The Performance Of Focused Error Control Codes

Fady Alajaji, *Student Member, IEEE*, and Thomas Fuja, *Member, IEEE*

Abstract— **Consider an additive noise channel with inputs and outputs in the field GF(q) where q > 2; every time a symbol is transmitted over such a channel, there are q−1 different errors that can occur, corresponding to the q−1 non-zero elements that the channel can add to the transmitted symbol. In many data communication/storage systems, there are some errors that occur much more frequently than others; however, traditional error correcting codes – designed with respect to the Hamming metric – treat each of these q−1 errors the same. Fuja and Heegard have designed a class of codes, called focused error control codes, that offer different levels of protection against "common" and "uncommon" errors; the idea is to define the level of protection in a way based not only on the number of errors, but the kind as well. In this paper, the performance of these codes is analyzed with respect to idealized "skewed" channels as well as realistic non-binary modulation schemes. It is shown that focused codes, used in conjunction with PSK and QAM signaling, can provide more than 1.0 dB of additional coding gain when compared with Reed-Solomon codes for small blocklengths.**

I. Introduction and Motivation

When a symbol from GF(q) is sent over a channel with additive noise, there are $q-1$ different non-zero noise symbols that can corrupt the transmitted field element. "Traditional" error control codes, designed with respect to the Hamming metric, treat each of these $q-1$ possibilities the same.

However, in many non-binary data transmission and storage channels, there are some errors that occur much more frequently than others. As an example, consider a modulation scheme in which data is mapped onto one of $M = 2^b$ signal points using a Gray code, so the most likely detection errors cause exactly one bit error per symbol. In such a system the most likely errors cause the received symbol to differ from the transmitted symbol in exactly one bit of their binary representation; thus, while there are $2^b - 1$ different *possible* errors, there are only b that are *likely*. A similar situation arises in byte-organized memory systems; while a code with "byte wide" symbols may be structurally appropriate for such a system, the dominant error types are often single-bit-per-byte failures. It is obviously inefficient to provide the same degree of protection against the uncommon errors as against the common ones.

Paper approved by Sergio Benedetto, the Editor for Signal Design, Modulation and Detection of the IEEE Communications Society. Manuscript received November 15, 1990; revised August 15, 1991 and October 15, 1991. Parts of this paper were presented at the 1990 International Symposium on Information Theory, January 14-19, 1990, San Siego, California.

This work was supported in part by the National Science Foundation Engineering Research Centers Program, NSFD CDR-8803012; also by NSF Grant NCR-8957623.

The authors are with the Department of Electrical Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742.

IEEE Log Number 9400909

This observation led Fuja and Heegard to develop the idea of a *focused* error control code [1]. Such codes are designed to give one level of protection against a specific set of common errors while maintaining another (lower) level of protection against uncommon errors. In [1] results are obtained regarding the existence and construction of such codes as well as bounds on their rates.

This paper analyzes the unique rate/performance trade-offs offered by focused codes. First, the pertinent results from [1] are reviewed; then, the performance of focused codes over an idealized skewed channel is analyzed and that analysis is applied to M-ary PSK and QAM signaling.

II. Background on Focused Codes

In this section the pertinent results from [1] are reviewed.

A. What is a Focused Code ?

For any $\mathbf{x} \in \mathrm{GF}(q)^n$ denote the *Hamming weight* of $\mathbf{x}$ by $||\mathbf{x}||$ – i.e., $||\mathbf{x}||$ denotes the number of non-zero components of $\mathbf{x}$. More generally, let $\mathrm{GF}(q)^*$ denote the set of non-zero elements of GF(q) and let $\mathbf{A} \subseteq \mathrm{GF}(q)^*$. Then for any $\mathbf{x} \in \mathrm{GF}(q)^n$ define the $\mathbf{A}$-*weight* of $\mathbf{x}$ as the number of components of $\mathbf{x}$ that lie in $\mathbf{A}$.

Definition: Let $\mathbf{B} \subseteq \mathrm{GF}(q)^*$ be a set of non-zero elements of GF(q). ($\mathbf{B}$ represents the set of common errors.) A code is (t_1, t_2)-focused on $\mathbf{B}$ if it can correct up to $t_1 + t_2$ errors provided at most t_1 of these errors lie outside $\mathbf{B}$. More precisely, such a code is a set $\mathcal{C}$ of n-tuples over GF(q) such that for any $\mathbf{x} \in \mathrm{GF}(q)^n$ there is at most one $\mathbf{c} \in \mathcal{C}$ satisfying *both* of the following conditions:

1. $d(\mathbf{x}, \mathbf{c}) \leq t_1 + t_2$;
2. $d_{\mathbf{B}^c}(\mathbf{x}, \mathbf{c}) \leq t_1$.

(Here $d(\mathbf{x}, \mathbf{y}) \triangleq ||\mathbf{x} - \mathbf{y}||$ and $d_{\mathbf{B}^c} \triangleq ||\mathbf{x} - \mathbf{y}||_{\mathbf{B}^c}$.)

Note that a $(t, 0)$-focused code is a "traditional" t-error correcting code, while a $(0, t)$-focused code is *completely* focused on $\mathbf{B}$ – i.e. it can correct up to t errors, provided they are *all* common.

B. Construction of Combined Focused Codes for Odd-Weight-per-Byte Errors

In [1] a method for constructing focused codes was presented. The codes thus constructed were called *combined focused codes,* and in this section we review that method.

Suppose the goal is to construct a code with blocklength n over GF(2^b) that is (t_1, t_2)-focused on the set of odd-weight symbols; that is, the common error set $\mathbf{B}$ consists of all the elements of GF(2^b) with a binary representation containing an odd number of 1's. (Note that this would include the set of single-bit errors.)

The construction from [1] is described as follows. For any n-tuple $\mathbf{x}$ over GF(2^b), let $\mathbf{b}(\mathbf{x})$ be the binary n-tuple obtained by taking the mod-two sum of each component of $\mathbf{x}$. For example, if $\mathbf{x}$=[0011,0100,1101,1010,1111], then $\mathbf{b}(\mathbf{x})$=[01100]. Let C_1 be an (n, nR_1) binary *inner* code with minimum distance $d_1 = 2t_1 + 2t_2 + 1$; let C_2 be an (n, nR_2) *outer* code over GF(2^{b-1}) with minimum distance $d_2 = 2t_1 + t_2 + 1$. To construct a codeword from the focused code, first take a codeword c_1 from C_1 and a codeword c_2 from C_2. Then add one bit to each symbol from c_2 such that $\mathbf{c}$, the resulting n-tuple over GF(2^b), satisfies $\mathbf{b}(\mathbf{c}) = c_1$.

To see that the code thus constructed is (t_1, t_2)-focused on $\mathbf{B}$, consider the following decoding algorithm. Given a received 2^b-ary n-tuple $\mathbf{r}$, compute $\mathbf{b}(\mathbf{r})$; find the codeword $\mathbf{x} \in C_1$ that is closest to $\mathbf{b}(\mathbf{r})$. As long as at most $t_1 + t_2$ odd-weight errors have occurred, $\mathbf{x}$ will be equal to $\mathbf{b}(\mathbf{c})$, where $\mathbf{c}$ is the codeword that was actually transmitted. Mark the locations where $\mathbf{x}$ differs from $\mathbf{b}(\mathbf{r})$ as erasures; strip off the last bit in each code symbol and pass the resulting 2^{b-1}-ary n-tuple plus erasure locations to a decoder for C_2.

Suppose ℓ_1 common and ℓ_2 uncommon errors occur during transmission; then as long as

$$\ell_1 \leq \lfloor (d_1 - 1)/2 \rfloor$$

and

$$\ell_2 \leq \lfloor (d_2 - \ell_1 - 1)/2 \rfloor$$

the above algorithm will correctly estimate the transmitted codeword. It is trivial to show that as long as $\ell_1 + \ell_2 \leq t_1 + t_2$ and $\ell_2 \leq t_1$ the above inequalities are met; thus, the code described above *is* (t_1, t_2)-focused.

Indeed, there are *other* error patterns – other values of ℓ_1 and ℓ_2 – that satisfy the above inequalities. As an example, suppose we wish to construct a $(0, 2)$-focused code over GF(2^b) using the above technique. Then we would need a binary inner code with minimum distance $d_1 = 5$ and an outer code over GF(2^{b-1}) with minimum distance $d_2 = 3$. Such a code would be able to correct any single *uncommon* error – as long as there were no common errors – in addition to the error patterns described by the "$(0, 2)$-focused" designation.

The overall rate of this code is $(1/b)R_1 + (b-1)R_2/b$. Furthermore, this technique can be generalized to cover a variety of common error sets; for details, refer to [1].

III. Performance of Focused Codes on an Idealized Channel

We now consider the performance of a (t_1, t_2)-focused code over an idealized "skewed" channel. One of the fundamental questions to be considered is: Under what conditions does a (t_1, t_2)-focused code perform identically to a $t_1 + t_2$-error correcting code ? Since a (t_1, t_2)-focused code can generally be constructed at a higher rate than a $t_1 + t_2$-error correcting code, answering this question will give insight into when focused codes might be appropriate for a particular application.

A. The Skewed Symmetric Channel

Consider the following model for a communication (or storage) channel. A character $X \in$ GF(q) is transmitted and the character $Y = X + Z \in$ GF(q) is received. Here, the noise Z is assumed to be i.i.d., independent of the input X, and distributed according to

$$\mathrm{P}(Z = z) = \begin{cases} 1 - \epsilon, & \text{if } z = 0; \\ \epsilon(1 - \gamma)/|\mathbf{B}|, & \text{if } z \in \mathbf{B}; \\ \epsilon\gamma/|\mathbf{B}^c|, & \text{if } z \in \mathbf{B}^c, \end{cases}$$

where

- $\mathbf{B}$ is a set of non-zero field elements.
- $\mathbf{B}^c$ consists of those non-zero elements that are *not* in $\mathbf{B}$.
- $\epsilon = \mathrm{P}(Z \neq 0)$ is the probability of symbol channel error.
- $\gamma = \mathrm{P}(Z \notin \mathbf{B} | Z \neq 0)$ is the probability that Z lies outside $\mathbf{B}$, given that $Z \neq 0$.

This channel – called the *skewed symmetric channel* (SSC) for the focus set $\mathbf{B}$ – was introduced in [1] as an idealized model of a channel that exhibits the "skewing" property that focused codes were designed to address. $\mathbf{B}$ represents the class of common errors and so $\gamma \ll 1$.

B. On the Decoding of Focused Codes

To analyze the performance of focused codes operating over a noisy channel we must first specify a decoding algorithm. The strategy we will use is the natural extension of bounded distance (or *incomplete*) decoding. Suppose a codeword from a (t_1, t_2)-focused code $\mathcal{C}$ is transmitted and the q-ary n-tuple $\mathbf{y}$ is received; then the decoder computes $f(\mathbf{y})$, its estimate of the transmitted codeword, according to the following rule:

$$f(\mathbf{y}) = \begin{cases} \mathbf{x}; & \text{if } \mathbf{x} \in \mathcal{C},\ d(\mathbf{x}, \mathbf{y}) \leq t_1 + t_2,\ \&\ d_{\mathbf{B}^c}(\mathbf{x}, \mathbf{y}) \leq t_1; \\ ?; & \text{if no } \mathbf{x} \in \mathcal{C} \text{ satisfies the above inequalities.} \end{cases}$$

Note that $f(\mathbf{y})$ is well-defined – i.e., for any $\mathbf{y} \in$ GF$(q)^n$ there is at most one codeword $\mathbf{x} \in \mathcal{C}$ that satisfies the two inequalities. The "?" indicates a detected error that cannot be corrected; depending on the application the decoder can ask for re-transmission or simply output $\mathbf{y}$ and raise a flag.

This algorithm will correctly decode the received codeword as long as no more than $t_1 + t_2$ errors occur during transmission and no more than t_1 of those errors lie outside $\mathbf{B}$. The decoding algorithm described in Section 2.2 is of this form.

In the following sections we will analyze the performance of this decoding algorithm when it is used with focused codes operating over the skewed symmetric channel. The two performance measures we will use are the probability of block decoding error – labeled P_d – and the probability of symbol error – labeled P_s.

Suppose we receive the n-tuple $\mathbf{y}$. Then

$$P_d \triangleq P(f(\mathbf{y}) \neq \mathbf{x} | \mathbf{x} \text{ was transmitted}).$$

(Note that P_d is defined as the probability that a correct decision is *not* made; thus it does not distinguish between detectable-but-uncorrectable errors and uncorrectable errors that cannot be detected.) Of course $\mathbf{y} = \mathbf{x} + \mathbf{z}$, where $\mathbf{z}$ is a random noise vector whose components are chosen independently according to the distribution given in Section 3.1; thus P_d is just the probability that $||\mathbf{z}|| > t_1 + t_2$ and/or $||\mathbf{z}||_{\mathbf{B}^c} > t_1$.

The symbol error probability P_s is defined as follows. Suppose the codeword $\mathbf{x} \in \mathcal{C}$ is transmitted and $\mathbf{y} \in \mathrm{GF}(q)^n$ is received. Define the random variable N to be the number of symbols where $\mathbf{x}$ and the decoder's estimate of $\mathbf{x}$ disagree – i.e., $N = d(\mathbf{x}, f(\mathbf{y}))$. Then

$$P_s \triangleq \frac{E[N]}{n},$$

where n is the blocklength of the code. By the above decoding algorithm, if the channel introduces at most $t_1 + t_2$ errors and at most t_1 of them are "uncommon", then $N = 0$. Furthermore we make the following pessimistic assumption: If an uncorrectable error occurs – i.e., if $d(\mathbf{x}, \mathbf{y}) > t_1 + t_2$ and/or $d_{\mathbf{B}^c}(\mathbf{x}, \mathbf{y}) > t_1$ – then the decoder introduces an *additional* $t_1 + t_2$ symbol errors. This assumption is pessimistic because the decoder output will always be within a distance $t_1 + t_2$ of its input; the worst-case scenario would be if the decoder changed $t_1 + t_2$ symbols that were correct, yielding $N = t_1 + t_2 + d(\mathbf{x}, \mathbf{y})$.

Finally, it should be noted that the decoding algorithm described above does *not* always produce the maximum likelihood estimate of the transmitted codeword for the skewed symmetric channel; thus this decoding algorithm is sub-optimal. For the SSC the maximum likelihood estimate of the transmitted codeword given a received vector $\mathbf{y}$ is the codeword that minimizes $Ad(\mathbf{x}, \mathbf{y}) + Fd_{\mathbf{B}^c}(\mathbf{x}, \mathbf{y})$ over all choices of $\mathbf{x} \in \mathcal{C}$; here A and F are non-negative constants determined by the channel parameters ϵ and γ. Thus while our decoding procedure is not optimal for the SSC, neither is bounded distance decoding – the strategy typically used with Hamming-metric codes.

C. Block Error Probability over a Skewed Symmetric Channel

Suppose a codeword from a (t_1, t_2)-focused code is transmitted over a skewed symmetric channel with parameters ϵ and γ. From Section 3.2, the probability of block decoding error is the probability that more than $t_1 + t_2$ errors occurred and/or more than t_1 uncommon errors occurred – i.e.,

$$P_d = \sum_{i=t_1+1}^{t_1+t_2} \sum_{j=t_1+1}^{i} \binom{n}{i} \binom{i}{j} \epsilon^i (1-\epsilon)^{n-i} \gamma^j (1-\gamma)^{i-j} + \sum_{i=t_1+t_2+1}^{n} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i}. \quad (1)$$

The first sum in (1) is the probability that there are at most $t_1 + t_2$ errors in a block of n transmitted symbols, but more than t_1 of them are uncommon; the second sum is the probability that there are more than $t_1 + t_2$ errors. Thus, the second sum is the probability of block decoding error for a "traditional" $t_1 + t_2$-error correcting code.

If $n\epsilon \ll 1$, we can approximate P_d by taking only the first terms in the sums in equation (1):

$$P_d \approx \binom{n}{t_1+1} \epsilon^{t_1+1} (1-\epsilon)^{n-t_1-1} \gamma^{t_1+1} + \binom{n}{t_1+t_2+1} \epsilon^{t_1+t_2+1} (1-\epsilon)^{n-t_1-t_2-1}. \quad (2)$$

Figure 1 shows P_d versus γ on a log-log scale for fixed $\epsilon = 10^{-3}$, blocklength $n = 50$ and for values of t_1 and t_2 such that $t_1 + t_2 = 4$. Each curve can be broken up into two distinct regions; for large values of γ, the graph is a straight line with slope $t_1 + 1$, whereas for small values of γ the graph has slope zero and coincides with the graph of P_d for a $(4, 0)$-focused code – i.e., a four-error correcting code. This is because, for large values of γ, the channel is not very focused – i.e., the uncommon errors are not *that* uncommon – and so the dominant cause of decoder error is the occurrence of $t_1 + 1$ uncommon errors; to put it more simply, when $\gamma \not\ll 1$, the first term in equation (2) is dominant. Similarly, when $\gamma \ll 1$ then the uncommon errors are *so* uncommon that the primary source of decoder error is the occurrence of $t_1 + t_2 + 1$ errors, which means that a (t_1, t_2)-focused code performs identically to a $t_1 + t_2$-error correcting code – i.e., the *second* term in equation (2) dominates.

So, partially answering the question that was posed at the beginning of this section: For fixed ϵ, a (t_1, t_2)-focused code has the same block decoding error probability as a $t_1 + t_2$-error correcting code when the second term in equation (2) is much larger than the first term. If we define (for fixed ϵ) γ_{crit} to be the value of γ for which the two terms in equation (2) are equal to one another, then

$$\log_{10} \gamma_{crit} = \frac{t_2}{t_1+1} \log_{10} \left(\frac{\epsilon}{1-\epsilon} \right) + \frac{1}{t_1+1} \log_{10} \left[\frac{\binom{n}{t_1+t_2+1}}{\binom{n}{t_1+1}} \right]. \quad (3)$$

It is observed from Figure 1 that γ_{crit} is a good approximation to the point at which a (t_1, t_2)-focused code begins to "match" the block decoding error probability of a $t_1 + t_2$-error correcting code; that is, for $\gamma < \gamma_{crit}$ the two codes perform equivalently.

A similar analysis can be performed if we assume that γ is held constant and ϵ is varied. Figure 2 shows P_d versus ϵ

for fixed $\gamma = 10^{-3}$, blocklength $n = 50$, and values of t_1 and t_2 such that $t_1 + t_2 = 4$. Analogous to the case described above, there is a critical value of ϵ – call it ϵ_{crit} – such that for $\epsilon > \epsilon_{crit}$ the block decoding error probability for a (t_1, t_2)-focused code is identical to that of a $t_1 + t_2$-error correcting code. Here, ϵ_{crit} is given by

$$\log_{10} \epsilon_{crit} = \frac{t_1+1}{t_2} \log_{10} \gamma + \frac{1}{t_2} \log_{10} \left[\frac{\binom{n}{t_1+1}}{\binom{n}{t_1+t_2+1}} \right]. \quad (4)$$

(Note: In obtaining equation (4) we have made the simplifying assumption that $\epsilon/(1-\epsilon) \approx \epsilon$.)

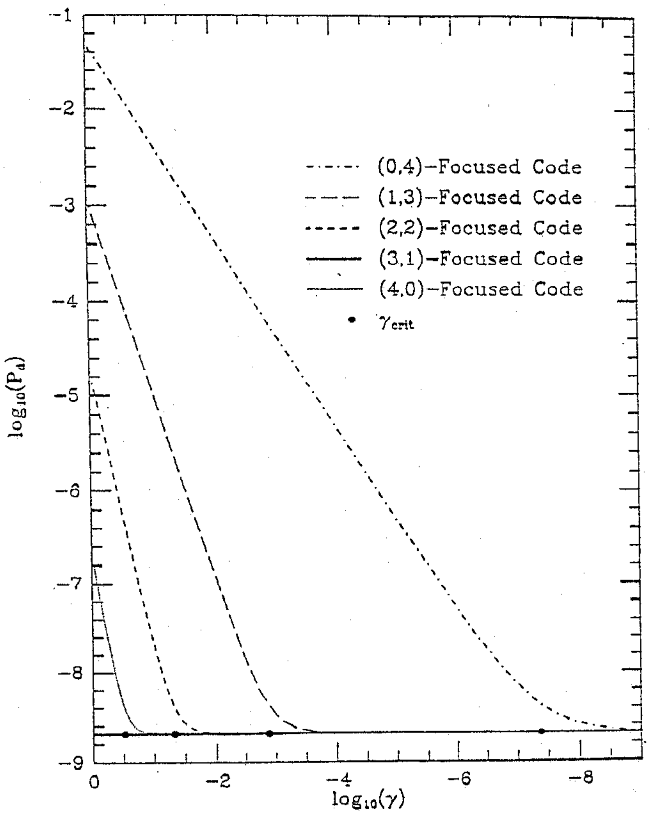

Figure 1: P_d versus γ for $t_1+t_2 = 4$, $n = 50$, and $\epsilon = 10^{-3}$.

Finally, we note that the above analysis implicitly assumes that ϵ and γ can be specified independently. As Section 4 shows, ϵ and γ are often both dependent on a third parameter – e.g., signal-to-noise ratio. If this is the case, we can guarantee "performance matching" by insuring that the second term in equation (2) is much larger than the first term; if we define a "benchmark" β by

$$\beta \triangleq \left[\frac{\gamma^{t_1+1}}{[\epsilon/(1-\epsilon)]^{t_2}} \right] \left[\frac{\binom{n}{t_1+1}}{\binom{n}{t_1+t_2+1}} \right], \quad (5)$$

then as long as $\beta \ll 1$ the block decoding error probability of a (t_1, t_2)-focused code will be the same as that of a t_1+t_2-error correcting code.

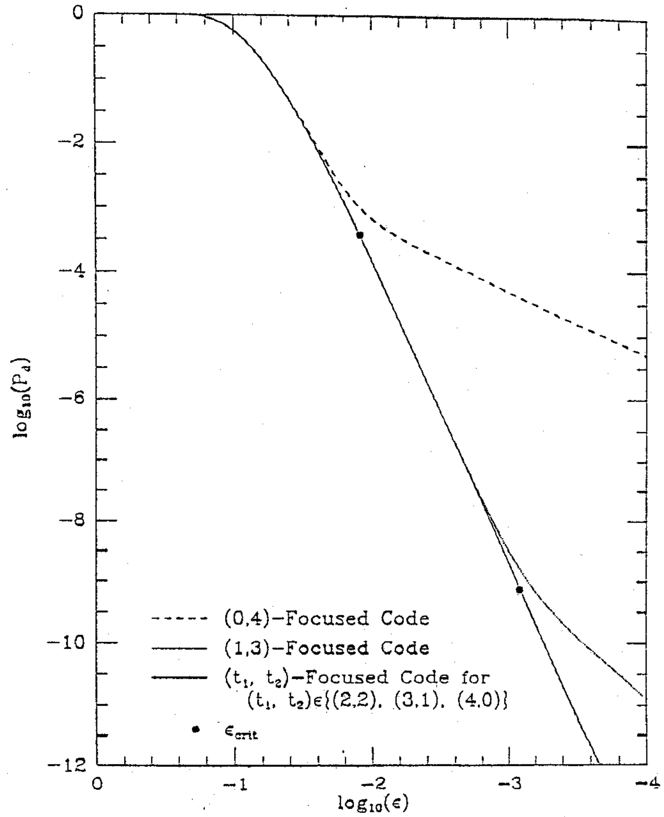

Figure 2: P_d versus ϵ for $t_1+t_2 = 4$, $n = 50$, and $\gamma = 10^{-3}$.

D. Symbol Error Probability Over a Skewed Symmetric Channel

We now compute the symbol error probability as defined in Section 3.2.

$$P_s = \frac{1}{n} \sum_{i=t_1+1}^{t_1+t_2} \sum_{j=t_1+1}^{i} \min(t_1+t_2+i, n) \binom{n}{i} \binom{i}{j} \cdot \epsilon^i (1-\epsilon)^{n-i} \gamma^j (1-\gamma)^{i-j} + \frac{1}{n} \sum_{i=t_1+t_2+1}^{n} \min(t_1+t_2+i, n) \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i}.$$

Note that for $n\epsilon \ll 1$ this can be approximated by

$$P_s \approx \frac{2t_1+t_2+1}{n} \binom{n}{t_1+1} \epsilon^{t_1+1} (1-\epsilon)^{n-t_1-1} \gamma^{t_1+1} + \frac{2t_1+2t_2+1}{n} \binom{n}{t_1+t_2+1} \cdot \epsilon^{t_1+t_2+1} (1-\epsilon)^{n-t_1-t_2-1}. \quad (6)$$

To guarantee that the *symbol* error probability of a (t_1, t_2)-focused code matches that of a t_1+t_2-error correcting code requires that the second term in equation (6) be dominant. Furthermore, note the similarity between equation (6) above and the formula for the block decoding error probability in equation (2); the two terms in (6) are the terms from (2) weighted by (slightly) different constants. For reasonable values of t_1 and t_2 these constants are close enough that matching the block decoding error probability of a (t_1, t_2)-focused code to that of a t_1+t_2-error correcting code is equivalent to doing the same for the symbol error probability.

E. Performance of "Combined" Focused Codes

Recall the technique for constructing "combined" (t_1, t_2)-focused codes from Section 2.2. Specifically, recall that the construction sometimes yields codes and decoding algorithms that are capable of correcting *more* errors than those indicated by the "(t_1, t_2)-focused" label – for instance, some error patterns that contain *more* than t_1 uncommon errors. (Thus this decoding algorithm will, for certain constructions, perform slightly better than the bounded distance decoder of Section 3.2.) In this section we show briefly how this enhanced capability can be taken into account when computing the performance of such a code.

Suppose that ℓ_1 common and ℓ_2 uncommon errors occur in a codeword during transmission. The decoding algorithm described in Section 2.2 will yield a correct estimate provided $\ell_1 \leq \lfloor (d_1-1)/2 \rfloor$ and $\ell_2 \leq \lfloor (d_2-\ell_1-1)/2 \rfloor$, where $d_1 = 2(t_1+t_2)+1$ and $d_2 = 2t_1+t_2+1$. Assume that such a combined code is used over a SSC with parameters ϵ and γ, and define the block decoding error P_d' to be the probability that $\ell_1 > \lfloor (d_1-1)/2 \rfloor$ and/or $\ell_2 > \lfloor (d_2-\ell_1-1)/2 \rfloor$ – i.e.,

$$P_d' = \sum_{\ell_1=0}^{\lfloor (d_1-1)/2 \rfloor} \sum_{\ell_2=\lfloor (d_2-\ell_1-1)/2 \rfloor+1}^{n-\ell_1} \binom{n}{\ell_1+\ell_2}\binom{\ell_1+\ell_2}{\ell_1} \cdot \epsilon^{\ell_1+\ell_2}(1-\epsilon)^{n-\ell_1-\ell_2}\gamma^{\ell_2}(1-\gamma)^{\ell_1} + \sum_{\ell_1=\lfloor (d_1-1)/2 \rfloor+1}^{n} \binom{n}{\ell_1}[\epsilon(1-\gamma)]^{\ell_1}[1-\epsilon(1-\gamma)]^{n-\ell_1}.$$

Similarly, the symbol error probability can be pessimistically approximated by

$$P_s' = \frac{1}{n}\sum_{\ell_1=0}^{\lfloor (d_1-1)/2 \rfloor} \sum_{\ell_2=\lfloor (d_2-\ell_1-1)/2 \rfloor+1}^{n-\ell_1} \min(\ell_1+\ell_2+\lfloor (d_2-\ell_1-1)/2 \rfloor, n) \cdot \binom{n}{\ell_1+\ell_2}\binom{\ell_1+\ell_2}{\ell_1}\epsilon^{\ell_1+\ell_2}(1-\epsilon)^{n-\ell_1-\ell_2}\gamma^{\ell_2}(1-\gamma)^{\ell_1} + \frac{1}{n}\sum_{\ell_1=\lfloor (d_1-1)/2 \rfloor+1}^{n} \sum_{\ell_2=0}^{n-\ell_1} \min(\ell_1+\ell_2+\lfloor (d_1-1)/2 \rfloor + \lfloor (d_2-\lfloor (d_1-1)/2 \rfloor-1)/2 \rfloor, n) \cdot \binom{n}{\ell_1+\ell_2}\binom{\ell_1+\ell_2}{\ell_1}\epsilon^{\ell_1+\ell_2}(1-\epsilon)^{n-\ell_1-\ell_2}\gamma^{\ell_2}(1-\gamma)^{\ell_1}. \quad (7)$$

The first double sum in (7) is due to uncorrectable errors in which the inner decoder is *not* overwhelmed; pessimistically, in such a case the ℓ_1 erasures passed onto the outer decoder would be decoded incorrectly, as would the ℓ_2 uncommon errors and $\lfloor (d_2-\ell_1-1)/2 \rfloor$ *additional* errors caused by the outer decoder, resulting in a total of $\ell_1+\ell_2+\lfloor (d_2-\ell_1-1)/2 \rfloor$ symbols decoded incorrectly. The second double sum is due to uncorrectable errors in which the inner decoder *is* overwhelmed. In addition to the ℓ_1 common and ℓ_2 uncommon errors that remain uncorrected, there can be as many as $\lfloor (d_1-1)/2 \rfloor$ symbols incorrectly identified as erasures by the inner decoder and $\lfloor (d_2-\lfloor (d_1-1)/2 \rfloor-1)/2 \rfloor$ additional symbols mistakenly "corrected" by the outer decoder.

IV. Performance of Focused Codes with Non-Binary Modulation Schemes

We now show how common non-binary modulation techniques can be approximated by the skewed symmetric channel with appropriate choice of parameters. Then the results from Section 3 will be used to analyze the performance of focused codes operating over an additive white Gaussian noise channel in conjunction with PSK and QAM modulation.

A. Parameters for PSK Modulation

If M-ary PSK modulation is used with a Gray code so that the difference between the binary representation of any two adjacent signals is one bit, then an additive Gaussian noise channel can be approximated by an M-ary SSC for the focus set **B** consisting of all elements of $GF(M)$ with a binary representation containing exactly one "1".

Let $\{(a_i, b_i) : i = 0, 1, \ldots, M-1\}$ denote the signals points in an M-ary PSK constellation with symbol energy E_s– i.e., $a_i^2 + b_i^2 = E_s$ and $b_i/a_i = \tan(2\pi i/M)$. (See Figure 3.) Thus every T seconds one of the signals in the set $\{s_i(t) = a_i\phi_1(t) + b_i\phi_2(t) : i = 0, 1, \ldots, M-1\}$ is transmitted. We assume further that the channel is zero-mean additive white Gaussian with power spectral density $S_z(f) = N_0/2$. Finally, it is assumed that hard-decisions are made at the demodulator, meaning that the received signal $r(t)$ is mapped onto the signal $s_i(t)$ that minimizes the Euclidean distance.

We now select ϵ and γ such that the resulting SSC describes the PSK modulation. Recall that ϵ is the probability of channel error while γ is the probability of an uncommon error, given that an error has occurred; in the PSK context, γ is the probability that the received signal lies outside of the decision regions *adjacent to* the one containing the transmitted signal, given that the received signal lies outside the region containing the transmitted signal. (See Figure 3.)

Thus ϵ is the probability that the received signal lies outside $s_i(t)$'s decision region, given $s_i(t)$ was transmitted. This is well known [2] to be approximated by

$$\epsilon \approx 2\,Q\left(\sqrt{2\,E_s/N_0}\,\sin\frac{\pi}{M}\right) \quad (8)$$

where $Q(y) \triangleq (1/\sqrt{2\pi})\int_y^{\infty} \exp(-t^2/2)dt$.

We now turn our attention to computing γ. For $M = 4$ this is easy: $P(\text{Uncommon Error}) = P(\sqrt{E_s/2} + n_1 <$

$0, \sqrt{E_s/2} + n_2 < 0)$ where n_1 and n_2 are independent zero-mean Gaussian random variables with variance $N_0/2$. Therefore $P(\text{Uncommon error}) = Q^2(\sqrt{E_s/N_0})$ and so

$$\gamma = \frac{P(\text{Uncommon error})}{P(\text{Error})} = \frac{Q^2\left(\sqrt{E_s/N_0}\right)}{2Q\left(\sqrt{E_s/N_0}\right)} = \frac{1}{2}Q\left(\sqrt{\frac{E_s}{N_0}}\right). \quad (9)$$

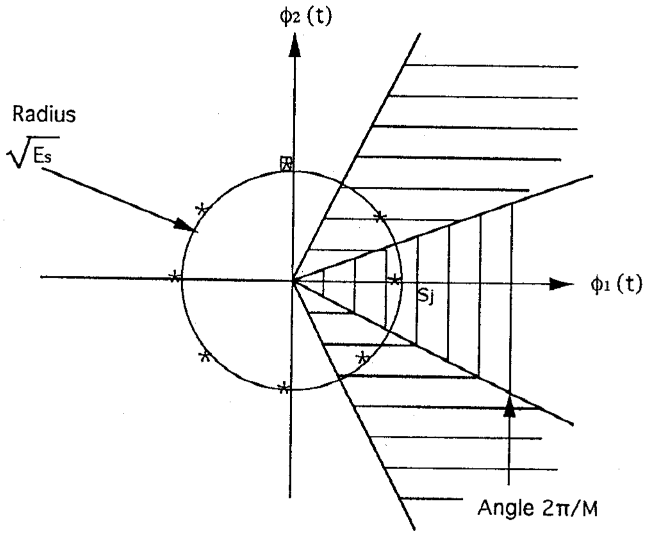

Figure 3: Decision regions corresponding to 8-PSK.

To obtain γ for $M \geq 8$ the following approximation for the pdf of the angle of the received signal [3] is useful:

$$f_\Theta(\theta) = \sqrt{E_s/(\pi N_0)} \cos\theta \exp[-(E_s/N_0)\sin^2\theta].$$

(This approximation is valid for $E_s/N_0 \gg 1$ and for small angle θ.) An uncommon error is made if the noise causes a phase displacement greater than $3\pi/M$ in absolute value. Using the above approximation for $f_\Theta(\theta)$, it can be shown that

$$P(\text{Uncommon error}) = 2Q\left(\sqrt{\frac{2E_s}{N_0}} \sin\frac{3\pi}{M}\right),$$

and so

$$\gamma = \frac{P(\text{Uncommon error})}{P(\text{Error})} = \frac{Q\left(\sqrt{2E_s/N_0}\,\sin(3\pi/M)\right)}{Q\left(\sqrt{2E_s/N_0}\,\sin(\pi/M)\right)}. \quad (10)$$

B. Parameters for a Square Constellation (QAM)

An additive white Gaussian noise (AWGN) channel and an M-ary square signal constellation can be approximated by an M-ary SSC channel with parameters ϵ and γ to be determined. For the square constellation, we assume that a two-dimensional Gray code is used so that the binary representation of two signals immediately adjacent to each other either horizontally or vertically differ in only one bit; thus, our set of common errors is (once again) the elements of GF(M) containing a single "1" in its binary representation. For example, when $M = 16$ as in Figure 4, if signal s_1 is sent then a common error occurs if the demodulator estimates the transmitted signal to be s_2, s_3, s_4 or s_5.

For a square ($\sqrt{M} \times \sqrt{M}$ where $\log_2 M$ is even) constellation of QAM signals, the coordinates of the signals with respect to the basis signals $\phi_1(t)$ and $\phi_2(t)$ are:

$$a_i = (2i + 1 - \sqrt{M})\frac{d}{2} \quad \text{and} \quad b_j = (2j + 1 - \sqrt{M})\frac{d}{2},$$

where i and j take on the values $0, 1, 2, \ldots, \sqrt{M} - 1$, and d is the constant horizontal or vertical distance between any two neighbors. The value of d is determined by the average symbol energy E_s and is given by $d = \sqrt{6E_s/(M-1)}$.

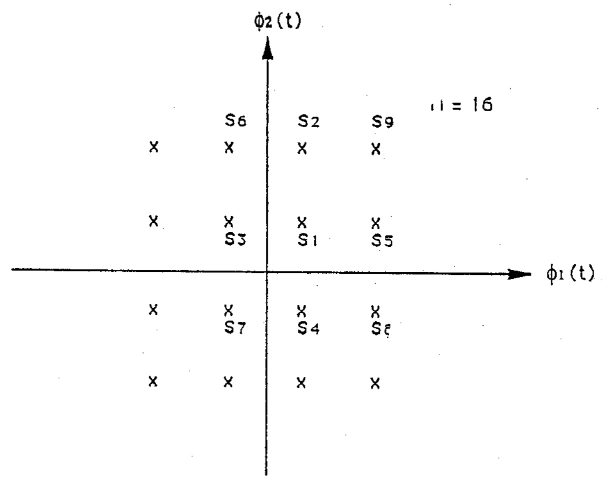

Figure 4: A 4×4 constellation displaying "common" and "uncommon" errors

In computing the parameters ϵ and γ a pessimistic approach will be used. It will be assumed that an error occurs whenever the received signal lies outside of a square with sides d in length centered on the transmitted signal; furthermore, an *uncommon* error occurs whenever the received signal lies outside of such a square *and* outside of the four squares immediately adjacent (horizontally and vertically) to it. (Such an assumption is pessimistic because the points on the *exterior* of the signal constellation will actually have lower probabilities of error than those indicated.)

It is well known [3] that ϵ – the probability of channel error – is given by

$$\epsilon = 2q - q^2 \tag{11}$$

where $q = 2Q(\sqrt{3E_s/N_0(M-1)})$.

Now consider the derivation of γ. Using the same pessimistic assumption used in deriving ϵ,

$$\begin{aligned} P(\text{Uncommon Error}) = \; & P(|n_1| > d/2, |n_2| > d/2) \\ & + P(|n_1| > 3d/2, |n_2| \le d/2) \\ & + P(|n_2| > 3d/2, |n_1| \le d/2), \end{aligned}$$

where n_1 and n_2 are independent zero-mean Gaussian random variables with variance $N_0/2$. If p_1 and p_2 are defined as

$$\begin{aligned} p_1 & \triangleq Q\left(d/\sqrt{2N_0}\right) = Q\left(\sqrt{\frac{3}{M-1}\frac{E_s}{N_0}}\right) \\ & = \frac{1}{2}P(|n_1| > d/2) \end{aligned}$$

and

$$\begin{aligned} p_2 & \triangleq Q\left(3d/\sqrt{2N_0}\right) = Q\left(3\sqrt{\frac{3}{M-1}\frac{E_s}{N_0}}\right) \\ & = \frac{1}{2}P(|n_1| > 3d/2) \end{aligned}$$

then

$$P(\text{Uncommon error}) = 4p_1^2 + 4p_2(1-2p_1).$$

This in turn implies

$$\gamma = \frac{p_1^2 + p_2(1-2p_1)}{p_1(1-p_1)}, \tag{12}$$

where p_1 and p_2 are as above.

C. Performance Matching for Focused Codes Used with PSK and QAM Modulation

Recall the question posed at the beginning of Section 3: Under what conditions does a (t_1, t_2)-focused code perform identically to a $t_1 + t_2$-error correcting code ? In Sections 3.1 and 3.2 this question was answered for a skewed symmetric channel; it was shown that the two have the same error rate as long as if $\beta \ll 1$, where β is defined in (5). Having shown in Sections 4.1 and 4.2 how M-ary PSK and QAM can be approximated by a SSC, we are now prepared to determine when a (t_1, t_2)-focused code operating in conjunction with these modulation schemes performs identically to a $t_1 + t_2$-error correcting code.

PSK: Equations (8)-(10) give the values of ϵ and γ that approximate M-ary PSK at a given signal-to-noise ratio; substituting these values into equation (5) and determining when $\beta \ll 1$ yields the following results.

- **M=8:** For octal PSK, a $(0, t)$-focused code performs identically to a t-error correcting code for $t = 1, 2, 3$ for all values of E_s/N_0 and all blocklengths $n \ge 7$.
- **M=16:** For 16-ary PSK, a $(0, t)$-focused code performs identically to a t-error correcting code for $t = 1, 2, 3, 4, 5, 6$ for all values of E_s/N_0 and all blocklengths $n \ge 10$.

These results indicate that a code capable of correcting t adjacent-region errors will perform identically to a code capable of correcting *any* t errors for many blocklengths and many values of t. It is interesting to compare the "focused approach" to PSK modulation with that taken by Lee-metric codes. A decoder for a t-error correcting Lee-metric code will correctly estimate the transmitted codeword provided that the total Lee distance between what is transmitted and what is received is no more than t; a received signal that is i regions away from the transmitted signal contributes a value i to the Lee distance. Thus, a 2-error correcting Lee-metric code, when used with M-ary PSK, can correct any two adjacent-region errors *and* it can correct any single error where the received symbol is *two* regions away from what was transmitted. The above results suggest that the added capability of Lee-metric codes – the ability to correct a (reduced) number of non-adjacent errors – often provides negligible performance improvement to PSK modulation.

Square Signal Sets: Equations (11) and (12) give the values of ϵ and γ that approximate M-ary QAM at a given signal-to-noise ratio; substituting these values into equation (5) and determining when $\beta \ll 1$ yields the following results.

- **M=64:** For an 8×8 constellation at all values of E_s/N_0 and all blocklengths $n \ge 7$:
  - A $(0, 1)$-focused code performs identically to a 1-error correcting code.
  - A $(1, 1)$-focused code performs identically to a 2-error correcting code.
  - A $(1, 2)$-focused code performs identically to a 3-error correcting code.
  - A $(2, 2)$-focused code performs identically to a 4-error correcting code.
  - A $(2, 3)$-focused code performs identically to a 5-error correcting code.
- **M=256:** For a 16×16 constellation at all values of E_s/N_0 and all blocklengths $n \ge 8$:
  - A $(0, 1)$-focused code performs identically to a 1-error correcting code.
  - A $(1, 1)$-focused code performs identically to a 2-error correcting code.
  - A $(1, 2)$-focused code performs identically to a 3-error correcting code.
  - A $(2, 2)$-focused code performs identically to a 4-error correcting code.

- A $(2,3)$-focused code performs identically to a 5-error correcting code.

Regarding Simulations: The results presented in this section maintain that a (t_1, t_2) focused code performs as well as a $t_1 + t_2$-error correcting code for various values of t_1 and t_2 and various signaling schemes. These results were derived with the analytical approximations from earlier sections; furthermore they have been supported by simulations. We now give a very brief description of the simulations carried out.

The noise afflicting a particular symbol in a codeword consists of two simulated independent zero-mean Gaussian random variables, each with variance $N_0/2$. If the noise afflicting a particular symbol is sufficient to "knock" the received signal into another decision region, an error is registered; if the noise is sufficient to knock the received symbol a *non-adjacent* decision region, then an uncommon error is registered. If more than $t_1 + t_2$ errors and/or more than t_1 uncommon errors occur in a codeword of n symbols, then it is assumed that a decoder error has occurred. By simulating many such codewords we arrive at an estimate of P_d.

Two things to note: First, for PSK the probability of an error of any type is independent of the signal transmitted; for QAM we made the pessimistic assumption that an interior point is always transmitted. Second, this simulation was capable only of estimating P_d – *not* P_s, since we did not try to implement decoders in software.

Nonetheless, the simulations supported the analytical claims. For instance, in simulating 10^5 7-tuples of octal PSK symbols over a channel with $E_s/N_0 = 11$ dB, we found that the fraction of errors that could not be corrected by a $(0,2)$-focused code was $4.96 \cdot 10^{-3}$; the fraction of 7-tuples containing errors that could not be corrected by a 2-error correcting code was $4.90 \cdot 10^{-3}$. Both of these figures closely approximate the number that equation (2) gives for the probability of block decoding error for this channel when used with a $(0,2)$-focused code – namely, $P_d = 4.91 \cdot 10^{-3}$. This and all the other simulations that were run support the conclusion that PSK and QAM modulation over an AWGN channel is well approximated by a skewed symmetric channel with parameters chosen appropriately.

D. Coding Gain of Focused Codes Used with PSK and QAM Modulations

In this section we compare the performance of some (t_1, t_2)-focused codes with that of $t_1 + t_2$-error correcting codes. The focused codes are constructed using the technique in Section 2.2, and the channel is additive white Gaussian noise with PSK and QAM signaling.

Our approach is to compare a (t_1, t_2)-focused code with a $t_1 + t_2$-error correcting code such that the two codes have identical probability of error; because the focused code can have a higher code rate than the "traditional" code, it will enjoy some coding gain. For example, Section 4.3 shows that a $(0,1)$-focused code over GF(8) with blocklength $n = 7$ will perform identically to a single-error correcting code of the same blocklength when used with octal PSK; however, it's possible to construct a $(0,1)$-focused code at a rate of $16/21 \approx 0.762$, whereas a single-error correcting Reed-Solomon code over GF(8) with blocklength $n = 7$ has rate $5/7 \approx 0.714$. The 6.7% rate improvement of the focused code translates into a constant coding gain of 0.28 dB.

Equation (6) gives the symbol error rate for a (t_1, t_2)-focused code operating over a skewed symmetric channel with parameters ϵ and γ; equation (7) gives the same when the code is one of the "combined" codes from Section 2.2. Equations (8), (9), and (10) give ϵ and γ for an AWGN channel employing PSK modulation; equations (11) and (12) give ϵ and γ for an AWGN channel when the modulation is QAM with a square signal set. Thus, for a given signal-to-noise ratio we can use equations (8)-(12) to compute ϵ and γ and then use (6)-(7) to yield P_s. To make a fair comparison between codes with different rates, P_s is computed as a function of E_b/N_0, where E_b is the energy per information bit – i.e., if R is the code rate, $E_b = E_s/(R \log_2 M)$.

Figure 5 shows P_s versus E_b/N_0 for 16-ary PSK used with various coding schemes. The solid line shows P_s for uncoded 16-ary PSK, while the dotted line shows P_s for an (8,4) two-error correcting shortened Reed-Solomon code over GF(16). The two dashed lines display the performance of a $(0,2)$-focused code over GF(16) with blocklength $n = 8$; the short-dashed line shows P_s as computed by equation (6), while the long-dashed line uses equation (7). (Since the code can correct any single uncommon error – a fact that equation (7) takes into account and (6) does *not* – we find that equation (7) shows the performance to be slightly improved over that suggested by (6).)

Figure 5 shows that, at a symbol error rate of $P_s = 10^{-6}$, the focused code provides approximately 1.13 dB of coding gain *above* that of the Reed-Solomon code; comparing the focused code with uncoded 16-ary PSK, we find a coding gain of approximately 2.48 dB.

Similarly, Figure 6 compares a blocklength $n = 11$ $(1,2)$-focused code with a 3-error correcting shortened Reed-Solomon code when used with an 8×8 square signal constellation. In this case the performance given by equations (6) and (7) were identical; the focused code provides a constant 0.91 dB of coding gain over the Reed-Solomon code and provides 2.38 dB of gain above uncoded 64-QAM at $P_s = 10^{-6}$.

In each of the two figures, the focused code is constructed according to Section 2.2 with the inner code being the highest-known-rate [4] binary code with minimum distance $d_{min} = 2(t_1 + t_2) + 1$ and the outer code being a shortened Reed-Solomon code. The gains are significant primarily for codes with blocklengths that are short relative to the field size; for instance, if the blocklength goes above $n = 20$ for 64-QAM, then the focused code construction of Section 2.2 provides a coding gain that is no more than 0.25 dB better than a shortened Reed-Solomon code. It would appear that this limitation is more a function of the particular construction than of focused codes in general.

References

[1] Tom Fuja and Chris Heegard, "Focused Codes for Channels with Skewed Errors", *IEEE Transactions on Information Theory,* Vol. 36, No. 9, pp.773-783, July 1990.

[2] Bernard Sklar, *Digital Communications,* Prentice Hall, 1988.

[3] S. Benedetto, E. Biglieri, V. Castellani, *Digital Transmission Theory,* Prentice Hall, 1987.

[4] T. Verhoff, "An Updated Table of Minimum-Distance Bounds for Binary Linear Codes", *IEEE Transactions on Information Theory,* Vol. 33, No.5, pp. 665-680, September 1987.

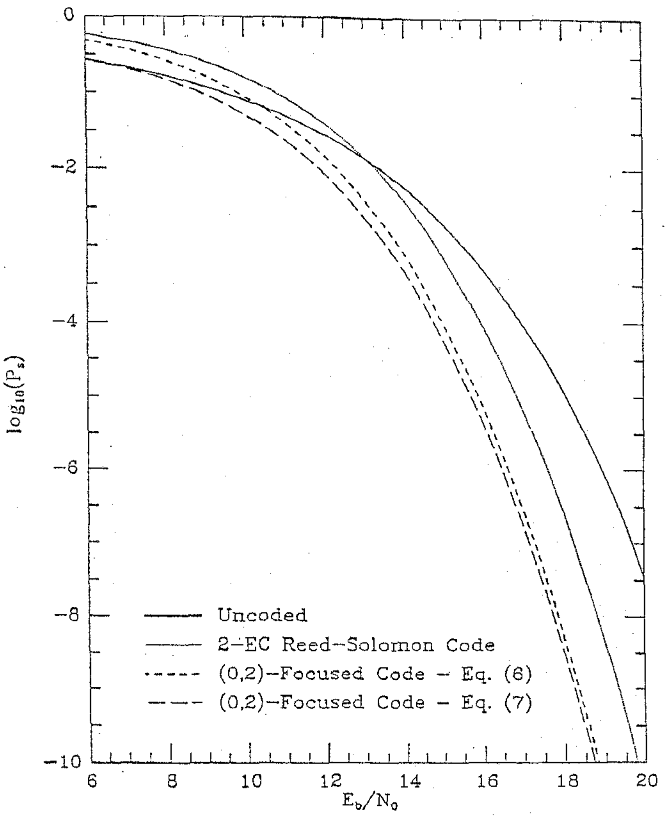

Figure 5: P_s for 16-ary PSK – uncoded, Reed Solomon, and focused code. (Blocklength $n = 8$ for both codes.)

Fady Alajaji (S'90) was born on May 1, 1966 in Beirut, Lebanon. He received the B.E. degree (with Distinction) in electrical engineering from the American University of Beirut, Lebanon, in 1988. In 1990, he received the M.S. degree in electrical engineering from the University of Maryland, College Park, MD.

From January 1990 to January 1992 he held a student Fellowship at the Institute for Systems Research (ISR), University of Maryland. During his graduate studies, he held Research Assistant and Teaching Assistant positions at the ISR and Electrical Engineering Department. He is the recepient of a Dissertation Fellowship from the Graduate School at the University of Maryland, for the academic year 93-94. He is currently working towards the completion of the Ph.D. degree in electrical engineering at the University of Maryland.

His research interests are in the areas of information theory, error control coding, combined source-channel coding and digital communication. He is a member of Phi Kappa Phi honor society.

Tom Fuja was born on August 15, 1959 in Durand, Michigan. He received his undergraduate education at the University of Michigan, graduating with a B.S.E.E. and a B.S.Comp.E. in 1981. He subsequently attended Cornell University, receiving the M.Eng. and the Ph.D. degrees in electrical engineering in 1983 and 1987, respectively.

In 1982 and 1983 Fuja worked as a Member of the Technical Staff at AT&T Bell Laboratories in Holmdel, NJ. In 1987 he joined the faculty of the University of Maryland, where he holds a joint appointment as an Associate Professor in the Department of Electrical Engineering and the Institute for Systems Research.

Fuja's research interests lie in the area of digital communication systems – specifically coding theory and applications and information theory. He is also interested in applying the techniques of these disciplines to the problem of data storage and recording.

In 1989 Dr. Fuja was the recipient of a Presidential Young Investigator award from the National Science Foundation; in 1991 he received the George Corcoran Memorial Award for his significant contribution to electrical engineering education at the University of Maryland at College Park. He has been active in the IEEE Information Theory Society, serving as its secretary in 1988 and its treasurer 1989-94.

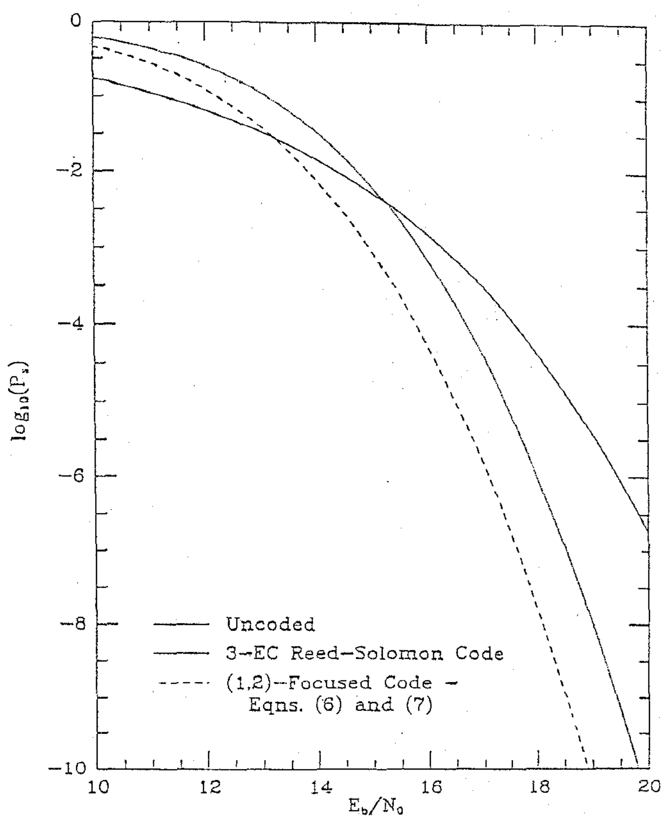

Figure 6: P_s for 8×8 QAM – uncoded, Reed Solomon, and focused code. (Blocklength $n = 11$ for both codes.)