

# On the Error Exponents of ARQ Channels with Deadlines

Praveen Kumar Gopala, Young-Han Nam and Hesham El Gamal\*

August 8, 2006

## Abstract

We consider communication over Automatic Repeat reQuest (ARQ) memoryless channels with deadlines. In particular, an upper bound  $L$  is imposed on the maximum number of ARQ transmission rounds. In this setup, it is shown that incremental redundancy ARQ outperforms Forney's memoryless decoding in terms of the achievable error exponents.

## 1 Introduction

In [1], Burnashev characterized the maximum error exponent achievable over discrete memoryless channels (DMCs) in the presence of perfect output feedback. Interestingly, Forney has shown that even one-bit feedback increases the error exponent significantly [2]. More specifically, Forney proposed a memoryless decoding scheme, based on the erasure decoding principle, which achieves a significantly higher error exponent than that achievable through maximum likelihood (ML) decoding without feedback [3]. In Forney's scheme, the transmitter sends codewords of block length  $N$ . After receiving each block of  $N$  symbols, the receiver uses a reliability-based erasure decoder and feeds back one ACK/NACK bit indicating whether it has accepted/erased the received block, respectively. If the transmitter receives a NACK message, it then re-transmits the same  $N$ -symbol codeword. After each transmission round, the receiver attempts to decode the message using **only** the latest  $N$  received symbols, and discards the symbols received previously. This process is repeated until the receiver decides to accept the latest received block and transmits an ACK message back to the transmitter.

It is intuitive to expect a better performance from schemes that do not allow for discarding the previous observations at the decoder, as compared with memoryless decoding. Our work here is concerned with one variant of such schemes, i.e., Incremental Redundancy Automatic Repeat reQuest (IR-ARQ) [4]. We further impose a deadline constraint in the form of an upper bound  $L$  on the maximum number of ARQ rounds. In the asymptotic case  $L \rightarrow \infty$ , we argue that IR-ARQ achieves the same error exponent as memoryless decoding, denoted by  $E_F(R)$ . On the other hand, for finite values of  $L$ , it is shown that IR-ARQ generally outperforms memoryless decoding, in terms of the achievable error exponents (especially at high rates and/or small values of  $L$ ). For example, we show that  $L = 4$  is enough for IR-ARQ to achieve  $E_F(R)$  for any binary symmetric channel (BSC), whereas the performance of memoryless decoding falls significantly short from this limit.

---

\*The authors are with the ECE Dept., The Ohio State University, Emails: {gopalap,namy,helgamal}@ece.osu.edu.

The rest of this correspondence is organized as follows. In Section 2, we briefly review the memoryless decoding scheme without any delay constraints, and argue that allowing for memory in decoding does not improve the error exponent. The performance of the memoryless decoder and the incremental redundancy scheme, under the deadline constraint, is characterized in Section 3. In Section 4, we consider specific channels (like the BSC, VNC and AWGN channels) and quantify the performance improvement achieved by incremental redundancy transmission. Finally, some concluding remarks are offered in Section 5.

## 2 The ARQ Channel

We first give a brief overview of the memoryless decoding scheme proposed by Forney in [2]. The transmitter sends a codeword  $\mathbf{x}_m$  of length  $N$ , where  $m \in \{1, \dots, M\}$ . Here  $M$  represents the total number of messages at the transmitter, each of which is assumed to be equally likely. The transmitted codeword reaches the receiver after passing through a memoryless channel with transition probability  $p(y|x)$ . We denote the received sequence as  $\mathbf{y}$ . The receiver uses an erasure decoder which decides that the transmitted codeword was  $\mathbf{x}_m$  iff  $\mathbf{y} \in \mathcal{R}_m$ , where

$$\mathcal{R}_m = \left\{ \mathbf{y} : \frac{p(\mathbf{y}|\mathbf{x}_m)}{\sum_{k \neq m} p(\mathbf{y}|\mathbf{x}_k)} \geq e^{NT} \right\}, \quad (1)$$

where  $T \geq 0$  is a controllable threshold parameter. If (1) is not satisfied for any  $m \in \{1, \dots, M\}$ , then the receiver declares an erasure and sends a NACK bit back to the transmitter. On receiving a NACK bit, the transmitter repeats the codeword corresponding to the same information message. We call such a retransmission as an ARQ round. The decoder discards the earlier received sequence and uses only the latest received sequence of  $N$  symbols for decoding (memoryless decoding). It again applies the condition (1) on the newly received sequence and again asks for a retransmission in the case of an erasure. When the decoder does not declare an erasure, the receiver transmits an ACK bit back to the transmitter, and the transmitter starts sending the next message. It is evident that this scheme allows for an infinite number of ARQ rounds. This scheme can also be implemented using only one bit of feedback (per codeword) by asking the receiver to only send back ACK bits, and asking the transmitter to keep repeating continuously until it receives an ACK bit. Since the number of needed ARQ rounds for the transmission of a particular message is a random variable, we define the error exponent of this scheme as follows.

**Definition 1** *The error exponent  $E(R)$  of a variable-length coding scheme is defined as*

$$E(R) = \limsup_{N \rightarrow \infty} - \frac{\log \Pr(\mathbf{E})}{\bar{\tau}}, \quad (2)$$

where  $\Pr(\mathbf{E})$  denotes the average probability of error,  $R$  denotes the average transmission rate, and  $\bar{\tau} = (\ln M/R)$  is the average decoding delay of the scheme, when codewords of block length  $N$  are used in each ARQ transmission round.

The probability of error of the decoder in (1), after each ARQ round, is given by [2]

$$\Pr(\varepsilon) = \sum_m \sum_{k \neq m} \sum_{\mathbf{y} \in \mathcal{R}_k} p(\mathbf{y}, \mathbf{x}_m),$$

and the probability of erasure is given by

$$\Pr(X) = \left( \sum_m \sum_{\mathbf{y} \notin R_m} p(\mathbf{y}, \mathbf{x}_m) \right) - \Pr(\varepsilon) .$$

It is shown in [2] that these probabilities satisfy

$$\Pr(X) \leq e^{-NE_1(R_1, T)} \quad \text{and} \quad \Pr(\varepsilon) \leq e^{-NE_2(R_1, T)} , \quad (3)$$

where  $R_1 = (\ln M/N)$  denotes the rate of the first transmission round,

$$E_2(R_1, T) = E_1(R_1, T) + T, \quad (4)$$

and  $E_1(R_1, T)$  is given at high rates by [2]

$$E_1(R_1, T) = \max_{0 \leq s \leq \rho \leq 1, \mathbf{p}} E_o(s, \rho, \mathbf{p}) - \rho R_1 - sT, \quad (5)$$

$$E_o(s, \rho, \mathbf{p}) = -\log \int \left( \int p(x)p(y|x)^{(1-s)} dx \right) \left( \int p(x)p(y|x)^{(s/\rho)} dx \right)^\rho dy , \quad (6)$$

and at low rates by

$$E_1(R_1, T) = \max_{0 \leq s \leq 1, \rho \geq 1, \mathbf{p}} E_x(s, \rho, \mathbf{p}) - \rho R_1 - sT, \quad (7)$$

$$E_x(s, \rho, \mathbf{p}) = -\rho \log \int \int p(x)p(x_1) \left( \int p(y|x)^{(1-s)} p(y|x_1)^s dy \right)^{(1/\rho)} dx dx_1 , \quad (8)$$

where  $\mathbf{p} = \{p(x), \forall x\}$  denotes the input probability distribution (We note that for discrete memoryless channels, the integrals in (6) and (8) are replaced by summations). The average decoding delay  $\bar{\tau}$  of the memoryless decoding scheme is given by

$$\begin{aligned} \bar{\tau} &= \sum_{k=1}^{\infty} kN \Pr(\text{Transmission stops after } k \text{ ARQ rounds}) \\ &= \sum_{k=1}^{\infty} kN [\Pr(X)]^{(k-1)} [1 - \Pr(X)] = \frac{N}{1 - \Pr(X)} , \end{aligned}$$

which implies that the average effective transmission rate is given by

$$R = \frac{\ln M}{\bar{\tau}} = \left( \frac{\ln M}{N} \right) [1 - \Pr(X)] = R_1 [1 - \Pr(X)] . \quad (9)$$

It is clear from (3) and (9) that  $R \rightarrow R_1$  as  $N \rightarrow \infty$  if  $E_1(R_1, T) > 0$ . The overall average probability of error can be now computed as

$$\Pr(\text{E}) = \sum_{k=1}^{\infty} [\Pr(X)]^{(k-1)} \Pr(\varepsilon) = \Pr(\varepsilon) [1 + o(1)] , \quad (10)$$

where the second equality follows from (3) when  $E_1(R_1, T) > 0$ . It is, therefore, clear that the error exponent achieved by the memoryless decoding scheme is

$$E(R) = \limsup_{N \rightarrow \infty} - \frac{\log(\Pr(\varepsilon)[1 + o(1)])}{\bar{\tau}} \geq E_2(R, T).$$

It is shown in [2] that choosing the threshold  $T$  such that  $E_1(R_1, T) \rightarrow 0$  maximizes the exponent  $E_2(R_1, T)$  while ensuring that  $R \rightarrow R_1$  as  $N \rightarrow \infty$ . This establishes the fact that the memoryless decoding scheme achieves the feedback error exponent  $E_F(R)$  defined as

$$E_F(R) \triangleq \lim_{E_1(R, T) \rightarrow 0} E_2(R, T) = \lim_{E_1(R, T) \rightarrow 0} T. \quad (11)$$

At this point, it is interesting to investigate whether a better error exponent can be achieved by employing more complex receivers which exploit observations from previous ARQ rounds in decoding (instead of discarding such observations as in memoryless decoding). Unfortunately, it is easy to see that this additional complexity does not yield a better exponent in the original setup considered by Forney [2]. The reason is that, as shown in (10), the overall probability of error in this setup is dominated by the probability of error  $\Pr(\varepsilon)$  in the first transmission round. So, while our more complex decoding rule might improve the probability of error after subsequent rounds, this improvement does not translate into a better error exponent. In the following section, however, we show that in scenarios where a strict deadline is imposed on the maximum number of feedback rounds, significant gains in the error exponent can be reaped by properly exploiting the received observations from previous ARQ rounds (along with the appropriate encoding strategy).

### 3 ARQ with a Deadline

In many practical systems, it is customary to impose an upper bound  $L$  on the maximum number of ARQ rounds (in our notation,  $L \geq 2$  since we include the first round of transmission in the count). Such a constraint can be interpreted as a constraint on the maximum allowed decoding delay or a deadline constraint. With this constraint, it is obvious that the decoder can no longer use the rule in (1) during the  $L^{\text{th}}$  ARQ round. Therefore, after the  $L^{\text{th}}$  round, the decoder employs the maximum likelihood (ML) decoding rule to decide on the transmitted codeword. We denote the probability of error of the ML decoder by  $\Pr^{(ML)}(\varepsilon)$ .

#### 3.1 Memoryless Decoding

The following theorem characterizes lower and upper bounds on the error exponent achieved by the memoryless decoding scheme, under the deadline constraint  $L$ .

**Theorem 2** *The error exponent  $E_{MD}(R, L)$  achieved by memoryless decoding, under a deadline constraint  $L$ , satisfies<sup>1</sup> (for  $0 \leq R \leq C$ )*

$$E_r(R) + (L - 1) \left[ \max_{0 \leq s \leq \rho \leq 1, \mathbf{p}} \left( \frac{E_o(s, \rho, \mathbf{p}) - \rho R - s E_r(R)}{1 + s(L - 2)} \right) \right] \leq E_{MD}(R, L) \leq L E_{sp}(R), \quad (12)$$

---

<sup>1</sup>We note that a tighter lower bound may be obtained by using the expurgated exponent  $E_{ex}(R)$  instead of the random coding exponent  $E_r(R)$  at low rates. This observation will be used when generating numerical results.

where  $E_r(R)$  and  $E_{sp}(R)$  denote the random coding and sphere packing exponents of the memoryless channel, and  $E_o(s, \rho, \mathbf{p})$  is as given in (6).

*Proof:* The average decoding delay of memoryless decoding is given by

$$\begin{aligned}
\bar{\tau} &= \left( \sum_{k=1}^{L-1} kN [\Pr(X)]^{(k-1)} [1 - \Pr(X)] \right) + LN[\Pr(X)]^{(L-1)} \\
&= \left( \sum_{k=0}^{L-1} (k+1)N[\Pr(X)]^k \right) - \left( \sum_{k=1}^{L-1} kN[\Pr(X)]^k \right) \\
&= N \left( \sum_{k=0}^{L-1} [\Pr(X)]^k \right) = N[1 + o(1)] , \tag{13}
\end{aligned}$$

where the last equality follows from (3) when  $E_1(R_1, T) > 0$ . Thus the average effective transmission rate is given by

$$R = \frac{\ln M}{\bar{\tau}} = \frac{\ln M}{N[1 + o(1)]} \rightarrow R_1,$$

as  $N \rightarrow \infty$  when  $E_1(R_1, T) > 0$ . The average probability of error is given by

$$\begin{aligned}
\Pr_{MD}(\mathbf{E}) &= \sum_{k=1}^{L-1} [\Pr(X)]^{(k-1)} \Pr(\varepsilon) + [\Pr(X)]^{(L-1)} \Pr^{(ML)}(\varepsilon) \\
&= \Pr(\varepsilon) [1 + o(1)] + [\Pr(X)]^{(L-1)} \Pr^{(ML)}(\varepsilon) \tag{14}
\end{aligned}$$

$$\leq e^{-N[E_1(R_1, T) + T]} [1 + o(1)] + e^{-N[E_r(R_1) + (L-1)E_1(R_1, T)]} , \tag{15}$$

where the inequality follows from (3) and the random coding upper bound on the ML decoding error probability [3]. Letting  $E_1(R_1, T) \rightarrow 0$  and maximizing  $T$  as before, we get the following error exponent

$$E_{MD}(R, L) = \limsup_{N \rightarrow \infty} - \frac{\ln \Pr_{MD}(\mathbf{E})}{\bar{\tau}} \geq \min\{E_F(R), E_r(R)\} = E_r(R),$$

since the feedback exponent  $E_F(R)$  is known to be greater than the random coding exponent  $E_r(R)$ . Thus by setting  $E_1(R_1, T) \rightarrow 0$ , as suggested by intuitive reasoning, we find that memoryless decoding does not give any improvement over ML decoding without feedback. However, we can get better performance by optimizing the expression in (15) w.r.t  $T$  without letting  $E_1(R_1, T) \rightarrow 0$ . From (15), it is clear that the optimal value of the threshold  $T^*$  is the one that yields

$$\begin{aligned}
E_1(R_1, T^*) + T^* &= E_r(R_1) + (L-1)E_1(R_1, T^*) \\
\Rightarrow T^* &= E_r(R_1) + (L-2)E_1(R_1, T^*) . \tag{16}
\end{aligned}$$

Using this optimal value of  $T^*$  in (5) and solving for  $E_1(R_1, T^*)$ , we get

$$E_1(R_1, T^*) = \max_{0 \leq s \leq \rho \leq 1, \mathbf{p}} \left( \frac{E_o(s, \rho, \mathbf{p}) - \rho R_1 - s E_r(R_1)}{1 + s(L-2)} \right) . \tag{17}$$

Since  $E_F(R_1) > E_r(R_1)$ , we have  $E_1(R_1, T^*) > 0$  and hence  $R \rightarrow R_1$  as  $N \rightarrow \infty$ . Thus the error exponent of memoryless decoding is lower bounded by

$$\begin{aligned} E_{MD}(R, L) &\geq E_2(R, T^*) = E_1(R, T^*) + T^* = E_r(R) + (L-1)E_1(R, T^*) \\ &= E_r(R) + (L-1) \left[ \max_{0 \leq s \leq \rho \leq 1, \mathbf{p}} \left( \frac{E_o(s, \rho, \mathbf{p}) - \rho R - sE_r(R)}{1 + s(L-2)} \right) \right]. \end{aligned} \quad (18)$$

Since  $E_1(R, T^*) > 0$ , it is clear that the optimal threshold  $T^*$  satisfies  $0 \leq T^* < E_F(R)$  and thus the lower bound on  $E_{MD}(R, L)$  in (18) is smaller than the feedback exponent  $E_F(R)$ .

We now derive an upper bound on  $E_{MD}(R, L)$  from (14) as follows.

$$\begin{aligned} \Pr_{MD}(\mathbf{E}) &= \Pr(\varepsilon) [1 + o(1)] + [\Pr(X)]^{(L-1)} \Pr^{(ML)}(\varepsilon) \\ &\geq [\Pr(X)]^{(L-1)} \Pr^{(ML)}(\varepsilon) \\ &\geq [\Pr(X)]^{(L-1)} (e^{-NE_{sp}(R_1)}) , \end{aligned} \quad (19)$$

where the last inequality follows from the sphere-packing lower bound on the ML decoding error probability [3]. It is easy to see that the probability of erasure  $\Pr(X)$  of the decoder in (1) decreases when the threshold parameter  $T$  is decreased. Thus the probability of erasure  $\Pr(X)|_{T=0}$  serves as a lower bound on  $\Pr(X)$  for any  $T > 0$ . In [5], upper and lower bounds on the erasure and error probabilities are derived using a theorem of Shannon *et al* [6]. From equations (10) and (11) in [5], we have

$$\frac{1}{4M} \sum_{m=1}^M \exp \left[ \mu_m(s) - s\mu'_m(s) - s\sqrt{2\mu''_m(s)} \right] < \Pr(X) + \Pr(\varepsilon) \leq \frac{1}{M} \sum_{m=1}^M \exp \left[ \mu_m(s) - s\mu'_m(s) \right] ,$$

and

$$\begin{aligned} \frac{1}{4M} \sum_{m=1}^M \exp \left[ \mu_m(s) + (1-s)\mu'_m(s) - (1-s)\sqrt{2\mu''_m(s)} \right] &< \Pr(\varepsilon) \\ &\leq \frac{1}{M} \sum_{m=1}^M \exp \left[ \mu_m(s) + (1-s)\mu'_m(s) \right] , \end{aligned}$$

where

$$\mu_m(s) = \ln \int p(\mathbf{y}|\mathbf{x}_m)^{(1-s)} \left[ \sum_{m_1 \neq m} p(\mathbf{y}|\mathbf{x}_{m_1}) \right]^s d\mathbf{y} .$$

It is clear from equation (8) in [5] that the threshold parameter  $T$  is related to the parameter  $\mu_m(s)$  by  $\mu'_m(s) = -NT$ . Thus the condition  $T = 0$  corresponds to the condition  $\mu'_m(s) = 0$ . Moreover, it is shown in [5] that  $\mu_m(s)$  and  $\mu''_m(s)$  are also proportional to  $N$ . Using this fact and the condition  $\mu'_m(s) = 0$  in the above expressions for the upper and lower bounds on  $\Pr(X)$  and  $\Pr(\varepsilon)$ , we get

$$\frac{1}{4M} \sum_{m=1}^M \exp \left[ \mu_m(s) \left( 1 + o \left( \frac{1}{\sqrt{N}} \right) \right) \right] < \Pr(X) + \Pr(\varepsilon) \leq \frac{1}{M} \sum_{m=1}^M \exp [\mu_m(s)] , \quad (20)$$

and

$$\frac{1}{4M} \sum_{m=1}^M \exp \left[ \mu_m(s) \left( 1 + o \left( \frac{1}{\sqrt{N}} \right) \right) \right] < \Pr(\varepsilon) \leq \frac{1}{M} \sum_{m=1}^M \exp [\mu_m(s)] . \quad (21)$$

It is clear from (20) and (21) that when  $T = 0$ , the exponents of the upper and lower bounds coincide as  $N \rightarrow \infty$ , and more importantly, the exponent of the erasure probability  $\Pr(X)$  is the same as that of the error probability  $\Pr(\varepsilon)$ . These exponents are further equal to the exponent of the ML decoding error probability since  $\Pr(\varepsilon) \leq \Pr^{(ML)}(\varepsilon) \leq \Pr(\varepsilon) + \Pr(X)$ . Using this fact and the sphere-packing lower bound on the ML decoding error probability in (19), we get

$$\Pr_{MD}(\mathbf{E}) \geq e^{-NLE_{sp}(R_1)} \Rightarrow E_{MD}(R, L) \leq LE_{sp}(R) ,$$

since  $R \rightarrow R_1$  as  $N \rightarrow \infty$ . □

From Theorem 2, it is clear that ARQ with memoryless decoding does not achieve Forney's error exponent  $E_F(R)$  when the maximum number of ARQ rounds  $L$  is constrained, at least at high rates for which  $LE_{sp}(R) < E_F(R)$ . As expected, when  $L \rightarrow \infty$ , the lower bound on the error exponent in (12) becomes

$$\lim_{L \rightarrow \infty} E_{MD}(R, L) \geq \max_{0 \leq s \leq \rho \leq 1, \mathbf{p}} \left( \frac{E_o(s, \rho, \mathbf{p}) - \rho R}{s} \right) = E_F(R).$$

### 3.2 Incremental Redundancy ARQ

We now derive a lower bound on the error exponent of incremental redundancy ARQ. In IR-ARQ, the transmitter, upon receiving a NACK message, transmits  $N$  **new** coded symbols (derived from the same message). Since our results hinge on random coding arguments, these new symbols are obtained as i.i.d. realizations from the channel capacity achieving distribution. The decoder does not discard the received observations in the case of an erasure and uses the received sequences of all the ARQ rounds jointly to decode the transmitted message. The following erasure decoding rule is employed by the receiver: After the  $k^{\text{th}}$  ARQ round, the decoder decides on codeword  $\mathbf{x}_m$  iff  $\mathbf{y} \in \mathcal{R}'_m$ , where

$$\mathcal{R}'_m = \left\{ \mathbf{y} : \frac{p(\mathbf{y}|\mathbf{x}_m)}{\sum_{i \neq m} p(\mathbf{y}|\mathbf{x}_i)} \geq e^{kNT_k} \right\} , \quad (22)$$

and  $\mathbf{y}$ ,  $\{\mathbf{x}_i\}$  are vectors of length  $kN$ , which contain the received sequences and transmitted codewords (respectively) corresponding to the  $k$  ARQ rounds. If no codeword satisfies the above condition, then an erasure is declared by the decoder. It is clear that our formulation allows for varying the threshold  $T_k$  as a function of the number of ARQ rounds  $k$ . Using thresholds  $\{T_k\}$  that decrease with the number of ARQ rounds  $k$  makes intuitive sense since the probability of error will be dominated by small values of  $k$  (initial ARQ rounds), and hence, one needs to use higher thresholds for these  $k$  values to reduce the overall probability of error. We let  $E_k$  denote the event that the decoder declares an erasure during *all* the first  $k$  ARQ rounds. We also let  $E_0 = \phi$  (the empty set). The probability of erasure and error of the decoder in the  $k^{\text{th}}$  ARQ round will thus be denoted by  $\Pr_{(k)}(X|E_{(k-1)})$  and  $\Pr_{(k)}(\varepsilon|E_{(k-1)})$ , respectively. Here the subscript  $(k)$  is used to highlight the fact that the decoder uses a received sequence of length  $kN$  for decoding in the  $k^{\text{th}}$  ARQ round. We are now ready to state our main result in this section.

**Theorem 3** The error exponent  $E_{IR}(R, L)$  achieved by IR-ARQ, under a deadline constraint  $L$ , is given by<sup>2</sup>

$$E_{IR}(R, L) \geq \min \{E_F(R), LE_r(R/L)\}, \quad 0 \leq R \leq C. \quad (23)$$

*Proof:* The average decoding delay for IR-ARQ is given by

$$\begin{aligned} \bar{\tau} &= \sum_{k=1}^L kN \Pr(\text{Transmission stops after } k \text{ ARQ rounds}) \\ &= \sum_{k=1}^{L-1} kN \left( \prod_{i=1}^{k-1} \Pr_{(i)}(X|E_{(i-1)}) \right) [1 - \Pr_{(k)}(X|E_{(k-1)})] \\ &\quad + LN \left( \prod_{i=1}^{L-1} \Pr_{(i)}(X|E_{(i-1)}) \right) \\ &= \sum_{k=0}^{L-1} (k+1)N \left( \prod_{i=1}^k \Pr_{(i)}(X|E_{(i-1)}) \right) - \sum_{k=1}^{L-1} kN \left( \prod_{i=1}^k \Pr_{(i)}(X|E_{(i-1)}) \right) \\ &= N \left[ 1 + \sum_{k=1}^{L-1} \left( \prod_{i=1}^k \Pr_{(i)}(X|E_{(i-1)}) \right) \right] \\ &\leq N \left[ 1 + \sum_{k=1}^{L-1} \Pr(X) \right] \leq N [1 + L\Pr(X)]. \end{aligned} \quad (24)$$

Since  $\Pr(X) \leq e^{-NE_1(R_1, T)}$ , it follows that  $\bar{\tau} \rightarrow N$  (and hence the average effective transmission rate  $R \rightarrow R_1$ ) as  $N \rightarrow \infty$  when  $E_1(R_1, T) > 0$ . The average probability of error of IR-ARQ is given by

$$\begin{aligned} \Pr_{IR}(E) &= \sum_{k=1}^L \Pr(\text{error in the } k^{\text{th}} \text{ ARQ round}) \\ &= \sum_{k=1}^{L-1} \Pr_{(k)}(\varepsilon, E_{(k-1)}) + \Pr_{(L)}^{(ML)}(\varepsilon, E_{(L-1)}) \\ &\leq \sum_{k=1}^{L-1} \Pr_{(k)}(\varepsilon) + \Pr_{(L)}^{(ML)}(\varepsilon), \end{aligned}$$

where  $\Pr_{(k)}(\varepsilon)$  refers to the probability of error when the decoder always waits for  $kN$  received symbols before decoding. Following the derivation in [2], it can easily be seen that for the thresholds  $\{T_k\}$  used in the decoding rule (22), we have

$$\Pr_{(k)}(X) \leq e^{-kNE_1(R_1/k, T_k)} \quad \text{and} \quad \Pr_{(k)}(\varepsilon) \leq e^{-kN[E_1(R_1/k, T_k) + T_k]}. \quad (25)$$

---

<sup>2</sup>Replacing the random coding exponent  $E_r(R)$  by the expurgated exponent  $E_{ex}(R)$  may yield a tighter lower bound at low rates.



Using this and the fact that  $\Pr_{(L)}^{(ML)}(\varepsilon) \leq e^{-LN E_r(R_1/L)}$ , the average probability of error of IR-ARQ can be upper bounded by

$$\Pr_{IR}(\mathbf{E}) \leq \sum_{k=1}^{L-1} e^{-kN[E_1(R_1/k, T_k) + T_k]} + e^{-LN E_r(R_1/L)}. \quad (26)$$

Thus the error exponent achieved by IR-ARQ is lower bounded by

$$E_{IR}(R, L) = \limsup_{N \rightarrow \infty} - \frac{\ln \Pr_{IR}(\mathbf{E})}{N} \geq \min \left( L E_r(R/L), \{ k[E_1(R/k, T_k) + T_k] \}_{k=1}^{L-1} \right).$$

Taking  $T_k = (T/k)$ ,  $\forall k \in \{1, \dots, (L-1)\}$ , we get

$$\begin{aligned} E_{IR}(R, L) &\geq \min \left( L E_r(R/L), \{ k E_1(R/k, T/k) + T \}_{k=1}^{L-1} \right) \\ &= \min (L E_r(R/L), E_1(R, T) + T), \end{aligned}$$

where the last equality follows from the fact that  $E_1(R/k, T/k)$  is an increasing function of  $k$ . Letting  $E_1(R, T) \rightarrow 0$  and maximizing  $T$ , we get

$$E_{IR}(R, L) \geq \min (L E_r(R/L), E_F(R)).$$

□

From Theorem 3, it is clear that if the deadline constraint  $L$  is large enough to satisfy

$$L E_r(R/L) \geq E_F(R), \quad (27)$$

then IR-ARQ achieves the feedback exponent  $E_F(R)$  at rate  $R$ . In the following section, we quantify the gains achieved by IR-ARQ, as compared with memoryless decoding, for specific channels.

## 4 Examples

### 4.1 The Binary Symmetric Channel (BSC)

Here, we compare the error exponents achievable by memoryless decoding and IR-ARQ over a BSC with crossover probability  $\epsilon$ . The bounds on the error exponents in (12) and (23) are plotted for a BSC with  $\epsilon = 0.15$  in Figs. 1(a) and 1(b) for  $L = 2$  and  $L = 4$ , respectively. The ML decoding error exponent (corresponding to the case  $L = 1$ ) and the feedback exponent  $E_F(R)$  are also plotted for comparison purposes. From Fig. 1(a), we find that when  $L = 2$ , memoryless decoding achieves an error exponent that is strictly sub-optimal to the feedback exponent  $E_F(R)$  for all  $R \geq 0.006$ . On the other hand, IR-ARQ achieves  $E_F(R)$  for  $0.18 \leq R \leq C$ . Moreover, it performs strictly better than memoryless decoding for all  $R \geq 0.057$ . When  $L = 4$ , from Fig. 1(b), we find that the error exponent for the memoryless decoder is strictly sub-optimal, as compared with  $E_F(R)$ , for  $R \geq 0.141$ , while IR-ARQ achieves  $E_F(R)$  for all rates below capacity. Finally, we note that even when  $L = 100$ , memoryless decoding is still strictly sub-optimal, as compared with IR-ARQ, for all rates  $0.38 \leq R \leq C = 0.39$ .

Now, we elaborate on our observation from Fig. 1(b) that  $L = 4$  is sufficient to achieve  $E_F(R)$  with IR-ARQ when  $\epsilon = 0.15$ . In particular, we wish to investigate the existence of a finite value for  $L$  such that  $E_F(R)$  is achieved by IR-ARQ universally (i.e., for all  $0 \leq \epsilon \leq 0.5$  and all rates below capacity). Towards this end, we derive an upper bound on the *minimum* required deadline constraint  $L_{req}$  for a given BSC( $\epsilon$ ). From (23), it is clear that  $L_{req}$  is upper bounded by the minimum value of  $L$  required to satisfy  $LE_r(R/L) \geq E_F(R)$  for all  $0 \leq R \leq C$ . We first prove the following result.

**Lemma 4** *A sufficient condition for ensuring that  $LE_r(R/L) \geq E_F(R)$  for all rates  $0 \leq R \leq C$  for a BSC is given by  $LE_r(0) \geq E_F(0)$ .*

*Proof:* It has been shown in [3] that both the random coding exponent  $E_r(R)$  and the feedback exponent  $E_F(R)$  are decreasing functions of  $R$ . Since

$$LE_r(R/L) = \max_{0 \leq \rho \leq 1} \{LE_o(\rho) - \rho R\} , \quad (28)$$

its slope at a given rate  $R$  is given by (following the steps in equations (5.6.28–5.6.33) in [3])

$$\frac{\partial (LE_r(R/L))}{\partial R} = -\rho^*(R) \geq -1 ,$$

where  $\rho^*(R)$  is the value of  $\rho$  that maximizes the RHS of (28) for rate  $R$ . For a BSC, it is shown in [2] that the feedback exponent can be expressed as

$$E_F(R) = (C - R) + \max_{\rho \geq 0} \{E_o(\rho) - \rho R\} . \quad (29)$$

Hence the slope of  $E_F(R)$  at a given rate  $R$  is given by

$$\frac{\partial E_F(R)}{\partial R} = -\left(1 + \rho'(R)\right) \leq -1 ,$$

where  $\rho'(R)$  is the value of  $\rho$  that maximizes the RHS of (29) for rate  $R$ . Hence it is clear that for any value of  $R$ , the rate of decrease of the feedback exponent  $E_F(R)$  is higher than that of  $LE_r(R/L)$ . It is shown in [2] that  $E_F(C) = E_r(C) = 0$ . Since  $E_r(R)$  is a decreasing function of  $R$ , we know that  $E_r(C/L) > E_r(C) = 0$ . Thus, when  $R = C$ , we have  $LE_r(C/L) > E_F(C)$ . Now, if the value of  $L$  is chosen such that  $LE_r(0) > E_F(0)$ , it is clear that the curve  $LE_r(R/L)$  lies strictly above the curve  $E_F(R)$  in the range  $0 \leq R \leq C$ . This directly follows from the fact that the feedback exponent  $E_F(R)$  decreases faster than  $LE_r(R/L)$ . Hence the condition  $LE_r(0) \geq E_F(0)$  is sufficient to guarantee that  $LE_r(R/L) \geq E_F(R)$  for all  $0 \leq R \leq C$ .  $\square$

The above lemma shows that for any BSC( $\epsilon$ ), an upper bound on  $L_{req}$  depends only on the values of  $E_F(R)$  and  $E_r(R)$  at  $R = 0$ . From the results in [3], it can be shown that

$$E_r(0) = \ln 2 - \ln \left(1 + 2\sqrt{\epsilon(1-\epsilon)}\right) \quad \text{and} \quad E_F(0) = C - \ln 2 - \ln \left(\sqrt{\epsilon(1-\epsilon)}\right) . \quad (30)$$

Using Lemma 4 and (30), we find that a deadline constraint of  $L = 4$  is enough to achieve the feedback exponent  $E_F(R)$  at all rates below capacity for *any* BSC with crossover probability  $0.05 \leq \epsilon \leq 0.5$ . However, the upper bound on  $L_{req}$ , derived using Lemma 4, becomes loose as  $\epsilon \rightarrow 0$ . To overcome this limitation, we use the expurgated exponent  $E_{ex}(R)$  [3] instead of the random coding exponent  $E_r(R)$  at low rates. Using numerical results, we find that the actual value of the minimum required deadline constraint is  $L_{req} = 3$  for all BSCs with  $\epsilon \leq 0.025$ , and  $L_{req} = 4$  otherwise.

## 4.2 The Very Noisy Channel (VNC)

As noted in [3], a channel is very noisy when the probability of receiving a given output is almost independent of the input, i.e., when the transition probabilities of the channel are given by

$$p_{jk} = \omega_j (1 + \epsilon_{jk}) ,$$

where  $\{\omega_j\}$  denotes the output probability distribution, and  $\{\epsilon_{jk}\}$  are such that  $|\epsilon_{jk}| \ll 1$  for all  $j$  and  $k$ , and  $\sum_j \omega_j \epsilon_{jk} = 0, \forall k$ . We plot the bounds on the error exponents given in (12) and (23), derived from the results in [2], in Figs. 2(a) and 2(b) for a VNC with capacity  $C = 1$  for  $L = 2$  and  $L = 4$  respectively. From the plots, it is clear that memoryless decoding is strictly sub-optimal to IR-ARQ for all rates  $R \geq 0.12$  (with  $L = 2$ ) and  $R \geq 0.25$  (with  $L = 4$ ). Moreover, it is evident that  $L = 4$  is sufficient for IR-ARQ to achieve the feedback exponent  $E_F(R)$  for all rates below capacity. This observation motivates the following result.

**Lemma 5** *For the very noisy channel, a deadline constraint of  $L = 4$  is enough for the proposed incremental redundancy scheme to achieve the feedback exponent  $E_F(R)$  for all rates  $0 \leq R \leq C$ .*

*Proof:* For a VNC, the random coding exponent is given by [2]

$$E_r(R) = \begin{cases} \left(\frac{C}{2} - R\right) , & 0 \leq R \leq \frac{C}{4} \\ (\sqrt{C} - \sqrt{R})^2 , & \frac{C}{4} \leq R \leq C \end{cases} . \quad (31)$$

Thus, under the deadline constraint  $L = 4$ , we have

$$4E_r(R/4) = 4 \left( \frac{C}{2} - \frac{R}{4} \right) = 2C - R , \quad 0 \leq R \leq C .$$

Also

$$E_F(R) = (C - R) + (\sqrt{C} - \sqrt{R})^2 \leq (C - R) + (\sqrt{C})^2 = 4E_r(R/4) .$$

Putting  $L = 4$  in (23), the error exponent of IR-ARQ is given by

$$E_{IR}(R, 4) \geq \min \{E_F(R) , 4E_r(R/4)\} = E_F(R) . \quad (32)$$

Thus, for a VNC, it is clear that a deadline constraint of  $L = 4$  is enough for IR-ARQ to achieve the feedback exponent  $E_F(R)$  at all rates below capacity.  $\square$

## 4.3 The Additive White Gaussian Noise (AWGN) channel

The random coding and expurgated exponents for an AWGN channel with a Gaussian input of power  $A$  and unit noise variance, are given in [3]. The sphere-packing exponent of the AWGN channel is derived in [7–9]. The parameter  $E_o(s, \rho, \mathbf{p})$  in the lower bound in (12) is replaced by  $E_o(s, \rho, t)$  which, following the steps in the derivation of the random coding exponent in [3], is given by

$$E_o(s, \rho, t) = (1 + \rho)tA + \left(\frac{1}{2}\right) \log(1 - 2tA) + \left(\frac{\rho}{2}\right) \log\left(1 - 2tA + \frac{sA}{\rho}\right) + \left(\frac{1}{2}\right) \log\left(1 + \frac{sA\left(1 - s - \frac{s}{\rho}\right)}{1 - 2tA + \frac{sA}{\rho}}\right) .$$

The feedback exponent for the AWGN channel is then given by [2, 3]

$$E_F(R) = \max_{0 \leq s \leq \rho \leq 1, t \geq 0} \left( \frac{E_o(s, \rho, t) - \rho R}{s} \right) .$$

We plot the bounds on the error exponents, given in (12) and (23), in Figs. 3(a) and 3(b) for an AWGN channel with signal-to-noise ratio  $A = 3$  dB for the deadline constraints  $L = 2$  and  $L = 4$  respectively. The plots clearly indicate that memoryless decoding is strictly sub-optimal to IR-ARQ for all rates  $R \geq 0.19$  (with  $L = 2$ ) and  $R \geq 0.46$  (with  $L = 4$ ). Moreover, when  $L = 4$ , the proposed IR-ARQ scheme achieves the feedback exponent  $E_F(R)$  for all rates below capacity (at the moment, we do not have a proof that this observation holds universally as in the case of BSCs).

## 5 Conclusions

We considered the error exponents of memoryless ARQ channels with an upper bound  $L$  on the maximum number of re-transmission rounds. In this setup, we have established the superiority of IR-ARQ, as compared with Forney's memoryless decoding. For the BSC and VNC, our results show that choosing  $L = 4$  is sufficient to ensure the achievability of Forney's feedback exponent, which is typically achievable with memoryless decoding in the asymptotic limit of large  $L$ . Finally, in the AWGN channel, numerical results also show the superiority of IR-ARQ over memoryless decoding, in terms of the achievable error exponent.

## References

- [1] M. V. Burnashev, “Data transmission over a discrete channel with feedback: Random transmission time,” *Problems of Information Transmission*, vol. 12, no. 4, pp. 250–265, 1976.
- [2] G. D. Forney Jr., “Exponential error bounds for erasure, list, and decision feedback schemes,” *IEEE Transactions on Information Theory*, vol. 14, pp. 206–220, March 1968.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, Inc., 1968.
- [4] G. Caire and D. Tuninetti, “The throughput of hybrid-ARQ protocols for the gaussian collision channel,” *IEEE Transactions on Information Theory*, vol. 47, pp. 1971–1988, July 2001.
- [5] A. J. Viterbi, “Error bounds for the white gaussian and other very noisy memoryless channels with generalized decision regions,” *IEEE Transactions on Information Theory*, vol. 15, pp. 279–287, March 1969.
- [6] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels I,” *Information and Control*, vol. 10, pp. 65–103, January 1967.
- [7] C. E. Shannon, “Probability of error for optimal codes in a gaussian channel,” *Bell Systems Technical Journal*, vol. 38, pp. 611–656, 1959.
- [8] A. Valembois and M. P. C. Fossorier, “Sphere-packing bounds revisited for moderate block lengths,” *IEEE Transactions on Information Theory*, vol. 50, pp. 2998–3014, December 2004.
- [9] S. Dolinar, D. Divsalar, and F. Pollara, “Code performance as a function of block size,” in *TMO Progress Report*, pp. 42–133, May 1998.

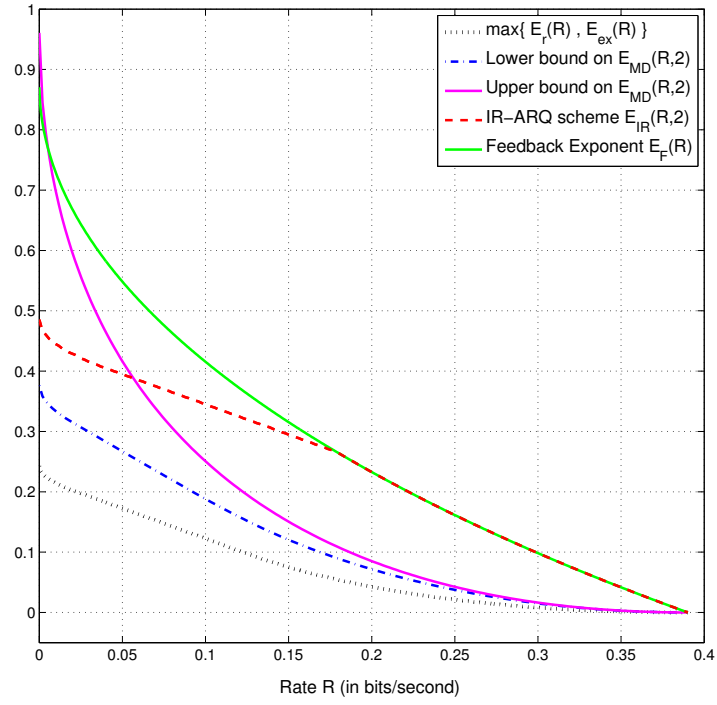


Figure 1(a): Comparison of the error exponents for a BSC with  $\epsilon = 0.15$  and  $L = 2$

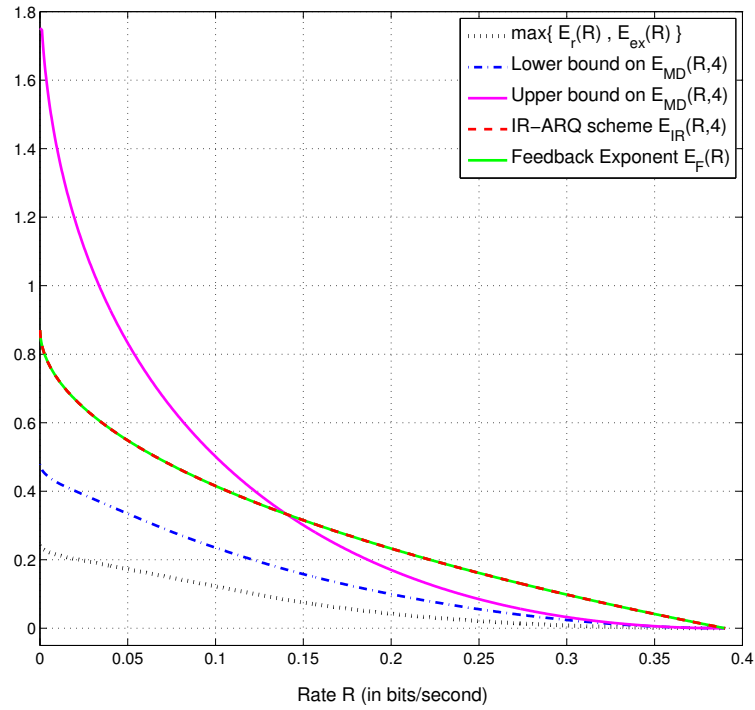


Figure 1(b): Comparison of the error exponents for a BSC with  $\epsilon = 0.15$  and  $L = 4$

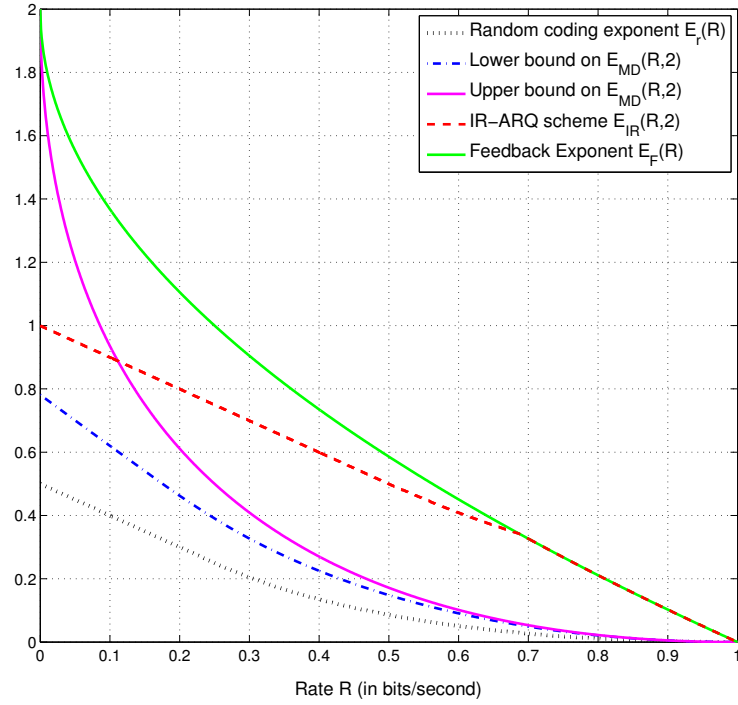


Figure 2(a): Comparison of the error exponents for a VNC with  $C = 1$  and  $L = 2$

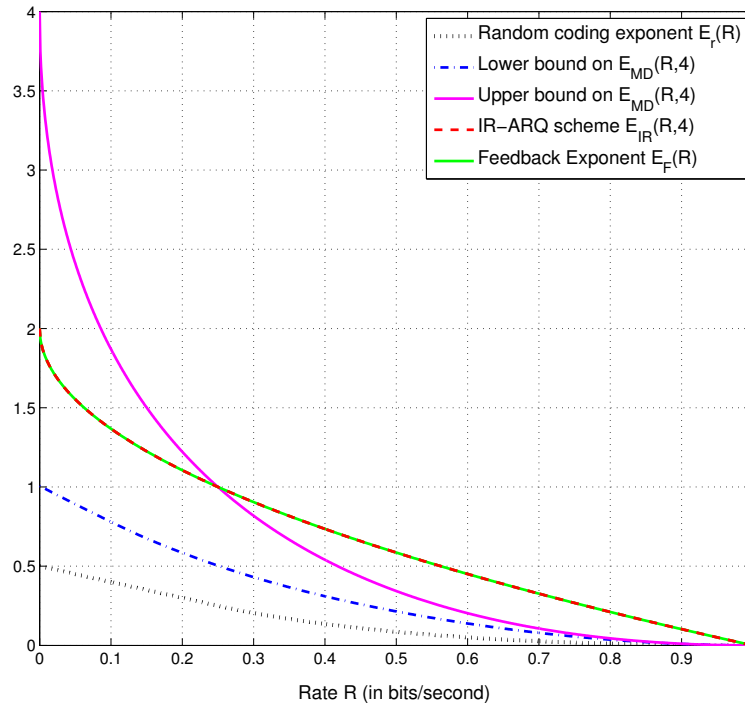


Figure 2(b): Comparison of the error exponents for a VNC with  $C = 1$  and  $L = 4$

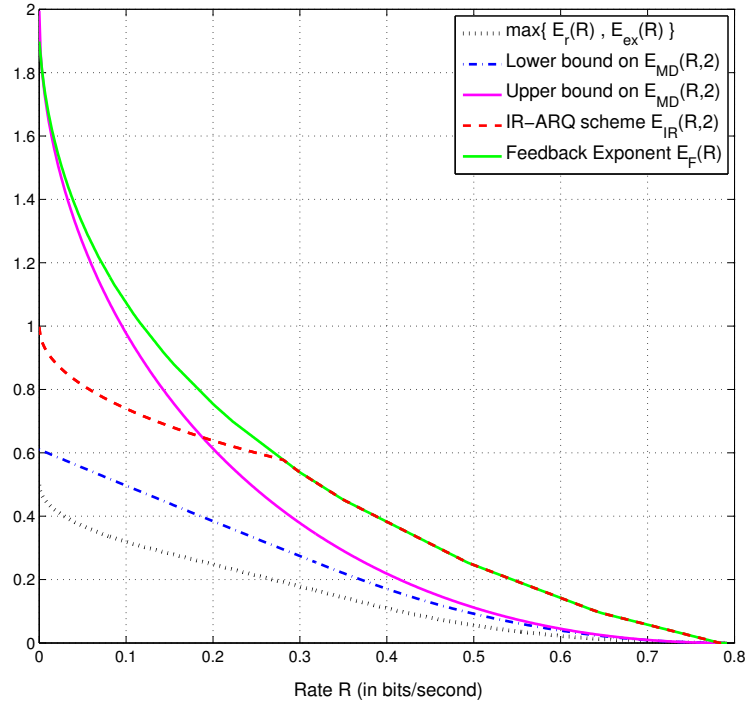


Figure 3(a): Comparison of error exponents for an AWGN channel with SNR = 3 dB and  $L = 2$

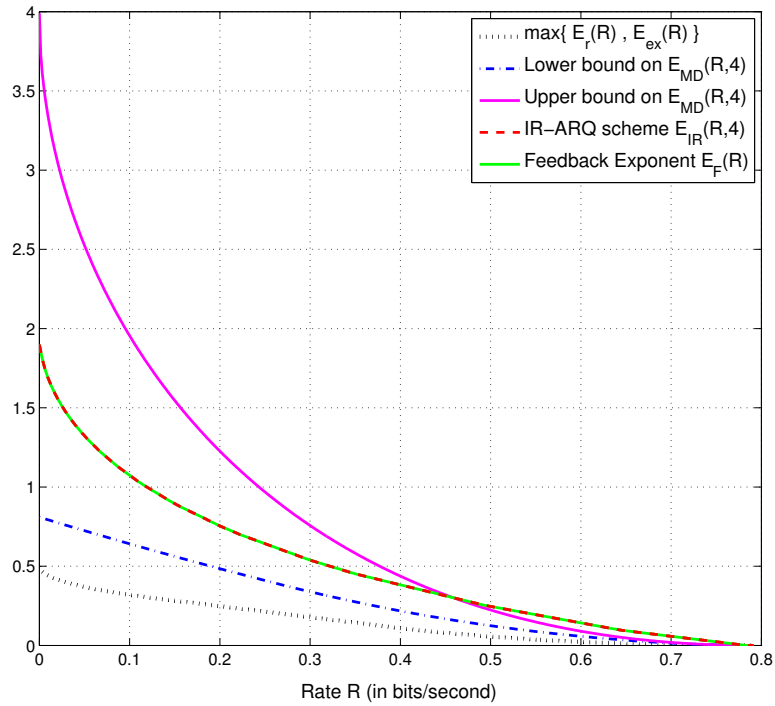


Figure 3(b): Comparison of error exponents for an AWGN channel with SNR = 3 dB and  $L = 4$