Coding Versus ARQ in Fading Channels: How reliable should the PHY be?

Peng Wu and Nihar Jindal
Department of Electrical and Computer Engineering
University of Minnesota
Email: {pengwu, nihar}@umn.edu

***Abstract*—This paper studies the tradeoff between *channel coding* and *ARQ* (automatic repeat request) in Rayleigh block-fading channels. A heavily coded system corresponds to a low transmission rate with few ARQ retransmissions, whereas lighter coding corresponds to a higher transmitted rate but more retransmissions. The optimum error probability, where optimum refers to the operating point that maximizes the average successful throughput, is derived and is shown to be a decreasing function of the average signal-to-noise ratio and of the channel diversity order. A general conclusion of the work is that the optimum error probability is quite large (e.g., 10% or larger) for reasonable channel parameters, and that operating at a very small error probability can lead to a significantly reduced throughput.**

I. Introduction

In contemporary wireless systems, simple ARQ (automatic repeat request) is used above the physical layer in order to provide reliable data communication. This leads to a natural tradeoff between the physical layer (PHY) transmitted rate and the probability of packet error/ARQ retransmissions. A large transmitted rate (i.e., the number of information bits per channel symbol) leads to many errors/retransmissions but also means each successfully received packet contains many information bits, whereas a small transmitted rate (i.e. heavy coding) corresponds to few error/retransmissions and packets with fewer information bits.

We investigate this tradeoff in the context of a wireless channel where the transmitter knows the channel statistics but does not have knowledge of the instantaneous channel (e.g., high velocity mobiles).[1] Assuming that the packet error probability is equal to the mutual information outage probability and that simple ARQ (incorrectly decoded packets are re-transmitted, and decoding is based only upon the most recently received packet) is used, we derive the optimum error probability in terms of an accurate approximation to the long-term average successful throughput rate (commonly referred to as goodput). The optimum error probability is shown to decrease with the average SNR and the diversity order. Furthermore, our results show that for reasonable operating conditions (in terms of average SNR and diversity order), the optimum reliability level is quite large (10% error rate is reasonably close to optimal for a wide range of channel parameters) and that choosing an overly reliable PHY can incur a significant goodput penalty.

[1]If instantaneous channel knowledge is available at the transmitter, the tradeoff is relatively trivial because the error probability vs. transmitted rate curve is generally close to a step function and thus the optimal rate lies slightly below this step.

A. Prior Work

In [1] the goodput of an IEEE 802.11 system is numerically optimized when transmission is limited to the finite rate modes available in the 802.11 standard. More recently, in [2] an online algorithm that dynamically adjusts the rate based upon the history of ACK/NACK errors is proposed. Our work considers the limiting, fast fading case where the channel is uncorrelated across packet transmissions, and thus the ACK/NACK history does not provide information and adaptation can only be performed with respect to the channel statistics in an off-line fashion. As we comment on further in Section III-E, our work is also closely connected to recent work that has considered rate/reliability optimization for systems employing outer erasure codes (e.g., rateless fountain codes) [3] [4].

To some extent, this work is "reverse engineering" of actual system designs, in which packet error rates around 10% are typical. In this light, the novelty of the present work is that we are able to derive analytical results that cleanly illustrate (a) the dependence of the optimizing reliability level upon the average SNR and the channel selectivity/diversity , and (b) the potentially significant goodput penalty incurred by choosing an overly reliable PHY operating point.

II. Preliminaries

We consider a Rayleigh block-fading channel which remains constant within each block but varies independently from one block to another. The t-th received symbol y_t during the ℓ-th block is given by:

$$y_t = \sqrt{\mathsf{SNR}}\ h_\ell x_t + z_t, \tag{1}$$

where SNR represents the average received signal-to-noise ratio, x_t is the unit-power transmitted symbol, $z_t \in \mathcal{CN}(0,1)$ is the additive noise (independent across channel uses and blocks), and the channel fading gain in the ℓ-th block $h_\ell \in \mathcal{CN}(0,1)$ is i.i.d. across blocks. Although we focus on Rayleigh fading and single antenna system, our results can be easily extended to other distributions and to MIMO systems.

Each packet (i.e., codeword) spans L fading blocks, and thus L represents the time/frequency selectivity experienced by each transmitted packet. If a strong channel code (with suitably long blocklength) is used, it is well known that the packet error probability is accurately approximated by the mutual information outage probability. We thus consider a system where the packet error probability ε for transmitted rate R (bits/symbol) is precisely equal to the outage probability:

$$\varepsilon = \mathbb{P}\left[\frac{1}{L}\sum_{\ell=1}^{L}\log_2\left(1+\mathsf{SNR}|h_\ell|^2\right) \le R\right], \quad (2)$$

where $h_1, \ldots, h_L$ are each $\mathcal{CN}(0,1)$ and i.i.d.. The error probability ε is clearly increasing in R.

We consider simple ARQ, which is the most basic form of ARQ. Upon reception of a packet, the receiver attempts to decode and determines correctness using a standard error detection method (e.g., CRC); perfect error detection is assumed, i.e., the receiver is able to perfectly determine whether or not decoding is successful. The receiver then feeds back a *one-bit* ACK/NACK to indicate success/failure of the decoding. If an ACK is received, the transmitter moves on to the next packet. If a NACK is received, the same packet is re-transmitted (with no limit on the number of retransmissions) and the receiver attempts to decode based only on this particular received packet; we do not consider hybrid ARQ schemes where the receiver attempts to decode on the basis of all transmissions of a particular packet. Consistent with the fast fading scenario of interest, it is assumed that packet errors are independent across transmissions.

Because ARQ is used, the number of transmission attempts needed for each packet is a random variable, which we denote by X. A simple calculation shows that the long-term average rate at which information is *successfully* received is [5]:

$$\eta = \frac{R}{\mathbb{E}\left[X\right]}, \quad (3)$$

where R is the transmitted rate (bits/symbol). The above quantity, which is also in units of information bits per channel symbol, is referred to throughout the paper as *goodput*. The specific dependence of $\mathbb{E}\left[X\right]$ upon R and other system parameters is specified in subsequent sections.

To summarize, our basic assumptions are as follows:

- The transmitted rate R and the packet error probability ε are related according to the mutual information outage probability in (2).
- Each receiver performs perfect error detection.
- Errors occur independently across packets.
- Simple ARQ is used with no limit on the number of packet retransmissions.[2].

We first analyze a system assuming error-free ACK/NACK feedback, but then generalize to allow for acknowledgement errors.

III. PHY Reliability with Perfect Acknowledgements

In this section we consider the scenario where acknowledgements (i.e., ACK/NACK) are assumed to be received error-free. In this setting, X, the random variable describing the number of transmissions per data packet, is geometrically distributed with parameter $1-\varepsilon$ and thus $\mathbb{E}[X] = \frac{1}{1-\varepsilon}$. Based upon (3), the goodput is therefore

$$\eta = R(1-\varepsilon). \quad (4)$$

Our objective is to maximize the goodput. Although this optimization can be phrased in terms of R or ε, we find it most insightful to deal with ε. Thus, the basic optimization problem to be solved is:

$$\varepsilon^\star(\mathsf{SNR}, L) = \arg\max_{\varepsilon} \; R(1-\varepsilon). \quad (5)$$

where $\varepsilon^\star(\mathsf{SNR}, L)$ is the packet error probability that maximizes goodput for particular values of SNR and L. The quantities R and ε are related according to (2), through which the dependence of the objective function upon SNR and L arises.

Although this optimization is easily solved numerically, it does not seem feasible to find an analytical solution because a closed-form expression for the outage probability exists only for $L = 1$.[3] However, we are able to gain insight by performing the optimization based upon the Gaussian approximation detailed in the following section.

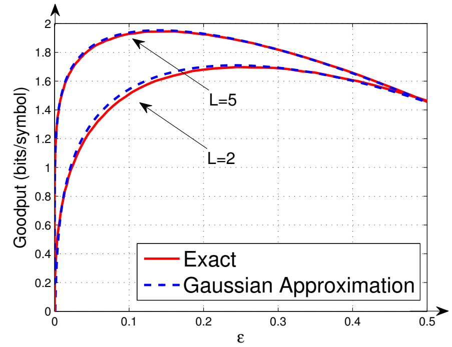


Fig. 1: Goodput (bits/symbol) vs. ε, $L = 2, 5$, $\mathsf{SNR} = 10$ dB

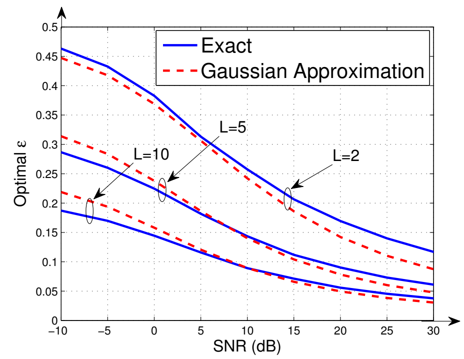
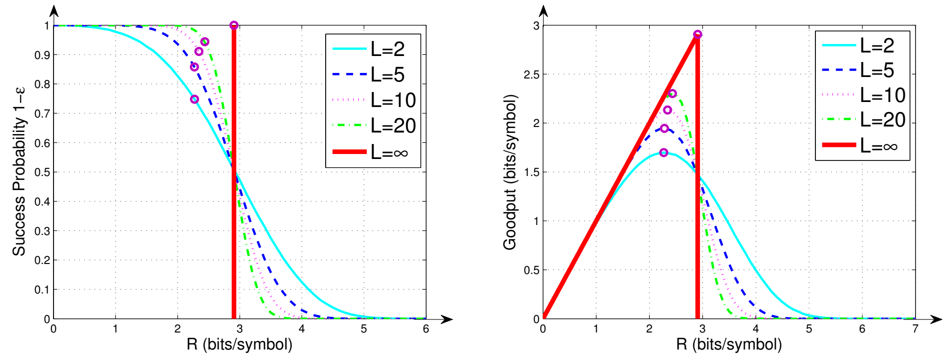


Fig. 2: Optimal ε vs. SNR (dB), $L = 2, 5, 10$

[2] This is quite different from [6], in which authors assume a delay limit and focus on high SNR behavior.

[3] Although not shown here, when $L = 1$ the optimizing $\varepsilon^\star$ can be expressed in closed form in terms of the Lambert W Function.



(a) $1-\varepsilon$ vs. R (bits/symbol)

(b) η (bits/symbol) vs. R (bits/symbol)

Fig. 3: Success probability $1-\varepsilon$ and η (bits/symbol) vs. R (bits/symbol), $\mathsf{SNR} = 10$ dB

A. Gaussian Approximation

The primary difficulty in solving the optimization in (5) stems from the fact that the outage probability in (2) can only be expressed as an L-dimensional integral, except for the special case $L = 1$. To circumvent this problem, we utilize a Gaussian approximation to the outage probability introduced in prior work [7] [8]. More specifically, we approximate the random variable $\frac{1}{L}\sum_{\ell=1}^{L}\log_2\left(1+\mathsf{SNR}|h_\ell|^2\right)$ by a Gaussian random variable with the same mean and variance $\mathcal{N}\left(\mu(\mathsf{SNR}), \frac{\sigma^2(\mathsf{SNR})}{L}\right)$, where $\mu(\mathsf{SNR})$ and $\sigma^2(\mathsf{SNR})$ are the mean[4] and the variance of $\log_2\left(1+\mathsf{SNR}|h|^2\right)$.

Based on this approximation we have

$$\varepsilon \approx Q\left(\frac{\sqrt{L}}{\sigma(\mathsf{SNR})}(\mu(\mathsf{SNR}) - R)\right), \tag{6}$$

where $Q(\cdot)$ is the tail probability of the standard normal. Solving this equation for R and plugging into (4) yields the following approximation for the goodput (denoted η_g):

$$\eta_g = \left(\mu(\mathsf{SNR}) - Q^{-1}(\varepsilon)\frac{\sigma(\mathsf{SNR})}{\sqrt{L}}\right)(1-\varepsilon), \tag{7}$$

where $Q^{-1}(\varepsilon)$ is the inverse of the Q function.

In Fig. 1, the numerically computed goodput η (solid line) and the approximation η_g (dotted line) are plotted vs. ε for $L = 2$ and $L = 5$ at $\mathsf{SNR} = 10$ dB, and the approximation is seen to be quite accurate. The figure also illustrates the basic tradeoff, as it is apparent that making the physical layer too reliable or too unreliable yields poor goodput.

B. Goodput Optimization

We can consider the optimization of η_g, which can be rewritten as

$$\eta_g = \mu(\mathsf{SNR})\left(1 - \kappa\cdot Q^{-1}(\varepsilon)\right)(1-\varepsilon), \tag{8}$$

where the constant $\kappa \in (0,1)$ is the μ-normalized standard deviation of the instantaneous received mutual information:

$$\kappa = \frac{\sigma(\mathsf{SNR})}{\mu(\mathsf{SNR})\sqrt{L}}. \tag{9}$$

[4]Note $\mu(\mathsf{SNR})$ is also the ergodic capacity at the given SNR level.

Notice that κ is decreasing in both SNR and L.

We define $\varepsilon_g^\star$ as the η_g-maximizing error probability.

$$\varepsilon_g^\star(\mathsf{SNR}, L) = \arg\max_{\varepsilon}\ \left(1 - \kappa\cdot Q^{-1}(\varepsilon)\right)(1-\varepsilon), \tag{10}$$

where we have pulled out the constant $\mu(\mathsf{SNR})$ from (8) because it does not affect the maximization. We can immediately see that $\varepsilon_g^\star$ depends on the channel parameters only through the constant κ. Furthermore, it can be shown that η_g is strictly concave in ε (see Appendix A for proof).

The derivative of η_g with respect to ε is:

$$\frac{d\eta_g(\varepsilon)}{d\varepsilon} = \mu\left[\kappa\left(Q^{-1}(\varepsilon) + (\varepsilon - 1)\frac{dQ^{-1}(\varepsilon)}{d\varepsilon}\right) - 1\right]. \tag{11}$$

Due to strict concavity, $\varepsilon_g^\star$ is the unique point where the first derivative is zero and thus is defined by the following fixed point equation:

$$\left(Q^{-1}(\varepsilon_g^\star) - (1-\varepsilon_g^\star)\frac{dQ^{-1}(\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=\varepsilon_g^\star}\right)^{-1} = \kappa. \tag{12}$$

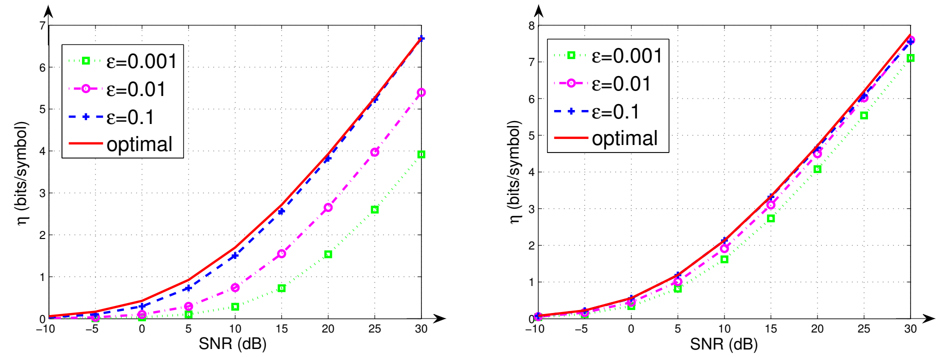
The above equation shows $\varepsilon_g^\star$ is only determined by κ.[5] Furthermore, the concavity of η_g implies that $\frac{d\eta_g(\varepsilon)}{d\varepsilon}$ is decreasing in ε and thus the left hand side of (12) is increasing in ε. As a result, the Gaussian-optimal error probability $\varepsilon_g^\star$ is an increasing function of κ. This in turn implies that $\varepsilon_g^\star$ decreases in L, the channel selectivity, and SNR.

In Fig. 2, the exact optimal $\varepsilon^\star$ and the approximate-optimal $\varepsilon_g^\star$ are plotted vs. SNR (dB) for $L = 2, 5$, and 10. The Gaussian approximation is reasonably accurate, and most importantly correctly captures the property that the optimal error probability is decreasing in L and SNR.

In order to gain an intuitive understanding of this optimization, in Fig. 3 the success probability $1-\varepsilon$ (left) and the goodput $\eta = R(1-\varepsilon)$ (right) are plotted versus the transmitted rate R for $\mathsf{SNR} = 10$ dB. For each L the goodput-maximizing operating point is circled.

First consider the curves for $L = 5$. For R up to approximately 1.5 bits/symbol the success probability is nearly one,

[5]The results extend to different fading distributions and multiple antenna systems by appropriately modifying $\mu(\mathsf{SNR})$ and $\sigma(\mathsf{SNR})$.



(a) $L = 2$ (b) $L = 10$

Fig. 4: η (bits/symbol) vs. SNR (dB), $\varepsilon = 0.001, 0.01, 0.1$, and $\varepsilon^\star$

i.e., $\varepsilon \approx 0$. As a result, the goodput η is approximately equal to R for R up to 1.5. When R is increased beyond 1.5 the success probability begins to decrease non-negligibly but the goodput nonetheless increases with R because the increased transmission rate makes up for the loss in success probability (i.e., for the ARQ retransmissions). However, the goodput peaks at $R = 2.3$ because beyond this point the increase in transmission rate no longer makes up for the increased retransmissions; visually, the optimum rate (for each value of L) corresponds to a point beyond which the success probability begins to drop off very sharply with the transmitted rate.

It is also useful to phrase this explanation in terms of decreasing rate/increasing reliability $(1-\varepsilon)$. If $R = \mu(\mathsf{SNR})$ the error probability is approximately 0.5 and $\eta \approx 0.5\mu$. By decreasing rate/increasing reliability $(1-\varepsilon)$, the goodput is increased because the increase in reliability $(1-\varepsilon)$ makes up for the decreased transmission rate R. However, as the rate is decreased below $R = 2.3$ (i.e., reliability increased above 0.85) the increase in reliability does not make up for the loss in rate; in other words, the PHY has become too reliable if one operates at a rate smaller than the optimum.

In order to understand the effect of increasing the selectivity order L, notice that increasing L leads to a steepening of the success probability-rate curve (towards a step function at the ergodic capacity $\mu(\mathsf{SNR})$) and also increases the largest rate for which the success probability is approximately one. This has the effect of moving the goodput curve closer and closer to the transmitted rate, and also leads to a larger optimum rate and a larger optimum reliability $(1-\varepsilon^\star)$.

To understand why $\varepsilon^\star$ decreases with SNR, based upon the rewritten version of η_g in (8) we see that the governing relationship is between the success probability $1-\varepsilon$ and the normalized, rather than absolute, transmission rate $R/\mu(\mathsf{SNR})$. Therefore, increasing SNR steepens the success probability-normalized rate curve (similar to the effect of increasing L) and thus leads to a smaller value of $\varepsilon^\star$.

Finally, it is important to note that the optimum error probabilities in Fig. 2 are quite large even for large selectivity and at high SNR levels. This follows from the earlier explanation that decreasing the error probability (and thus the rate) beyond a certain point is inefficient because the decrease in ARQ retransmissions does not make up for the loss in transmission rate. An extreme example of making the PHY too reliable would be to choose a transmission rate in the region where the rate-success probability curve is essentially flat, e.g., $R = 1$ for $L \geq 5$ in Fig. 3(a), because at such a point increasing R leads to only a negligible decreases in the success probability.

To underscore the importance of not operating the PHY too reliably, in Fig. 4 goodput is plotted versus SNR (dB) for $L = 2$ and $L = 10$ for the optimum error probability, i.e., $\eta(\varepsilon^\star)$, as well as for $\varepsilon = 0.1$, 0.01, and 0.001. Choosing $\varepsilon = 0.1$ leads to near-optimal performance for both selectivity values. On the other hand, there is a significant penalty if $\varepsilon = 0.01$ or 0.001 when $L = 2$; this penalty is reduced in the highly selective channel ($L = 10$) but is still non-negligible.

C. Scaling with L

Although we earlier saw that $\varepsilon^\star$ decreases with L, it is of interest to quantify how quickly this decrease occurs and also to study the behavior of the optimized goodput as $L \to \infty$. In Appendix B, we use the Gaussian approximation and the fixed-point characterization in (12) to show

$$\varepsilon^\star \approx \frac{\sigma}{\mu\sqrt{L \log L}} \tag{13}$$

$$\mu(\mathsf{SNR}) - \eta(\varepsilon^\star) \approx \sigma \cdot \sqrt{\frac{\log L}{L}}. \tag{14}$$

It is worth noticing that the gap between ergodic capacity and optimized goodput goes to zero on the order of $O\left(L^{-1/2}\log L\right)$ rather than $O\left(L^{-1/2}\right)$, the latter of which is the speed at which the transmitted rate R approaches the ergodic capacity for any fixed ε [8].

D. Scaling with SNR

It is also possible to more precisely quantify the behavior of $\varepsilon^\star$ and $\eta(\varepsilon^\star)$ in the limit as $\mathsf{SNR} \to \infty$. In this case the pre-log factor $(1-\varepsilon)$ dominates the behavior of the goodput, and in order to achieve a pre-log of 1 it is necessary to have

$\varepsilon^\star \to 0$ as $\text{SNR} \to \infty$. Using the same approach as in the previous section, $\varepsilon^\star$ can be approximated as

$$\varepsilon^\star \approx \frac{1.85}{\sqrt{2L \log\log_2(\text{SNR})} \cdot \log_2 \text{SNR}}, \quad (15)$$

where the constant in the numerator comes from the fact that $\lim_{\text{SNR}\to\infty} \sigma(\text{SNR}) \approx 1.85$ (see [8]). The optimizing transmission rate $R^\star$ and the optimized goodput $\eta(\varepsilon^\star)$ both behave as $\log_2 \text{SNR} - o(\log \text{SNR})$, where the sub-logarithmic $o(\log \text{SNR})$ term goes to infinity. This implies that the optimized goodput achieves a pre-log of one but does not have a bounded high-SNR offset, which is consistent with the high-SNR goodput results for MIMO channels in [9].

E. Applicability to Rateless Coding

If rateless coding is used (e.g., fountain coding without ARQ), the important quantity is the number of successfully received packets/information bits. When a very large number of packets are transmitted, the number of successfully received packets converges (by the law of large numbers) to the product of the number of transmitted packets and the per-packet success probability. This is simply a constant multiple of goodput, and thus the optimizing rate/reliability quantified in this work is also optimal for rateless coding when a very large number of packets are transmitted.

IV. PHY Reliability with Imperfect Feedback

We now remove the simplifying assumption of perfect acknowledgement feedback and instead consider the more realistic scenario where the one-bit acknowledgement is received incorrectly (i.e., an ACK is interpreted by the receiver as a NACK, or vice versa) with probability θ.

If an ACK→NACK error occurs, instead of moving on to the next packet the transmitter re-transmits. It is assumed that the receiver can detect that the same packet is being transmitted – this could be accomplished by simply correlating the received packet with the previous packet – in which case the receiver does not try to decode the packet (since it has already been successfully decoded); at the end of the unnecessary retransmission another ACK is sent. This process is continued until the ACK is correctly received and the transmitter finally moves on to the next packet.

On the other hand, when a NACK→ACK error occurs the transmitter erroneously moves on to the next data packet even though the previous one has not been correctly received. In this case we assume that such an error does not cause any harm to the next packet (i.e., the receiver also moves on to the next packet), and that the transmitter eventually retransmits the packet which experienced a NACK→ACK error (e.g., some higher layer mechanism indicates that the NACK→ACK packet has not yet been received and eventually triggers a retransmission). Since the packet is eventually retransmitted, a NACK→ACK error turns out to have no effect.[6]

[6]Since the eventual retransmission may occur much later, a NACK→ACK error increases the delay of that particular packet. However, such delays do not affect goodput.

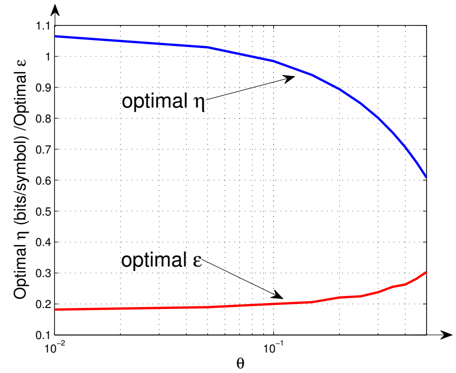


Fig. 5: $\eta(\varepsilon^\star)$ (bits/symbol) and $\varepsilon^\star$ vs. θ, $L = 5$, $\text{SNR} = 5$ dB

Based upon these assumptions, the number of rounds (a round corresponds to a packet transmission followed by an acknowledgement) needed for each packet is

$$X = (\text{rounds to decode}) + (\text{rounds to correctly ACK}) - 1. \quad (16)$$

This relationship holds because the packet must first be successfully decoded, after which the acknowledgement must be successfully received. One round is deducted because the acknowledgement occurs in the same round as the transmission. The number of transmissions required for decoding is again geometric with parameter $1 - \varepsilon$, whereas the number of acknowledgements required for success is geometric with parameter $1 - \theta$. Therefore, the expected number of rounds per packet is

$$\mathbb{E}[X] = \frac{1}{1-\varepsilon} + \frac{1}{1-\theta} - 1 = \frac{1-\theta\varepsilon}{(1-\varepsilon)(1-\theta)}. \quad (17)$$

and, according to (3), the goodput is

$$\eta = R(1-\varepsilon) \cdot \frac{1-\theta}{1-\theta\varepsilon}. \quad (18)$$

Comparing this with (4) we see that acknowledgement errors reduce goodput by a factor of $(1-\theta)/(1-\theta\varepsilon)$, which is less than 1 for positive ε and θ and is decreasing in θ. The goodput maximization now becomes

$$\varepsilon^\star(\text{SNR}, L, \theta) = \arg\max_{\varepsilon} R(1-\varepsilon) \cdot \frac{1-\theta}{1-\theta\varepsilon}, \quad (19)$$

where we now explicitly note the dependence of $\varepsilon^\star$ on θ.

Although we have not yet been able to characterize $\varepsilon^\star$ using the Gaussian approximation, we can show (details are omitted for brevity) that $\varepsilon^\star$ is an increasing function of θ (for fixed values of SNR and L). In other words, as the feedback channel becomes less reliable, the PHY should also be made less reliable. This becomes clear if we re-examine the expression for $\mathbb{E}[X]$ in (17): if $\varepsilon << \theta$, then $\mathbb{E}[X]$ is dominated by the $(1-\theta)^{-1}$ term and choosing a larger ε leads to a larger transmission rate R while barely increasing $\mathbb{E}[X]$.

To verify this point, in Fig. 5 $\varepsilon^\star$ and the optimized throughput $\eta(\varepsilon^\star)$ (bits/symbol) are plotted versus the probability of feedback error θ for $L = 5$ and $\text{SNR} = 5$ dB. When $\theta \leq 0.1$

it has little effect on $\varepsilon^\star$ and a relatively minor effect on $\eta(\varepsilon^\star)$; when $\theta > 0.1$, however, the effect on both quantities is significant.

V. Conclusion

In this paper we have conducted a detailed study of the optimum physical layer reliability when simple ARQ is used to retransmit incorrectly decoded packets. When the role of ARQ is not considered, it seems reasonable to make the physical layer very reliable by decreasing the transmission rate. However, our findings show that when a cross-layer perspective is taken and ARQ is accounted for, it is optimal to use a rather unreliable physical layer (e.g., a packet error probability of 10% for a wide range of channel parameters). Furthermore, we find that making the physical layer too reliable can actually lead to a significant throughput penalty. The role of diversity is also examined, and the optimum reliability level is shown to increase with the diversity order. Indeed, this leads to the general message that the physical layer should not be made very reliable when doing so is rather difficult, e.g., when diversity is lacking.

Appendix A
Proof of Concavity of η_g

For any invertible function $f(\cdot)$, the following holds:

$$(f^{-1})'(a) = \frac{1}{f'(f^{-1}(a))}. \tag{20}$$

By combining this with $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, we get

$$(Q^{-1})'(\varepsilon) = -\sqrt{2\pi} e^{\frac{(Q^{-1}(\varepsilon))^2}{2}}, \tag{21}$$

which is strictly negative. For the second derivative of $\eta_g(\varepsilon)$,

$$\begin{aligned} &\frac{d^2\eta_g(\varepsilon)}{d\varepsilon^2} \\ =\ & \kappa\mu \left(2(Q^{-1})'(\varepsilon) + (\varepsilon - 1)\left(-\sqrt{2\pi} e^{\frac{(Q^{-1}(\varepsilon))^2}{2}}\right)'\right) \\ =\ & \kappa\mu (Q^{-1})'(\varepsilon)\left(2 + (1-\varepsilon)\sqrt{2\pi} e^{\frac{(Q^{-1}(\varepsilon))^2}{2}} Q^{-1}(\varepsilon)\right) \end{aligned}$$

Because $\kappa(Q^{-1})'(\varepsilon) < 0$, in order to prove $\frac{d^2\eta_g(\varepsilon)}{d\varepsilon^2} < 0$ we only need to show that the expression inside the bracket of the last line is strictly positive, i.e.,

$$(\varepsilon - 1) e^{\frac{(Q^{-1}(\varepsilon))^2}{2}} Q^{-1}(\varepsilon) < \sqrt{\frac{2}{\pi}} \tag{22}$$

Define $x = Q^{-1}(\varepsilon)$ so that $Q(x) = \varepsilon$. If we substitute ε with $Q(x)$ in the above inequality, then we only need to prove

$$(Q(x) - 1) e^{\frac{x^2}{2}} x < \sqrt{\frac{2}{\pi}} \tag{23}$$

When $x \geq 0$, the left hand side is negative (because $Q(x) \leq 1$) and the inequality holds. When $x < 0$, the left hand side becomes $Q(-x) e^{\frac{x^2}{2}}(-x)$. From [10], $Q(-x) < \frac{1}{\sqrt{2\pi}(-x)} e^{-\frac{x^2}{2}}$, so if $x < 0$,

$$(Q(x) - 1) e^{\frac{x^2}{2}} x < \frac{1}{\sqrt{2\pi}(-x)} e^{-\frac{x^2}{2}} e^{\frac{x^2}{2}} (-x) = \frac{1}{\sqrt{2\pi}} < \sqrt{\frac{2}{\pi}}.$$

As a result, the second derivative of $\eta_g(\varepsilon)$ is strictly smaller than zero and thus η_g is concave in ε.

Appendix B
Proof of (13) and (14)

According to (12) and (21), the following holds:

$$Q^{-1}(\varepsilon_g^\star) + (1 - \varepsilon_g^\star)\sqrt{2\pi} e^{\frac{(Q^{-1}(\varepsilon_g^\star))^2}{2}} = \frac{1}{\kappa}. \tag{24}$$

As $L \to \infty$, $\varepsilon_g^\star \to 0$, so $Q^{-1}(\varepsilon_g^\star) \to \infty$. Then $Q^{-1}(\varepsilon_g^\star)$ in the left hand side of (24) is negligible compared to the exponential term, and $1 - \varepsilon_g^\star$ goes to 1. Next, simple algebra shows

$$\sqrt{2 \log \frac{1}{\kappa\sqrt{2\pi}}} \approx Q^{-1}(\varepsilon_g^\star). \tag{25}$$

If we let $Q^{-1}(\varepsilon_g^\star) = x$, then $Q(x) = \varepsilon_g^\star$. Notice as $x \to \infty$, it can be shown that $Q(x) \to \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}}$. Combine this with (25), we get

$$\varepsilon_g^\star \approx \frac{\kappa}{\sqrt{2 \log \frac{1}{\kappa\sqrt{2\pi}}}} \tag{26}$$

Plug $\kappa = \frac{\sigma}{\mu\sqrt{L}}$ into (26), we obtain an approximation to $\varepsilon^\star$ and complete the proof of (13). Next, we know

$$\begin{aligned} \mu - \eta(\varepsilon^\star) &\approx \mu - \mu\left(1 - \kappa \cdot Q^{-1}(\varepsilon_g^\star)\right)(1 - \varepsilon_g^\star) \\ &= \mu\left(\varepsilon_g^\star + (1 - \varepsilon_g^\star) \cdot \kappa Q^{-1}(\varepsilon_g^\star)\right) \end{aligned} \tag{27}$$

Plug (26) into (27), the proof of (14) is then complete.

References

[1] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptation for IEEE 802.11 a wireless LANs," *IEEE Trans. Mobile Computing*, vol. 1, no. 4, pp. 278–292, 2002.

[2] R. Aggarwal, P. Schniter, and C. Koksal, "Rate adaptation via link-layer feedback for goodput maximization over a time-varying channel," *submitted to IEEE Trans. Wireless Commun.*, 2008.

[3] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Trans. Broadcasting*, vol. 53, no. 1 Part 2, pp. 235–246, 2007.

[4] T. A. Courtade and R. D. Wesel, "A cross-layer perspective on rateless coding for wireless channels," *to appear at Proc. of IEEE Int'l Conf. in Commun. (ICC'09)*.

[5] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1971–1988, Jul. 2001.

[6] H. E. Gamal, G. Caire, and M. E. Damen, "The MIMO ARQ channel: diversity-multiplexing-delay tradeoff," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.

[7] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," *Proc. of IEEE Int'l Conf. in Commun. (ICC'02)*, pp. 406–410, Apr. 2002.

[8] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: a fixed outage probability analysis," *submitted to IEEE Trans. Commun.*, 2008.

[9] N. Prasad and M. Varanasi, "Throughput analysis for MIMO systems in the high SNR regime," *Proc. of IEEE Int'l Symp. on Inform. Theory (ISIT'06)*, pp. 1954–1948, 2006.

[10] N. Kingsbury, "Approximation formula for the Gaussian error integral, Q(x)," http://cnx.org/content/m11067/latest/.