Optimal Allocation of Redundancy Between Packet-Level Erasure Coding and Physical-Layer Channel Coding in Fading Channels

Thomas A. Courtade, *Student Member, IEEE,*
and Richard D. Wesel, *Senior Member, IEEE*

Abstract

For a block-fading channel, this paper optimizes the allocation of redundancy between packet-level erasure coding (which provides additional packets to compensate for packet loss) and physical layer channel coding (which lowers the probability of packet loss). After some manipulation, standard optimization techniques determine the trade-off between the amount of packet-level erasure coding and physical-layer channel coding that minimizes the transmit power required to provide reliable communication. Our results indicate that the optimal combination of packet-level erasure coding and physical-layer coding provides a significant benefit over pure physical-layer coding when no form of channel diversity is present within a packet transmission. However, the benefit of including packet-level erasure coding diminishes as more diversity becomes available within a packet transmission. Even with no diversity within a packet transmission, this paper shows that as the total redundancy becomes large the optimal redundancy for packet-level erasure coding reaches a limit while the optimal redundancy for physical-layer coding continues to increase. Hence providing limitless redundancy at the packet-level with rateless codes such as fountain codes is not the best use of limitless redundancy for block-fading channels.

Index Terms

Cross-layer coding, cross-layer optimization, selection diversity, rateless codes, Rayleigh fading channels.

I. Introduction

WIRELESS channels require physical-layer coding in order to combat the interference and noise faced by every packet and some form of packet-level erasure coding to deal with packets lost to severe fading. Because both forms of coding draw from a common pool of available redundancy, the transmitter should optimize the allocation of redundancy between these two coding mechanisms. This paper provides a convex-optimization approach to determine the optimal allocation for wireless block fading channels. Applying this approach provides insight into how the allocation of redundancy between

The authors are with the Department of Electrical and Engineering, University of California, Los Angeles, CA, 90024 USA e-mail: {tacourta, wesel}@ee.ucla.edu.

This paper was presented in part at the IEEE International Conference on Communications, Dresden, Germany, June 2009.

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

Manuscript revised October 29, 2010.

these two coding mechanisms evolves as more redundancy becomes available and how available diversity affects the value of packet-level erasure coding.

Several recent works [1], [2], and [3] consider a similar system model and jointly optimize the code rates at the physical layer and packet level. In [3], the authors restrict their attention to ARQ for the packet-level code and optimize the physical-layer rate in order to maximize overall throughput. In [1] and [2], the authors consider maximizing throughput (and the equivalent dual problem of minimizing outage probability [1]) instead of optimizing the system to work over the widest possible range of SNR's by minimizing the required transmit power as this paper does. Neither [1] nor [2] considers the effect of diversity on the optimal solution.

Luby et al. noticed the need to consider cross-layer effects in the simulation-based analysis provided in [5] for 3GPP. Vehkapera and Medard also note the importance of cross-layer coding in [7]. In a similar vein to the present work, the authors of [14] determine the total number of expected channel uses required to transmit a packet if retransmission is allowed. However, the packet-level coding in [14] is restricted to simple retransmission. In this regard, this paper generalizes their results. Xiao [15, Ch. 5] considers the problem of jointly balancing linear network coding with packet-length in multi-hop networks to minimize expected delay (between the source and sink nodes), but neither fading nor diversity are considered in [15, Ch. 5].

In recent years, "rateless" packet-level erasure coding schemes such as Raptor coding [4] and fountain codes have gained widespread popularity. These schemes incrementally add redundancy as required; in principle they can add a limitless amount of redundancy.

In a pure erasure channel where packets are received noiselessly unless they are erased, no physical layer coding is necessary and it suffices to use a packet-level erasure code designed for the particular probability of erasure. In scenarios where this erasure probability is not known exactly, rateless codes are frequently employed.

In contrast, on a channel where packets are never erased but always distorted by AWGN with some known SNR that remains constant from packet to packet, physical-layer channel coding is essential and no packet-level erasure coding is necessary. If the physical-layer code rate does not exceed capacity we can expect to decode all packets reliably. On the other hand, if the physical-layer code rate exceeds capacity, all packets will be declared erasures overwhelming the erasure code so that, again, it provides no benefit.

Wireless block fading channels are at neither of the two extremes discussed above. Both packet-level and physical-layer coding mechanisms play important roles, and the allocation optimization presented in this paper becomes important to obtaining the best performance.

It has been noted (see for example [6]) that rateless coding can be used to drive the outage probability to zero in a fading environment. A fundamental question is how to allocate such a stream of incremental redundancy. Should it all be applied to packet-level erasure coding as a "rateless" code typically would do? Perhaps surprisingly, a key result of this paper is that as the amount of available redundancy grows large, additional redundancy should be allocated entirely to the *physical-layer* channel code for block-fading wireless channels. Thus, packet-level "rateless" erasure codes are not the best approach for these channels.

Diversity is another important factor affecting the optimal allocation of redundancy between the packet level and the physical layer. This paper shows that packet-level erasure coding provides essentially the same benefit as a form of diversity. Specifically, our results show that the optimal amount of redundancy that should be applied to packet-level erasure coding decreases as either the selection diversity order or the number of block fades per packet increases. Either form of diversity (or their combination) can become large enough that packet-level erasure coding provides only a small benefit to the communication system.

To summarize, this paper studies the optimal allocation of redundancy between a packet-level erasure code and a physical-layer channel code for a Rayleigh block-fading channel with or without selection diversity. The paper is organized as follows: Section II introduces the communication model. Section III expresses the tradeoff between the physical-layer channel code rate, R_C, and the packet-level erasure code rate, R_E, as an explicit optimization problem with a clear solution path. Section IV discusses several special cases of interest. The three subsections of Section IV each reduce one of these special cases to a form that may be solved explicitly using standard optimization methods. Section V gives numerical results. Section VI shows that in the limit of large redundancy, additional redundancy should be allocated entirely to the physical-layer channel code. Thus, packet-level "rateless" erasure codes are not optimal for fading channels. Section VII delivers the conclusions.

II. Communication Model

This paper considers a communication model where a single transmitter attempts to deliver a message consisting of m packets to a receiver employing N-fold selection diversity. In other words, the receiver has N independent "looks" at each transmitted packet and decodes the best one (See [13, p. 208] for a detailed discussion on selection diversity). One application for this model is a coded message broadcast to N distinct but cooperating receivers through N i.i.d. channels. The receivers recover the original message by sharing at the packet-level [8].

Let the transmission time be T channel uses. The optimal allocation of redundancy varies with T. For a given transmission time T, the transmitter uses a packet-level erasure code to encode the original m information packets (each with information content k nats/packet[1]) into m/R_E erasure-coded packets. The quantity R_E is defined to be the rate of the erasure code. The value of $1/R_E$ need not be an integer, but m/R_E will always be an integer in practice. We also have $R_E \leq 1$. Each of the m/R_E erasure-coded packets is then encoded for transmission over a wireless channel using a physical-layer channel code with rate R_C [nats/channel-use]. For an elapsed transmission time of T channel uses, R_E and R_C must be chosen to satisfy $T = mk/(R_E R_C)$. We can also write the relationship between R_E and R_C as

$$mk/T = R_E R_C \,. \tag{1}$$

Since the LHS of (1) is constant, there exists a tradeoff between the amount of erasure coding, evident from R_E, and the amount of channel coding, evident from R_C, that can be applied to a message for a fixed value of T. Indeed, R_E and R_C are inversely proportional to one another. A particular allocation of redundancy selects a point in this tradeoff between R_E and R_C.

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

The model assumes N-fold selection diversity, i.e., the receiver receives each of the m/R_E packets through N independent channels and chooses the best signal for decoding on a packet-by-packet basis.

Each channel is a block-fading Rayleigh channel with additive white Gaussian noise. That is, $y = ax+n$ where y is the received symbol, x is the transmitted symbol, a is the Rayleigh fading coefficient, and n is additive white Gaussian noise (AWGN) with variance σ^2.

The SNR for given fading coefficient a is $\frac{a^2\mathbb{E}[x^2]}{\sigma^2}$, where $\mathbb{E}[x^2] = P$ is the transmit power. Without loss of generality, we assume that $\mathbb{E}[a^2] = 1$ and $\sigma^2 = 1$. Thus, the SNR at the receiver for the Rayleigh fading cases treated in this paper follows an exponential distribution with parameter $1/P$ (see [13, Ch. 6] or similar for a derivation). Specifically, $\Pr(\text{SNR} < x) = 1 - e^{-x/P}$. Note that the average SNR is P.

We characterize the block fading nature of the channel by the number of fading blocks F. F is the number of *independent* fades that each packet experiences in each channel. For example, if $F = 1$, a packet is received at a single SNR through each channel. If $F = 2$, one half of the packet is received at one SNR and the other half at another (independent) SNR through each channel.

Due to the N-fold selection diversity, the receiver has N independent attempts to decode each of the m/R_E channel-coded packets and determines whether the decoding was successful based on an indicator mechanism such as a cyclic redundancy check. If a particular channel-coded packet was successfully decoded (i.e. if the CRC passes for at least one of the N decoding attempts), we say that this packet was successfully received.

The receiver successfully recovers the original message if it successfully decodes a sufficiently large subset of the original m/R_E erasure-encoded packets. To be more precise, the original message can be successfully recovered if $\hat{m} = (1+\delta)m$ of the original m/R_E erasure-encoded packets were successfully received by the receiver, where δ is the *overhead* of the packet-level erasure code. For Reed-Solomon codes this overhead is zero. For Raptor codes $\delta \approx 0.038$ for $m = 65536$ (see [4]). It suffices to consider only $\hat{m}$ in our analysis because, for a known m and δ (which are design parameters), $\hat{m}$ can be computed for a specific erasure code.

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

Remark 1: This paper does not discuss the details of physical-layer or packet-level code design since we are interested in the optimal allocation strategy, not the actual codes that can be used to implement it. The existence of good packet-level erasure codes and physical-layer channel codes is well established.

III. Optimizing the Tradeoff between R_E and R_C

This section presents the optimization problem at the heart of this paper. Define the optimal redundancy allocation as follows:

Definition 1: An allocation strategy (selection of R_E and R_C) is **optimal** if it minimizes the transmitter power P required to ensure that the probability of the message not being correctly received is below a specified threshold κ_1.

For given parameters T, m, and k, (1) shows that R_C implies R_E and vice versa. Thus, optimization can focus on finding the optimal R_C. Note also that $R_C \geq k\hat{m}/T$. This lower bound on R_C follows from requirement that the transmission of the encoded message must fit into the number of channel uses in the transmission time T.

The probability that the message is not recovered is the probability that the receiver does not successfully decode the required $\hat{m}$ packets. The largest tolerable message error probability, denoted κ_1, is a design parameter and will vary depending on the reliability required for a particular application. Noting that N is the selection diversity order and that the number of transmitted packets (assumed to be an integer) is $R_C T k^{-1}$, we can express this constraint on message error probability as

$$\sum_{i=0}^{\hat{m}-1} \begin{pmatrix} R_C T k^{-1} \\ i \end{pmatrix} (1-p_e^N)^i (p_e^N)^{R_C T k^{-1} - i} \leq \kappa_1 \,. \tag{2}$$

In (2), p_e is the packet-erasure probability, or the probability that a particular packet is not successfully decoded during a particular one of the N attempts. This is the probability that the instantaneous channel capacity C is not sufficient to support reliable transmission of the packet, i.e. $C < R_C(1+\epsilon)$. The instantaneous capacity, which is a random variable, is defined as follows:

$$\mathcal{C} = \frac{1}{F}\sum_{j=1}^{F}\frac{1}{2}\log(1+\gamma_j)\,.$$

Note that p_e is independent of packet index i because in our model the block fading channel is i.i.d. and memoryless from packet to packet. However, p_e does depend on R_C (which we make explicit by introducing the notation $p_e(R_C)$). Specifically:

$$p_e(R_C) = \Pr\left(cFR_C > \sum_{j=1}^{F}\log(1+\gamma_j)\right)\,, \tag{3}$$

where $c = 2(1+\epsilon)$ is a constant introduced for notational convenience, and γ_j is the SNR experienced at the receiver during fade $j \in \{1,\ldots,F\}$. In (3), $p_e(R_C)$ is the outage probability of a Rayleigh block-fading channel [12, p. 105].

Constraint (2) is difficult to manipulate directly, but fortunately we can use a Gaussian approximation which is tight if $\hat{m}$ is relatively large. Define S_j to be a random variable indicating the success or failure of transmission of the j^{th} packet over a particular channel. Let $S_j = 1$ if packet j is received and $S_j = 0$ otherwise.

The number of packets successfully received over a particular channel is $S = S_1 + S_2 + \cdots + S_{R_C T k^{-1}}$. Thus Constraint (2) may be rewritten as $\Pr\left(S \leq (\hat{m}-1)\right) \leq \kappa_1$.

The Central Limit Theorem allows approximation of S by a Gaussian random variable with Mean(S) = $R_C T k^{-1}(1-p_e^N)$ and Var(S) = $R_C T k^{-1} p_e^N (1-p_e^N)$. This Gaussian approximation yields the following optimization problem:

$$\begin{aligned} \textbf{minimize} \quad & P \\ \textbf{subject to:} \quad & p_e(R_C) = \Pr\left(cFR_C > \sum_{i=1}^{F} \log(1+\gamma_i)\right) && (4) \\ & \Phi\left(\frac{(\hat{m}-1) - R_C T k^{-1}(1-p_e^N)}{\sqrt{R_C T k^{-1} p_e^N (1-p_e^N)}}\right) \le \kappa_1 && (5) \\ & R_C \ge \frac{k\hat{m}}{T}\,, && (6) \end{aligned}$$

where $\Phi(x)$ is the CDF of a standard normal random variable.

Note that the minimization takes place over the transmit power P, which parameterizes the distributions of the γ_i's. This optimization minimizes the transmit power required for the system to perform with the desired reliability κ_1. Equivalently, for a system in which P is fixed, this optimization provides the rate allocation that minimizes the SNR at which the desired reliability κ_1 is achieved.

In the above problem formulation, we can rewrite Constraint (5) as

$$(\hat{m}-1) - R_C T k^{-1}(1-p_e^N) + \kappa_2 \sqrt{R_C T k^{-1} p_e^N (1-p_e^N)} \le 0\,, \tag{7}$$

where $\kappa_2 = -\Phi^{-1}(\kappa_1) > 0$. Introducing the change of variables $X = \sqrt{R_C T k^{-1}(1-p_e^N)}$ and $Y = \sqrt{p_e^N}$, Constraint (7) reduces to

$$Y \le \frac{1}{\kappa_2}\left(X - \frac{\hat{m}-1}{X}\right)\,. \tag{8}$$

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

$$1 = \frac{k}{R_C T} X^2 + Y^2\,. \tag{9}$$

Because p_e is a decreasing function of P when R_C is constant and Y is a monotonically increasing function of p_e, Y is a monotonically decreasing function of P for a fixed R_C. Therefore, maximizing Y for a fixed R_C also minimizes P and therefore ensures optimality.

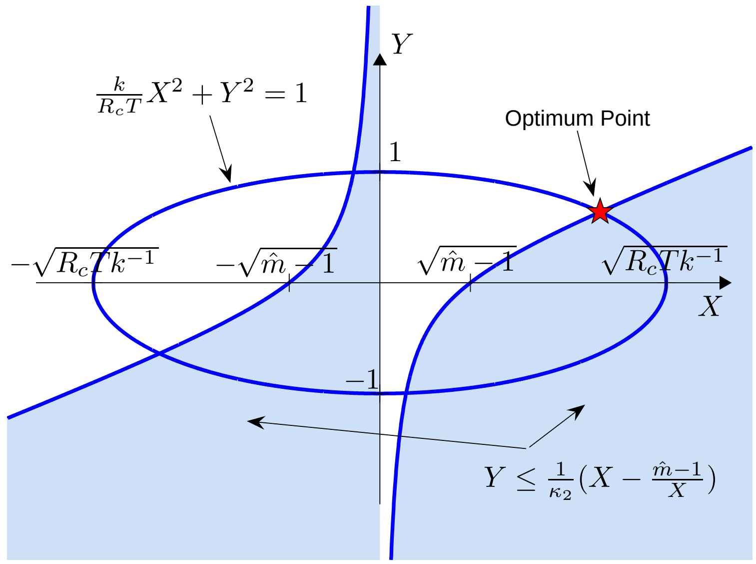

Fig. 1. Geometric interpretation of Constraint (2) using the change of variables $X = \sqrt{R_C T k^{-1}(1 - p_e^N)}$ and $Y = \sqrt{p_e^N}$. The optimum point is used to compute $p_e^*(R_C)$.

Figure 1 shows graphically how Y is maximized. The non-negativity of X and Y restricts our attention to the nonnegative quadrant of Figure 1. Constraint (8) restricts our attention to the shaded region in Figure 1, and Constraint (9) indicates that our solution must lie on the ellipse in Figure 1. Thus the maximum (and therefore optimal) Y is the star in the nonnegative quadrant of Figure 1 at the intersection of the ellipse described by (9) and the curve where (8) achieves equality.

Another change of variables reduces the fourth-order equation obtained by combining (8) and (9) to produce a quadratic equation that preserves the solution of interest as follows: Let $Z = (1 - Y^2)$ to obtain the following equation, where $\kappa_3 = \kappa_2^2$:

$$Z^2 R_C T k^{-1}(R_C T k^{-1} + \kappa_3) - Z R_C T k^{-1}(2(\hat{m} - 1) + \kappa_3) + (\hat{m} - 1)^2 = 0\,. \qquad (10)$$

Solving (10) for Z (we are interested in the larger root) gives p_e as a function of R_C since $Z = 1 - p_e^N(R_C)$. This yields the following optimization problem:

$$\begin{aligned}
\textbf{minimize} \quad & P \\
\textbf{subject to:} \quad & p_e(R_C) = \Pr\left(cFR_C > \sum_{i=1}^{F} \log(1+\gamma_i)\right) && (11) \\
& Z^2 R_C T k^{-1}(R_C T k^{-1} + \kappa_3) - Z R_C T k^{-1}(2(\hat{m}-1) + \kappa_3) + (\hat{m}-1)^2 = 0 && (12) \\
& Z = 1 - p_e^N(R_C) && (13) \\
& R_C \geq \frac{k\hat{m}}{T}\,. && (14)
\end{aligned}$$

Denote as $p_e^*(R_C)$ the value of $p_e(R_C)$ that solves (13) with Z taken to be the larger root of (12). This effectively combines Constraints (11)-(14) into a single constraint which yields the following compact expression for our optimization problem:

$$\begin{aligned}
\textbf{minimize} \quad & P \\
\textbf{subject to:} \quad & p_e^*(R_C) = \Pr\left(cFR_C > \sum_{i=1}^{F} \log(1+\gamma_i)\right) && (15) \\
& R_C \geq \frac{k\hat{m}}{T}\,. && (16)
\end{aligned}$$

In general, no closed form expression exists for the probability distribution that is required to evaluate (15) accurately. We therefore resort to analyzing several special cases for which we can either enforce Constraint (15) by computing the distribution of $\sum_{i=1}^{F} \log(1+\gamma_i)$ exactly or by using an approximation.

The analysis of this section may be summarized as follows: The choice of channel-coding rate R_C affects two things. First, it determines the packet erasure probability as a function of P through (3). Second, it determines the message error probability as a function of the packet erasure probability through (2). The optimization seeks the value of R_C that balances these two effects so that reliable communication (acceptably low message error probability) occurs at the minimum possible P. That is, it balances the redundancy used to lower the packet-erasure probability with the redundancy left available for erasure correction.

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 (15) can be rewritten as:

$$\begin{aligned} p_e^*(R_C) &= \Pr\left(e^{cR_C} - 1 > \gamma\right) \\ &= 1 - e^{-\left(e^{cR_C}-1\right)/P} . \end{aligned} \tag{17}$$

Solving for P yields

$$P = -\frac{e^{cR_C} - 1}{\log\left(1 - p_e^*(R_C)\right)} . \tag{18}$$

Where (17) and (18) follow because γ is an exponential random variable with mean P. In this case, the optimization problem is now to minimize (18) subject to the inequality constraint (16).

Differentiating P with respect to R_C allows us to solve this case by standard optimization techniques which we mention briefly at the beginning of Section V. Several more details are given in the Appendix. See [10] for a comprehensive review of standard optimization techniques.

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 (15). Specifically, (15) becomes:

$$\sqrt{F}\frac{cR_C - \mu(P)}{\sqrt{\mathrm{Var}(P)}} = \Phi^{-1}\left(p_e^*(R_C)\right) . \tag{19}$$

Where $\mu(P) = \text{Mean}(\log(1+\gamma))$ and $\mathrm{Var}(P) = \mathrm{Var}(\log(1+\gamma))$ for γ an exponential random variable with mean P. $\Phi(x)$ is, again, the CDF for a standard normal random variable.

If we introduce the auxiliary equations:

$$\alpha(P) = \int_{P^{-1}}^{\infty} \frac{1}{t} e^{-t} dt \quad \text{and} \tag{20}$$

$$\beta(P) = \int_{P^{-1}}^{\infty} \frac{\log(t)}{t} e^{-t} dt \,, \tag{21}$$

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

$$\begin{aligned} \mu(P) &= e^{1/P}\alpha(P) \\ \text{Var}(P) &= 2e^{1/P}\beta(P) + 2e^{1/P}\log(P)\alpha(P) - e^{2/P}\alpha^2(P) \,. \end{aligned}$$

The LHS of (19) must be monotonic in P, because for fixed R_C, the probability of an erased packet decreases as P increases. Therefore, for a fixed R_C, we can solve (19) for P numerically by bisection. The optimization problem can then be solved by standard methods using the derivatives of P with respect to R_C obtained by differentiating (19) implicitly.

Remark 2: It should be pointed out that high quality numerical methods exist for computing the exponential integral (20). However, the authors are unaware of any accurate methods for computing the log-exponential integral (21) when P is small. Therefore, (19) can only be reliably solved for P when P is relatively large (i.e. when the average SNR is high). The next subsection provides a solution for low average SNR.

C. Case 3: Low Average SNR

When the average SNR is low, we can use the approximation $\log(1+\gamma) \simeq \gamma$ which is tight when γ is small. In this case, (15) simplifies to:

$$p_e^*(R_C) = P_\Gamma\left(cFR_C | F, P^{-1}\right) \,. \tag{22}$$

In (22), $P_\Gamma(x|F, P^{-1})$ is the CDF for a Gamma random variable with parameters F and P^{-1}. Explicitly, this function can be written as:

$$P_\Gamma(x|F, P^{-1}) = \frac{P^{-F}}{(F-1)!} \int_0^x t^{F-1} e^{-t/P} dt \,. \tag{23}$$

Using a change of variables, we can solve for P as:

$$P = \frac{cFR_C}{P_\Gamma^{-1}(p_e^*(R_C)|F, 1)} \,. \tag{24}$$

Where P_Γ^{-1} is the inverse of the Gamma CDF defined by:

$$P_\Gamma^{-1}(y|F, 1) = \{x : P_\Gamma(x|F, 1) = y\} \,. \tag{25}$$

We can now solve the optimization problem using the derivatives obtained from (24).

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. Each of the three optimization problems are solved via standard optimization techniques. A complete description of these techniques is beyond the scope of this paper but can be found in any standard optimization text. The interested reader is directed to [10, Ch. 11] for a description of the barrier method used here.

It should be noted that the results in this section are analytical results obtained by solving the optimization problems described in the previous section and are *not* obtained via Monte Carlo simulation.

Figure 2 plots typical curves of the minimum average operating SNR, which is equal to $10\log_{10}(P)$, as a function of R_E^{-1}. Since these curves are calculated using $F = 1$, they do not involve the approximations introduced in Cases 2 and 3 of Section IV. The first characteristic to note is that the curves are convex with a global optimum. This confirms our intuition that there should be an optimal tradeoff between R_C and R_E. The next characteristic to note is that the optimal minimum average operating SNR is decreasing as

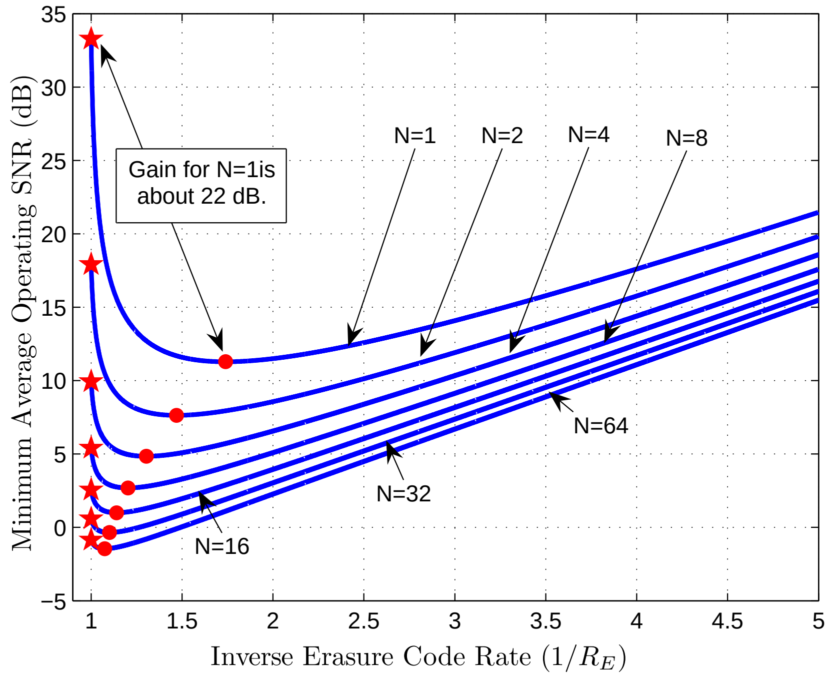

Fig. 2. Typical curves of the minimum operating SNR ($10 \log_{10} P$) as a function of R_E^{-1} for selection diversity ranging from $N = 1$ to $N = 64$. These curves were generated for an overall code rate of $1/2$ ($Tk^{-1} = 200$, $m = \hat{m} = 100$), $\epsilon = 0.05$, $F = 1$, and $\kappa_1 = 10^{-3}$. Solid circles represent the system performance with optimal cross-layer coding. Solid stars show performance of pure physical-layer coding with no packet-level erasure coding.

N increases. Again, this confirms our intuition that selection diversity allows a message to be successfully communicated over a channel with a lower average SNR. The final interesting point is that the optimizing R_E^{-1} is decreasing and approaching 1 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 such as the selection diversity studied here.

Figure 2 also shows that the gain achievable by introducing more diversity 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 (i.e. a combined packet-level and physical-layer coding scheme) effectively gives the same performance as using a physical-layer-coding-only scheme with more selection diversity. If diversity is not present, cross-layer coding allows us to create diversity through erasure coding which has the effect of adding *virtual* independent channels to the system. In the case studied in Figure 2, cross-layer coding with no diversity is almost as good as pure physical-layer coding with 4-fold selection diversity. Note that in this case cross-layer coding provided a 22 dB gain over pure physical-layer coding.

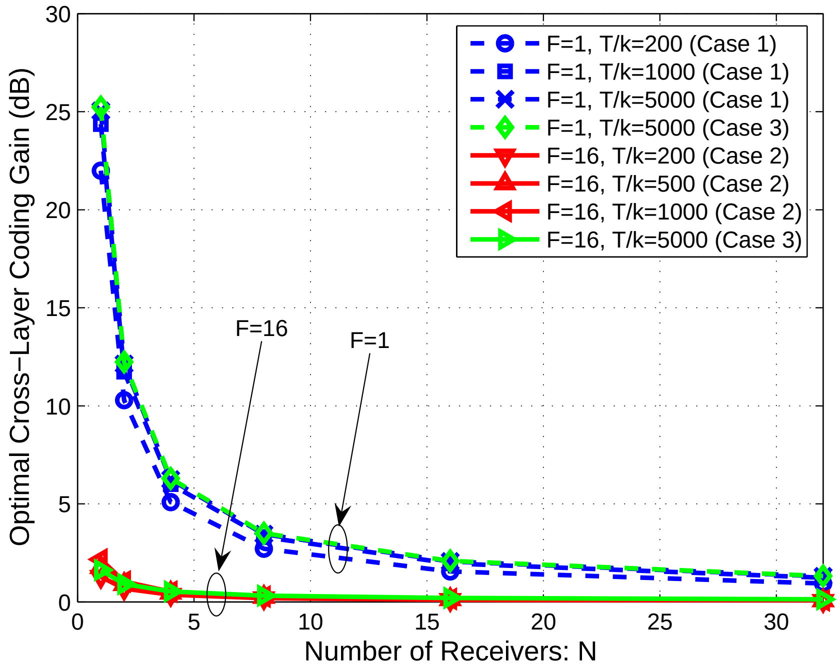

Fig. 3. Achievable gain by using the optimal cross-layer coding strategy versus a pure channel coding ($R_E = 1$) strategy as a function of N. All curves were generated using $m = \hat{m} = 100$, $\epsilon = 0.05$, and $\kappa_1 = 10^{-3}$. The corresponding optimization case (and solution method) defined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

Figure 3 plots the gain achieved by using the optimal cross-layer coding strategy instead of a pure physical-layer coding strategy. For $N = 1$ and $F = 1$, we see that optimal cross-layer coding provides a benefit of over 20 dB as compared to pure physical-layer coding. However, Figure 3 also shows that the achievable gain decreases dramatically as diversity is introduced either by increasing the selection diversity order N or the block fading diversity per packet F. This is an important result because it suggests that near-optimal performance does not require any packet-level erasure coding when sufficient diversity is present. We will comment more on the practical implications of this observation at the end of this section.

For $F = 1$, Figure 4 explores the evolution of the optimal transmission strategy as a function of instantaneous rate km/T of the overall cross-layer code for a range of selection diversity orders from $N = 1$ to $N = 64$. Figure 4 shows that the optimal erasure code rate R_E^* is an increasing function of selection diversity order N. 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 independent channels.

Figure 4 also shows the limit points for R_E^* and R_C^* as the instantaneous overall code rate approaches

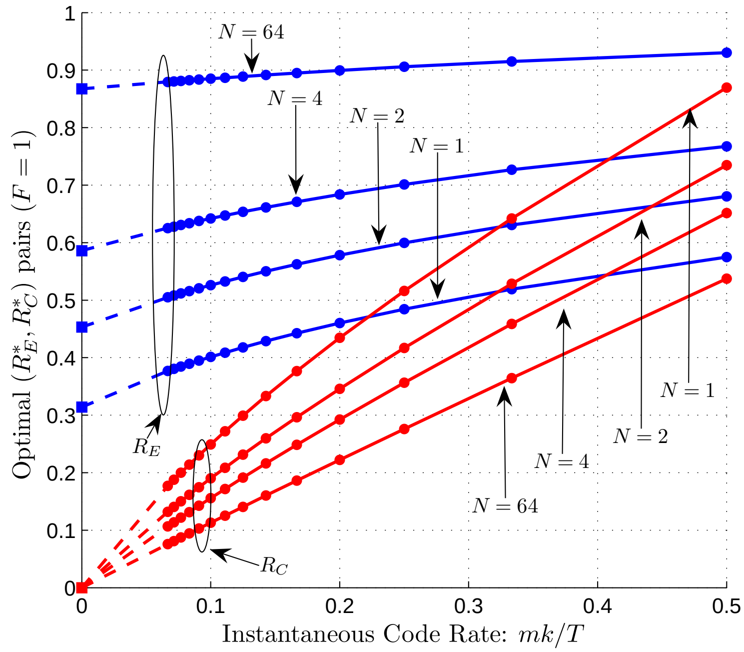

Fig. 4. Evolution of the optimal packet-level erasure coding rate R_E^* and the optimal physical-layer channel coding rate R_C^* as a function of the instantaneous rate of the cross-layer code. Curves for different values of selection diversity order N are given for the case when $F = 1$. All curves 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 $km/T \to 0$ (i.e. as the elapsed transmission time $T \to \infty$).

zero. R_E^* converges to nonzero limit points, but R_C^* converges to *zero*. This indicates that as T grows very large (and instantaneous overall code rate km/T becomes very small) the optimal physical-layer code rate continues to decrease towards zero but the optimal packet-level erasure code rate does not. Section VI presents the derivation of the limit points and further discusses their interpretation. We further note that R_E^* is insensitive to changes in T when N is large.

Although not shown in Figure 4, similar computations reveal that R_E^* is nearly one in all cases when F is large. This implies that the diversity provided by the multiple fades is sufficient, and very little redundancy is required in the form of erasure protection. Note that if there are sufficiently many fades per packet ergodicity makes it unlikely that the observed fading characteristics of one packet would be markedly different from those of another packet. Therefore, severely faded packets are no longer a key obstacle. Thus all (or almost all) available coding resources should go into the physical channel code which does not assist with recovery from severely faded 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 sufficiently large, a nearly optimal strategy is 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 both small, we may be able to take advantage of the insensitivity of R_E^* with respect to T to implement another simple, yet nearly optimal, coding design. This is discussed in the next section.

Remark 3: Our analysis has treated F as a parameter independent of R_C. In a practical application, F is likely to increase linearly with the transmission time of a packet in a time-varying channel. In such a scenario, F is inversely proportional to R_C. This would make R_E^* nearly one in most situations based on the results given in this section.

VI. Asymptotic Results and Rateless Coding

This section shows that the optimal erasure coding rate R_E^* changes very little after the elapsed transmission time, T, becomes sufficiently large. In other words, the limit of R_E^* as $km/T \to 0$ exists and is bounded away from zero. This is a particularly interesting result because it indicates that rateless packet-level erasure codes (e.g. Raptor codes) are not well-suited for fading channel applications.

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 R_E^* is independent of T when T is large.

For fixed T, $P \propto PT$ so that minimizing P is equivalent to minimizing PT. Using the low SNR approximation (24),

$$PT = \frac{cFR_CT}{P_\Gamma^{-1}(p_e^*(R_C)|F,1)}\,. \tag{26}$$

Introducing the auxiliary variable $\ell = R_CT$, we can rewrite (10), (16), and (26) all in terms of ℓ:

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

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

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

Because the optimization problem involves R_C and T only through their product ℓ, changing T does not change the optimum ℓ. Rather, it changes the required value of R_C to maintain the optimal ℓ value. Thus, as $T \to \infty$, $R_C \to 0$. Figure 4 shows this behavior in which R_C always converges to 0 as $T \to \infty$ (or $mk/T \to 0$).

Note from (1), $R_E = km/\ell$. In this large-T, high-SNR region, increasing T does not change ℓ so it does not change R_E. Again, Figure 4 shows this behavior in which R_E converges to a finite positive rate while R_C converges to zero.

This implies that, after the elapsed transmission time becomes large, the ever-increasing redundancy of a rateless coding architecture should be applied to the physical-layer channel coding.

Remark 4: Separate from the above analysis, R_E is insensitive to T when F (or N) is large since, regardless of T, little or no erasure coding is needed producing $R_E \approx 1$.

A practical transmission strategy that would achieve near-optimal performance across all scenarios would have the transmitter select an erasure code rate R_E based on some estimate of the the average channel SNR (and therefore the required transmission time T). The transmitter would then use a rateless *physical-layer* channel code (See e.g. [11]) to encode the erasure-coded packets and send additional physical-layer coded symbols for each packet on a round-robin basis. This method would guarantee that the transmission fills the entire transmission window, that R_C would be as low as possible (which lowers $p_e^*(R_C)$), and that each erasure coded packet is transmitted at nearly the same rate.

VII. Conclusion

In this paper, we developed techniques to determine the optimal tradeoff of redundancy allocation between packet-level erasure coding and physical-layer channel coding for wireless fading channels with a specified degree of selection diversity and a specified number of fades per packet. We provided results for N-fold selection diversity schemes with N ranging from 1 to 64. The optimal (R_C, R_E) point for each situation permits reliable message reception while minimizing transmit power.

Three formulations of the general optimization problem provide the optimal (R_C, R_E) pairs for a wide range of interesting cases. The results demonstrate the relationship between the system parameters and the optimal solution. The largest amount of packet-level erasure coding in the cases studied corresponded approximately to a rate-1/3 erasure code. Little erasure coding is needed when diversity is provided by a high degree of selection diversity or a large number of independent fades per packet. Conversely, when little or no diversity is available, packet-level erasure coding combined with physical-layer channel coding offers a significant gain over pure physical-layer channel coding. In the cases we studied, gains of up to 25 dB were achieved using the optimal combination instead of pure physical-layer channel coding.

We show that the optimal amount of packet-level erasure code redundancy is partially determined by the amount of diversity available in the system. The more diversity available through multiple independent fading instances or multiple independent channels, the less erasure coding is needed.

We also found that the optimal packet-level erasure coding rate approaches a non-zero limit as the elapsed transmission time grows (or equivalently as the amount of available redundancy grows). In contrast, the optimal physical-layer channel coding rate approaches zero. This demonstrates that traditional raptor-style rateless packet-level codes are not the best choice in a wireless fading channel. Instead, it is the physical-layer channel code that needs to be able to continuously lower its rate with incremental redundancy. See [11] for a family of capacity-approaching turbo-codes that can be used to generate incremental redundancy at the physical layer.

Appendix

Implementation of the Optimization Routine when $F = 1$

All three cases of the optimization problem presented in Section IV were solved using the Barrier Method. For details on implementing the Barrier Method, readers are directed to the excellent reference, [10], by Boyd and Vandenberghe.

The barrier method provides primal-dual feasible points at each iteration of the optimization procedure. The dual problem is always convex, and any dual feasible point provides a lower bound on the primal optimal solution. In our standard implementation of the barrier method, the algorithm terminates when the duality gap is less than some specified tolerance. This proves that our numerical results are indeed optimal regardless of the convexity of the optimization problem.

In order to actually do the optimization, one has to derive the first and second derivatives of Z, $p_e^*(R)$, and P with respect to the optimization variable, R_C. This can be quite tedious, so the results are provided in this appendix for convenience. For brevity, we only provide the complete set of derivatives for the case when $F = 1$. The other cases require a similar amount of effort to derive.

In order to evaluate the derivatives given by (27)-(32) at a particular value of R, one needs the values of Z, $p_e^*(R_C)$, and P. Z can be computed by solving (10) for the larger root, then one can compute $p_e^*(R_C) = (1-Z)^{1/N}$, and consequently P from (18). In what follows, let $\lambda = 1/P$.

$$\frac{\partial Z}{\partial R_C} = -\frac{Z}{R_C}\frac{(\hat{m}-1) - Z\ell + (1-Z)\frac{\kappa_3}{2}}{(\hat{m}-1) - Z\ell + (1-2Z)\frac{\kappa_3}{2}} \tag{27}$$

$$\frac{\partial p_e^*(R_C)}{\partial R_C} = -\frac{\partial Z}{\partial R_C}\frac{(1-Z)^{1/N-1}}{N} \tag{28}$$

$$\frac{\partial^2 Z}{\partial R_C^2} = \frac{1}{R_C}\left(\frac{Z^2\ell R_C^{-1} + (\frac{\partial Z}{\partial R_C})^2 R_C(\ell+\kappa_3)}{(\hat{m}-1) - Z\ell + (1-2Z)\frac{\kappa_3}{2}} - 2\frac{\partial Z}{\partial R_C}\frac{(\hat{m}-1) - 2Z\ell + (1-Z)\frac{\kappa_3}{2}}{(\hat{m}-1) - Z\ell + (1-2Z)\frac{\kappa_3}{2}}\right) \tag{29}$$

$$\frac{\partial^2 p_e^*(R_C)}{\partial R_C^2} = \frac{(1-Z)^{1/N-1}}{N}\left(\frac{1-N}{N(1-Z)}\left(\frac{\partial Z}{\partial R_C}\right)^2 - \frac{\partial^2 Z}{\partial R_C^2}\right) \tag{30}$$

$$\frac{\partial \lambda}{\partial R_C} = \frac{1}{1-e^{cR_C}}\left(c\lambda e^{cR_C} - \frac{\partial p_e^*(R_C)}{\partial R_C}e^{\lambda(e^{cR_C}-1)}\right) \tag{31}$$

$$\begin{aligned}\frac{\partial^2 \lambda}{\partial R_C^2} = {} & \frac{ce^{cR_C}}{1-e^{cR_C}}\left(\frac{\partial \lambda}{\partial R_C} + c\lambda\left(1 + \frac{e^{cR_C}}{1-e^{cR_C}}\right)\right) - \\ & \frac{e^{\lambda(e^{cR_C}-1)}}{1-e^{cR_C}}\left(\frac{\partial^2 p_e^*(R_C)}{\partial R_C^2} + \left(\frac{\partial p_e^*(R_C)}{\partial R_C}\right)^2 e^{\lambda(e^{cR_C}-1)} + \frac{\partial p_e^*(R_C)}{\partial R_C}\frac{ce^{cR_C}}{1-e^{cR_C}}\right)\end{aligned} \tag{32}$$

As a final note, the authors found that it was more convenient to minimize the objective function $f_0 = -\log(\lambda) = \log(P)$ rather than minimize P directly. The reason for this is that the optimization routine converged faster when optimizing $\log(\lambda)$. The derivatives of this new objective function are easily obtained as a function of derivatives (27)-(32).

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] C. Berger, S. Zhou, Y. Wen, P. Willett, and K. Pattipati, "Optimizing Joint Erasure- and Error-Correction Coding for Wireless Packet Transmissions". IEEE Transactions on Wireless Communications, Vol. 7, No. 11, November 2008, pp. 4586-4595.

[2] T. Cui, T. Ho, and J. Kliewer, "Achievable Rate and Optimal Physical Layer Rate Allocation in Interference-Free Wireless Networks". ISIT 2009, Seoul, Korea, June 28-July 3 2009.

[3] P. Wu and N. Jindal, "Coding versus ARQ in Fading Channels: How reliable should the PHY be?". arXiv:0904.0226v2 [cs.IT], March 30, 2010.

[4] A. Shokrollahi, "Raptor Codes", IEEE Trans. Inform. Theory, Vol. 52, No. 6, pp. 2551-2567, June 2006.

[5] 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.

[6] J. Castura and Y. Mao, "Rateless Coding over Fading Channels", IEEE Comm. Letters, Vol. 10, No. 1, Jan. 2006.

[7] M. Vehkapera and M. Medard, "A throughput-delay trade-off in packetized systems with erasures", Proc. International Symposium on Information Theory, 2005. ISIT 2005.

[8] T. A. Courtade, R. D. Wesel, "Efficient Universal Recovery in Broadcast Networks". Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing: Monticello, IL, Sept. 29 Oct. 1, 2010.

[9] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991.

[10] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.

[11] 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.

[12] E. Biglieri, "Coding for Wireless Channels". Springer, 2005.

[13] A. Goldsmith, "Wireless Communications". Cambridge University Press, 2005.

[14] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Coding vs. Packet Retransmission over Noisy Channels". 40th Annual Conference on Information Sciences and Systems, 2006.

[15] M. Xiao, "Efficient Coding Techniques for Networks & Channels with Transmission Errors". Ph.D. Thesis, Chalmers University of Technology, Sweden, Nov. 2007.