Algebraic Geometric Codes

M. Shahshahani
Communications Systems Research Section

This article discusses the performance characteristics of certain algebraic geometric codes. Algebraic geometric codes have good minimum distance properties. On many channels they outperform other comparable block codes; therefore, one would expect them eventually to replace some of the block codes used in communications systems. This article suggests that it is unlikely that they will become useful substitutes for the Reed-Solomon codes used by the DSN in the near future. However, they may be applicable to systems where the signal-to-noise ratio is sufficiently high so that block codes would be more suitable than convolutional or concatenated codes.

I. Introduction

In their 1982 paper [1], Tsfasman, Vladut, and Zink showed that, by using algebraic curves, one can construct codes that lie above the Varshamov-Gilbert bound. These codes perform better than other comparable block codes on many channels. This important discovery led to a resurgence of interest in geometric, or Goppa, codes. Most of the research in recent years has focused on developing practical decoding algorithms. For an account of these efforts, refer to [2,3]. This article discusses some of the performance characteristics of certain algebraic geometric codes. For the reader's convenience, some of the basic ideas are developed in Section II as an introduction to the more technical papers in this field.

II. Definition and Basic Properties

Let $\mathbf{F}_q$ be the field of q elements ($q = p^s$, p prime) and $\mathbf{F}_q{}^\times$ be the set of nonzero elements of $\mathbf{F}_q$. To construct the projective space P^r over $\mathbf{F}_q$, let $V = V_{r+1}$ be the vector space of $(r+1)$-tuples of elements of $\mathbf{F}_q$, and $V^* = V \backslash 0$ that is, V with the origin removed). On V^* define the equivalence relation

$$[x_o, \ldots, x_r] \sim [y_o, \ldots, y_r]$$

if $y_j = \lambda x_j$, for all j and some $\lambda \in \mathbf{F}_q{}^\times$. The quotient space $V^*/\sim$, where equivalent elements are identified, is the projective space $\mathbf{P}^r$ over $\mathbf{F}_q$. A point $x \in \mathbf{P}^r$ is represented by $x = [x_o, \ldots, x_r]$ (of course, not uniquely), and x_i's are called the homogeneous coordinates of x. For many geometric problems this space is more convenient to use than the Euclidean space $\mathbf{F}_q{}^r$. There is a way of translating statements about subsets of Euclidean space into those of the projective space, which is described next.

First, embed $\mathbf{F}_q{}^r$ into V^* by

$$i(x_1, \ldots, x_r) = [1, x_1, \ldots, x_r]$$

Then, to a subset S of $\mathbf{F}_q{}^r$, assign the image $\mathbf{S}$ of $i(S)$ in P^r. This means that, in terms of the homogeneous coordinates, i.e., in V^*, $\mathbf{S}$ is represented by the cone

$$S^* = \{[\lambda, \lambda x_1, \ldots, \lambda x_r] \mid \lambda \in \mathbf{F}_q^{\times}, \text{and } [x_1, \ldots, x_r] \in S\}$$

Thus, statements about the subsets of Euclidean space can be translated into statements about those of the projective space. In particular, notice that points in Euclidean space become rays through the origin and that lines in Euclidean space become two-dimensional planes through the origin in V^*.

To see the value of this translation, consider the special case $r = 2$, i.e., the projective plane. P^2 may be regarded as ${\mathbf{F}_q}^2$ with a line and a point at infinity added to it. The line is the set $\{[0,1,x_2]\}$ and the point is [0,0,1]. Let S and T be two parallel lines in the plane ${\mathbf{F}_q}^2$, as defined by the equations

$$S : ax_1 + bx_2 = c \quad \text{and} \quad T : ax_1 + bx_2 = c'$$

Then, the intersection of S^* and T^* in V^* is the line defined by

$$ax_1 + bx_2 = 0 \quad \text{and} \quad x_o = 0$$

Thus, S and T intersect at the point $[0,1,-a/b]$ of P^2. So an important difference between Euclidean space and the projective space is that in the latter space, all lines intersect. Since many problems in geometry can be reduced to problems of intersections, it makes more sense to work in the projective space and avoid the exceptional case of nonintersecting or parallel lines.

Let $f(x_1, x_2) = \sum a_{ij} {x_1}^i {x_2}^j$ be a polynomial in two variables of degree e. Then, homogenize this polynomial by adding the variable x_o and considering the homogeneous polynomial

$$F(x_o, x_1, x_2) = \sum a_{ij} {x_o}^{e-i-j} {x_1}^i {x_2}^j$$

For the set $Z(f) = \{(x_1, x_2) | f(x_1, x_2) = 0\}$, the procedure of going from Euclidean space to the projective space (or to V^*) amounts to going from $Z(f)$ to $Z(F) = \{[x_o, x_1, x_2] | F(x_o, x_1, x_2) = 0\}$. It is convenient for this application to consider only polynomials F satisfying a certain technical property (called nonsingularity) that will be described at the end of this section.

Let F and G be homogeneous polynomials of degrees e and m in three variables. For a subset S, denote its cardinality by $|S|$. Then $|Z(F) \cap Z(G)|$ is bounded by em. It is actually equal to em if the intersections are counted with multiplicities (for example, tangency has multiplicity two, etc.) and allow points to have coordinates in the algebraic closure of $\mathbf{F}_q$. These more technical points will not be discussed here. Note, however, that if F or G is a product of linear polynomials, then the assertion that $|Z(F) \cap Z(G)|$ is bounded by em follows from the fundamental theorem of algebra.

Now assume that a linear space L (over $\mathbf{F}_q$) of functions on the subset $Z(F)$ of P^2 and a subset $S = \{\xi_1, \ldots, \xi_n\}$ of $Z(F)$ are specified. Consider the mapping

$$\mu : L \rightarrow {\mathbf{F}_q}^n \quad \text{where} \quad \mu(G) = \big(G(\xi_1), \ldots, G(\xi_n)\big)$$

Then, the image of μ is a linear subspace of ${\mathbf{F}_q}^n$, and is, therefore, a code. In order to analyze this code, some control over the linear space L must be exercised. Here algebraic geometry provides "naturally" defined linear spaces L, and the parameters of the corresponding code may be evaluated. Note also that certain Reed-Solomon codes may be defined in a similar manner. In fact, if L is the space of all polynomials of degree less than k and $\mathbf{F}_q = \{\xi_1, \ldots, \xi_q\}$, then an extended Reed-Solomon code is the image of the map $\mu : L \rightarrow {\mathbf{F}_q}^q$, where $\mu(f) = \big(f(\xi_1), \ldots, f(\xi_q)\big)$. This code has parameters $(q, k, q - k + 1)$ and is a maximum distance separable code.

Let $\mathbf{R} = \mathbf{F}_q[x_o, x_1, x_2]$ be the vector space of polynomials in three variables with coefficients in $\mathbf{F}_q$, and let $\mathbf{R}_m$ be the subspace spanned by the homogeneous polynomials of degree m. In this case (i.e., $r = 2$), $Z(F)$ is called a plane curve. It is necessary to construct a linear space of functions on $Z(F)$ from $\mathbf{R}_m$. Notice that for $G \in \mathbf{R}_m$

$$G(\lambda x_o, \lambda x_2, \lambda x_2) = \lambda^m G(x_o, x_2, x_2)$$

so that G is not a well-defined function on $Z(F)$ or P^2. There are two ways of avoiding this difficulty:

(1) Define the value of G at a point x of $Z(F)$ or P^2 to be $G(x_o, x_2, x_2)$ where the representative $[x_o, x_2, x_2]$ of x is selected so its first nonzero coordinate is 1.

(2) Fix a homogeneous polynomial H of degree m with the property that $H(\xi_i) \neq 0$ for all $\xi_i \in S$. Then G/H is a well-defined function on $Z(F)$.

With either alternative, $\mathbf{R}_m$ may be regarded as a linear space L_m of functions on $Z(F)$, and, therefore, the code is denoted by $C(F, S)$. One can determine the parameters (n, k, d) of this code under some additional hypotheses.

Two polynomials G and $G' \in \mathbf{R}_m$ determine the same function on $Z(F)$, i.e., the same element of L_m if, and only if, their difference is a multiple of F. Assuming that the degree e of F is less than m, the dimension of L_m is expected to be

$$\dim (L_m) = \dim (\mathbf{R}_m) - \dim (\mathbf{R}_{m-e})$$

since multiplication by F maps $\mathbf{R}_{m-e}$ into $\mathbf{R}_m$. It is easy to see that $\dim (\mathbf{R}_m) = (1/2)(m+1)(m+2)$. Substituting and simplifying yields

$$\dim (L_m) = c - g + 1 \tag{1}$$

where the quantities $c = em$ and $g = (1/2)(e-1)(e-2)$ are called the degree of L_m and the genus of the plane curve $Z(F)$, respectively. Formula (1) is a very special case of the celebrated Riemann-Roch theorem. The above discussion should take some of the mystery out of this useful formula.

Next, assume that $n > em$. To determine the parameters of the code $C(F,S)$, suppose that $\mu(G) = 0$, then the intersection $Z(F) \cap Z(G)$ has at least $n > em$ points. But since G has degree m and F has degree e, $|Z(F) \cap Z(G)| \leq em$. Therefore, $G = 0$ and the map μ is one to one. This implies that the code $C(F,S)$ has rate

$$\rho = k/n = \{em - (1/2)(e-1)(e-2) + 1\}/n$$

The minimum distance d of the code is the minimum number of nonzero entries of $(G(\xi_1), \ldots, G(\xi_n))$ as G ranges over the nonzero elements of L_m. As noted, $|Z(F) \cap Z(G)|$ does not exceed em, and therefore

$$d \geq n - em$$

Having defined the code $C(F,S)$, it is natural to try to understand its dual code $C(F,S)^*$ with parameters (n, k^*, d^*). The computation of the parameters of $C(F,S)^*$ involves introducing more algebraic geometry, and will not be discussed in detail. The result is:

$$k^* = n - c + g - 1 = n - k \text{ and } d^* \geq c - 2g + 2$$

Here, only a restricted class of algebraic geometric codes was considered. While there are more general constructions, the special case considered will suffice for the problems at hand.

Finally, the nonsingularity property mentioned earlier must be clarified. For a homogeneous polynomial F, $Z(F)$ is nonsingular if, for every i, the equations

$$\{F_i = 0 \text{ and } \partial F_i/\partial x_j = 0 \text{ for } j \neq i\}$$

do not have a solution. Here, F_i is the polynomial obtained from F by setting $x_i = 1$. Since for each i, this is a set of three equations in two unknowns, the nonsingularity condition is generically satisfied. For example, if the polynomial F satisfies this condition, then it cannot be written in the form $F = GH$ with G and H homogeneous polynomials of positive degree. In fact, the nonsingularity condition will not be satisfied, since $F_i = 0$ and $\partial F_i/\partial x_j = 0$ on $Z(F) \cap Z(G)$. On the other hand, $Z(F)$, where $F(x_o, x_1, x_2) = {x_o}^t + {x_1}^t + {x_2}^t$, is nonsingular for those values of t and p which are relatively prime.

III. Construction and Performance of Certain Codes

It is clear from the expressions for k and d that, to construct "good" codes, one should find polynomials with $|Z(F)|$ as large as possible, so that d can be large, while the symbol size is fairly small. [Recall that $|Z(F)|$ means cardinality of $Z(F)$ as a subset of P^2 and not V^*.] There is an important inequality (the Weil-Serre bound) relating the size of $|Z(F)|$ to e, namely,

$$|Z(F)| \leq q + 1 + g[2\sqrt{q}] \tag{2}$$

where $[y]$ denotes the largest integer not exceeding y. This bound is sharp in the sense that there are plane curves for which the equality in Formula (2) is achieved. In order to understand the basic properties of the algebraic geometric code $C(F,S)$, compute $|Z(F)|$. In the following example, $|Z(F)|$ is computed for a class of homogeneous polynomials in three variables:

Example: Let $q = p^{ab}$ and $r = p^a$, then $q-1 = (r-1)t$ for some integer t. Consider the Fermat curve defined by

$$F(x_o, x_1, x_2) = {x_o}^t + {x_1}^t + {x_2}^t \tag{3}$$

For such F

$$|Z(F)| = 3t + (r-2)t^2$$

Consider the mapping $\chi(\zeta) = \zeta^t$, which is a homomorphism of ${\mathbf{F}_q}^\times$ into itself. Since $\zeta^{tr} = \zeta^{q-1+t} = \zeta^t$, ζ^t lies in ${\mathbf{F}_r}^\times$. (Note $\mathbf{F}_q \supset \mathbf{F}_r$.) Therefore, χ is actually onto ${\mathbf{F}_r}^\times$ and its kernel has order t. First, consider solutions to $F = 0$ with $x_o = 0$, then one may assume $x_1 = 1$. It follows that there are exactly $3t$ solutions with exactly one coordinate zero. Next, set $x_o = 1$, and seek solutions where all the coordinates are nonzero. For $-1 \neq \alpha \in {\mathbf{F}_r}^\times$, the equation $\chi(\zeta) = -\alpha - 1$ has t solutions in ${\mathbf{F}_q}^\times$. Since $\eta^t = \alpha$ also has t solutions, $(r-2)t^2$ solutions were obtained with all the coordinates nonzero. Hence, there are $3t + (r-2)t^2$ solutions to $F = 0$ in $\mathbf{F}_q$. For $a = b$ and $t = r + 1$, $|Z(F)| = r^3 + 1$. Since $g = r(r-1)/2$, and the

right-hand side of Formula (2) is also r^3+1, the inequality of Formula (2) is sharp in the sense described earlier.

This example will suffice for investigating some of the properties of algebraic geometric codes. The general phenomenon is that "good" algebraic geometric codes, comparable in rate and word size to Reed-Solomon codes, have larger minimum distance and smaller symbol size. For example, setting $a = b$ and $p = 2$ in the example above, one obtains algebraic geometric codes of rate ρ with minimum distance d and word length L ($= n \times$ symbol size) given by (approximately)

$$d \approx (1-\rho)2^{3a} \text{ and } L \approx (2a)2^{3a}$$

In fact, note that $em \gg g$, $n \approx |Z(F)| \approx 2^{3a}$, and $d \approx (1-\rho)n$ to obtain the above estimates. For a Reed-Solomon code of rate ρ

$$d' \approx (1-\rho)2^N \text{ and } L' \approx n2^N$$

from which the claim follows.

To estimate the parameters of several more specific codes, one considers $C(F,S)$ where $F(x_o,x_1,x_2)$ is as in the above example, $p = 2$, $a = b = 3$, and $|Z(F)| = 513$. Consider the following codes:

(1) $C(1)$: $n = 504$, $e = 9$, and $m = 17$, so that $k/n = 1/4$ and $d = 351$

(2) $C(2)$: $n = 504$, $e = 9$, and $m = 31$, so that $k/n = 1/2$ and $d = 225$

(3) $C(3)$: $n = 504$, $e = 9$, and $m = 52$, so that $k/n = 7/8$ and $d = 36$

The performance of these codes has been studied and compared with that of certain Reed-Solomon codes. Assume that the communication channel is a binary symmetric one, which models a binary-phase shift keyed (BPSK) modulation system over an additive white Gaussian noise channel with a hard limiter (hard decision). It is well-known that a good approximation to the output bit error probability is

$$P \approx (p/s)\sum_{j=u+1,\ldots,n} C(n,j)s^j(1-s)^{n-j}$$

where $n = 504$ for the codes $C(1)$, $C(2)$, and $C(3)$, $C(n,j)$ is the binomial coefficient n choose j, $u = (1+d)/2$, and $s = 1-(1-p)^l$ with l the symbol size. Recall also that the information bit signal-to-noise ratio E_b/N_0 is related to the bit error probability p by the formula

$$p = (1/2)Erfc\left[(kE_b/nN_0)^{1/2}\right]$$

where

$$Erfc(x) = (2/\pi)\int_x^\infty e^{-t^2}dt$$

The performance of $C(1)$ and $C(2)$ is compared with Reed-Solomon codes $RS(511,127)$ and $RS(511,255)$, respectively. Note that the comparison is between codes of similar rates, but that the symbols for the Reed-Solomon codes are longer—4599 bits, as compared with 3024 bits for $C(1)$ and $C(2)$. While $C(1)$ performs somewhat better than $RS(511,127)$ (about 1/3 dB better), $C(2)$ and $RS(511,255)$ perform almost identically. The algebraic geometric code $C(3)$ was compared with the Reed-Solomon code $RS(255,223)$, which is part of the concatenated code used for the Galileo spacecraft. Here again, the performance of the codes is very close, but the Reed-Solomon code has shorter symbol length. Also, note that optimal decoding was assumed in these comparisons. The results are given in Figs. 1–3.

The output bit error was also computed as a function of the channel symbol error probability for $C(1)$. The results are given in Fig. 4.

Practical encoding procedures for algebraic geometric codes are not known at this time, and their decoding is more difficult than that of Reed-Solomon codes. While much progress has been made on the decoding of these codes, they cannot always be optimally decoded with algorithms having acceptable complexity (see [2,3]).

IV. Conclusion

Algebraic geometric codes are "good" block codes. They are the first codes to beat the Varshamov-Gilbert bound, and on many channels outperform other comparable block codes. However, their performance characteristics are such that they are unlikely to be useful for the DSN in the near future. Their most likely application is to systems where the signal-to-noise ratio is sufficiently high so that block codes would be generally more suitable than trellis, convolutional, or concatenated code systems. Like other block codes, algebraic geometric codes can only be hard decoded at this time. When their soft decoding becomes possible, these conclusions will have to be revised.

Acknowledgments

The author wishes to thank D. Divsalar, S. Dolinar, and F. Pollara for helpful conversations on coding theory. A memo (I.O.M. 331-90.2-042) by Cheung, Dolinar, and Pollara comparing the performance of a rate 1/4 algebraic geometric code with certain Reed-Solomon codes was made available to the author.

References

[1] M. A. Tsfasman, S. C. Vladut, and T. Zink, "Modular Curves, Shimura Curves, and Goppa Codes Better Than Varshamov-Gilbert Bound," *Math. Nachr.*, vol. 109, pp. 21–28, 1982.

[2] J. Justensen, K. J. Larsen, H. E. Jensen, A. Havemose, and T. Hoholdt, "Construction and Decoding of a Class of Algebraic Geometric Codes," *IEEE Transactions on Information Theory*, vol. 35, no. 4, pp. 811–821, July 1989.

[3] A. N. Skorobogatov and S. G. Vladut, "On the Decoding of Algebraic Geometric Codes," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 1051–1060, September 1990.

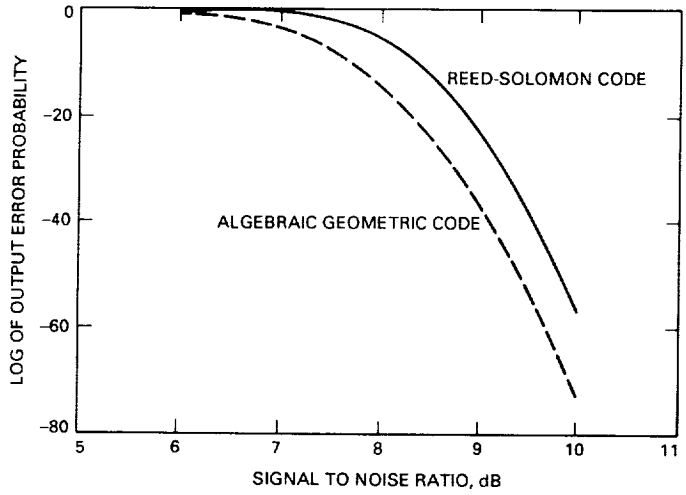

Fig. 1. Comparison of algebraic geometric code C (1) and Reed-Solomon code (511, 127).

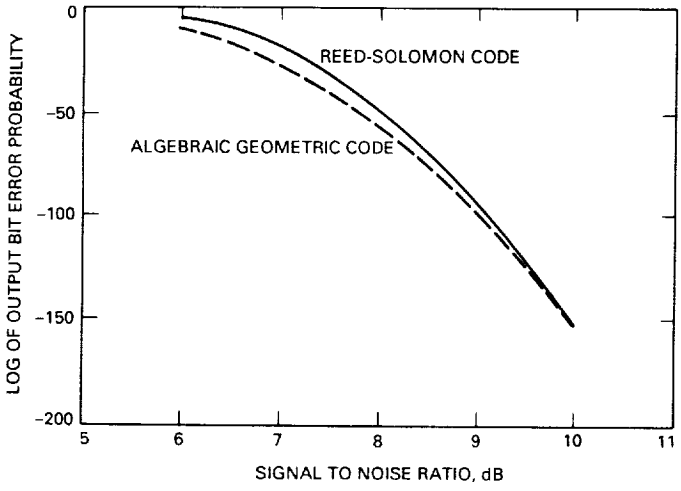

Fig. 2. Comparison of algebraic geometric code C (2) and Reed-Solomon code (511, 255).

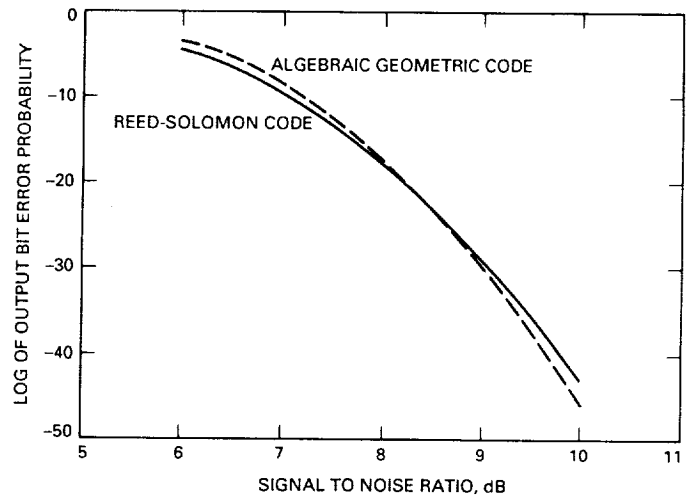

Fig. 3. Comparison of algebraic geometric code C (3) and Reed-Solomon code (255, 223).

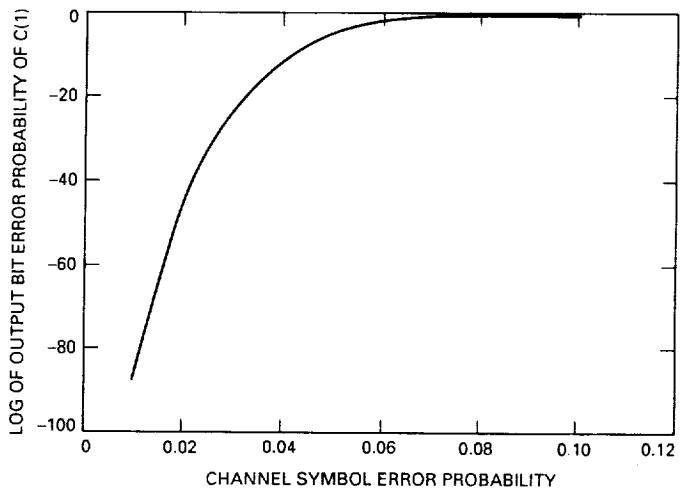

Fig. 4. Output of C (1).