

An Efficient ARQ Scheme with SNR Feedback

Uri Erez
Tel Aviv University
Ramat Aviv, Israel
Email: uri@eng.tau.ac.il

Mitchell Trott
Hewlett-Packard Laboratories
Palo Alto, CA 94304
Email: mitchell.trott@hp.com

Gregory Wornell
Dept. EECS, MIT
Cambridge, MA
Email: gww@mit.edu

Abstract— We consider the problem of transmission over an unknown time-varying Gaussian channel, whose signal-to-noise ratio (SNR) is constant within a block but varies arbitrarily between blocks. For this scenario, we develop efficient automatic repeat request (ARQ) protocols in the form of low-complexity rateless codes, which encode a message into a sequence of incremental redundancy blocks. Following the receipt of each redundancy block, the receiver feeds back to the transmitter the SNR experienced by that block, which the encoder makes use of in structuring subsequent blocks. The resulting architecture, which involves layered repetition encoding and successive cancellation decoding, is capacity-achieving, enabling the message to be recovered with the minimum possible number of blocks for the realized channel.

I. INTRODUCTION

This paper considers the problem of efficient communication over an additive Gaussian noise channel of time-varying signal-to-noise ratio (SNR). We consider rateless codes for this channel, whereby a message of finite length is encoded into a sequence of incremental redundancy blocks. During the transmission of each block, the channel SNR is constant, but varies arbitrarily from block to block. Channel state information is available to the encoder only with delay. Specifically, at the end of the transmission of each incremental redundancy block, the encoder is informed of the realized SNR for that block. We stress that our treatment imposes no statistical model on the channel variation. This work generalizes the constructions developed in [1] for additive white Gaussian noise (AWGN) channels of unknown but fixed SNR.

For this channel, we develop an efficient (hybrid) ARQ protocol (see, e.g., [2] and references therein for relevant recent works). The encoder produces a sequence of redundancy blocks, but instead of simple ACK/NACK feedback from the receiver after each block, the receiver feeds back the SNR of the most recent block. From this information, the transmitter can determine whether a further redundancy block is required, and, if so, what form it should take. We remark in advance that while such a scheme might appear to be sensitive to delay in the feedback path, in practice, delay issues can be avoided by a suitable multiplexing strategy of interleaving multiple transmissions.

Our protocols take the form of low-complexity capacity-approaching rateless codes for this scenario. In particular,

This work was supported in part by HP under the MIT/HP Alliance, NSF Grant No. CCF-0515122, Draper Laboratory, ONR MURI Grant No. N00014-07-1-0738, and by MITRE Corp..

we consider layered encoding and successive cancellation decoding. Moreover, redundancy takes the form of dithered repetition, which is used in conjunction with minimum mean-square error (MMSE) combining.

Our construction transforms the time-varying channel into an equivalent AWGN channel, to which standard base codes can be applied. We show that with capacity-achieving base codes, our architecture can be perfect, i.e., capacity-achieving rateless codes can result for such channels. In particular, for the case of two redundancy blocks, we show that three layers are sufficient to obtain a capacity-achieving code for all rates below a certain threshold.

We further show that for arbitrary numbers of redundancy blocks, at least asymptotically-perfect rateless codes exist in the limit of a large number of layers in the construction. Our existence proof exploits simple symbol-by-symbol random dithering at the encoder, together with simple maximal ratio combining (MRC) at the decoder.

II. CHANNEL AND SYSTEM MODEL

The channel of interest takes the form

$$\mathbf{y}_m = \beta_m \mathbf{x}_m + \mathbf{z}_m, \quad m = 1, 2, \dots, \quad (1)$$

where the $\{\beta_m\}$ are a sequence of complex channel gains, \mathbf{x}_m is a vector of N input symbols, \mathbf{y}_m is the vector of channel output symbols, and \mathbf{z}_m is a noise vector of N independent, identically distributed (i.i.d.) complex, circularly-symmetric Gaussian random variables of (without loss of generality) unit variance, independent across blocks $m = 1, 2, \dots$. The channel input is limited to average power P per symbol. In our model, the channel gains β_m are known a priori at the receiver but not at the transmitter. The block length N is assumed to be large, but otherwise plays no important role in the analysis.

The encoder transmits a message w by generating a sequence of incremental redundancy blocks $\mathbf{x}_1(w)$, $\mathbf{x}_2(w)$, \dots . The receiver accumulates sufficiently many received blocks \mathbf{y}_1 , \mathbf{y}_2 , \dots to recover w . Immediately following the transmission of block \mathbf{x}_m , the encoder is notified of β_m , for $m = 1, 2, \dots$. Thus, knowledge of β_1, \dots, β_m can be used in the construction of the redundancy block $\mathbf{x}_{m+1}(w)$.

A code for such a scenario is parameterized by the ceiling rate R , which is the (realized) rate of the code if the message can be decoded from the first incremental redundancy block alone, and the range M , which is the maximum number of redundancy blocks generated by the code. A *perfect* rateless

code is then one in which capacity is achieved for any number $m = 1, \dots, M$ of redundancy blocks, i.e., whenever the (realized) channel gains are such that for some $1 \leq m \leq M$,

$$R = \sum_{m'=1}^m \log(1 + P|\beta_{m'}|^2), \quad (2)$$

is effectively satisfied, the message can be recovered (with high probability).

For values of m such that the right side of (2) is less than R , it is convenient to define *target* channel gains α_{m+1} required for successful decoding once block $m+1$ is obtained. In particular, α_{m+1} is defined via

$$R = \log(1 + P|\alpha_{m+1}|^2) + \sum_{m'=1}^m \log(1 + P|\beta_{m'}|^2), \quad (3)$$

whenever $|\alpha_m| > |\beta_m|$.

III. PERFECT RATELESS CODES

The rateless code constructions we pursue are natural generalizations of those in [1]. First, we choose the range M , the ceiling rate R , the number of layers L , and finally the associated base codebooks $\mathcal{C}_1, \dots, \mathcal{C}_L$. We assume a priori that the base codebooks all have equal rate R/L .

Given codewords $\mathbf{c}_l \in \mathcal{C}_l$, $l = 1, \dots, L$, the redundancy blocks $\mathbf{x}_1, \dots, \mathbf{x}_M$ take the form

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} = \mathbf{G} \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_L \end{bmatrix}, \quad (4)$$

where \mathbf{G} is an $M \times L$ matrix of complex gains, and where \mathbf{x}_m for each m and \mathbf{c}_l for each l are row vectors of length N . The power constraint enters by limiting the rows of \mathbf{G} to have squared norm P and by normalizing the codebooks to have unit power. Note that with this notation, the m th row of \mathbf{G} consists of the weights used in constructing the m th redundancy block from the L codewords. In the sequel we use g_{ml} to denote the (m, l) th entry of \mathbf{G} , and $\mathbf{G}_{m,l}$ to denote the upper-left $m \times l$ submatrix of \mathbf{G} . Where needed, we adopt the convention that $\mathbf{G}_{m,0} = \mathbf{0}$.

We emphasize that the m th row of \mathbf{G} will in general be a function of the (realized) channel gains $\beta_1, \dots, \beta_{m-1}$. Specifically, the m th row is designed for the M -block channel matrix

$$\mathbf{B}_m = \text{diag}(\beta_1, \dots, \beta_{m-1}, \alpha_m). \quad (5)$$

With this construction, each redundancy block contains a repetition of the codewords used in the earlier blocks, but with a different complex scaling factor, which can be interpreted as a “dither.” The code structure may therefore be viewed as a hybrid of layering and repetition.

In addition to the layered code structure, there is additional decoding structure, namely that the layered code be successively decodable. Specifically, to recover the message, we first decode \mathbf{c}_L , treating $\mathbf{G}[\mathbf{c}_1^T \dots \mathbf{c}_{L-1}^T]^T$ as (colored) noise, then decode \mathbf{c}_{L-1} , treating $\mathbf{G}[\mathbf{c}_1^T \dots \mathbf{c}_{L-1}^T]^T$ as noise, and so on.

Our aim is to select \mathbf{G} so that the code is perfect in the sense defined in Section II. Both the layered repetition encoding structure (4) and the successive decoding constraint impose requirements on \mathbf{G} in order to have a perfect code. In particular, from the encoding structure we require, as in [1], that the rows of \mathbf{G} be orthogonal, while from the decoding structure, we have the requirement

$$\frac{R}{L} = \log \frac{\det(\mathbf{I} + \mathbf{B}_m \mathbf{G}_{m,l} \mathbf{G}_{m,l}^\dagger \mathbf{B}_m^\dagger)}{\det(\mathbf{I} + \mathbf{B}_m \mathbf{G}_{m,l-1} \mathbf{G}_{m,l-1}^\dagger \mathbf{B}_m^\dagger)}, \quad (6)$$

for all $l = 1, \dots, L$ and $m = 1, \dots, M$.

As the simplest example, for the case of $M = 2$ redundancy blocks and $L = 3$ layers these constraints can be met, i.e., a perfect rateless code is possible, provided R is not too large, as we now develop.

In this case, we determine our gain matrix

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \end{bmatrix} \quad (7)$$

as a function of the ceiling rate R , where the second row also depends in general on the realized channel gain β_1 experienced by the first incremental redundancy block.

As in [1], we may without loss of generality take the first row and column to be real and nonnegative. Assume, also without loss of generality, that $|\alpha_1|^2 = 1$. Then the first row of \mathbf{G} , which corresponds to the first redundancy block, is computed exactly as in [1]. In particular, from (6) with $m = 1$, it must satisfy

$$R/3 = \log(1 + g_{11}^2) \quad (8)$$

$$2R/3 = \log(1 + g_{11}^2 + g_{12}^2) \quad (9)$$

$$3R/3 = \log(1 + g_{11}^2 + g_{12}^2 + g_{13}^2) \quad (10)$$

together with the power constraint

$$P = g_{11}^2 + g_{12}^2 + g_{13}^2. \quad (11)$$

Thus, with $x \triangleq 2^{R/6}$ we have

$$P = 2^R - 1 = x^6 - 1 \quad (12)$$

and

$$g_{11}^2 = 2^{R/3} - 1 = x^2 - 1, \quad (13)$$

$$g_{12}^2 = 2^{R/3}(2^{R/3} - 1) = x^2(x^2 - 1), \quad (14)$$

$$g_{13}^2 = 2^{2R/3}(2^{R/3} - 1) = x^4(x^2 - 1). \quad (15)$$

The derivation now departs from [1]. Recall that β_1 is the realized channel gain for the first block. A second redundancy block is thus needed when $|\beta_1| < |\alpha_1|$. The minimum required value of the channel gain $|\alpha_2|^2$ necessary for decoding is the solution to [cf. (3)]

$$R = \log(1 + P|\beta_1|^2) + \log(1 + P|\alpha_2|^2), \quad (16)$$

which is

$$|\alpha_2|^2 = \frac{1 - |\beta_1|^2}{1 + P|\beta_1|^2}. \quad (17)$$

Using (6) for $m = 2$ and $l = 1$ yields

$$R/3 = \log(1 + |\beta_1|^2 g_{11}^2 + |\alpha_2|^2 g_{21}^2). \quad (18)$$

Substituting the previously computed expressions (13) and (17) for g_{11}^2 and $|\alpha_2|^2$ and solving for g_{21} yields

$$g_{21}^2 = (x^2 - 1)(1 + P|\beta_1|^2). \quad (19)$$

As in [1], to solve for the rest of the second row of \mathbf{G} we use (6) with $m = l = 2$ together with the requirement that the first and second rows be orthogonal. It is useful at this stage to switch to polar coordinates, i.e., $g_{22} = |g_{22}|e^{j\theta_1}$ and $g_{23} = |g_{23}|e^{j\theta_2}$.

Orthogonality of the first and second rows means that

$$0 = g_{11}g_{21} + g_{12}|g_{22}|e^{j\theta_1} + g_{13}|g_{23}|e^{j\theta_2}. \quad (20)$$

The three terms in the above expression may be viewed as the legs of a triangle, so by the law of cosines

$$2g_{11}g_{21}g_{12}|g_{22}|\cos\theta_1 = g_{13}^2|g_{23}|^2 - g_{11}^2g_{21}^2 - g_{12}^2|g_{22}|^2. \quad (21)$$

We now use (6) with $m = 2$ and $l = 1, 2$ to infer that

$$2^{2R/3} = x^4 = \det(\mathbf{I} + \text{diag}(|\beta_1|^2, |\alpha_2|^2)\mathbf{G}_{2,2}\mathbf{G}_{2,2}^\dagger). \quad (22)$$

To expand this expression, we compute

$$\mathbf{G}_{2,2}\mathbf{G}_{2,2}^\dagger = \begin{bmatrix} g_{11}^2 + g_{12}^2 & g_{11}g_{21} + g_{12}|g_{22}|e^{-j\theta_1} \\ (*) & g_{21}^2 + |g_{22}|^2 \end{bmatrix}, \quad (23)$$

where $(*)$ is the complex conjugate of the upper right entry, from which we find

$$\begin{aligned} \det(\mathbf{I} + \text{diag}\{|\beta_1|^2, |\alpha_2|^2\}\mathbf{G}_{2,2}\mathbf{G}_{2,2}^\dagger) = \\ |\beta_1|^2|\alpha_2|^2(g_{11}^2|g_{22}|^2 + g_{12}^2g_{21}^2 - 2g_{11}g_{21}g_{12}|g_{22}|\cos\theta_1) \\ + |\beta_1|^2(g_{11}^2 + g_{12}^2) + |\alpha_2|^2(g_{21}^2 + |g_{22}|^2) + 1. \end{aligned} \quad (24)$$

Substituting (21) into (24) and using (22) yields

$$\begin{aligned} x^4 = |\beta_1|^2|\alpha_2|^2 \\ \cdot (g_{11}^2|g_{22}|^2 + g_{12}^2g_{21}^2 - g_{13}^2|g_{23}|^2 + g_{11}^2g_{21}^2 + g_{12}^2|g_{22}|^2) \\ + |\beta_1|^2(g_{11}^2 + g_{12}^2) + |\alpha_2|^2(g_{21}^2 + |g_{22}|^2) + 1. \end{aligned} \quad (26)$$

Finally, substituting the expressions for g_{11}^2 , g_{12}^2 , g_{13}^2 , g_{21}^2 , and $|\alpha_2|^2$ computed above, using the power constraint

$$|g_{23}|^2 = P - |g_{22}|^2 - g_{21}^2, \quad (27)$$

solving for $|g_{22}|^2$, and simplifying terms, we arrive at

$$\begin{aligned} |g_{22}|^2 = \frac{x^2 - 1}{1 + (x^6 - 1)|\beta_1|^2} \\ \cdot (x^2 + |\beta_1|^2(x^{10} + x^8 - 2x^6 + x^4 - 2x^2 + 1) \\ - |\beta_1|^4(x^{12} - x^{10} + x^8 - 2x^6 + x^4 - x^2 + 1)). \end{aligned} \quad (28)$$

Evidently, a necessary condition for the existence of a solution for \mathbf{G} is that $g_{21}^2 + |g_{22}|^2 < P$. It can be shown that

the sum of the powers on the first two layers is maximized when $|\beta_1| \rightarrow 1$, and then the necessary condition simplifies to

$$2^{R+1} - 2^{2R/3+1} < 2^R - 1, \quad (29)$$

which may be shown to hold for all $R < \log(2 + \sqrt{5}) \approx 2.08$ bits per complex symbol.

The final step, which we omit due to space constraints, is to apply the triangle inequality to (20) to prove that the required triangle exists, and thus also the phases θ_1 and θ_2 .

Establishing the existence of perfect rateless codes for larger values of M and/or L requires more effort. As an alternative, we next demonstrate that in the limit of a large number of layers L , asymptotically perfect codes for all values of M are possible via essentially the same construction, even using suboptimal encoding and decoding.

IV. ASYMPTOTICALLY PERFECT RATELESS CODES

Our construction is a slight generalization of the corresponding construction in [1]. First, as in Section III we fix M , R , L , and the associated codebooks $\mathcal{C}_1, \dots, \mathcal{C}_L$ each of rate R'/L for some $R' < R$ to be determined. Using $c_l(n)$ and $x_m(n)$ to denote the n th elements of codeword \mathbf{c}_l and redundancy block \mathbf{x}_m , respectively, we have [cf. (4)]

$$\begin{bmatrix} x_1(n) \\ \vdots \\ x_M(n) \end{bmatrix} = \mathbf{G}(n) \begin{bmatrix} c_1(n) \\ \vdots \\ c_L(n) \end{bmatrix} \quad (30)$$

for $n = 1, \dots, N$. The value of M plays no role in our development and may be taken arbitrarily large. Moreover, as before, the power constraint enters by limiting the rows of $\mathbf{G}(n)$ to have a squared norm P and by normalizing the codebooks to have unit power.

A. Power Allocation

A suitable power allocation for our construction is obtained as that which is optimum for a slightly different construction, which we now develop. In this section, different (independent) codebooks are used for different redundancy blocks, and we take $\mathbf{G}(n)$ to be independent of n , so that $\mathbf{G}(n) = \mathbf{P}$, where

$$\mathbf{P} = \begin{bmatrix} \sqrt{p_{1,1}} & \cdots & \sqrt{p_{1,L}} \\ \vdots & \ddots & \vdots \\ \sqrt{p_{M,1}} & \cdots & \sqrt{p_{M,L}} \end{bmatrix}. \quad (31)$$

The mutual information in the l th layer of the m th block is then

$$I_{m,l} = \log(1 + \text{SNR}_{m,l}(\beta_m)). \quad (32)$$

where

$$\text{SNR}_{m,l}(\beta_m) = \frac{|\beta_m|^2 p_{m,l}}{|\beta_m|^2 (p_{m,1} + \cdots + p_{m,l-1}) + 1}. \quad (33)$$

is the associated per-layer SNR experienced during successive decoding.

We now obtain the elements of \mathbf{P} recursively. We proceed from the first block $m = 1$ to block M , where in each block m

we start by determining $P_{m,1}$ and proceed up through $P_{m,L}$. By definition of α_1 , we have

$$\log \left(1 + |\alpha_1|^2 \sum_{l=1}^L P_{1,l} \right) = R.$$

Viewing the layering as superposition coding for a multi-access channel, it is clear that any rate vector is achievable as long as its sum-rate is R . We may therefore obtain an equal rate per layer by taking $P_{1,1}, \dots, P_{1,L}$ such that

$$\log(1 + P_{1,l}|\alpha_1|^2) = R/L, \quad l = 1, \dots, L. \quad (34)$$

Upon receiving knowledge of $|\beta_1|$ we proceed to determine the power allocation for block $m = 2$. More generally, suppose the power allocations through block $m - 1$ have been determined and we have now acquired channel state knowledge through β_{m-1} . To determine the allocation for block m , we first compute the mutual information shortfall in layer l as

$$\Delta_{m,l} = \frac{R}{L} - \sum_{m'=1}^{m-1} \log(1 + \text{SNR}_{m',l}(\beta_{m'})). \quad (35)$$

By the induction hypothesis, had the realized channel gain been $|\beta_{m-1}| = |\alpha_{m-1}|$, then $\Delta_{m,l}$ would be zero for all $l = 1, \dots, L$. Now since we have $|\beta_{m-1}| < |\alpha_{m-1}|$, clearly the shortfall is positive for all layers. By definition of α_m , we also have

$$\Delta_m = \sum_{l=1}^L \Delta_{m,l} = \log(1 + P|\alpha_m|^2). \quad (36)$$

We then solve for $P_{m,1}, \dots, P_{m,L}$, in order, via

$$\log(1 + \text{SNR}_{m,l}(\alpha_m)) = \Delta_{m,l}. \quad (37)$$

The resulting power allocation ensures that the aggregate mutual information per layer is at least R/L if $|\beta_m| > |\alpha_m|$ when i.i.d. Gaussian codebooks for all layers and blocks. However, we wish to use the same set of L codebooks for all redundancy blocks, to keep decoding complexity low. We return to this problem next, but in doing so will exploit this power allocation.

B. Dithered Encoding

We restrict our attention to $\mathbf{G}(n)$ of the form

$$\mathbf{G}(n) = \mathbf{P} \odot \mathbf{D}(n), \quad (38)$$

where \odot denotes elementwise multiplication, where \mathbf{P} is the matrix (31) developed in Section IV-A, and where $\mathbf{D}(n)$ is a (random) phase-only “dither” matrix of the form

$$\mathbf{D}(n) = \begin{bmatrix} d_{1,1}(n) & \cdots & d_{1,L}(n) \\ \vdots & \ddots & \vdots \\ d_{M,1}(n) & \cdots & d_{M,L}(n) \end{bmatrix}. \quad (39)$$

In our analysis, the $d_{m,l}(n)$ are all i.i.d. in m, l , and n , and are drawn independently of all other random variables, including noises, messages, and codebooks. It is sufficient for $d_{m,l}(n)$ to take on only values ± 1 , and with equal probability.

C. Decoding

Since $\mathbf{G}(n)$ is drawn i.i.d., the overall channel is i.i.d., and thus we may express the channel model in terms of an arbitrary individual element in the block. Specifically, assume that the channel gain for block m is the minimal required $\beta_m = \alpha_m$, then our received symbol can be expressed as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \mathbf{B}_m \mathbf{G} \begin{bmatrix} c_1 \\ \vdots \\ c_L \end{bmatrix} + \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix},$$

where $\mathbf{G} = \mathbf{P} \odot \mathbf{D}$, with \mathbf{G} denoting the arbitrary element in the sequence $\mathbf{G}(n)$, and where $y_{m'}$ is the corresponding received symbol from redundancy block m' (and similarly for $c_{m'}, z_{m'}, \mathbf{D}$).

It is sufficient to employ successive cancellation decoding with simple maximal ratio combining (MRC) of the redundancy blocks. In decoding the l th layer, the MRC decoder not only treats the undecoded layers $1, \dots, l - 1$ as noise, but also treats the dither in those layers as a process with known statistics but unknown realization. As done in [1], it is easy to show that the effective SNR at which this l th layer is decoded from m blocks via such MRC decoding is

$$\text{SNR}_{\text{MRC}} = \sum_{m'=1}^m \text{SNR}_{m',l}. \quad (40)$$

D. Efficiency Analysis

To show that the resulting scheme is asymptotically perfect, we begin by noting that the mutual information $I'_{m,l}$ when random dither encoding, MRC decoding, and capacity-achieving base codes are used, we have

$$I'_{m,l} \geq \log(1 + \text{SNR}_{\text{MRC}}) \quad (41)$$

with SNR_{MRC} as in (40), and where (41) is an inequality because the effective noise is not necessarily Gaussian.

The efficiency of our scheme ultimately depends on the choice of our power allocation matrix (31). Note that we may further bound $I'_{m,l}$ for all m by

$$I'_{m,l} \geq \frac{1}{\ln 2} \log \left(1 + \ln 2 \frac{R}{L} \right), \quad (42)$$

where we have applied the inequality $u \geq \ln(1 + u)$ valid for $u \geq 0$. Thus, if we choose the rate of the base code in each layer to be

$$\frac{R'}{L} = \frac{1}{\ln 2} \log \left(1 + \ln 2 \frac{R}{L} \right), \quad (43)$$

then (42) ensures decodability after m blocks are received when the channel gain satisfies $|\beta_m| \geq |\alpha_m|$, as required. Moreover, the efficiency R'/R can be made as close as desired to one by taking L sufficiently large.

REFERENCES

- [1] U. Erez, M. Trott, and G. Wornell, “Rateless coding for Gaussian channels,” *IEEE Trans. Inform. Theory*, 2007, submitted for publication. Also available on arXiv; see <http://arxiv.org/abs/0708.2575v1>.
- [2] S. Sesia, G. Caire, and G. Vivier, “Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes,” *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004.