quantizers have a deeper role to play in determining the ultimate performance limits of such systems.

Also of interest would be to evaluate the performance of our schemes over more general topoligies and traffic scenarios, where there is more inherent mixing of information streams. Similar to the benefits provided by network coding for wired networks, further significant gains over routing/scheduling strategies are expected to emerge in such settings.

REFERENCES

[1] S. Avestimehr, S. N. Diggavi and D. Tse, "Wireless network information flow: a deterministic approach", *IEEE Trans. on Infor. Theory*, vol. 57, no. 4, pp. 1872-1905, Apr. 2011.
[2] T. Richardson, R. Urbanke, "The Capacity of Low-Density Parity-Check Codes Under Message-Passing Decoding", *IEEE Trans. on Infor. Theory*, vol. 47, no. 2, pp. 599-618, Feb. 2001.
[3] V. Nagpal, I Wang, M. Jorgovanovic, D. Tse, B. Nikolic, "Quantize-map-and-forward relaying: coding and system design", *Proc. of the Annual Allerton Conf. on Comm., Control and Computing*, Sept. 2010.
[4] B. E. Schein, "Distributed Coordination in Network Information Theory", *Ph.D thesis*, Massachusetts Institute of Technology, Sept. 2001.