Multiterminal Communication Systems

[Jeff Castura and Yongyi Mao]

Rateless Coding and Relay Networks

[Benefits, challenges, and considerations for designing with rateless codes]

A framework for coding over relay channels using rateless codes is the intersection of two active areas of research in communications; namely relay networks and rateless coding. We demonstrate that there is a very natural and useful fit between these two areas and describe some design challenges and implementation considerations for this framework.

The use of relays in wireless communication networks provide a new dimension to the design space of wireless networks that promises enhancements to both the coverage and throughput of the network. In its simplest form, a relay network is a collection of terminals that are able to transmit, receive, and possibly assist the reliable delivery of information from source terminals to destination terminals. Thus, communication of data through a wireless relay network is not required to be direct; it may pass through a number of other terminals, though direct communication from source to destination is not precluded. In fact, it is possible to simultaneously use single-hop, i.e., direct, and multihop communications paths.

A question then arises: How does one code for and coordinate the various transmissions that various relays may make? This becomes a particularly difficult challenge when channel information is unavailable at the transmitting terminal, as is typically the case with time-varying, wireless channels.

As will be explained in detail in the following sections, the use of rateless codes in this setting is a promising strategy; it provides some answers to the question above and overcomes many of the problems that typically arise in relay networks.

Digital Object Identifier 10.1109/MSP.2007.904814

In general, a rateless code is a code that has a rate determined by the number of transmitted symbols required before the decoder is able to decode. The rate then is not known a priori as it is in typical fixed-rate block codes. Existing rateless codes, namely classes of fountain codes, exhibit the property of naturally adapting to channel conditions without requiring channel knowledge at the transmitter. This alone suggests their usefulness in a relay network, and is explored in detail in the following sections.

[FOUNTAIN CODES HAVE CAPACITY-ACHIEVING OR CAPACITY-APPROACHING PERFORMANCE FOR SEVERAL CLASSES OF CHANNEL MODELS.]

The remainder of the article is organized as follows. The following section provides a survey of relay networks and introduces rateless codes in detail. Factors affecting code performance and insight into the properties of rateless codes are provided. Then, the system model utilizing rateless codes for relay channels is presented. An achievable rate region for the system we consider is provided and analyzed. Results of a Monte Carlo simulation of a simple system, demonstrating that gains in throughput and outage probability can be achieved simultaneously follows and extensions and practical implementation considerations are discussed. The article concludes with some final remarks.

BACKGROUND

RELAY CHANNELS AND RELAY NETWORKS

Traditional wireless networks have predominantly used direct point-to-point or point-to-multipoint (e.g. cellular) topologies. The fundamentally different mode of transport, possible uncertainty in terminal geographical location, and difficulty in theoretical analysis of relay networks have kept them mainly to academic realms. Occasional attempts by industry to reap the benefits of relaying have been made, such as those by Ricochet Networks in the late 1990's, but have met with limited success. (Riccochet Networks is still providing service based on a relay network topology. Other vendors, such as Bel-Air Networks, are also introducing similar relay-based networks.) However, there has been a number of recent theoretical results that may spur the use of relaying techniques in practical networks. The key behind these advances are mainly a result of research in multiple antenna systems.

Multiple antenna systems, or multiple-input, multiple-output (MIMO) systems have seen remarkable growth in recent years and can deliver significant throughput and coverage gains over wireless channels compared with single antenna systems. Wireless standards such as the IEEE 802.16e and IEEE 802.11n depend heavily on MIMO principles to achieve promised throughput and reliability targets that are being demanded in the marketplace.

These systems, by using multiple antennas at the source and destination nodes, are able to achieve spatial diversity in addition to temporal diversity that traditional coding provides, as introduced in [1]. It is also possible to exploit the multiplicity of spatial channels in MIMO systems to increase the throughput beyond what would be possible in single antenna systems.

A fundamental observation to make with respect to wireless relay networks is that—with appropriate coordination or cooperation—communication between two terminals in the network can be viewed as a type of MIMO system. This may be achieved in a number of different ways, however all require some level of cooperation within the network. The relaying strategy used may impose limits on the similarity to a MIMO system. Also, constraints on the system—such as the half-duplex constraint, i.e., no terminal may receive and transmit simultaneously—may limit the degree to which communication through a relay network behaves like a MIMO system.

There are three classic relaying strategies that are commonly considered: amplify-and-forward (AF), decode-and-forward (DF) and compress-and-forward (CF). In AF, a relay is a repeater, amplifying and transmitting the received signal. In DF, the relay attempts to decode received signals. If successful, it re-encodes the information and transmits it. Finally, CF attempts to generate an estimate of the received signal. This is then compressed, encoded and transmitted in the hope that the estimate may assist in decoding the original codeword at the destination.

In their two-part paper [2], [3], Sendonaris, Erkip and Aazhang introduce and examine the concept of user cooperation diversity. Here, the authors demonstrate that simple cooperation between transmitting users can increase throughput and coverage simultaneously. The implemented strategy uses a pair of transmitting, full-duplex users who cooperate in sending independent data from both users to a common destination. This is accomplished by a multiperiod transmission process. Initially, each user sends its own data, while listening to the other users' transmission. After some time, each user will allocate some amount of power to send an estimate of what it received from the other. In essence, each user is acting as a relay for the other and using the AF relaying strategy. This approach demonstrates some of the gains that may be had through cooperation, though it relies on some channel information at the transmitter.

The DF and CF strategies are thoroughly examined for wireless channels in [4]. In addition to providing a thorough survey of relay networks, they show, under certain conditions, that the DF strategy is capable of achieving rates up to the ergodic capacity of the channel. Furthering these results, Lai, Liu and Gamal in [5] combine the use of DF and CF to achieve the same bounds, but with fewer restrictions on the system.

Dohler et al. introduce virtual antenna arrays in [6]. Here, groups of terminals cooperate to form a virtual MIMO system and exploit the spatial diversity that results. This is a similar concept to user cooperation, but focuses on different design aspects, such as link budget impact. Further work done in the area of signal design for relay networks is by Nabar, Bolcskei and

Kneubuhler in [7], where the authors consider code design aspects of DF and AF strategies.

A natural extension of the basic implementation described in [2] is to use coded cooperation and this is described in [8] by Hunter and Nosratinia with further analysis and implementation details for wireless channels given in [9]. Further approaches using coded cooperation are given in [10], [11], where the authors propose a DF-based scheme with many opportunistically cooperating terminals, and show that diversity gains scale in the number of potential relays rather than the actual number of participating relays.

As the concept of coded cooperation has grown, implementations of the concept have begun to appear. In [12], Zhang, Bahceci, and Duman present a strategy based on turbo codes for communication on relay channels. There, they design the code and iterative decoder, and show that the performance can be close to an achievable rate bound.

In works that foreshadow the use of rateless codes for relay channels, Caire and Tuninetti in [13], and Zhao and Valenti in [14] propose and analyze the use of Hybrid-ARQ for relay channels. In [13], hybrid-ARQ protocols for the Gaussian collision channel is studied. Notably, the results translate to relay channels. This is noted in [14], where the authors propose the use of a hybrid-ARQ-type protocol for relay networks using orthogonal signaling slots with the half-duplex constraint. They demonstrate that their protocol provides significant improvements in throughput and average transmission delay.

Also foreshadowing the application of rateless codes, Mitran, Ochiai, and Tarokh present in [15] a two-phase communication scheme for wireless devices in a network with the half-duplex constraint along with an information-theoretic performance analysis. The two phases are the listening phase, in which the source node broadcasts and other nodes listen, and the collaboration phase, in which multiple nodes cooperate to transmit to the destination. It is assumed that the channel state information is not available at the transmitters but is available at the receivers. The results suggest that such a collaborative communication scheme can lead to significant diversity enhancement compared to direct communication between source and destination.

From an information theoretic perspective, the work by Cover and El Gamal in [16] remains the foundational treatment on the relay channel. There, the authors present a number of theorems for this channel under different conditions. Unfortunately, the capacity is not solved for the general case; only an achievable rate is provided. Capacity results are provided by the authors for certain degraded channels. A partial converse to the general case was communicated in [17], though the full solution to this problem remains open.

As the idea of relay networks attracted attention, researchers began to investigate their information theoretic aspects. In particular, determining achievable rate regions was investigated under a number of different assumptions. As a fairly general case, Gupta and Kumar in [18] and [19] present an achievable rate region for arbitrary relay networks. Further results are available in [5] and [20]–[22] under different network assumptions and topologies.

More recently, attention has turned to generalizations of relay networks, particularly the MIMO relay network. In this area, Wang et al. in [23] present bounds on the capacity of MIMO relay channels. This is furthered by Tang et al. in [24].

As the connections between relay channels and MIMO systems matured, it became clear that the fundamental properties of MIMO channels are also applicable to the relay case. The well-known diversity-multiplexing trade-off (DMT) presented by Zheng and Tse in [25] was extended to the relay case by Yuksel and Erkip in [26]. When constrained to half-duplex channels, Azarian, El Gamal and Schniter demonstrated in [27] a novel AF scheme that achieves a DMT bound for AF strategies, and provides insight into some of the obstacles facing practical relay network implementations. Finally, [28] demonstrates that he CF strategy is the only known strategy capable of achieving the full duplex relay DMT.

RATELESS CODING

Rateless codes are codes that encode a finite number of messages but have an infinitely long block length and are thus parameterized by a single number k, the length in bits of the information block. Comparatively, fixed-rate block codes are parameterized by the pair (k, n), where n defines the codeword length. The transmission of a rateless codeword is terminated when the receiver decodes the message and uses a feedback channel to communicate an acknowledgement (`ACK`) to the transmitter. As indicated by its name, a rateless code does not have a fixed rate, but rather the rate is determined on the fly by the time at which the receiver decodes the message.

As the first efficient class of rateless codes, fountain codes not only have low complexity but also have capacity-achieving or capacity-approaching performance for several classes of channel models. LT Codes, introduced by Luby in [29] were shown to achieve capacity for any BEC. In practice, LT codes are prone to have a noticeable error floor for small k. Raptor codes were introduced by Shokrollahi in [30] and use LT codes as an inner code with a high-rate LDPC outer code. Raptor codes were also shown to be capacity achieving for the BEC, but also to have the beneficial properties of little or no error floor for small block lengths, and a linear-time encoding and decoding computational complexity.

Other rateless codes have been proposed, and their performance on other channels have been investigated, see e.g., [31]–[33], and very good performance has been found for many other channels including the AWGN channel and various types of fading channels. Additionally, rateless codes can operate universally over classes of channels, adapt their rates to the channel realization, and require no knowledge of channel state information or even channel statistics at the transmitter. The performance and properties of these codes make them well suited for space-time collaboration over relay channels.

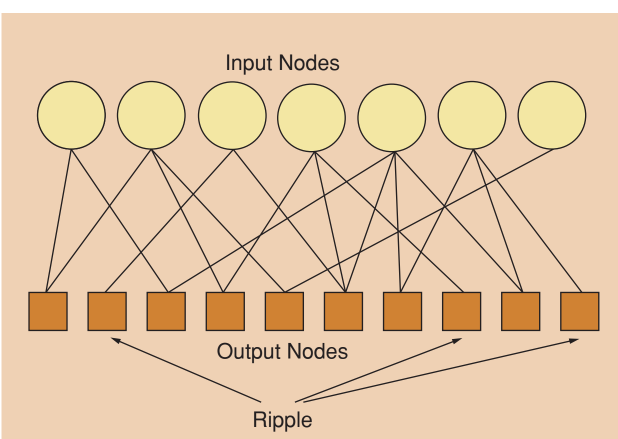
An example of a fountain code over GF (2) operating over the binary erasure channel is described here, with these restrictions chosen for clarity of exposition. The transmitter and receiver are assumed to be synchronized in some manner. This may be as simple as sharing a common clock source. Both transmitter and receiver are initialized at time zero. Given some block of k bits of information, the transmitter consecutively generates codeword symbols, and each symbol transmission corresponds to a single time-step.

THE USE OF RATELESS CODES FOR RELAY NETWORKS NATURALLY ACCOUNTS FOR COMMUNICATION EFFICIENCY AND SYSTEM ROBUSTNESS SIMULTANEOUSLY.

The generation of a codeword symbol is a two-step process. The first step requires the transmitter to draw a pseudo-random number from some a priori known distribution Ω over the positive integers. The receiver also knows this distribution and is assumed to be able to independently draw the identical value as the transmitter. Given this value $d \leq k$, the transmitter pseudorandomly selects d distinct information bits uniformly. These bits are XORed together, and the result is transmitted over the channel. This operation is equivalent to pseudo-random, on-the-fly generation of the generator matrix $\mathbf{G}$.

The receiver is capable of reproducing the identical random realization that is used to generate codeword symbols at the transmitter and in this way is able to maintain track of $\mathbf{G}$. The receiver also tracks, in some manner, the amount of information it has received. At some point it determines that it may be possible to successfully decode. It then attempts to decode. The algorithm used to decode may depend on the implementation, though the iterative belief propagation algorithm is often acceptable [34]. In the case of the binary erasure channel, belief propagation amounts to a graph pruning procedure.

Following the terminology of [29], [30], define the factor graph associated with a rateless code at time t to consist of k variable nodes called the input nodes, and x check nodes called the output nodes. The number of check nodes x is equal to the number of nonerased symbols received at the decoder at time t. Each of the x output nodes has edges connecting it to every input node that was used to generate it, based on the distribution Ω.



[FIG1] Factor graph decoding of a fountain code.

An example factor graph for a rateless code is presented in Figure 1 for $x = 10$. Decoding attempts to prune the graph to coverall of the input nodes. An input node that is covered is a node for which there is no uncertainty about the value it must take, and occurs whenever it is connected to a single output node. The decoding process operates as follows: Define the set of all degree-one output nodes to be the ripple. If this set is empty, terminate decoding. Choose a node from the ripple and cover the unique, connected input node with the value of this node. Remove both the output node and input node from the graph. If no input nodes remain in the graph, decoding has completed successfully and the received message is given by the value of each of the covered input nodes.

If decoding is successful, the decoder sends an ACK to the transmitter using a feedback channel to terminate transmission. Otherwise, the decoder will simply collect some further number of codeword symbols and attempt decoding again. Using this approach, it is almost always possible for the receiver to successfully decode the transmitted codeword.

The choice of distribution Ω is critical to decoding performance. It was shown in [29] that for the binary erasure channel there exists an Ω, which results in a rateless code that achieves the capacity of any erasure rate on the channel. This remarkable distribution is termed the soliton distribution for its resemblance to soliton pulses. Unfortunately this distribution is only optimal for $k \to \infty$. However, Luby also defined the robust-soliton distribution, a slight modification of the soliton distribution that works very well for finite k, and can achieve rates arbitrarily close to capacity on the binary erasure channel.

Over other channels, there is no known distribution that is universal in the sense that the soliton distribution is for the binary erasure channel. In fact, it was shown in [32] that there does not exist such a distribution for the AWGN channel. Despite this fact, excellent results over large ranges of channel parameters have been shown for rateless codes over the AWGN and binary symmetric channel [31], [32] as well as a number of fading channels [33].

Given these properties, the use of rateless codes for wireless relay channels seems to be a synergistic match. Building on the ideas developed in [15], Castura and Mao in [35] present a framework for rateless coding over relay channels that achieves the predicted diversity gains. Liu in [36] presents an extension of the work of [35], demonstrating a rateless code based protocol that can achieve even better efficiencies combining DF and CF concepts.

Another approach that utilizes rateless coding for relay networks is described by Molisch et al. in [37] and [38]. Focusing on multirelay networks, the authors present two protocols utilizing rateless coding for a large numbers of potential relays. The first protocol requires some fixed number of relays to receive the source message in the listening phase before the collaboration begins. The second is a natural extension of [35] where relays independently begin collaboration as soon as they have decoded the source message. Practical implementation issues relating to the coordination of relays are addressed by the authors, and simulations of the proposed protocols are presented demonstrating the benefits of the approach, including an analysis of the expected transmission energy.

MOTIVATION FOR RATELESS-CODED RELAY NETWORKS

Approaching the communication problem from a coding-theoretic perspective, one seeks to design a practical coding framework that effectively implements a collaboration strategy to maximize achievable rate and minimize outage probability. (One may also, depending on the ultimate application, wish to minimize transmission latency, transmit power, etc.)

It appears that no fixed-rate coding system is capable of driving the outage probability to zero without channel state information at the source. Also, unless operating at a low efficiency (i.e., low rate), no fixed-rate coding system is robust to the variation of channel statistics typical in a wireless setting. With the use of a feedback channel it is possible to signal channel information and thereby overcome many of these limitations. However, for large relay networks this overhead becomes significant compared to the small number of bits needed for acknowledgments in a rateless system, particularly for time-varying channels or moving terminals. Further analysis comparing fixed-rate and rateless approaches in the relay network setting are given in [37]. In particular, the benefits of a rateless system for large relay networks are highlighted.

Usual solutions to this problem have been to utilize ARQ or hybrid-ARQ methods such as described in [14]. Like rateless coding, feedback channels are required to signal successful reception of codewords at the destination. In fact, rateless coding may be considered a form of continuous incremental redundancy compared to the block-based incremental redundancy provided by hybrid-ARQ.

The use of rateless codes for relay networks naturally accounts for communication efficiency and system robustness simultaneously. We review some theoretical bounds resulting from this framework and show the diversity and throughput gains that may be had. We present a simulated implementation of this framework based on fountain codes as a tutorial to demonstrate these advantages. Finally, we discuss a number of practical issues and limitations of this system, suggesting alternatives, such as those provided in [37].

RATELESS RELAY SYSTEM MODEL

To demonstrate the usefulness of rateless codes in relay networks, we will focus on the base system model presented in [15] and extended for rateless codes in [35].

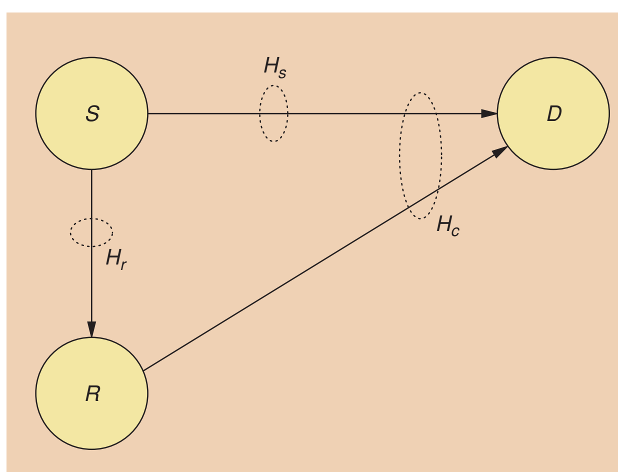
Consider the system shown in Figure 2 with three wireless devices. This configuration forms the building block for all relay networks and so is useful to first understand before considering general networks. The source s wishes to communicate with the destination d, possibly with the help of the third device, relay r. We consider the half-duplex scenario, where the relay does not receive from the source and transmit to the destination at the same time. We assume that each transmitted symbol from a source antenna or from a relay antenna has the same average energy E_s, though in general this need not be the case. Let $X[i]$ and $U[i]$ be respectively the symbol vectors at time i transmitted from the source antennas and the relay antennas. Consider the quasi-static Rayleigh fading model and use H_r, H_s and H_c to denote the channel gain matrices for the source-to-relay channel, source-to-destination channel,and the compound channel from the combined antennas of source and relay to the antennas of the destination. Let $Y[i]$ and $Z[i]$ be the received signals at time instant i at the relay and destination respectively. During each codeword transmission

$$Y[i] = H_r X[i] + N_Y[i],$$
$$Z[i] = H_c[X[i]^T \, U[i]^T]^T + N_Z[i],$$

where N_Y and N_Z are zero-mean (vector-valued) white complex Gaussian noise processes received at the relay and the destination respectively. We note that H_s is a submatrix of H_c and if the relay does not assist the source at time instant i, the received signal $Z[i]$ reduces to

$$Z[i] = H_s X[i] + N_Z[i].$$

We assume that the entries of H_r and H_c are drawn independent identically distributed (i.i.d.), respectively, from complex Gaussian distributions with variances G and 1. This restricts the source-to-destination channel and relay-to-destination channel to have the same fading statistics. Such a restriction plays no essential role in the applicability of this work, and merely serves as a simplification assumption, also making these results comparable with those of [15]. The values H_r and H_c are assumed to be known at the respective receivers but unknown at each transmitter. The variance of N_Y and N_Z are both $N_0/2$, also known at the receivers. It is also assumed that the relay is



[FIG2] Three node wireless relay network.

capable of synchronizing with the source at the symbol level. Synchronization issues will be discussed in the following.

For such a setup, an information-theoretic analysis of block-coded schemes is presented in [15], which may be summarized as follows. The source selects a rate R to transmit a block of k-bits of information using a block code of length $n := k/R$. The relay, aware of the channel state H_r, decides a number $f \in (0, 1]$ such that after listening for $n_1 := fk/R$ symbols, the relay is able to decode the information block and collaboratively transmits to the destination.

[COMMUNICATION BETWEEN TWO TERMINALS IN THE NETWORK CAN BE VIEWED AS A TYPE OF MIMO SYSTEM.]

The case that $f = 0$ is operationally equivalent to the situation where the relay has the knowledge of source codeword noncausally, and therefore equivalent to a MIMO system where the antennas of the source and those of the relay jointly form the set of transmit antennas. On the other hand, the case that $f = 1$, is equivalent to the situation where the relay is not able to decode before the destination and thus does not collaborate.

Under this scheme, Theorem 1 of [15] shows that given a channel realization (H_r, H_c), and given a choice of f, any rate R satisfying

$$R < fC(H_s, \gamma) + (1 - f)C(H_c, \gamma) \quad (1)$$

and

$$R < fC(H_r, \gamma) \quad (2)$$

simultaneously, or satisfying

$$R < C(H_s, \gamma) \quad (3)$$

is achievable, where $C(H, \gamma) := \log_2 \det(I + \gamma HH^T)$ is the MIMO capacity formula under equal power allocation across antennas, and γ in our setup equals E_s/N_0. Since H_r and H_c are both random variables, the authors of [15] argue that for any fixed R and any f, the outage probability may be defined as the probability that R is outside the interval prescribed by (1), (2), and (3), under the channel law for H_r and H_c. Furthermore, they show that there exists an optimal choice $\hat{f}$ of f—in the sense of minimizing the outage probability for a fixed rate R. Such a choice can also be seen as improving the diversity order, as shown in [27].

This fixed-rate strategy is challenged by practical issues. Without channel knowledge, as long as the source selects a transmission rate R, there is no opportunity for the relay to improve the rate, even when the channel supports much higher rates. The only benefit that the relay offers is a decreased outage probability. It is also notable that no matter what rate is chosen, the outage probability is bounded above zero for quasi-static or slow fading channels. With the use of feedback in the system, it is possible to signal a suitable R to the source, though knowledge and coordination of relays remains a difficult obstacle. In fact, the optimal choice of rate requires that the source be aware of the existence and participation of the relay as well as channel knowledge.

To handle efficiency and robustness simultaneously, rateless codes may be used in place of the fixed-rate codes. The source encodes a block of k information bits using a rateless code with parameter k and sequentially broadcasts the codeword symbols to both the destination and the relay. The relay attempts to decode the information until it succeeds. At this time, if the destination has not decoded the information, the relay then collaborates with the source by transmitting to the destination using another rateless code. Starting from time instant 1, the destination also attempts to decode the source information, and whenever it can decode, it sends an ACK back to the source and the relay in order to terminate the current transmission. Here we note that the 1-b feedback for acknowledgment entails little implementation difficulty as long as the feedback channel exists. Note that it is straightforward to generalize this rateless coding concept to multiple relay, multiple antenna systems and to other channel models.

In this strategy, the source does not need to be aware that a relay exists. The rateless nature of the coding scheme allows the source to communicate with the destination at a rate adapted to the channel conditions, to the availability of the relay, and to the collaborating strategy of the relay. Furthermore, although we are dealing with fading channels, the outage probability can be made arbitrarily small. This is because the successful decoding almost surely occurs at a time corresponding to a rate supported by the channel.

A fundamental question then arises: What rates are achievable with the presented scheme? Here, we present an achievable rate result built on the work of [15].

Let n be the time needed for the destination to decode a message. Similarly, we denote by n_1 the time needed for the relay to decode a message. We define the realized rate R of a transmission by $R := k/n$ b/channel use. Notice that both n_1 and n are random variables depending on channel realization (H_r, H_c) and noise realizations. Therefore the realized rate R is also a random variable. We say that a rate R is achievable for a given channel realization (H_r, H_c) if there exists a family of rateless codes (each parameterized by a different k) such that after k/R channel uses, the decoding error at the destination can be made arbitrarily small for sufficiently large k.

It is shown in [35] that there is an optimal choice $\tilde{f}$ for f in the sense of maximizing achievable rate, and this is given by

$$\tilde{f} := \begin{cases} \frac{C(H_c,\gamma)}{C(H_c,\gamma)+C(H_r,\gamma)-C(H_s,\gamma)} & \text{if } C(H_r,\gamma)>C(H_s,\gamma) \\ 1 & \text{otherwise.} \end{cases} \quad (4)$$

We then define

$$\tilde{R} := \tilde{f}C(H_s, \gamma) + (1 - \tilde{f})C(H_c, \gamma) \quad (5)$$

to be the rate achieved using $\tilde{f}$. This result can be argued geometrically following Figure 3. The shaded region in the figure is the set of all (R, f) pairs satisfying (1) and (2). To find the supremum of R satisfying constraints (1) and (2) simultaneously or satisfying (3), we need not consider (3) due to the fact $C(H_r, \gamma) > C(H_s, \gamma)$. It is clear that the supremum $\tilde{R}$ is defined by the intersection of the boundary constraint of (1) and the boundary constraint of (2). Such an intersecting point always exists for some $f \in [0, 1]$ since $C(H_c, \gamma)$is strictly greater than $C(H_s, \gamma)$. Solving for the intersecting point, we get $(\tilde{R}, \tilde{f})$ as shown in the figure. Then if source chooses rate $\tilde{R} - \delta$ and the relay starts collaboration at the time that is $\tilde{f}$ fraction of the code block length, with standard information-theoretic arguments, the decoding error at the destination can be made arbitrarily small for sufficiently large block length.

Some insights may be obtained by distinguishing $\tilde{f}$ with $\hat{f}$. Given a rate R, $\hat{f}$ minimizes outage probability resulting from channel uncertainty, and such minima are strictly bounded above zero. However, knowing channel realizations at the source, $\tilde{f}$ maximizes the achievable rate dictated by (1), (2), and (3) and allows virtually zero outage probability.

Although this achievable rate is only shown to be valid for block coding schemes and for the case in which the channel is known at the transmitters, we demonstrate in the following section that rateless codes can also achieve this rate without channel knowledge at the transmitter.

RATELESS CODE IMPLEMENTATION

The presented rateless coding scheme may be implemented with fountain codes. We implement such a scheme for illustrative purposes using a Raptor code as an outer code concatenated with a space-time inner code to form a rateless code. The presented system makes simplifying assumptions that real systems need to address. These are discussed in the following.

This architecture is tested over a simplified channel model, where the source, relay and destination each have only one antenna, and are time-synchronized at the symbol level. A Raptor code with $k = 9,500$ is used as the inner code both at the source and the relay. The LDPC component code of the Raptor code has rate 0.95. We use the degree distributions as in [30] for the inner (LDPC) and outer (LT) components of the Raptor code. Consecutive output symbols from the Raptor code are then QPSK modulated, chosen for simplicity.

The rateless code used by the relay is the same Raptor code also with QPSK modulation. When in the collaboration phase, the relay aligns its output symbols in time to correspond to the source's output symbols.

The output symbols of the source and relay are input to the distributed space-time inner code. The space-time inner code uses the Alamouti scheme [39] and works as follows. During both the listening and collaboration phases, the source simply passes through the input symbols. The relay, once in the collaboration phase, acts as the secondary antenna in which consecutive pairs of input symbols are transformed according to the Alamouti scheme.

This distributed space-time code is received and decoded at the destination. During the listening phase the destination receives symbols only from the source, and during the collaboration phase the destination performs standard Alamouti decoding.

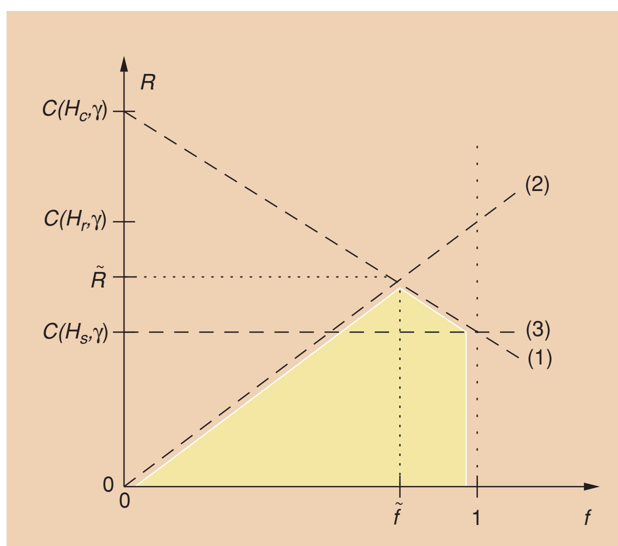
For each transmitted codeword, belief propagation decoding is attempted periodically at intervals of 100 channel uses. During each iteration of decoding, the decoder examines whether hard decisions on the messages form a codeword. When this occurs, the transmission of the codeword is terminated. If the hard decisions do not form a codeword within 100 iterations, the current decoding attempt is stopped and the decoder waits for the next decoding attempt. For the purposes of the simulation we assume that the receiver knows whether it decodes correctly. In practice, this may correspond to the case where CRC bits are used in the message. (We note that the entailed rate loss is negligible for large k.)

SIMULATION RESULTS

We perform a Monte Carlo simulation of the presented rateless code implementation over a range of SNR from −5 dB to 15 dB. For the purpose of analyzing our results, we denote for each codeword transmission the decoding times at the relay and destination by n_1 and n, respectively. The realized fraction of time that the relay was listening is defined as $f := \min\{1, n_1/n\}$. The maximum achievable realized rate for a given realized f is defined as

$$R^* := fC(H_s, \gamma) + (1 - f)C(H_c, \gamma), \tag{6}$$

which is prescribed by (1), (2), and (3). Then the gap between $\tilde{R}$ and R^* indicates the suboptimality of the realized f compared to the optimal $\tilde{f}$ for a given channel realization. To an extent, this gap also reflects the spectral efficiency limitation associated with the modulation scheme. The realized rate of a single codeword transmission is denoted by $R := k/n$ and if decoding fails after



[FIG3] **Sketch of the relationship between (1), (2) and (3) for $C(H_r, \gamma) > C(H_s, \gamma)$.**

200,000 channel uses or an incorrect message is decoded we regard $R = 0$. The gap between R^* and R indicates the coding loss due to the suboptimality of the Raptor code.

Figure 4 presents the curves of R, R^* and $\tilde{R}$ versus SNR, as well as the standard 2×1 MIMO capacity bound as a function of receive SNR, for $G = 15$ dB. Notice from the figure that R^* converges with $\tilde{R}$ in the low SNR regime, which indicates that the realized f is nearly optimal. As SNR increases, the R^* curve reaches the asymptote of 2-b/channel use, governed by the QPSK modulation scheme. There is an approximately constant fraction of rate loss across all SNR when comparing R with R^*. This holds true for a wide range of values for G. Clearly, at high SNR, both R^* and R can be increased by using higher-order modulations.

Verifying that the diversity order of the implemented system matches that of a 2×1 MIMO system is straightforward, which is expected based on the results of [15]. This also agrees with the results of [27] with respect to the DMT, where fixed-rate transmission is considered.

EXTENSION TO MULTIPLE RELAYS

The presented framework may naturally be extended to include an arbitrary number of relays, corresponding to the asynchronous protocol in [37]. All terminals in the network that are neither the source or destination terminal may act as a receiver for the codeword transmission from the source and other relays.

Each relay accumulates information from all other current collaborators. As soon as it is able to decode, it begins collaboration with the source and other relays already collaborating. This approach can match the achievable rate region for arbitrary relay networks found in [19].

In addition to this approach, other options are detailed in [37]. Specifically, one may coordinate relays in the network to not collaborate until a sufficient number have successfully decoded the source message. Tradeoffs between these two extensions of the basic framework are discussed in detail in [37].

PRACTICAL CONSIDERATIONS

Some aspects of this framework have been idealized or otherwise simplified for the sake of convenience. In practical settings, any such collaboration scheme must consider at least the coordination and synchronization of relays, and provide an operational feedback channel.

Coordination of relays, and of the system as a whole, is of fundamental importance. By coordination, we mean the allocation of unique resources to collaborating relays. For example, if using space-time codes for collaboration, how should a relay choose which column (in the case of space-time block codes) or component of the code to use? Do the source or relays need to be notified of the participation of other relays? The work in [40] presents one possible option that the authors term as opportunistic large arrays, in which collaborating relays ‘flood’ the network. The random distribution of relays and channel realizations result in both space and time diversity gains. Other alternatives utilizing space-time codes may be found in [7] and [41].

A different approach found in [37] uses CDMA spreading codes allocated uniquely per relay, which, at the expense of bandwidth expansion, provides orthogonality at the destination. A Rake receiver is used for recovery of the desired symbols from the collaborators. Finally, the work in [27] suggests the use of an artificial ISI channel instead of an orthogonalization method, which they show can achieve the bounds of the DMT.

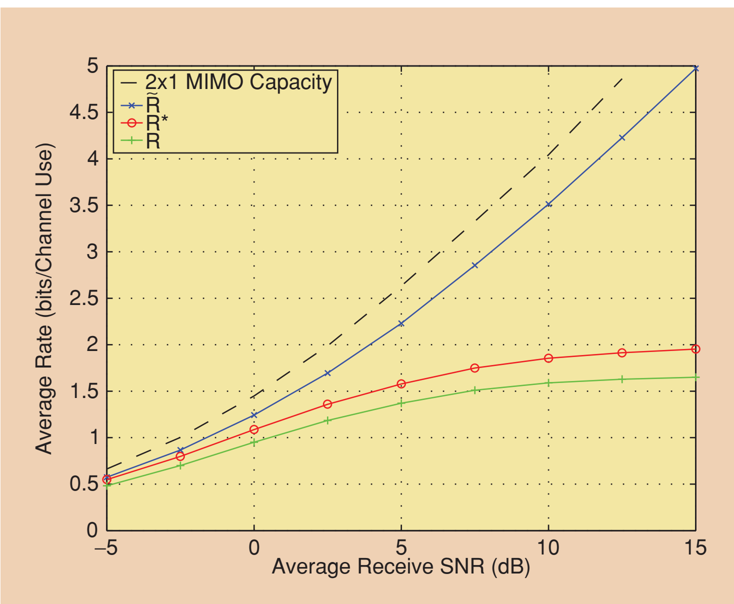
In the presented system, it is assumed that symbol-level synchronization is maintained between received signals. In relay networks with unknown or variable distances between terminals, this may not be simple to achieve, although a number of potential solutions may be used. For example, at the expense of bandwidth expansion, OFDM may be applied, effectively reducing the symbol rate (for the same data rate) to a suitable degree. Synchronization in not vital for operation of the system, though lack of it may limit achievable rates [37].

Finally, a feedback channel is an essential feature of any rateless system. Practically, the feedback channel must exist orthogonally to the forward channel. This requirement may have cost or complexity implications for the terminals in the network. Natural options are to use either TDD of FDD between forward and feedback channels.

CONCLUSIONS

We have presented a framework for collaboration over wireless relay channels based on rateless codes that is simultaneously robust against outage and efficient in rate. In particular, we have shown that this system can be implemented using fountain codes and is capable of performing at rates approaching theoretical limits across a wide variety of channel configurations.

The combination of the ideas of rateless codes and



[FIG4] Average system rates versus average receive SNR, $G = 15$ dB.

wireless relay networks provides for a method of increasing the throughput and reliability of wireless networks. Rateless codes, with their ability to adapt to the conditions of the channel—without requiring channel knowledge at the transmitter—are well suited for the channel conditions that wireless relay networks encounter.

Relay networks, with their ability to exploit the spatial diversity that physically separated terminals provide, are capable of providing increased coverage and throughput in the system. Together, these two technologies complement each other, and promise to help multiterminal networks to become economically viable means for wireless communications.

AUTHORS

Jeff Castura (jcastura@site.uottawa.ca) received his Bachelor of Applied Science in engineering physics from Queen's University, Canada. He completed his Master of Applied Science degrees in electrical engineering from the University of Toronto. After a number of years in the wireless telecommunications industry, Jeff joined the University of Ottawa in 2004, where he is currently a Ph.D. candidate in electrical engineering. Jeff's current research interests include information theory, wireless systems, machine learning and computational finance.

Yongyi Mao (yymao@site.uottawa.ca) received his Bachelor of Engineering degree at the Southeast University, Nanjing, China in 1992, his medical degree at Nanjing Medical University, China in 1995, his Master of Science degree at the University of Toronto, in the Department of Medical Biophysics in 1998, and his Ph.D. in electrical engineering at the University of Toronto in 2003, when he joined the faculty of School of Information Technology and Engineering at the University of Ottawa. His research interests include communications, information theory, and statistical inference.

REFERENCES

[1] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, Part I: System description," *IEEE Trans. Commun.*,vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. part II. implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[4] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[5] L. Lai, K. Liu, and H. El-Gamal, "The three-node wireless network: Achievable rates and cooperation strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 805–828, Mar. 2006.

[6] M. Dohler, E. Lefranc, and H. Aghvami, "Virtual antenna arrays for future mobile communication systems," in *Proc. IEEE ICT 2002*, Beijing, China, June 2002 [CD-Rom].

[7] R. Nabar, H. Bolcskei, and F. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[8] T.E. Hunter and A. Nosratinia, "Cooperation diversity through coding," in *Proc. IEEE Int. Symp. Inform. Theory*, 2002, p. 220.

[9] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: Space-time transmission and iterative decoding," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 362–371, Feb. 2004.

[10] J.N. Laneman and G.W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[11] J.N. Laneman, D.N.C. Tse, and G.W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[12] Z. Zhang, I. Bahceci, and T.M. Duman, "Capacity approaching codes for relay channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2004, p. 2.

[13] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1971–1988, July 2001.

[14] B. Zhao and M.C. Valenti, "Practical relay networks: A generalization of hybrid-ARQ," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 7–18, Jan. 2005.

[15] P. Mitran, H. Ochiai, and V. Tarokh, "Space-time diversity enhancements using collaborative communications," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2041–2057, June 2005.

[16] T. Cover and A.E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[17] Z. Zhang, "Partial converse for a relay channel (corresp.)," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1106–1110, Sept. 1988.

[18] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[19] P. Gupta and P.R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[20] L. Xie and P. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[21] S. Kulkarni and P. Viswanith, "A deterministic approach to throughput scaling in wireless networks," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1041–1049, June 2004.

[22] M. Gastpar and M. Valenti, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.

[23] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[24] T. Tang, C.-B. Chae, R.H. Jr., and S. Cho, "On achievable sum rates of a multiuser MIMO relay channel," in *Proc. IEEE Int. Symp. Inform. Theory*, 2006, pp. 1026–1030.

[25] L. Zheng and D.N.C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[26] M. Yuksel and E. Erkip, "Diversity-multiplexing tradeoff in multiple-antenna relay systems," in *Proc. IEEE Int. Symp. Inform. Theory*, 2006, pp. 1154–1158.

[27] K. Azarian, H.E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[28] M. Yuksel and E. Erkip, "Multi-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective," *IEEE Trans. Inform. Theory*, 2007, to be published.

[29] M. Luby, "LT codes," in *Proc. 43rd Ann. IEEE Symp. Found. Comp. Sci.*, 2002, pp. 271–280.

[30] A. Shokrollahi, "Raptor codes," in *Proc. IEEE Int. Symp. Inform. Theory*, 2004, p. 36.

[31] R. Palanki and J.S. Yedidia, "Rateless codes on noisy channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2004, p. 37.

[32] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.

[33] J. Castura and Y. Mao, "Rateless coding over fading channels," *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 46–48, Jan. 2006.

[34] K. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[35] J. Castura and Y. Mao, "Rateless coding for wireless relay channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2005, pp. 810–814.

[36] Y. Liu, "A low complexity protocol for relay channels employing rateless codes and acknowledgement," in *Proc. IEEE Int. Symp. Inform. Theory*, 2006, pages 1244–1248.

[37] A.F. Molisch, B. Neelesh, J.S. Yedidia, and J. Zhang, "Cooperative relay networks using fountain codes," in *Proc. Global Telecommun. Conf.*, Nov. 2006, pp. 1 – 6.

[38] A.F. Molisch, B. Neelesh, J.S. Yedidia, and J. Zhang, "Cooperative relay networks with mutual-information accumulation," submitted for publication.

[39] S.M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[40] A. Scaglione and Y. Hong, "Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances," *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2082–2092, Aug. 2003.

[41] B. Mergen and A. Scaglione, "Randomized space-time coding for distributed cooperative communication," in *Proc. IEEE Int. Conf. Commun.*, June 2006, vol. 10, pp. 4501–4506. [SP]