

# A Cross-Layer Perspective on Rateless Coding for Wireless Channels

Thomas A. Courtade and Richard D. Wesel

Department of Electrical Engineering, University of California, Los Angeles, CA 90095

Email: {tacourta, wesel}@ee.ucla.edu

**Abstract**—Rateless coding ensures reliability by providing ever-increasing redundancy, traditionally at the packet level (i.e. the application layer) through erasure coding. This paper explores whether additional redundancy for wireless channels is most helpful at the packet level through erasure coding or at the physical layer through lower-rate channel coding.

This cross-layer trade-off is explored in a traditional wireless setting where the communication of a message consisting of a fixed number of packets takes place over a Rayleigh fading channel. The examined scenarios include both a single receiver and multiple cooperating receivers allowing the results to be extended to situations where selection diversity is available in the system.

For several interesting scenarios, this paper determines the optimal trade-off between the amount of packet-level erasure coding and physical-layer channel coding required to provide reliable communication over the widest range of operating SNR's. Our results indicate that packet-level erasure coding can provide a significant benefit when no other form of diversity is available. In many cases, the amount of redundancy that should be allocated to such erasure coding is nearly constant, and further redundancy (i.e. any rateless coding) should be applied to the physical layer.

## I. INTRODUCTION

In recent years, packet-level erasure coding schemes, such as Raptor coding introduced in [1], have gained widespread popularity as an efficient and reliable means of message delivery over erasure channels because of their rateless nature. Erasure channels are commonly found in wired computer networks, however wireless links can also be viewed as packet-level erasure channels when a cyclic redundancy check, or similar mechanism, is employed. In fact, most modern wireless computer networks are packet-based and therefore have an erasure channel component that should be considered in the system model.

Unlike wired networks, wireless channels typically require significant physical-layer coding in order to combat interference and fading. Therefore, in a wireless application of rateless coding, we should consider the effects of the packet-level and physical-layer coding jointly. Luby et al. noticed the need to consider cross-layer effects in the simulation-based analysis provided in [2] for 3GPP.

As the name implies, rateless codes operate by adding ever-increasing redundancy to a particular codeword until it is successfully decoded at the receiver. Therefore, the key issue to address in cross-layer rateless coding is whether there exists an

optimal allocation of this ever-increasing redundancy between the packet-level and physical-layer codes.

In an AWGN channel, it is intuitive that the optimal solution is to use no erasure coding. In other words, we should use a rateless physical-layer code in an AWGN channel and use no packet-level coding. In this way, because we expect little variation in SNR from packet to packet, if the rate does not exceed capacity we can expect to decode all packets reliably.

The optimal solution is less clear when we consider a fading channel. In a fading channel, intuition suggests that we may want a cross-layer scheme with a slightly higher channel-code rate used jointly with some packet-level erasure coding to take advantage of the channel when the SNR is high and minimize the loss incurred by dropping packets during deep fades. It has been noted (see for example [3]) that rateless coding can be used to drive the outage probability to zero in a fading environment. However, a cross-layer perspective has been missing from this discussion.

In this paper, we study the optimal allocation of redundancy between layers for a Rayleigh block-fading channel. This paper is organized as follows. Section II introduces the definitions and notation used throughout the paper. Section III expresses the tradeoff of channel-code rate  $R$  and erasure-code depth  $D$  as an explicit optimization problem with a clear solution path. Section IV discusses several special cases of interest. Each subsection of Section IV reduces one of these special cases to a form that may be solved explicitly using standard optimization methods. Section V gives numerical results, and Section VI describes the limiting behavior of the optimal coding strategy. Section VII delivers the conclusions.

## II. DEFINITIONS AND PRELIMINARIES

This paper considers a communication model where a single transmitter attempts to deliver a message consisting of  $m$  packets to a set of  $N$  receivers that cooperate at the packet level to recover a message. In other words, the system we consider has  $N$ -fold selection diversity - an equivalent model is one with  $N$  transmitters and one receiver that selects the best signal for decoding.

In this model, a single transmitter uses a packet-level erasure code, such as a Raptor or Reed-Solomon code, to encode the original  $m$  packets (each with information content  $k$  nats/packet) into  $Dm$  erasure-coded packets. The quantity  $D$  is defined to be the erasure code *depth* (or the inverse of the

This research was supported by Rockwell Collins through contract #4502769987.

erasure code rate). The value of  $D$  need not be an integer, but  $Dm$  will always be an integer in practice. We also have  $D \geq 1$ . Each of these  $Dm$  erasure-coded packets is then encoded for transmission over a wireless channel using a physical-layer channel code of rate  $R$  [nats/channel-use]. After the entire encoding process,  $T = Dmk/R$  uses of the channel are required to send the encoded message.

*Definition 1:* We define a transmission strategy, or selection of  $D$  and  $R$ , to be optimal if it ensures reliable communication over the widest operating region (i.e. it ensures communication at the lowest possible average SNR).

In order to study the relationship between  $D$  and  $R$ , we find it convenient in our analysis to introduce the quantity  $T$ , which is defined to be the the number of channel uses during the elapsed transmission time. We call  $T$  the *transmission window duration*. For a particular value of  $T$ , we have a linear relation between  $D$  and  $R$ , given by:

$$Dmk = RT. \quad (1)$$

It is clear from (1) that there exists a tradeoff between the amount of erasure coding and the amount of channel coding that can be applied to a message for each particular value of  $T$ .

In general, one can expect the optimal allocation of erasure coding and physical-layer coding to change as the elapsed transmission time increases (i.e., as the instantaneous code rate decreases) in a rateless scheme. We study the evolution of the optimal transmission strategy by optimizing the allocation across a range of  $T$ .

In our chosen model, the  $N$  receivers receive the  $Dm$  packets through  $N$  independent channels. Each channel is a block-fading Rayleigh channel with additive white Gaussian noise. We characterize the channel by three parameters: the Rayleigh parameter  $\alpha$ , the noise power  $\sigma^2$ , and the number of block fades  $F$ .  $F$  is the number of *independent* fades that each packet experiences. For example, if  $F = 1$ , a packet is received at a single SNR. If  $F = 2$ , one half of the packet is received at one SNR and the other half at another (independent) SNR.

The distribution of the SNR is determined by the parameters  $\alpha$  and  $\sigma^2$ . For the Rayleigh fading cases treated in this paper, the SNR follows an exponential distribution with parameter  $\lambda = \sigma^2/2\alpha^2$ . Specifically,  $\Pr(\text{SNR} < x) = 1 - \exp(-\lambda x)$ . Note that the average SNR is  $1/\lambda$ .

Each of the  $N$  receivers independently attempts to decode each of the  $Dm$  channel-coded packets and determines whether the decoding was successful based on an indicator mechanism such as a cyclic redundancy check. If the channel-coded packet was successfully decoded (i.e. if the CRC passes), we say that this packet was successfully received.

The receivers cooperate by sharing their successfully received packets with one another. As a group, the receivers are successful in recovering the initial message if they have successfully received a sufficiently large subset of the original  $Dm$  erasure-encoded packets. We quantify the *sufficiently large subset* by saying that the original message can be successfully

recovered if  $\hat{m} = (1+\delta)m$  of the original  $Dm$  erasure-encoded packets were successfully received by the collection of  $N$  receivers. We say that  $\delta$  is the *overhead* of the erasure code. For Reed-Solomon codes this overhead is zero. For Raptor codes  $\delta \approx 0.038$  for  $m = 65536$  (see [1]). The overhead for Raptor codes increases as the message size,  $m$ , decreases. Note that it suffices to consider only  $\hat{m}$  in our analysis because, for a known  $m$  and  $\delta$  (which are design parameters),  $\hat{m}$  can be computed explicitly for a specific erasure code.

Furthermore, we characterize our channel code by the parameter  $\epsilon$  and say that a packet is successfully received at receiver  $i$  if  $(1 + \epsilon)R < \mathcal{C}_i$ , where  $\mathcal{C}_i$  is the capacity of the channel from the transmitter to receiver  $i$ . Here we make the implicit assumption that the physical-layer codeword length,  $kR^{-1}$ , is sufficiently large so that reliable communication is possible within the given distance,  $\epsilon$ , relative to channel capacity.

*Remark 1:* This paper does not discuss the physical-layer code design since we are interested in the optimal transmission strategy, not the actual codes that can be used to implement it. An example of a turbo code that provides near-optimal performance at many rates to support rateless channel coding is presented in [6].

*Remark 2:* For convenience, we use the natural logarithm throughout this paper, and therefore all information quantities are in terms of *nats* rather than bits.

### III. OPTIMIZING THE TRADEOFF OF $R$ AND $D$

For given parameters  $\alpha$ ,  $T$ ,  $\hat{m}$ , and  $N$ , when choosing the optimal code rate  $R$  (and therefore  $D$  following the relation in (1)) we would like to minimize the SNR required for reliable transmission of the message. This gives the widest possible operating region. Therefore, we would like to maximize  $\sigma^2$ , or equivalently maximize  $\lambda$ , subject to a constraint on the probability that the message is not recovered successfully by the receivers.

The probability that the message is not recovered is the probability that the collection of receivers does not receive the required  $\hat{m}$  packets. This probability, denoted  $\kappa_1$ , is a design parameter and will vary depending on the reliability required for a particular application. Noting that  $RTk^{-1}$  is equal to the number of transmitted packets, we can express this constraint as:

$$\sum_{i=0}^{\hat{m}-1} \binom{RTk^{-1}}{i} (1 - p_e^N)^i (p_e^N)^{RTk^{-1}-i} \leq \kappa_1. \quad (2)$$

In the above constraint,  $p_e$  is the probability of erasure, or the probability that a single node does not successfully receive a specified packet. Specifically:

$$p_e = \Pr \left( cFR > \sum_{i=1}^F \log(1 + \gamma_i) \right), \quad (3)$$

where  $c = 2(1 + \epsilon)$  is a constant introduced for notational convenience, and  $\gamma_i$  is the SNR experienced during fade  $i$ . The block fades are orthogonal in time, so the capacity of the block

fading channel is simply the average of the capacities of each block [4]:

$$C = \frac{1}{F} \sum_{i=1}^F \frac{1}{2} \log(1 + \gamma_i).$$

Constraint (2) is difficult to manipulate directly, but fortunately we can use a Gaussian approximation if  $\hat{m}$  is relatively large. Define  $P_i$  to be a random variable indicating the success or failure of transmission of the  $i^{\text{th}}$  packet to a specified receiver node. Let  $P_i = 1$  if packet  $i$  is received by the specified node in the network and  $P_i = 0$  otherwise. Then, the number of packets successfully received at the specified node is  $P = P_1 + P_2 + \dots + P_{RTk-1}$ . We now invoke the Central Limit Theorem and can approximate  $P$  by a Gaussian random variable with  $\text{Mean}(P) = RTk^{-1}(1 - p_e^N)$  and  $\text{Var}(P) = RTk^{-1}p_e^N(1 - p_e^N)$ . This allows Constraint (2) to be rewritten as:

$$(\hat{m} - 1) - RTk^{-1}(1 - p_e^N) + \frac{\kappa_2 \sqrt{RTk^{-1}p_e^N(1 - p_e^N)}}{\kappa_2} \leq 0, \quad (4)$$

where  $\kappa_2 = -\Phi^{-1}(\kappa_1)$ , and  $\Phi(x)$  is the CDF of a standard normal random variable.

Furthermore, if we introduce the change of variables  $X = \sqrt{RTk^{-1}(1 - p_e^N)}$  and  $Y = \sqrt{p_e^N}$ , then Constraint (4) reduces to

$$Y \leq \frac{1}{\kappa_2} \left( X - \frac{\hat{m} - 1}{X} \right), \quad (5)$$

and since  $(1 - p_e^N) + p_e^N = 1$  we also have

$$1 = \frac{k}{RT} X^2 + Y^2. \quad (6)$$

Because  $p_e$  is an increasing function of  $\lambda$  when  $R$  is constant,  $Y$  is a monotonically increasing function of  $\lambda$  for a fixed  $R$ . Therefore, maximizing  $Y$  for a fixed  $R$  ensures optimality. We can see that the optimum  $Y$  lies at the intersection of the ellipse described by (6) and the curve defined where (5) achieves equality. There are four such intersections, but we are only interested in the one in the non-negative quadrant due to the implicit constraint that  $0 \leq p_e^N \leq 1$ . The optimum point is shown in Figure 1. Note also the implicit constraint,  $R \geq k\hat{m}/T$ , which is a trivial lower bound on  $R$ . In words, it says that the transmission of the encoded message must fit in the transmission window of length  $T$ .

Finally, we introduce another change of variables that reduces the fourth order equation obtained by combining (5) and (6) to produce a quadratic equation that preserves the solution that we are interested in. If we let  $Z = (1 - Y^2)$ , we obtain the following equation, where  $\kappa_3 = \kappa_2^2$ :

$$Z^2 RTk^{-1}(RTk^{-1} + \kappa_3) - ZRTk^{-1}(2(\hat{m} - 1) + \kappa_3) + (\hat{m} - 1)^2 = 0. \quad (7)$$

We can solve (7) for  $Z$  (we are interested in the larger root) which in turn gives  $p_e$  as a function of  $R$ . We denote

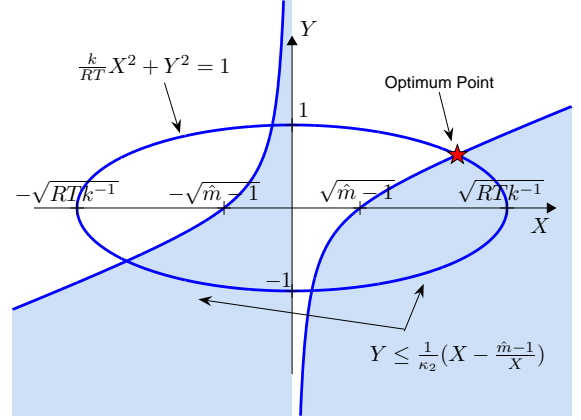


Fig. 1. Geometric interpretation of Constraint (2) using the change of variables  $X = \sqrt{RTk^{-1}(1 - p_e^N)}$  and  $Y = \sqrt{p_e^N}$ . The optimum point is used to compute  $p_e^*(R)$ .

this solution  $p_e^*(R)$ . This effectively combines Constraints (2) and (3) into a single constraint which yields the following optimization problem:

$$\begin{aligned} & \text{maximize} \quad \lambda \\ & \text{subject to:} \quad p_e^*(R) = \Pr \left( cFR > \sum_{i=1}^F \log(1 + \gamma_i) \right) \end{aligned} \quad (8)$$

$$R \geq \frac{k\hat{m}}{T}. \quad (9)$$

In general, no closed form expression exists for the probability distribution that is required to evaluate (8) accurately. We therefore resort to analyzing several special cases for which we can either enforce Constraint (8) by computing the distribution exactly or by using an approximation.

The analysis of this section may be understood as follows: The choice of channel-coding rate  $R$  affects two things. First, it determines the actual erasure probability as a function of  $\lambda$  through (3). Second, it controls the remaining redundancy available for erasure coding through (1). The optimization seeks the value of  $R$  that balances these two effects so that reliable communication occurs at the highest possible  $\lambda$ . That is, it balances the redundancy  $R$  leaves available for erasure correction with the actual probability of erasure provided by  $R$  faced with  $\lambda$ .

#### IV. EXPLICIT SOLUTION OF THREE CASES

We now apply the results of the previous section to three special cases.

##### A. Case 1: Single Fade per Packet

In this case,  $F = 1$ , so Constraint (8) can be rewritten as:

$$\begin{aligned} p_e^*(R) &= \Pr(e^{cR} - 1 > \gamma) \\ &= 1 - e^{-\lambda(e^{cR} - 1)} \end{aligned} \quad (10)$$

$$\Rightarrow \lambda = -\frac{\log(1 - p_e^*(R))}{e^{cR} - 1}. \quad (11)$$

Where (10) and (11) follow because  $\gamma$  is an exponential random variable with parameter  $\lambda$ . In this case, the optimization problem is now to maximize (11) subject to the inequality constraint (9).

Differentiating  $\lambda$  with respect to  $R$  allows us to solve this case by standard methods.

### B. Case 2: Many Fades per Packet

If there are many fades per packet (i.e.  $F \gg 1$ ), we can use a Gaussian approximation in place of Constraint (8). Specifically, (8) becomes:

$$\sqrt{F} \frac{cR - \mu(\lambda)}{\sqrt{\text{Var}(\lambda)}} = \Phi^{-1}(p_e^*(R)). \quad (12)$$

Where  $\mu(\lambda) = \text{Mean}(\log(1 + \gamma))$ ,  $\text{Var}(\lambda) = \text{Var}(\log(1 + \gamma))$ , and  $\Phi(x)$  is the CDF for a standard normal random variable.

If we introduce the auxiliary equations:

$$\text{Ei}(\lambda) = \int_{\lambda}^{\infty} \frac{1}{t} e^{-t} dt \quad \text{and} \quad (13)$$

$$\text{LEi}(\lambda) = \int_{\lambda}^{\infty} \frac{\log(t)}{t} e^{-t} dt, \quad (14)$$

it can be shown, after some intermediate calculus, that:

$$\begin{aligned} \mu(\lambda) &= e^{\lambda} \text{Ei}(\lambda) \\ \text{Var}(\lambda) &= 2e^{\lambda} \text{LEi}(\lambda) - 2e^{\lambda} \log(\lambda) \text{Ei}(\lambda) - e^{2\lambda} \text{Ei}^2(\lambda). \end{aligned}$$

Intuitively, the LHS of (12) must be monotonic in  $\lambda$ , because for fixed  $R$ , the probability of an erased frame should be increasing as  $\lambda$  increases. Therefore, for a fixed  $R$ , we can solve (12) for  $\lambda$  numerically by bisection.

The optimization problem can then be solved by standard methods using the derivatives of  $\lambda$  with respect to  $R$  obtained by differentiating (12) implicitly.

*Remark 3:* It should be pointed out that high quality numerical methods exist for computing the exponential integral (13). However, the authors are unaware of any accurate methods for computing the log-exponential integral (14) when  $\lambda$  is large. Therefore, (12) can only be reliably solved for  $\lambda$  when  $\lambda$  is relatively small (i.e. when the average SNR is high).

### C. Case 3: Low Average SNR

When the average SNR is low, we can use the approximation  $\log(1 + \gamma) \simeq \gamma$ . In this case, (8) simplifies to:

$$p_e^*(R) = P_{\Gamma}(cFR|F, \lambda). \quad (15)$$

In (15),  $P_{\Gamma}(x|F, \lambda)$  is the CDF for a Gamma random variable with parameters  $F$  and  $\lambda$ . Explicitly, this function can be written as:

$$P_{\Gamma}(x|F, \lambda) = \frac{\lambda^F}{(F-1)!} \int_0^x t^{F-1} e^{-\lambda t} dt. \quad (16)$$

Using a simple change of variables, we can solve for  $\lambda$  as:

$$\lambda = \frac{P_{\Gamma}^{-1}(p_e^*(R)|F, 1)}{cFR}. \quad (17)$$

Where  $P_{\Gamma}^{-1}$  is the inverse of the Gamma CDF defined by:

$$P_{\Gamma}^{-1}(p|F, 1) = \{x : P_{\Gamma}(x|F, 1) = p\}. \quad (18)$$

We can now solve the optimization problem using the derivatives obtained from (17).

## V. NUMERICAL RESULTS

This section presents and discusses the results obtained by solving instances of the three optimization problems presented in the previous section. The barrier method of optimization produced these results. (See [5] for a description of the barrier method.)

Figure 2 plots typical curves of the minimum average operating SNR, which is equal to  $10 \log_{10}(1/\lambda)$ , as a function of  $D$ . Since these curves are calculated using  $F = 1$ , they do not involve any of the approximations introduced in Cases 2 and 3 of Section IV. The first characteristic to note is that the curves are quasiconvex with a global optimum. This confirms our intuition that there should be an optimal tradeoff between  $R$  and  $D$ . The next characteristic to note is that the optimal minimum average operating SNR, denoted  $\gamma^* = 1/\lambda^*$ , is decreasing as  $N$  increases. Again, this confirms our intuition that a larger collection of cooperative receivers should be able to perform over a wider range of operating SNR's than a smaller set of receivers. The final interesting point is that  $D^*$  is decreasing as  $N$  increases. This suggests that the redundancy required to tolerate packet erasures can be in the form of redundant packets, created by an erasure code, or in the form of diversity created by multiple cooperative receivers.

Figure 2 also shows that the gain achievable by adding a small number of receivers is most significant when  $N$  is small. This is essentially a standard diversity result, however it is interesting to note that using the optimal cross-layer coding scheme effectively gives us the same performance as using more conventional [physical-layer coding only] receivers. When  $N$  conventional receivers cooperate to recover the message, we obtain the performance of an  $N$ -fold packet-level selection diversity scheme because we only use the successfully decoded packets to recover the transmitted message. Cross-layer coding allows us to create diversity through erasure coding which has the effect of adding *virtual* cooperative users to the system. In the case studied in Figure 2, cross-layer coding with one receiver is almost as good as pure physical-layer coding with three-way selection diversity.

In Figure 3 we plot the gain that can be achieved by using the optimal cross-layer coding strategy instead of a pure physical-layer coding strategy. This figure shows two important trends. The first is that the achievable gain is a decreasing function of  $N$ . The second is that the achievable gain is a decreasing function of  $F$ . We note that the only time when the achievable gain is significant, say greater than 1dB, is when both  $F$  and  $N$  are small. This is an attractive result because it suggests that we can achieve near-optimal performance without using any packet-level erasure coding in many cases. We will comment more on this at the end of this section.

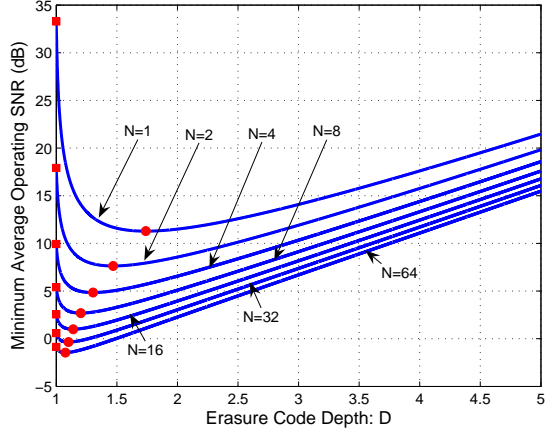


Fig. 2. Typical curves of the minimum operating SNR ( $-10 \log_{10} \lambda$ ) as a function of  $D$ . These curves were generated for an overall code rate of  $1/2$  ( $Tk^{-1} = 200$ ,  $m = \hat{m} = 100$ ),  $\epsilon = 0.05$ ,  $F = 1$ , and  $\kappa_1 = 10^{-3}$ . Solid circles represent the system performance with  $N$ -user optimal cross-layer coding. Solid squares are  $N$ -user selection diversity performance.

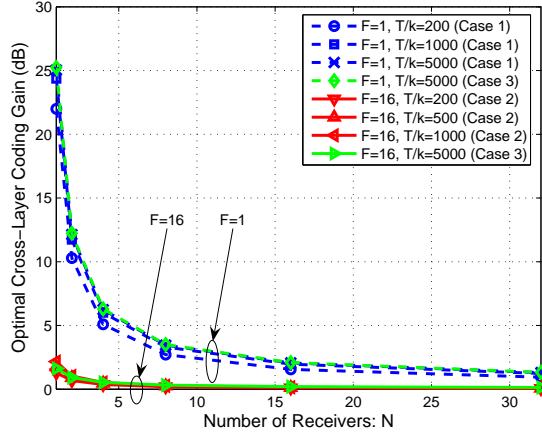


Fig. 3. Achievable gain by using the optimal cross-layer coding strategy versus a pure channel coding ( $D = 1$ ) strategy as a function of  $N$ . All curves were generated using  $m = \hat{m} = 100$ ,  $\epsilon = 0.05$ , and  $\kappa_1 = 10^{-3}$ . Color is used to identify the corresponding optimization case (and solution method) defined in Section IV.

Figure 4 shows the evolution of the optimal transmission strategy as a function of  $N$  and the instantaneous rate of the cross-layer rateless code. This figure also suggests two important trends. The first is that  $D^*$  is a decreasing function of  $N$ . As we said earlier, this implies that the redundancy created by the erasure code becomes less useful as we gain diversity in terms of the independent observations possible through multiple receivers.

Figure 4 also suggests a second important behavior: for a fixed  $N$ ,  $D^*$  has a limit point as the instantaneous code rate approaches zero. In fact, we have calculated these limit points and included them in the figure. The fact that these limit points exist means that  $D^*$  is relatively insensitive to very large changes in  $T$  when the instantaneous code rate is already

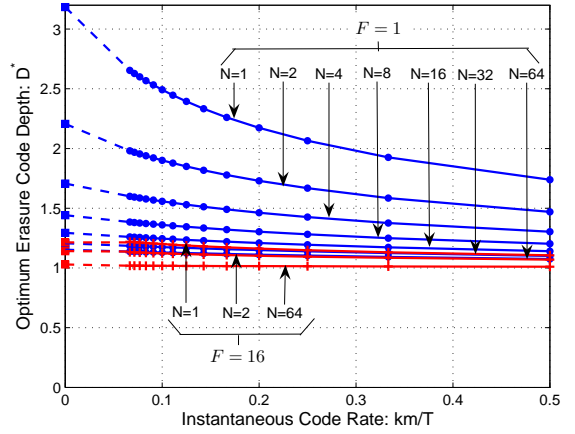


Fig. 4. Evolution of the optimal erasure coding depth,  $D^*$ , as a function of the instantaneous rate of the cross-layer rateless code. Curves for different values of  $N$  are given for  $F = 1$  and  $F = 16$ . All curves were generated using  $m = \hat{m} = 100$ ,  $\epsilon = 0.05$ , and  $\kappa_1 = 10^{-3}$ . The dashed lines in the figure extend the curves to their respective limits as  $km/T \rightarrow 0$  (i.e. as the elapsed transmission time  $T \rightarrow \infty$ ).

sufficiently small (as would happen when the average SNR is low). We discuss this phenomenon in detail in Section VI. We further note that  $D^*$  is insensitive to changes in  $T$  when either  $N$  or  $F$  is large. This implies that, in these three cases, we don't need to know  $T$  accurately in order to closely approximate optimality.

Elaborating on our previous observation,  $D^*$  is nearly one in all cases when  $F$  is large. This implies that the diversity provided by the multiple fades is sufficient so that no additional (or very little) redundancy is required in the form of erasure protection. In other words, there are sufficiently many fades per packet that there is no reason to assume that the observed fading characteristics of one packet would be markedly different from those of another packet. Therefore, all (or almost all) available coding resources should go into the channel code which does not assist with recovery from erased packets, but rather reliably makes the number of erased packets negligible.

In summary, the gains achieved by having packet-level erasure coding are marginal except when both  $N$  and  $F$  are small. If either  $N$  or  $F$  is not small, a nearly optimal strategy would be pure physical-layer coding. This is important from a practical standpoint because a gain on the order of a fraction of a dB may not merit the cost of implementing the packet-level erasure coding mechanism. If  $N$  and  $F$  are small, we may be able to take advantage of the insensitivity of  $D^*$  with respect to  $T$  to implement another simple, yet nearly optimal, coding design. This is discussed in the next section.

*Remark 4:* In a practical application,  $F$  may be a function of  $R$ . In such a case, we would expect  $F$  to be inversely proportional to  $R$ , because as we increase the transmission time required to send a packet by lowering  $R$ , we are likely to experience a greater number of fades. If this indeed were the case, we would expect that  $D^*$  would be nearly one in almost every situation based on the results given in this section.

## VI. INSENSITIVITY TO THE TRANSMISSION WINDOW SIZE

As the results of Section V suggest, the optimal erasure coding depth changes very little after the elapsed transmission window,  $T$ , becomes sufficiently large. In other words, the limit of  $D^*$  as  $km/T \rightarrow 0$  exists. This is a particularly attractive result because, in practical rateless applications,  $T$  is unknown.

Note that as more transmission time elapses and  $T$  becomes large, the instantaneous coding rate,  $km/T$ , becomes very small, allowing reliable communication at a low average SNR. Therefore, we can use the low SNR approximation to show that  $D^*$  is independent of  $T$  when  $T$  is large. Starting with the low SNR approximation (17), we can write:

$$\lambda \propto \frac{P_{\Gamma}^{-1}(p_e^*(R)|F, 1)}{cFRT}. \quad (19)$$

Introducing the auxiliary variable  $\ell = RT$ , we can rewrite (7), (9), and (19) all in terms of  $\ell$ :

$$\begin{aligned} & \text{maximize} && \frac{P_{\Gamma}^{-1}(p_e^*(\ell)|F, 1)}{cF\ell} \\ & \text{subject to:} && \ell \geq k\hat{m}. \end{aligned}$$

Where  $p_e^*(\ell) = (1 - Z)^{1/N}$ , and  $Z$  is the larger root of the quadratic equation:

$$\begin{aligned} & Z^2 \ell k^{-1} (\ell k^{-1} + \kappa_3) - \\ & Z \ell k^{-1} (2(\hat{m} - 1) + \kappa_3) + (\hat{m} - 1)^2 = 0. \end{aligned}$$

Therefore, we can express the optimization problem as a maximization strictly in terms of  $\ell$ , which means that the optimal erasure code depth,  $D = \ell/km$ , is *independent* of  $T$ . In the above analysis, changing  $T$  does not change the optimum  $\ell$ , and therefore does not change the optimum  $D$ . Rather, it changes the required value of  $R$ . This implies that, after the elapsed transmission time becomes large, the ever-increasing redundancy of a rateless approach should be applied to the physical-layer channel coding.

*Remark 5:* Separate from the above analysis,  $D$  is insensitive to  $T$  when  $F$  (or  $N$ ) is large since, regardless of  $T$ , little or no erasure coding is needed producing  $D \approx 1$ .

A practical transmission strategy that would achieve near-optimal performance across all scenarios would have the transmitter select an erasure code depth  $D$  based on some estimate of the transmission window size. The transmitter would then use a rateless *physical-layer* channel code to encode the erasure-coded packets and send additional coded bits (generated by the rateless code) for each packet on a round-robin basis. This method would guarantee that the transmission fills the entire transmission window, that  $R$  would be as low as possible (which lowers  $p_e^*(R)$ ), and that each erasure coded packet is transmitted at nearly the same rate. This strategy would also work if the transmission window is unconstrained. The transmitter would continue sending redundant information until an ACK is received. This would, in effect, maximize the rate at which the message is communicated to the receivers and allow us to drive outage probability to zero.

## VII. CONCLUSION

In this paper, we developed techniques to determine the optimal tradeoff between packet-level erasure coding and physical-layer channel coding for wireless fading channels with any specified number of cooperating receivers. We provided numerical results for 1-64 receivers. The optimal tradeoff point for each situation allows reliable message reception over the widest possible range of SNRs. Three formulations of the general optimization problem allow us to solve for the optimal tradeoff for a wide range of interesting cases. We presented numerical results that demonstrate the relationship between the system parameters and the optimal solution. The largest amount of erasure coding we saw in the cases we studied was  $D = 3.18$  corresponding approximately to a rate-1/3 erasure code. Little erasure coding is needed when diversity is provided by a large number of cooperating independent receivers or a large number of independent fades. Conversely, when little or no diversity is available, packet-level erasure coding offers a significant gain over pure channel coding. In the cases we studied, gains of up to 25 dB were achieved using packet-level erasure coding.

We also found that the optimal erasure coding depth approaches a limit as the elapsed transmission time grows. This allows for the implementation of a practical and nearly optimal cross-layer coding scheme when the transmission window size is unknown or unconstrained as in the case of rateless coding.

Moreover, the amount of packet-level erasure code redundancy is determined by the amount of diversity otherwise available. The more diversity available through multiple independent fading instances or multiple independent receivers, the less erasure coding is needed. Whatever the appropriate amount of erasure-code redundancy is, it does not change significantly as the transmission window increases if diversity is available. Thus, in a practical implementation the ever-increasing redundancy of a rateless code should be applied to the physical-layer channel code rather than the packet-level erasure code. See [6] for a family of capacity-approaching turbo-codes that can provide rateless physical-layer coding.

## ACKNOWLEDGEMENTS

The authors are grateful to Shubha Kadambe, Kent Benson, and Rob Frank of Rockwell Collins for proposing the investigation that led to this paper. The authors also thank Professor Lieven Vandenberghe, whose excellent courses in optimization provided tools that were essential to this work.

## REFERENCES

- [1] A. Shokrollahi, "Raptor Codes", IEEE Trans. Inform. Theory, Vol. 52, No. 6, pp. 2551-2567, June 2006.
- [2] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, "Reliable Multimedia Download Delivery in Cellular Broadcast Networks", IEEE Trans. on Broadcasting, Vol. 53, No. 1, March 2007.
- [3] J. Castura and Y. Mao, "Rateless Coding over Fading Channels", IEEE Comm. Letters, Vol. 10, No. 1, Jan. 2006.
- [4] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991.
- [5] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.
- [6] M. Griot, A. I. Vila Casado, and R. D. Wesel, "On the Design of Arbitrarily Low-Rate Turbo-Codes". Globecom 2007, Washington, D.C., Nov. 2007.