The Art of Signaling: Fifty Years of Coding Theory

A. R. Calderbank, *Fellow, IEEE*

(Invited Paper)

***Abstract*— In 1948 Shannon developed fundamental limits on the efficiency of communication over noisy channels. The coding theorem asserts that there are block codes with code rates arbitrarily close to channel capacity and probabilities of error arbitrarily close to zero. Fifty years later, codes for the Gaussian channel have been discovered that come close to these fundamental limits. There is now a substantial algebraic theory of error-correcting codes with as many connections to mathematics as to engineering practice, and the last 20 years have seen the construction of algebraic-geometry codes that can be encoded and decoded in polynomial time, and that beat the Gilbert–Varshamov bound. Given the size of coding theory as a subject, this review is of necessity a personal perspective, and the focus is reliable communication, and not source coding or cryptography. The emphasis is on connecting coding theories for Hamming and Euclidean space and on future challenges, specifically in data networking, wireless communication, and quantum information theory.**

***Index Terms*—Algebraic, information and coding theory, quantum and space–time codes, trellis.**

I. A Brief Prehistory

BEFORE Shannon [187] it was commonly believed that the only way of achieving arbitrarily small probability of error on a communications channel was to reduce the transmission rate to zero. Today we are wiser. Information theory characterizes a channel by a single parameter; the channel capacity. Shannon demonstrated that it is possible to transmit information at any rate below capacity with an arbitrarily small probability of error. The method of proof is random coding, where the existence of a good code is shown by averaging over all possible codes. Now there were codes before there was a theory of coding, and the mathematical framework for decoding certain algebraic codes (Bose-Chaudhuri-Hocquengham (BCH) codes) was written down in the late 18th century (see Wolf [227] and Barg [5]). Nevertheless, it is fair to credit Shannon with creating coding theory in that he established fundamental limits on what was possible, and presented the challenge of finding specific families of codes that achieve capacity.

Classical coding theory is concerned with the representation of information that is to be transmitted over some noisy channel. There are many obstacles to reliable communication, including channel estimation, noise, synchronization, and interference from other users, but there are only two resources available to the code designer; memory and redundancy. The proper allocation of these resources to the different obstacles is fertile ground for information theory and coding, but for the past 50 years the focus of coding theory in particular has been reliable communication in the presence of noise. This general framework includes the algebraic theory of error-correcting codes, where codewords are strings of symbols taken from some finite field, and it includes data transmission over Gaussian channels, where codewords are vectors in Euclidean space. Compact disk players [168], [113], hard-disk drives [152], and high-speed modems [83] are examples of consumer products that make essential use of coding to improve reliability. The importance of these applications has served to focus the coding theory community on the complexity of coding techniques, for it is entirely appropriate that performance of a code should be valued as a function of delay and decoding complexity. Ever since Shannon's original paper, coding theorists have attempted to construct structured codes that achieve channel capacity, but this problem remains unsolved. It is in fact tempting to ask a slightly different question; to fix the complexity of decoding and to ask for the maximum transmission rate that is possible. There is a sense in which the journey is more important than the goal, for the challenge of coming close to capacity has generated many important coding techniques.

The notion of combined source/channel coding is present in the telegraph codebooks that were used from 1845 until about 1950 (see [120, Ch. 22]). These books, arranged like dictionaries, would list many useful phrases, or even sentences, each with its corresponding codeword. They were compiled by specialists who competed on the basis of compression (the ability to capture a specialist vocabulary in few words), ease of use, and resistance to errors (exclusion from the codebook of words obtained from codewords by single letter substitution or transposition of adjacent letters). An important motivation was the price per word on undersea cablegrams which was considerable (about $5 per word on a trans-Atlantic cable message in 1867, falling to 25 cents per word by 1915). The addition of adjacent transpositions to Hamming errors means that the universe of words makes for a more complicated metric space, so that determining efficiency or even optimality of a particular code is extremely complicated. This framework did not encourage the creation of coding theory but it did not prevent telegraph code makers from using linear codes over a variety of moduli, and from realizing that the more parity-check equations were used, the greater the minimum distance would be.

Manuscript received June 1, 1998; revised August 15, 1998.

The author is with the Information Sciences Research Center, AT&T Labs., Florham Park, NJ 07932 USA.

Publisher Item Identifier S 0018-9448(98)06887-4.

In exploring the beginnings of coding theory, it is important to be mindful of intent. In the early 1940's the famous statistician Fisher [71], [72] discovered certain remarkable configurations in Hamming space through his interest in factorial designs. Consider $2^N - 1$ factors taking values ± 1 that influence the yield of a process, and suppose pairwise interactions do not affect yield. We are led to an expression

$$f(X) = \sum_{a \in \mathbb{F}_2^N} \lambda_a X_a + E$$

where E captures error and imprecision in the model. We look to determine the coefficients λ_a by measuring $f(X)$ for a small number of binary vectors X called *experiments*. Further, we are interested in a collection of experiments that will allow us to distinguish the effect of factor X_a from that of $X_{a'}$; in the language of statistical design, these factors are not to be *confounded*. The 2^N experiments $X_a = (-1)^{a \cdot v}$, for $v \in \mathbb{F}_2^N$, have this property, and correspond to codewords in the binary simplex code. The assertion that main effects X_a are not confounded is simply that the minimum weight in the Hamming code is at least 3. In classical statistical design the experiments are taken to be a linear code C, and large minimum weight in the dual code $C^\perp$ is important to ensure that potentially significant combinations of factors are not confounded. Since coding theory and statistical design share a common purpose we can understand why Fisher discovered the binary simplex code in 1942, and the generalization to arbitrary prime powers in 1945. However, it is important to remember his intention was not the transmission of information.

On an erasure channel, a decoding algorithm interpolates the symbols of a codeword that are not erased. In an algebraic error-correcting code the information in each encoded bit is diffused across all symbols of a codeword, and this motivates the development of decoding algorithms that interpolate. This notion is fundamental to the Berlekamp–Massey algorithm that is used for decoding a wide class of cyclic codes, and to the new list decoding algorithm of Sodan [203]. However Wolf [227] observed that as far back as 1795, de Prony [58] considered the problem of solving over the real field, the system of equations

$$\sum_{i=1}^{v} e_i X_i^k = S_i, \qquad k = 1, 2, \cdots, 2t$$

for the coefficients e_i, in the case where $v = t$. In algebraic coding theory this system of equations appears in the decoding of t-error-correcting BCH codes, but the underlying field is finite, the index v ($v \leq t$) is the number of errors, and the coefficients e_i are the error values. Nevertheless, the solutions proposed by de Prony [58] and Peterson [170], Gorenstein and Zierler [104] have the same form: all solve for the coefficients $\sigma_1, \cdots, \sigma_v$ of the *error-locator polynomial*

$$\sigma(x) = \prod_{i=1}^{v} (1 - xX_i)$$

by analyzing the recurrence relation

$$\sigma_1 S_{j+v-1} + \cdots + \sigma_v S_j = -S_{j+v}, \qquad j = 1, \cdots, v.$$

Algebraic coding theory calculates the determinant of this linear system for $v = t, t-1, \cdots$. It is zero if v exceeds the number of errors that occurred and nonzero if equality holds. Once the error-locator polynomial is known Chien search [40] can be used to find the error locations X_i, and then finding the errors e_i is simple linear algebra. By contrast, de Prony used Lagrange interpolation, and this corresponds to the refinement of the basic algorithm for decoding BCH codes that was suggested by Forney [73]. Berlekamp ([11, Ch. 7]) and Massey [153] expressed the problem of finding the coefficients of the error-locator polynomial as that of finding the shortest linear feedback shift register that generates the syndrome sequence. The Berlekamp–Massey algorithm has recently been generalized to more than one dimension, and used to decode algebraic-geometry codes. This story is told in more detail by Barg [5], but even this outline reveals considerable synergy between the discrete and the Euclidean world. This synergy is one of the strengths of the text by Blahut [14] and there is reason to resist any balkanization of coding theory into algebraic codes and codes for the Gaussian channel.

II. An Introduction to Hamming Space

Let $\mathbb{F}_q$ denote the finite field with q elements, and let $\mathbb{F}_q^N$ denote the set of N-tuples $(a_1, \cdots, a_N)$, where $a_i \in \mathbb{F}_q$. The *Hamming weight* $\mathrm{wt}(x)$ of a vector $x \in \mathbb{F}_q^N$ is the number of nonzero entries. The *Hamming distance* $D(x, y)$ between two vectors $x, y \in \mathbb{F}_q^N$ is the number of places where x and y differ. Thus $D(x, y) = \mathrm{wt}(x + y)$. An (N, M, D) *code* C over the alphabet $\mathbb{F}_q$ is a collection of M vectors from $\mathbb{F}_q^N$ (called *codewords*) such that

$$D(x, y) \geq D, \qquad \text{for all distinct } x, y \in C$$

and D is the largest number with this property. The parameter D is called the *minimum distance* of the code.

Vector addition turns the set $\mathbb{F}_q^N$ into an N-dimensional vector space. A *linear code* is just a subspace of $\mathbb{F}_q^N$. The notation $[N, k, D]$ indicates a linear code with blocklength N, dimension k, and minimum distance D. The next result is both fundamental and elementary.

Theorem: The minimum distance of a linear code is the minimum weight of a nonzero codeword.

It is possible to describe any code by just listing the codewords, and if the code has no structure, then this may be the only way. What makes a linear code easier to discover is that it is completely determined by any choice of k linearly independent codewords. Perhaps ease of discovery is the main reason that coding theory emphasizes linear codes.

A *generator matrix* G for an $[N, k]$ linear code C is a $k \times N$ matrix with the property that every codeword of C is some linear combination of the rows of G. Given an $[N, k]$ linear code C, the *dual code* $C^\perp$ is the $[N, N-k]$ linear code given by

$$C^\perp = \left\{ x \in \mathbb{F}_q^N \,\middle|\, (x, c) = 0 \text{ for all } c \in C \right\}$$

where

$$((x_1,\cdots,x_N),(y_1,\cdots,y_N))=\sum_{i=1}^{N}x_iy_i$$

is the standard inner product. An $[N,k]$ linear code C is also completely determined by any choice of $N-k$ linearly independent codewords from $C^{\perp}$. A *parity-check matrix* H for an $[N,k]$ linear code C is an $(N-k)\times N$ matrix with the property that a vector $x\in\mathbb{F}_q^N$ is a codeword in C if and only if $Hx^T=0$. Thus a generator matrix for C is a parity-check matrix for $C^{\perp}$ and vice versa. A linear code C is said to be *self-orthogonal* if $(x,y)=0$ for all $x,y\in C$. If C is self-orthogonal, then $C\subseteq C^{\perp}$ and we can construct a parity-check matrix for C by adding rows to a generator matrix. If $C=C^{\perp}$, then C is said to be *self-dual*. In this case, a single matrix serves as both a generator matrix and a parity-check matrix.

It is interesting to look back on Blake [15] which is an annotated selection of 35 influential papers from the first 25 years of algebraic coding theory and to distinguish two larger themes; geometry and algorithms. Here the early work of Slepian [196]–[198] on the internal structure of vector spaces provides a geometric framework for code construction. By contrast, the emphasis of work on cyclic codes is on the decoding algorithm. In the last 25 years, the fear that good codes might turn out to be very difficult or impossible to decode effectively ("messy") has been proved to be unfounded.

Hamming distance is not changed by *monomial transformations* which consist of permutations of the coordinate positions followed by diagonal transformations $\mathrm{diag}[\lambda_1,\cdots,\lambda_N]$ that multiply coordinate i by the nonzero scalar λ_i. Monomial transformations preserve the Hamming metric and we shall say that two codes C_1 and C_2 are *equivalent* if one is obtained from the other by applying a monomial transformation. In her 1962 Harvard dissertation, MacWilliams [146] proved that two linear codes are equivalent if and only if there is an abstract linear isomorphism between them which preserves weights. Extensions of this result to linear codes over finite rings and to different weight functions (for example, Lee weight) have been derived recently by Wood [228].

A. The Sphere-Packing Bound

The sphere $S_e(a)$ of radius e centered at the vector $a\in\mathbb{F}_q^N$ is the set

$$S_e(a)=\{x\in\mathbb{F}_q^N \mid D(x,a)\le e\}.$$

Since there are $q-1$ ways to change an individual entry we have

$$|S_e(a)|=\sum_{i=0}^{e}\binom{N}{i}(q-1)^i.$$

Let C be a code in $\mathbb{F}_q^N$ with minimum Hamming distance D and let $e=\lfloor (D-1)/2\rfloor$. *The sphere-packing bound*

$$|C|\left(\sum_{i=0}^{e}\binom{N}{i}(q-1)^i\right)\le q^N.$$

expresses the fact that spheres of Hamming radius e centered at the codewords of C are disjoint, and the union of these spheres is a subset of $\mathbb{F}_q^N$. An e-error-correcting code C for which equality holds in the sphere-packing bound is said to be *perfect*. For perfect single-error-correcting *linear* codes, the sphere-packing bound gives

$$|C|(1+(q-1)N)=q^N.$$

Since C is linear, there is a dual code $C^{\perp}$ satisfying $|C^{\perp}|=q^N/|C|=q^s$ for some s, and so $N=(q^s-1)/(q-1)$. The columns h_i, $i=1,2,\cdots,N$ in a parity-check matrix H for C are vectors in $\mathbb{F}_q^s$. If $\lambda h_i=h_j$ for some $\lambda\in\mathbb{F}_q$, then $(e_i-\lambda e_j)H^T=0$. This means $e_i-\lambda e_j\in C$, which contradicts the fact that C is a code with minimum Hamming weight $D=3$. Hence different columns of H must determine different one-dimensional subspaces of $\mathbb{F}_q^s$. Since there are exactly $N=(q^s-1)/(q-1)$ distinct one-dimensional subspaces of $\mathbb{F}_q^s$, we must choose exactly one vector from each subspace. Note that given s, any two codes of length $(q^s-1)/(q-1)$ obtained in this way are equivalent. This completes the classification of perfect single-error-correcting linear codes, but even perfect single-error-correcting nonlinear codes are not yet completely understood.

It is natural to start the search for other perfect codes by looking for instances where $\sum_{i=0}^{e}\binom{N}{i}(q-1)^i$ is a power of q. For $e=2$, $q=3$, $N=11$ we find

$$3^6\left(1+11\cdot 2+\binom{11}{2}\cdot 4\right)=3^{11}$$

and for $e=3$, $q=2$, $N=23$ we find

$$2^{12}\left(1+23+\binom{23}{2}+\binom{23}{3}\right)=2^{23}.$$

In each case there was a code waiting to be found; the $[11,6,5]$ ternary Golay code, and the $[23,12,7]$ binary Golay code.

The ternary Golay code was discovered by Virtakallio in 1947 and communicated in issues 27, 28, and 33 of the Finnish football pool magazine *Veikaaja*. The ternary alphabet is associated with the possible outcomes of a soccer match (win, lose, or draw), and Virtakallio's aim was to approximate closely an arbitrary vector in Hamming space (the ternary Golay code has the property that given any $x\in\mathbb{F}_3^{11}$ there is a unique codeword c such that $d_H(x,c)\le 2$).

The Golay codes [102] were discovered by Golay in 1949, but their rich algebraic structure was not revealed until much later. The $[24,12,8]$ binary Golay code is obtained from the perfect $[23,12,7]$ code by adding an overall parity check, and it is a most extraordinary code. The codewords of any given weight form beautiful geometric configurations that continue to fascinate combinatorial mathematicians. The symmetry group of this code plays a central role in finite group theory, for it is the Mathieu group M_{24}, which is perhaps the most important of the 26 sporadic simple groups.

In a perfect e-error-correcting code, the spheres of radius e about the codewords are disjoint and they cover the whole space. MacWilliams [146], [147] proved that an e-error-correcting linear code is perfect if and only if there are exactly e nonzero weights in the dual code. For example, the $[11,6,5]$

ternary Golay code is perfect, and nonzero codewords in the dual code have weight 6 or 9. Uniformly packed codes are a generalization of perfect codes that were introduced by Semakov, Zinoviev, and Zaitzev [183] in which the spheres of radius $e+1$ about the codewords cover the whole space, and these spheres overlap in a very regular way. There are constants λ and μ (with $\lambda < (n-e)(q-1)/e+1$) such that vectors at distance e from the code are in $\lambda+1$ spheres and vectors at distance $e+1$ from the code are in μ spheres. If the restriction on λ were removed, a perfect code would also be uniformly packed. Goethals and van Tilborg [101] showed that an e-error-correcting linear code is uniformly packed if and only if there are exactly $e+1$ nonzero weights in the dual code. For example, the $[24, 12, 8]$ binary Golay code is uniformly packed with $\lambda = 0$ and $\mu = 6$, and is self-dual with nonzero weights 8, 12, 16, and 24.

The connection between the metric properties of a linear code and the weight spectrum of the dual code is just one facet of the structural framework for algebraic coding theory that was introduced by Delsarte [48] in his Ph.D. dissertation, and this dissertation might well be the most important publication in algebraic coding theory over the past 30 years. The framework is that of association schemes derived from a group-theoretic decomposition of the Hamming metric space, and it will be described briefly in Section IV. The concept of an association scheme appears much earlier in the statistics literature, and Delsarte was able to connect bounds on orthogonal arrays from statistics with bounds for codes.

Of course, perfect codes are best possible since equality holds in the sphere-packing bound. However, Tietäväinen [212], van Lint [138], and Zinoviev and Leontiev [231] have shown that the only perfect multiple-error-correcting codes are the binary and ternary Golay codes, and the binary repetition codes. Critical to these classification results is a remarkable theorem of Lloyd [141] which states that a certain polynomial associated with a group-theoretic decomposition of the Hamming metric space must have integral zeros (for a perfect linear code these zeros are the weights that appear in the dual code).

B. The Gilbert–Varshamov Bound

We fix the transmission rate R, and we increase the block-length N in order to drive the error probability to zero. If the symbol error probability is p, then the average number of errors in a received vector of length N is Np. The minimum distance D must grow at least as fast as $2Np$. This explains the importance of the quantity $\alpha(\delta)$ which measures achievable rate, given by

$$\alpha(\delta) = \limsup_{N\to\infty} \frac{\log_q A_q(N, \delta N)}{N},$$

where $A_q(N, \delta N)$ is the maximum size of a code with minimum distance δN. To study $\alpha(\delta)$ we need to estimate the number of vectors $V_q(N, e)$ in a sphere of radius e in $\mathbb{F}_q^N$. If $0 \le \lambda \le (q-1)/q$, then

$$\log_q \frac{V_q(N, \lfloor \lambda N \rfloor)}{N} = H_q(\lambda)$$

where $H_q(x)$ defined on $[0, (q-1)/q]$ is the appropriate generalization of the binary entropy function, and is given by

$$H_q(0) = 0$$
$$H_q(x) = x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x), \quad \text{for } 0 < x \le \tfrac{q-1}{q}.$$

Independently, Gilbert [95] and Varshamov [218] derived a lower bound on achievable rate that is surprisingly difficult to beat. In fact, Varshamov proved there exist linear codes C with

$$|C| \sum_{i=0}^{d-2} \binom{N-1}{i} (q-1)^i \ge q^N$$

which for particular values N, d is sometimes stronger.

Theorem (The Gilbert–Varshamov Bound): If $0 \le \delta \le (q-1)/q$, then

$$\alpha(\delta) \ge 1 - H_q(\delta).$$

Proof: It is sufficient to prove

$$A_q(N, D) \ge q^N / V_q(N, D-1).$$

Let C be an (N, M, D) code in $\mathbb{F}_q^N$, where $M = A_q(N, D)$. Then, by definition, there is no vector in $\mathbb{F}_q^N$ with Hamming distance D or more to all codewords in C. This means that

$$\mathbb{F}_q^N = \bigcup_{c\in C} S_{D-1}(c)$$

which implies $|C| V_q(N, D-1) \ge q^N$. □

The proof shows it is possible to construct a code with at least $q^N/V_q(N, D-1)$ codewords by adding vectors to a code with minimum distance D until no further vectors can be added. What is essential to the Gilbert–Varshamov (G-V) argument is an ensemble of codes, where for each vector v that appears in some code, we have control over the fraction λ_v of codes from the ensemble that contain v. In the original G-V argument, the ensemble consists of all linear codes of a certain dimension. The group of nonsingular linear transformations preserves this ensemble (though linear transformations do not, in general, preserve Hamming weight) and acts transitively on nonzero vectors, so that $\lambda_v = \lambda$ is constant. The G-V argument applies to more restrictive ensembles of codes, for example, to binary self-orthogonal codes with all Hamming weights divisible by 4 [149]. Here the function $Q(v) = \mathrm{wt}(v)/2$ defines a quadratic form on the space of all binary vectors with even Hamming weight. Self-orthogonal codes correspond to totally singular subspaces and transitivity of the underlying orthogonal group leads to the G-V bound. Similar arguments provide lower bounds for quantum error-correcting codes [34] and for the minimum norm of certain lattices (see [142]), and there is a sense in which the classical bounds of Conway and Thompson are also obtained by averaging.

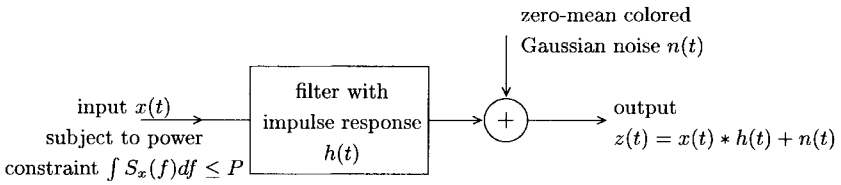

Fig. 1. The Gaussian channel model.

III. Euclidean Space

A *Gaussian channel* combines a linear filter with additive Gaussian noise as shown in Fig. 1. In the time domain the output $z(t)$ is given by

$$z(t) = x(t) * h(t) + n(t)$$

where $x(t)$ is the input waveform, $h(t)$ is the *channel impulse response*, $x(t) * h(t)$ is the convolution of $x(t)$ with $h(t)$, and $n(t)$ is zero-mean-colored Gaussian noise.

The Fourier transform of $h(t)$ is the *frequency response* $H(f)$ of the channel, and the *power spectrum* $S_h(f)$ is given by $S_h(f) = |H(f)|^2$. In the frequency domain the signal $x(t)$ and the noise $n(t)$ are characterized by their Fourier transforms $X(f)$ and $N(f)$, respectively, and by their power spectra $S_x(f)$ and $S_n(f)$. An essential feature of the model is a power constraint

$$\int S_x(f)df \le P$$

on the power spectrum $S_x(f)$ of the input waveform $x(t)$. The *channel signal-to-noise function* $\mathrm{SNR}_h(f)$ is given by $\mathrm{SNR}_h(f) = S_h(f)/S_n(f)$, and is measured in decibels by taking $10\log_{10}\mathrm{SNR}_h(f)$.

The model is limited in that the output $z(t)$ is assumed to depend linearly on the input $x(t)$, and to be time-invariant. In magnetic-recording applications, this linearity assumption becomes less valid once the recording density exceeds a certain threshold. In modem applications, the noise $n(t)$ starts to depend on the input $x(t)$ once the transmission rate exceeds a certain threshold. However, these caveats should not subtract from the importance of the basic model.

We think of the input $x(t)$ and the output $z(t)$ as random variables. The mutual information between $x(t)$ and $z(t)$ is the conditional entropy of $z(t)$ given $x(t)$. Channel capacity results from maximizing mutual information. Information-theoretic "waterfilling arguments" show that there is a constant K and a frequency band $\mathcal{W} = \{f \mid K \ge 1/\mathrm{SNR}_h(f)\}$, such that the capacity achieving input power spectrum $S_x^*(f)$ is given by

$$S_x^*(f) = \begin{cases} K - 1/\mathrm{SNR}_h(f), & \text{if } f \in \mathcal{W} \\ 0, & \text{if } f \notin \mathcal{W}. \end{cases}$$

The sampling theorem of Nyquist and Shannon allows us to replace a continuous function limited to the frequency band $\mathcal{W}$ by a discrete sequence of W equally spaced samples, without loss of any information. This allows us to convert our continuous channel to a discrete-time channel with signaling interval $T = 1/W$. The input $x(t)$ is generated as a filtered sequence $\sum x_k p(t - kT)$, where x_k is complex and the pulse p has power spectrum proportional to $S_x^*(f)$ on $\mathcal{W}$. The output $z(t)$ is sampled every T seconds and the decoder operates on these samples.

Opportunity for coding theorists is a function of communications bandwidth. The capacity-achieving bandwidth of an optical fiber is approximately 10^9 Hz, which is too large for sophisticated signal processing. By contrast, the capacity achieving bandwidth of a telephone channel is approximately 3300 Hz. If a modem is to achieve data rates of 28.8 kb/s and above, then every time we signal, we must transmit multiple bits. Mathematics now has a role to play because there is time for sophisticated signal processing.

An *ideal band-limited Gaussian channel* is characterized by a "brickwall" linear filter $H(f)$ that is equal to a constant over some frequency band of width W hertz and equal to zero elsewhere, and by white Gaussian noise with a constant power spectrum over the channel bandwidth. The equivalent discrete-time ideal channel represents the complex output sequence z_k as

$$z_k = x_k + n_k$$

where (x_k) is the complex input sequence and (n_k) is a sequence of independent and identically distributed (i.i.d.) complex zero-mean Gaussian random variables. We let S_x denote the average energy of the input samples (x_k), and we let S_n denote the average energy of the noise samples. Shannon proved that the channel capacity of this ideal channel is given by

$$C = \log_2(1 + S_x/S_n) \text{ bits/Hz}$$

or

$$\tilde{C} = CW = W\log_2(1 + S_x/S_n) \text{ bits/s.}$$

We may transmit m bits per hertz by selecting x_k from a fixed constellation of 2^m points from the integer lattice $\mathbb{Z}^2$ in the complex plane. This method of signaling is called 2^m-*Quadrature Amplitude Modulation* (2^m-QAM), and this is uncoded transmission since there is no redundancy. There is a gap between capacity of this ideal channel and the rate that can be achieved by uncoded QAM transmission. The size of this gap varies with channel SNR and for sufficiently high SNR it is approximately 3 bits/Hz. This can also be expressed as a gap in SNR of approximately 9 dB since the extra rate changes S_x to $S_x/8$ and $10\log_{10} 8 \approx 9$ dB.

Shannon recognized that signals input to a Gaussian channel should themselves be selected with a Gaussian distribution; the statistics of the signals should match that of the noise. We start by choosing a lattice Λ in real N-dimensional space $\mathbb{R}^N$.

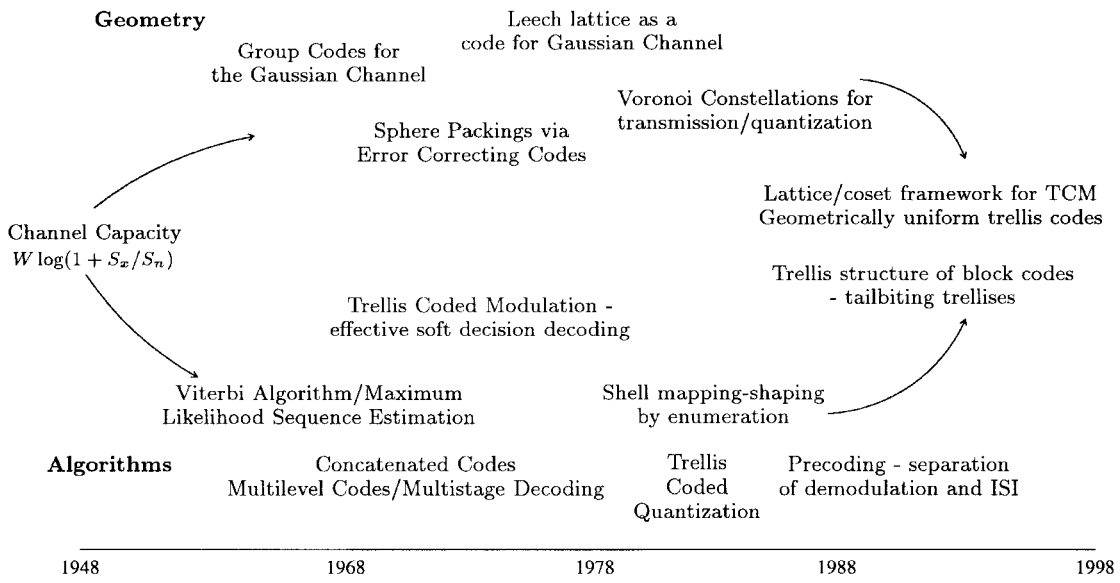

Fig. 2. Fifty years of information theory and coding for the power-constrained Gaussian channel.

Here the text by Conway and Sloane [44] is a treasury of information about sphere packings, lattices, and multidimensional Euclidean geometry. The signal constellation Ω consists of all lattice points within a region $\mathcal{R}$. The reason we consider signal constellations drawn from lattices is that signal points are distributed regularly throughout N-dimensional space. This means that the average signal power P of the constellation Ω is approximately the average power $P(\mathcal{R})$ of a probability distribution that is uniform within $\mathcal{R}$ and zero elsewhere. This approximation is called the *continuous approximation* and we shall use it extensively. If we fix the size of the signal constellation, then the average signal power depends on the choice of lattice and on the shape of the region that bounds the constellation. We obtain a Gaussian distribution by choosing the bounding region to be an N-dimensional sphere.

From the time that Shannon derived the capacity of the Gaussian channel there has been a divide between coding theory and coding practice. The upper track in Fig. 2 is the world of geometry and the lower track is the world of algorithms. We shall illustrate the differences by following an example, but a very positive development over the last five years is that these two tracks are converging.

A. Lattices

We begin with geometry. Formally, a *lattice* Λ in real N-dimensional space is a discrete additive subgroup of $\mathbb{R}^N$. A basis for the lattice Λ is a set of m vectors $v_1, \cdots, v_m$ such that

$$\Lambda = \left\{ \sum_{i=1}^{m} \lambda_i v_i \mid \lambda_i \in \mathbb{Z},\ i = 1, \cdots, m \right\}.$$

The lattice Λ is said to be m-dimensional and usually we have $m = N$. If $w_1, \cdots, w_m$ is another choice of basis then there exists a unimodular integral matrix Q such that $w_i = Qv_i$ for all $i = 1, \cdots, m$. The Gosset lattice E_8 was discovered in the last third of the nineteenth century by the Russian mathematicians A. N. Korkin and E. I. Zolotaroff, and by the English lawyer and amateur mathematician Thorold Gosset:

$$\begin{aligned} E_8 = \{(z_1, \cdots, z_8) \mid\ & z_i \in \mathbb{Z},\ i = 1, \cdots, 8 \\ & \text{or } z_i \in \mathbb{Z} + 1/2,\ i = 1, \cdots, 8, \\ & \text{and } z_1 + z_2 + \cdots + z_8 \in 2\mathbb{Z}\}. \end{aligned}$$

A *fundamental region* $\mathcal{R}$ for a lattice Λ is a region of $\mathbb{R}^N$ that contains one and only one point from each equivalence class modulo Λ. In the language of mathematics, $\mathcal{R}$ is a complete system of coset representatives for Λ in $\mathbb{R}^N$. If $v_1, \cdots, v_m$ are a basis for a lattice Λ then the parallelotope consisting of the points

$$\mu_1 v_1 + \cdots + \mu_m v_m \qquad (0 \le \mu_i < 1)$$

is an example of a fundamental region of Λ. This region is called a *fundamental parallelotope*. If $\Lambda \subseteq \mathbb{R}^N$ is a lattice, and $y \in \Lambda$ is a lattice point, then the *Voronoi region* $\mathcal{R}(y)$ consists of those points in $\mathbb{R}^N$ that are at least as close to y as to any other $y' \in \Lambda$. Thus

$$\mathcal{R}(y) = \{x \in \mathbb{R}^N \mid \|x - y\|^2 \le \|x - y'\|^2 \text{ for all } y' \in \Lambda\}.$$

The interiors of different Voronoi regions are disjoint though two neighboring Voronoi regions may share a face. These faces lie in the hyperplanes midway between two neighboring lattice points. Translation by $y \in \Lambda$ maps the Voronoi region $\mathcal{R}(w)$ to the Voronoi region $\mathcal{R}(w + y)$, so that all Voronoi regions are congruent.

A maximum-likelihood decoding algorithm for the lattice Λ finds the Voronoi region $\mathcal{R}(y)$ that contains the received vector $v \in \mathbb{R}^N$. The Voronoi regions $\mathcal{R}(y)$ are the decision regions for this algorithm. We may create a fundamental region for the lattice Λ by deleting faces from a Voronoi region. Different ways of deleting faces correspond to different rules for resolving ties in a maximum-likelihood decoding algorithm.

Given a lattice $\Lambda \subseteq \mathbb{R}^N$, there are many ways to choose a fundamental region, but the volume of the fundamental

region is uniquely determined by the lattice Λ. This volume is called the *fundamental volume* and we denote it by $V(\Lambda)$. There is a simple formula for the fundamental volume. Let $v_i = (v_{i1}, \cdots, v_{iN})$, $i = 1, \cdots, m$ be a basis for $V(\Lambda)$, and let $A = [v_{ij}]$. The fundamental volume $V(\Lambda)$ is given by $V(\Lambda)^2 = \det A^T$. It is easily verified that the fundamental volume of the Gosset lattice E_8 is equal to 1, the same as the integer lattice $\mathbb{Z}^8$.

Let Ω be an N-dimensional signal constellation consisting of all points from a lattice Λ that lie within a region $\mathcal{R}$, with centroid the origin. If signals are equiprobable, then the average signal power P is approximately the average power $P(\mathcal{R})$ of a continuous distribution that is uniform within $\mathcal{R}$ and zero elsewhere. Thus

$$P \approx P(\mathcal{R}) = \frac{1}{NV(\mathcal{R})} \int_{\mathcal{R}} ||x||^2 \, dv \approx G(\mathcal{R}) V(\mathcal{R})^{2/N}$$

where

$$V(\mathcal{R}) = \int_{\mathcal{R}} dv$$

is the volume of the region $\mathcal{R}$, where

$$G(\mathcal{R}) = \frac{\int_{\mathcal{R}} ||x||^2 \, dv}{NV(\mathcal{R})^{1+2/N}}$$

is the normalized or dimensionless second moment. The second moment $G(\mathcal{R})$ results from taking the average squared distance from a point in $\mathcal{R}$ to the centroid, and normalizing to obtain a dimensionless quantity.

We see that the average signal power P depends on the choice of lattice, and on the shape of the region that bounds the signal constellation. The formula $P \approx G(\mathcal{R})V(\mathcal{R})^{2/N}$ separates these two contributions. The volume $V(\mathcal{R}) = |\Omega| V(\Lambda)$, so that the second factor is determined by the choice of lattice. Since different lattices require different volumes to enclose the same number of signal points, it is possible to save on signal power by choosing the lattice appropriately. Since the second moment $G(\mathcal{R})$ is dimensionless, it is not changed by scaling the region $\mathcal{R}$. Therefore, the first factor $G(\mathcal{R})$ measures the effect of the shape of the region $\mathcal{R}$ on average signal power.

It is natural to compare the performance of E_8 as a code for the Gaussian channel with uncoded QAM transmission (the integer lattice $\mathbb{Z}^8$). Since the fundamental volumes coincide we may use the same region to bound both signal constellations. Performance gain is then determined by the minimum squared Euclidean distance $d^2(E_8)$ between two distinct points in the lattice E_8. We have $d^2(E_8)/d^2(\mathbb{Z}^8) = 2$ which corresponds to a coding gain of 3 dB.

B. Trellis Codes Based on Lattices and Cosets

Next we turn to algorithms. In 1976, Ungerboeck [215] constructed simple trellis codes for the Gaussian channel that provided coding gains of between 3 and 6 dB. His original paper has transformed the subject of coding for the Gaussian channel. Calderbank and Sloane [36] then abstracted the idea of redundant signaling based on lattices and cosets. The signal

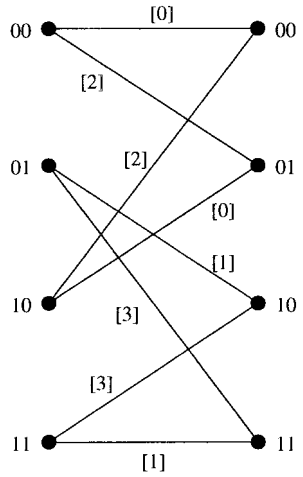

Fig. 3. Labeling edges by cosets in $[\mathbb{Z} : 4\mathbb{Z}]$.

points are taken from an N-dimensional lattice Λ, and the signal constellation contains an equal number of points from each coset of a sublattice Λ'. One part of the binary data stream selects cosets of Λ' in Λ, and the other part selects points from these cosets. All the redundancy is in the coset-selection procedure, and the bits that select the signal point once the coset has been chosen are referred to as *uncoded bits*. Forney [79], [80] coined the name *coset code* to describe redundant signaling based on lattices and cosets, and this name captures the essential property of these signaling schemes. Coset coding provides a level of abstraction that makes it possible for a code designer to handle complicated codes and large signal constellations.

Switching from uncoded transmission using the integer lattice Λ to coded transmission using a coset code C based on the lattice partition Λ/Λ' requires that the N-dimensional signal constellation be expanded by a factor $2^{N\rho(C)}$, where $\rho(C)$ is the redundancy of the coset code C. Note that all redundancy is in the method of selecting cosets, so this quantity is easy to calculate. We assume that the constellation is expanded by scaling a bounding region, so that the power penalty incurred by expansion is $4^{\rho(C)}$. The coding gain $\gamma(C)$ of the coset code C is then given by

$$\gamma(C) = d^2(C) 4^{-\rho(C)}.$$

This is the gain over uncoded transmission using the integer lattice (QAM signaling).

We introduce the method of trellis coding by means of an example where the lattice Λ is the integer lattice $\mathbb{Z}$, and the sublattice Λ' is $4\mathbb{Z}$. Fig. 3 shows the encoder trellis where the edges have been relabeled by the four residue classes modulo 4. All the redundancy is in the coset (residue class modulo Λ') selection procedure; one bit chooses from four cosets. The symbol $[i]$ represents the coset $\{z \mid z \equiv i \,(\mathrm{mod} 4)\}$. For transmission all cosets are translated by $-1/2$. Since all redundancy is in the coset-selection procedure, we can achieve any transmission rate by just increasing the number of uncoded bits.

The power and simplicity of the lattice/coset viewpoint comes from viewing the signal constellation as a finite subset

of an infinite lattice. By focusing on the infinite lattice, we eliminate the influence of constellation boundary effects on code structure and code performance.

It is not hard to prove that the minimum squared distance $d^2(C)$ between different signal sequences is given by $d^2(C) = 9$. To calculate the redundancy $\rho(C)$, we observe that every one-dimensional signaling interval, one input bit selects half of the integer lattice. The redundancy $\rho(C) = 1$, and the nominal coding gain $\gamma(C)$ is given by

$$\gamma(C) = 10\log_{10}\frac{9}{4^1} = 3.3 \text{ dB}.$$

There is, however, a difference between the nominal coding gain calculated above and the coding gain observed in practice. For channels with high SNR the performance of a trellis code C is determined by the minimum squared distance $d^2(C)$ between output sequences corresponding to distinct input sequences. For coset codes this minimum squared distance is determined by the minimum nonzero norm in the sublattice Λ' and by the method of selecting cosets. For channels with moderate SNR (symbol error probability $\sim 10^{-6}$) performance is determined by the minimum squared distance $d^2(C)$, and by the number of nearest neighbors or path multiplicity. A telephone channel is an example of a channel with moderate SNR. Here Motorola Information Systems has proposed a rule of thumb that reducing the path multiplicity by a factor of two produces a coding gain of 0.2 dB. The result of discounting nominal coding gain by path multiplicity in this way is called *effective coding gain*.

Every lattice point in E_8 has 240 nearest neighbors; the neighbors of the origin (the point 0^8) are the 112 points $(\pm 1)^2 0^6$, and the 128 points $(\pm 1/2)^8$ where the number of minus signs is even. This means that E_8 offers a way of arranging unit spheres in eight-dimensional space so that 240 spheres touch any given sphere. Levenshtein [134] and Odlyzko and Sloane [165] proved that it is impossible to exceed this. We can start to appreciate that the lattice E_8 is a fascinating mathematical object, and this large *kissing number* contributes to its allure. When we apply the discounting rule to the lattice E_8 the path multiplicity (per dimension) is $240/8 = 30$, whereas for the trellis code the path multiplicity is 4. The difference is an important reason why high-speed modems employ trellis codes based on lattices and cosets, rather than lattices in their natural state.

Before the invention of trellis-coded modulation by Ungerboeck [215] researchers designed codes for the Gaussian channel using heuristics that approximated Euclidean distance. For example, Nakamura [161] designed codes for phase modulation by restricting the congruence of signals modulo 8. This approach was also used for QAM transmission by Nakamura, Saito, and Aikawa [162]. Their measure of distance was Lee distance, which is computed entry by entry as a sum of Lee weights. The *Lee weight* $\mathrm{wt}_L([i])$ of a coset in $[\mathbb{Z} : 8\mathbb{Z}]$ is the smallest absolute value $|x|$ of an integer x congruent to i modulo 8. This amounts to designing codes for the L^1 metric. The assumption that noise is Gaussian makes it more appropriate to follow Ungerboeck and work with the L^2 metric directly.

One reason that trellis-coded modulation has had an enormous impact on communications practice is that around 1982 digital electronics were sufficiently advanced to implement codes of the type proposed by Ungerboeck. And when it is not possible to build circuits the only recourse is geometry. A second reason, also very important, is that consumers were waiting for new products, like high-speed modems, that this invention made possible. With all the benefits of hindsight we may look back and find the principles of set partitioning in earlier mathematical work by Leech [131] at a time when digital electronics were not ready for this innovation. However, Leech's work lacked any vision of communications practice, and Ungerboeck made the link explicit between his mathematical theory of set partitioning and the transmission of information.

C. Sphere Packings and Codes

Leech [131] showed how to use error-correcting codes to construct dense sphere packings in N-dimensional space. The idea is to specify a set of vectors with integer entries by constraining the binary expansion of those entries.

The *Leech coordinate array* of a vector $x = (x_1, \cdots, x_N)$ with integer coordinates is obtained by writing the binary expansion of the coordinates x_i in columns starting with the least significant digit. The first row of the coordinate array is the 2^0 row, the second row is the 2^1 row, the third row is the 2^2 row, and so on. To find the binary expansion (a_l) of a negative number $-a$, simply write

$$-a = \sum_{l \geq 0} a_l 2^l$$

and for $i = 1, 2, \cdots$ solve the equation

$$-a \equiv \sum_{l \geq 0} a_l 2^l \pmod{2^i}.$$

In row 2^0, the entry 1 represents an odd integer, and the entry 0 represents an even integer. We define subsets of the integer lattice $\mathbb{Z}^N$ by constraining the first L rows of the coordinate array. Given L binary codes $C_1, \cdots, C_L$ with blocklength N, the sphere packing $\Lambda(C_1, \cdots, C_L)$ consists of all vectors $x \in \mathbb{Z}^N$ for which the ith row of the coordinate array of x is a codeword in C_i. If $L = 1$, and if C_1 is a binary linear code, then

$$\Lambda(C_1) = \{x \in \mathbb{Z}^N \mid x \equiv c \pmod 2, \text{ for some } c \in C_1\}.$$

Here $\Lambda(C_1)$ is a lattice, since it is closed under addition. This construction is described by Leech and Sloane [132], where it is called *Construction A*, though Forney [80] uses the term *mod 2 lattice* to distinguish lattices constructed in this way. In general $\Lambda(C_1, \cdots, C_L)$ is not a lattice.

We make contact again with the Gosset lattice E_8 by taking C_1 to be the extended $[8, 4, 4]$ Hamming code C. The fundamental volume $V(\Lambda(C)) = 16$, and the minimum squared distance $d^2(\Lambda(C)) = 4$. The code C contains the zero vector, 14 codewords of weight 4, and the all-one vector 1 of weight 8. There are 14×2^4 vectors in $\Lambda(C)$ of type $(\pm 1)^4 0^4$, and 16 vectors in $\Lambda(C)$ of type $(\pm 2)0^7$. This gives 240 vectors

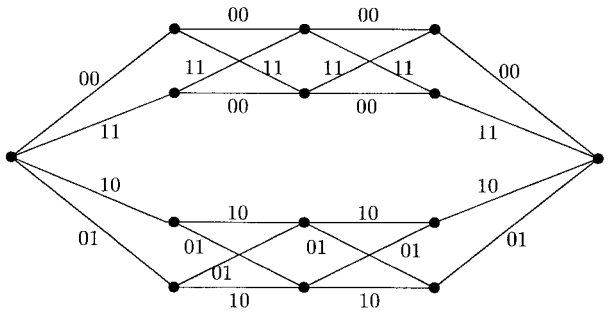

Fig. 4. A decoding trellis for the $[8,4,4]$ Hamming code.

in $\Lambda(C)$ with minimum norm 4, and it is easily seen that there are no others. This second appearance of the number 240 is not happenstance. The lattice $\Lambda(C)$ is a realization of the Gosset lattice E_8 on a different scale. There is a norm-doubling linear transformation $\Phi : \mathbb{R}^8 \to \mathbb{R}^8$ satisfying $||\Phi(x)||^2 = 2||x||^2$ that transforms the original realization of E_8 into $\Lambda(C)$.

Conway and Sloane [44] describe more sophisticated variants of Construction A, but it may be more interesting to apply the original construction to codes defined over the ring $\mathbb{Z}_{2^a}$ of integers modulo 2^a. For example, extended cyclic codes over $\mathbb{Z}_{2^a}$ obtained from certain binary cyclic codes by Hensel lifting determine even unimodular lattices via Construction A. The binary Golay code determines the Leech lattice in this way, and this is perhaps the simplest construction for this remarkable lattice that is known. For more details see [108], [19], and [28].

D. Soft-Decision Decoding

The origin of the term *trellis code* is that the graph of state transitions looks like the structures used by gardeners to support climbing plants. Codewords are represented as paths through this trellis.

The decoder has a copy of the trellis. It processes the noisy samples and tries to find the path taken by the binary data. The decoding algorithm was proposed by Viterbi [219] and later shown to be a variant of dynamic programming. Every trellis stage, the decoder calculates and stores the most likely path terminating in a given state. The decoder also calculates the path metric, which measures distance from the partial received sequence to the partial codeword corresponding to the most likely path. Fig. 4 shows a decoding trellis for the $[8,4,4]$ Hamming code or for the lattice E_8 (in this interpretation the digits 0, 1 represent the cosets $2\mathbb{Z}, 2\mathbb{Z}+1$ and the metric for an edge labeled $\epsilon\epsilon'$ is determined by the distances from the received signals r, r' to $2\mathbb{Z}+\epsilon$, $2\mathbb{Z}+\epsilon'$). At time $\ell = 4$ in Fig. 4, the decoder only needs to update two path metrics and make one comparison to determine the most likely path terminating in a given state.

Viterbi [219] originally introduced this decoding method only as a proof technique, but it soon became apparent that it was really useful for decoding trellis codes of moderate complexity. The importance of this application is the reason the decoding method is called the Viterbi algorithm by communication theorists. Forney [77] recognized that the Viterbi algorithm is a recursive optimal solution to the problem of estimating the state sequence of a discrete time finite state Markov process observed in memoryless noise. Many problem in digital communication can be cast in this form.

TABLE I-A
MAXIMUM-LIKELIHOOD DECODING OF THE BINARY GOLAY CODE

When	Who	How	How many
1986	Conway and Sloane	128 cosets of $(1111) \oplus P_6$	1614
1986	Be'ery and Snyders	Fast Hadamard Transform	1551
1988	Forney	64-state trellis decoder via $\|(8,7)/(8,4)/(8,1)\|^3$	1351
1989	Snyders and Be'ery	generalized Wagner rule	827
1991	Vardy and Be'ery	ML decoding of $C'_{24} \cup C''_{24}$ using ML hexadecoder	651

TABLE I-B
MAXIMUM-LIKELILHOOD DECODING OF THE LEECH LATTICE

When	Who	How	How many
1986	Conway and Sloane	Turyn construction from E_8	55,968
1988	Forney	256-state trellis decoder via $\|E_8/RE_8/2E_8\|^3$	15,167
1989	Lang and Longstaff	Wagner decoding rule	~10,000
1989	Be'ery, Shahar, Snyders	generalized Wagner rule and look-up tables	6,129
1993	Vardy and Be'ery	ML decoding of $4 \times Q_{24}$ using ML hexadecoder	3,595

Decoding algorithms are assembled from basic binary operations such as real addition, real subtraction, comparing two real numbers, and taking an absolute value. For simplicity, we might assign unit cost to each of these operations, and we might neglect the complexity of say multiplication by 2 (since this can be accomplished by merely shifting a binary expansion). It is then possible to compare different algorithms, and to show, for example, that the iterative decoding procedure for Reed–Muller codes based on the $|u \mid u+v|$ construction is less complex than the standard procedure using the fast Hadamard transform (see [80]). Quite recently there has been substantial interest in effective trellis-based decoding of codes and lattices. Tables I-A and I-B follow the progress that has been made in reducing the number of operations required for maximum-likelihood decoding of the Golay code and the Leech lattice (see [216] and [217] for details and additional references).

Decoding complexity can be reduced still further through bounded-distance decoding. Here the decoder corrects all error patterns in the Euclidean sphere of radius ρ about the transmitted point, where ρ is the packing radius of the code or lattice. This means that the error exponent of the bounded-distance decoder is the same as that of a maximum-likelihood decoder. Forney and Vardy [87] have shown that bounded-distance decoding of the binary Golay code and Leech lattice requires only 121 and 331 operations, respectively. The overall

degradation in performance is about 0.1 dB over a wide range of SNR's.

It was Conway and Sloane [43] who revived the study of the complexity of soft-decision decoding algorithms for block codes and lattices. Their paper served to inspire a great deal of work, including the results reported in Table I. However, it is fair to say that this work was specific to particular families of codes, and fundamental asymptotic questions seemed out of reach. That changed with Tarokh's 1995 thesis [208] showing that decoding complexity grows exponentially with coding gain. The lower bound on complexity is established by means of an ingenious argument involving a differential equation, and the upper bound uses a sophisticated tensor product construction. Together the results show that the lower bound is asymptotically exact.

It is instructive to look back at the work of Slepian [199] who constructed codes for the Gaussian channel by taking a finite group of $N \times N$ matrices, and applying each matrix to a fixed vector in $\mathbb{R}^N$. It is remarkable that Ungerboeck codes are examples of Slepian signal sets (see [81]). One minor difference is that the group of isometries has become infinite. A more important difference is the emphasis today on the complexity of the group. This was not an issue that concerned Slepian, but it is of paramount importance today, because it determines the complexity of soft-decision decoding.

E. Multilevel Codes and Multistage Decoding

The coded-modulation schemes proposed by Ungerboeck make use of a partition Γ_L of the signal constellation into 2^L subsets sometimes corresponding to L levels in the Leech coordinate array. A rate $(L-1)/L$ convolutional code selects the subset, and the remaining uncoded bits select a signal from the chosen subset. Instead of coding across all levels at once, we might directly allocate system redundancy level by level, an idea that first appeared in the context of binary codes. In 1977, Imai and Hirakawa [112] presented their multilevel method for constructing binary block codes. Codewords from the component codes form the rows of a binary array, and the columns of this array are the codewords in the multilevel code. Imai and Hirakawa also described a multistage bounded-distance decoding algorithm, where the bits are decoded in order of decreasing sensitivity, starting with the bits protected by the most powerful error-correcting code. Subsequently, Calderbank [22] and Pottie and Taylor [173] described simple multilevel coset codes for the Gaussian channel, and quantified the performance/complexity advantages of multistage decoding over full maximum-likelihood decoding. Here the purpose of the parity check is to provide immunity against single symbol errors. Concerning theoretical limits, Wachsmann and Huber [220] have shown that multilevel codes with turbo code components come within 1 dB of the Shannon limit.

F. The Broadcast Channel

The flexibility inherent in multilevel coding and multistage decoding makes it easy to introduce unequal error protection when some bits are extremely sensitive to channel errors and others exhibit very little sensitivity. For example, Code Excited Linear Prediction (CELP) is a method of transmitting speech by first communicating a model of the vocal tract specified by parameters that depend on the speaker, and then exciting the model. This model includes pitch information, and an error here has much more impact on the reproduced speech quality, than an error at the input to the model. Specific speech/channel coding schemes for wireless channels are described by Cox, Hagenauer, Seshadri, and Sundberg [47]. This matching of speech and channel coding has become standard practice in the engineering of cellular voice services.

A second example is digital High-Definition Television (HDTV) that has been made possible by recent advances in video compression. Digital broadcast differs from digital point-to-point transmission in that different receivers have different signal-to-noise ratios, which decrease with distance from the broadcast transmitter. One concern with digital broadcast is its sensitivity to small variations in SNR at the various receiver locations. This sensitivity is manifested as an abrupt degradation in picture quality, which is generally considered unacceptable by the TV broadcast industry.

It is possible to achieve more graceful degradation by means of joint source and channel coding. There are algorithms for compressing video signals that output coarse information and fine information. The coarse information is sensitive because it provides a basic TV picture, and the fine information is less sensitive because it adds detail to the coarse picture. The channel-coding scheme is designed to provide greater error protection for the coarse information, so that the distant receiver always has access to the coarse picture. Receivers that are closer to the broadcast transmitter can obtain both the coarse picture, and the fine detail, so that, indeed, there is a more graceful decline in the quality of reception.

This philosophy of joint source and channel coding has its roots in the information-theoretic work of Cover [46] on broadcast channels. He considered a typical broadcast environment where a source wishes to transmit information over a Gaussian channel to a strong receiver with SNR S_2, and a weak receiver with SNR S_1. Cover established the efficiency of *superimposing information*; that is, broadcasting so that the detailed information meant for the stronger user includes the coarse information meant for the weaker user. The geometry of the achievable rate region makes it apparent that it is possible to achieve close to capacity $C_2 = \frac{1}{2}\log(1+S_2)$ for the strong receiver at the cost of reducing the achievable rate for the weaker receiver only slightly below capacity $C_1 = \frac{1}{2}\log(1+S_1)$. Specific multilevel codes that can be used in terrestrial broadcasting of HDTV to provide unequal error protection are described by Calderbank and Seshadri [34]. The data rate for HDTV is about 20–25 Mb/s in 6-MHz bandwidth, corresponding to transmission of 4 bits/symbol. It is possible to provide virtually error-free transmission (greater than 6-dB coding gain) for some fraction (for example, 25%) of the data, while providing a modest gain of 1–2 dB for the remaining data with respect to uncoded transmission. The connection with the information-theoretic work of Cover on broadcast channels is described by Ramchandran, Ortega, Uz, and Vetterli [175] in the context of their multiresolution joint source/channel coding scheme for this same application. Their

paper proposes a complete system, and describes a particular source-coding algorithm that delivers bits with different sensitivity to channel errors.

G. Methods for Reducing Average Transmitted Signal Power

We consider signal constellations that consist of all lattice points that fall within some region $\mathcal{R}$. If the region $\mathcal{R}$ is an N-cube with faces parallel to the coordinate axes, then the induced probability distribution on an arbitrary M-dimensional projection is uniform. Changing the shape of the region $\mathcal{R}$ induces a nonuniform probability distribution on this M-dimensional projection. Thus gains derived from shaping a high-dimensional constellation can be achieved in a low-dimensional space by nonequiprobable signaling. The asymptotic shaping gain is $\pi e/6$ or 1.53 dB.

The problem of addressing a signal constellation is that of mapping a block of input data to a signal point. This problem enters into the design of both encoder and decoder; for the decoder needs to invert the mapping in order to recover the data stream corresponding to the estimate for the transmitted sequence of signals. The N-cube is a particularly simple Cartesian product for which the addressing problem is trivial, but here there is no shape gain. Spheres optimize the shape gain available in a given dimension but are hard to address. Conway and Sloane [42] proposed the use of Voronoi constellations based on a lattice partition Λ/Λ_S—the constellation consists of points from a translate of Λ that fall within a Voronoi region for the shaping lattice Λ_S. They showed how to use a decoding algorithm for Λ_S to address the constellation. Unfortunately, the ratio of peak-to-average power for Voronoi constellations (and spheres) is very high, precluding their use.

Calderbank and Ozarow [31] introduced the method of shaping on rings, where the region $\mathcal{R}$ is partitioned into T subregions so as to obtain T equal subconstellations with increasing average power. A shaping code then specifies sequences of subregions, and it is designed so that subconstellations with lower average power are more frequent. The purpose of the shaping code is to create a good approximation to the desired Gaussian distribution, and it is important to minimize the complexity of the shaping code. The shell mapping algorithm used in the V.34 modem standard enumerates all points in the Cartesian product of a basic two-dimensional constellation that are contained in a higher dimensional sphere. Laroia, Farvardin, and Tretter [130] show that it is possible to construct a 64-dimensional constellation from a 384-point two-dimensional constellation that supports uncoded transmission at 8 bits/symbol with a shaping gain of 1.20 dB and a peak-to-average power ratio (PAR) of 3.76. Alternatively, it is possible to achieve a shaping gain of 1 dB with a PAR of 2.9 (for comparison, the PAR of the two-dimensional sphere is 2).

1) Shaping by Searching a Trellis: A trellis code is an ensemble of codewords that can be searched efficiently. This search can be carried out with respect to any nonnegative measure that is calculated on a symbol-by-symbol basis. In Viterbi decoding this measure is distance from the received sequence. Here the measure is signal energy, but many other applications are possible, for example, the reduction of peak to average power in OFDM systems. Trellis shaping is a method proposed by Forney [82] that selects a sequence with minimum power from an equivalence class of sequences, by means of a search through the trellis diagram of a code. The signal constellation is divided into rings labeled by the possible outputs of a binary convolutional code. Shaping information is transmitted by choosing a coset of the convolutional code, and a decoder selects the minimum-norm vector in the coset for transmission. Now data is transmitted in blocks of about 1000 symbols by periodically terminating the convolutional code. The delay would be unacceptable if it were only possible to recover information carried by the shaping code on a block-by-block basis. However, it is possible to specify cosets on a symbol-by-symbol basis using the theory of syndrome formers, developed by Forney [75] as part of his algebraic theory of convolutional codes. Forney ([75], [77], [78], [81]) has explored the algebraic structure of convolutional codes, and the connections with linear systems theory in some depth. Forney and Trott [85] have since shown that most of this structure theory extends to trellis codes based on lattices and cosets.

H. Precoding for ISI Channels

We begin with a brief account of the evolution in signal processing for magnetic-recording channels. Until quite recently, virtually all magnetic recording systems employed peak detection, where one sampled output is used to estimate the value of one symbol recorded on the disk. The reliability of peak detection depends on the minimum spacing between transitions. If two transitions are too close, the peaks are reduced in amplitude and shifted. Binary sequences input to magnetic recording systems that employ peak detection are required to meet certain runlength constraints in order to improve linear density and to improve system reliability. The (d,k) constraint requires that adjacent 1's be separated by at least d 0's and by at most k 0's. Here it is important to recall that in NRZI (nonreturn-to-zero-interleaved) recording the symbol 0 represents no transition, and the symbol 1 represents a transition. Long runs of 0's correspond to long stretches of constant magnetization. When the binary input satisfies a (d,k) constraint, it is possible to signal $(d+1)$ times as fast while maintaining the same spacing between transitions. If the code rate is R then the increase in linear density is given by the product $R(d+1)$. The k constraint aids timing recovery since timing is derived from transitions in the recorded data. Note that increasing the speed of circuitry is not without its challenges.

Peak detection looks at a signal sequence with respect to itself, not with respect to other signal sequences that could have been transmitted. The idea of using maximum-likelihood sequence estimation in magnetic-recording systems was suggested in 1971 by Kobayashi and Tang [125]. However, it has only recently become possible to implement partial response maximum likelihood (PRML) detection at sufficiently high speeds. PRML detection provides increases in linear density of about 30% by eliminating the d constraint. The resulting intersymbol interference (ISI) is equalized at the output of the

channel to some tractable response such as PRIV $(1-D^2)$ or EPRIV $((1-D)(1+D)^2)$. Maximum-likelihood (Viterbi) decoding is accomplished by tracking the state of the channel, as described in Kobayashi [124] or Forney [76].

A basic feature of telephone channels and certain optical memories (see [27]) is that they are linear subject to a peak constraint, and support a continuum of recording levels. This is fundamentally different from conventional magnetic-recording channels which are inherently nonlinear and where, to force linearity, the write current in the recording head has to be sufficient to ensure positive or negative saturation of the magnetic medium. Hence it is only possible to record the levels ± 1. The ability to write a continuum of levels at the input to this channel makes it possible to employ precoding techniques such as the one developed by Tomlinson [213], and by Miyakawa and Harashima [157], for Gaussian channels subject to ISI. The philosophy behind this precoding technique is that since the channel is known, it is possible to anticipate and correct for the effects of the channel at the input, so that a very simple decoder can be used at the output. It is not possible to use Tomlinson–Harashima precoding on conventional magnetic- and optical-recording systems where it is only possible to record a small discrete number of levels.

We consider transmission of A equally spaced analog levels $a_i \in \{0, 1, \cdots, A-1\}$ over a discrete time channel with causal impulse response q_i, $i \geq 0$ for which $q_0 = 1$. The output v_i is given by

$$v_i = a_i + \sum_{j \geq 1} a_{i-j} q_j.$$

Tomlinson–Harashima precoding [157], [213] is a nonlinear method of precoding the data a_i that renders the output of the $Q(z)$ channel effectively free of intersymbol interference, and allows instantaneous symbol-by-symbol decoding of the data.

The Tomlinson filter does not transmit the data a_i directly, but instead transmits precoded data a'_i, where

$$a'_i = a_i - \sum_{j \geq 1} a'_{i-j} q_j + A m_i$$

where m_i is the unique integer such that $a'_i \in [0, A]$. Now the output v_i is given by

$$\begin{aligned} v_i &= a_i - \sum_{j \geq 1} a'_{i-j} q_j + A m_i + \sum_{j \geq 1} a'_{i-j} q_j \\ &= a_i + A m_i \end{aligned}$$

and instantaneous symbol-by-symbol decoding is possible via congruence modulo A.

Precoding is a part of the V.34 modem standard [116] for communication over bandlimited Gaussian channels and variants thereof. In telephone-line modem applications it is important that the statistics of the channel symbols are Gaussian, so they match the statistics of the noise. Here Tomlinson–Harashima precoding is not appropriate since reduction modulo A seems to produce channel symbols a'_i that are uniformly distributed over the interval $[0, A]$. The ISI precoder [129] that forms a part of the V.34 standard is a more sophisticated alternative to Tomlinson–Harashima precoding. It achieves significant shaping gain (the saving in average transmitted signal power provided by a Gaussian input distribution over a uniform distribution) without increasing the complexity of trellis decoding much beyond that of the baseline memoryless channel. The key idea is to separate the problem of decoding in the presence of additive white Gaussian noise (AWGN) from that of resolving intersymbol interference. This is captured geometrically in Fig. 5. Precoding modifies the input just enough to ensure that the output of the channel $Q(D)$ is a trellis codeword. A Viterbi decoder takes care of the noise, and inversion of the channel provides an approximation to the original input. The original input can be recognized from the approximation, since both lie in a common Voronoi region. There is a small power penalty connected with the power of the sequence that modifies the original input, but this penalty can be made insignificant. Running this precoded transmission system "backward" provides a system for quantizing an individual source with memory (cf. trellis-coded quantization [150]).

I. The AWGN Channel and the Public Switched Telephone Network

Trellis codes provide effective coding gains of about 4.5 dB on the AWGN channel, and a further 1 dB is available through shaping schemes of moderate complexity. Forney and Ungerboeck [86] observe that the cutoff role of a high-SNR channel corresponds to an effective coding gain (without shaping) of about 5.7 dB at error probabilities of about 10^{-6}. This is as high an effective coding gain as anyone has achieved with moderate complexity trellis codes.

The coset codes described in this paper select signal points from uniformly spaced constellations. When harmonic distortion and PCM noise (logarithmic quantization noise) are significant channel impairments it can be advantageous to distort the uniform spacing. Testing of high-speed voiceband modems has revealed a significant increase in distortion for points near the perimeter of a QAM signal constellation. This distortion increases with distance from the center of the constellation and limits performance at data rates above 19.2 kb/s. The perimeter distortion can be reduced by transforming the signal constellation so that points near the center are closer together, and points near the perimeter are further apart. When the channel SNR is high, such a transformation reduces immunity to Gaussian noise because points near the center of the transformed constellation are closer together than in a uniformly spaced constellation with the same average power. Betts, Calderbank, and Laroia [13] have demonstrated theoretically that for channel SNR's of practical interest, there is actually a small gain in immunity to Gaussian noise. In fact, an appropriate coded-modulation scheme can produce gains of about 0.25 dB. Experiments support the intuition that it is advantageous to employ trellis codes for which the dominant error is a trellis path error, and the longer that error the better.

In fact, the Public Switched Telephone Network is evolving toward the point where an analog voiceband channel will consist of short analog end links connected to PCM codes, with no intermediate tandem D/A or A/D transformations. This observation inspired development of the V.90 modem

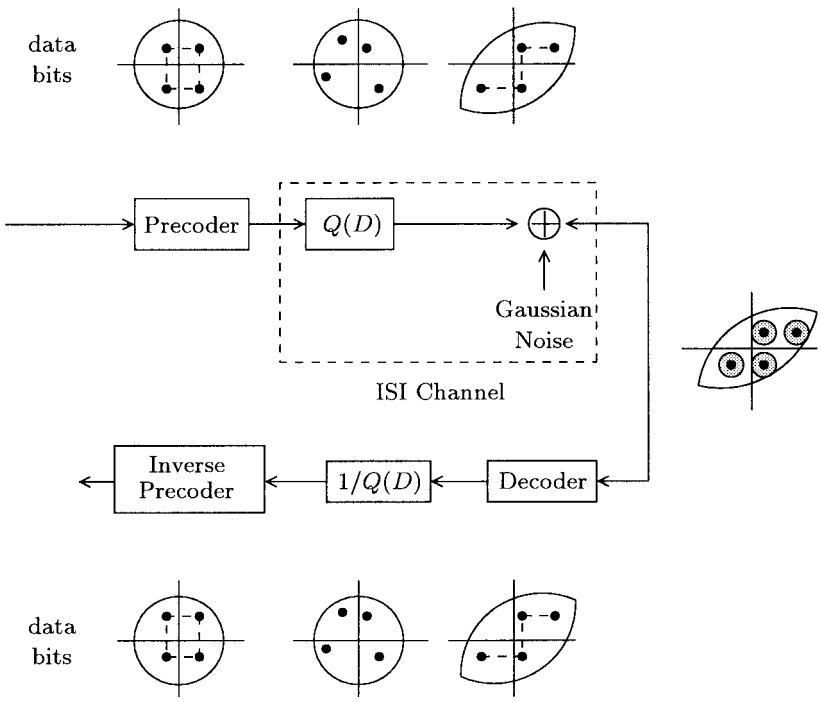

Fig. 5. A geometric rendering of a precoded transmission system.

standard, concluded in February 1998 which promises 56 kb/s downstream and delivers V.34 upstream.

J. The Potential of Iterated Decoding

In algebraic error-correcting block codes the information in each encoded bit is distributed across all symbols in a codeword. Information carried by concatenated codes is segmented, with some bits determining the encoding of a particular component subblock, and other bits the linkage between the different components. In decoding concatenated codes, initial estimates for component codewords are combined by decoding the code that links components together. In turbo codes the information in each bit is replicated in two different localities of a codeword. The original construction is to produce a first parity sequence by encoding an information sequence using a rate $1/2$ recursive systematic encoder, to permute the information bits using a very long interleaver (10^4–10^5 bits), and to produce a second parity sequence by encoding the permuted sequence using a second encoder of the same type as the first (possibly identical). Decoding is an iterative process where bit level soft decisions produced by one decoder are used to improve (hopefully) the decisions produced by the other decoder at the next step. The potential of this combination of local encoding and iterative decoding was revealed by Berrou, Glavieux, and Thitmajshima [12] who demonstrated that a 16-state rate $1/2$ turbo code can operate at an SNR 0.7 dB greater than capacity of the AWGN channel, with a decoded bit-error rate of 10^{-5}. For comparison, the Big Viterbi Decoder [41] designed to decode a 16 384-state convolutional code requires 2.4 dB to achieve the same bit-error rate. Like many revelations there was a period of initial scepticism, but now there are no doubts that this is a spectacular achievement. It is interesting to observe that the search for theoretical understanding of turbo codes has transformed coding theorists into experimental scientists. One empirical discovery is the existence of an *error floor* at low error rates that depends on the size of the interleaver. Perhaps the most interesting theoretical connection is that between the *forward–backward algorithm* [6] (a.k.a. the *BCJR algorithm* [3]) used in decoding convolutional codes, and belief propagation in Bayesian networks [167], a technique used for training and system identification in the neural network community [90], [145].

The ideas of local encoding and iterative decoding were present in a classic paper of Gallager [91], [92] written some 30 years before the discovery of turbo codes. A low-density parity-check (LDPC) matrix is a binary array where the number of 1's in each row and column is kept small. Gallager suggested using the adjacency matrix of a randomly chosen low-degree bipartite graph as the parity-check matrix. Decoding is again an iterative process where bit-level soft decisions obtained at one stage are used to update bit-level soft decisions about a particular bit at the next stage by means of the parity-check equations involving that bit. Gallager distinguished two different types of information, *intrinsic* and *extrinsic*, and understood that only extrinsic information is useful for iterative decoding. He developed the geometric picture of a *support tree* where the influence of a bit fans out across all symbols in a controlled way as the iterations progress. Gallager was not able to show correctness of the proposed iterative algorithm but he showed long LDPC codes can achieve rates up to capacity on the binary-symmetric channel with maximum-likelihood decoding. Subsequently, Zyablov and Pinsker [232] showed that with high probability over the choice of graph, the codes proposed by Gallager could be decoded in $\log n$ rounds, where each decoding round

TABLE II
A GENERATOR MATRIX FOR THE $[24, 12, 8]$ BINARY GOLAY CODE

11	00	11	11	11	00	00	00	00	00	00	00
00	11	01	01	01	11	00	00	00	00	00	00
00	00	11	10	10	11	11	00	00	00	00	00
00	00	00	11	11	01	01	11	00	00	00	00
00	00	00	00	11	10	11	10	11	00	00	00
00	00	00	00	00	11	11	11	00	11	00	00
00	00	00	00	00	00	11	01	11	01	11	00
00	00	00	00	00	00	00	11	01	11	01	11
11	00	00	00	00	00	00	00	11	11	10	01
11	11	00	00	00	00	00	00	00	11	01	10
11	10	11	00	00	00	00	00	00	00	11	01
01	11	10	11	00	00	00	00	00	00	00	11

removes a constant fraction of errors. More recently, MacKay and Neal [145] demonstrated near Shannon-limit performance of LDPC codes with iterative decoding. If the art of simulation had been more advanced in 1963, the history of coding theory might look very different today.

Sipser and Spielman [192] only discovered Gallager's paper after deriving asymptotically good linear error-correcting codes with decoding complexity $O(\log N)$-linear time only under the uniform cost model where the complexity of adding two N-bit binary vectors is independent of N. The combinatorial objects at the heart of the Sipser–Spielman construction are *expander graphs* in which every vertex has an unusually large number of neighbors, and these codes are of the type proposed by Gallager. The machinery of expander graphs enabled Sipser and Spielman to prove that the sequential decoding algorithm proposed by Gallager was in fact correct for these expander codes, something Gallager had not been able to do 30 years earlier.

The idea that graphical models for codes provide a natural setting in which to describe iterative decoding techniques is present in Tanner [207] but has undergone a revival in recent years [221], [222]. One way this school of coding theory connects with the classical theory is through the study of tailbiting trellises for binary block codes. Solomon and van Tilborg [200] demonstrated that a tailbiting trellis for a binary block code can in fact have fewer states than a conventional trellis. Table II shows a generator matrix of the $[24, 12, 8]$ binary Golay code that provides a 16-state, 12-section tailbiting trellis [26], whereas a conventional trellis must have 256 states at its midpoint [158]. The specific discovery was motivated by a suggestion of Wiberg [221, Corollary 7.3], and by the general result that the number of states in a tailbiting trellis can be as few as the square root of the corresponding number for a conventional trellis at the midpoint [222]. The time axis for a tailbiting trellis is defined most naturally on the circle, though it can also be defined on an interval with the added restriction that valid paths begin and end in the same state. The *span* of a generator is the interval from the first to the last nonzero component, and the generator is said to be *active* in this interval. For the Golay code, we see from Table II that at every time slot only four generators are active, hence the $2^4 = 16$ states in the tailbiting trellis (see [26] for details). It is quite possible that other extremal self-dual block codes (notably the $[48, 24, 12]$ Quadratic Residue code) will also have generator matrices that correspond to low-complexity tailbiting representations.

In iterative decoding the focus is on understanding the domain of attraction for a codeword rather than understanding the boundaries of a Voronoi region. In the future we might well see a shift in emphasis within coding theory from static geometry to dynamical systems. Certainly it would be interesting to have a counterpart of turbo codes in the world of algebraic error-correcting codes.

K. On Duality Between Transmission and Quantization

The theory of communication and that of quantization overlap significantly, but there has been less cross pollination between the two communities than might be expected. Nevertheless, it is commonly understood that the problems of coding and quantization are in some sense dual.

The lattice-decoding algorithms described in previous sections can be used to represent a source sequence x as the sum of a lattice point v, and an error sequence $e = (e_i)$. In quantization the objective is the lattice point v, and the expected value $E(e_i^2)$ is the *mean-squared error* (mse) normalized per dimension. By contrast, the objective in transmission is not the lattice point v, but the error sequence e. The idea is to choose a suitable discrete set of source sequences x, so that the entries of the error sequence e have a distribution that is approximately Gaussian.

The error sequence e is distributed over the Voronoi region $\mathcal{R}$ of the lattice, and if this distribution is uniform, then the mean-squared error $E(e_i^2)$ is equal to the second moment $G(\mathcal{R})$. In quantization, the quantity $\gamma(\mathcal{R}) = 1/12G(\mathcal{R})$ is called the *granular gain*, and it measures the reduction in mean-squared error that comes from choosing the shape of the quantization cell. The baseline for comparison is uniform scalar quantization (using the integer lattice) where the quantization cell is the N-cube C_N with second moment $G(C_N) = 1/12$. Table III presents a correspondence between quantities of interest in communications and in quantization (with respect to Gaussian channels/sources). Successive refinement is a

TABLE III
CORRESPONDENCE BETWEEN QUANTITIES OF INTEREST IN CODING AND QUANTIZATION

Coding	Quantization
Transmission over AWGN channel (memoryless) subject to a power constraint	Quantization of a memoryless Gaussian source (mmse distortion) [143]
Coding Gain: The Voronoi cells should cover a region of appreciable noise probability	Boundary Gain: For a given mse the spheres of appropriate radius about the codewords should cover a region of appreciable source probability
Shaping Gain: Bounding region that minimizes second moment leads to minimum average transmitted signal power	Granular Gain: mse distortion favors Voronoi cells with the smallest second moment
Trellis Coded Modulation [215]	Trellis Coded Quantization [150]
Broadcast Channel [46]	Successive Refinement [127], [65]
Precoding for ISI Channels [129]	Quantization of Sources with Memory
Multiple Access Channels	Quantization of Correlated Sources [200]

particular case of multiple descriptions, where two channels connect the source to the destination (see [166], [64], and [214]). Either channel may fail and this failure is known to the decoder but not the encoder. The objective is to obtain good performance when both channels work and to degrade gracefully if either channel fails. The two channels may be considered equally important, and this is different in spirit from layered coding (successive refinement) where a high-priority channel transports important bits. The emergence of wireless systems employing multiple antennas, and of active networking in lossy packet networks represent an opportunity for the multiple descriptions coding paradigm.

L. The Notion of Frequency Domain

This is the idea of using constraints in the frequency domain to separate codewords in the time domain. We begin by considering integer valued sequences $p = (p_0, p_1, \cdots, p_{N-1})$ which we represent as polynomials $p(D) = \sum_{i=0}^{N-1} p_i D^i$. We shall say that the sequence $p(D)$ has a Kth-order spectral null at $\theta = 2\pi\ell/M$, if $p(D)$ is divisible by $(e^{i\theta} - D)^K$. A collection of sequences with this property is called a *spectral null code*. To show that it is possible to separate vectors in Euclidean space by placing spectral constraints in the frequency domain, we consider the case $\theta = 0$. We say that the sequence $p(D)$ has *a sign change at position* u if $p_u \neq 0$, and $\mathrm{sign}\,(p_u) = -\mathrm{sign}\,(p_t)$, where $t = \max\{i < u \mid p_i \neq 0\}$.

Theorem (Descartes Rule of Signs): Let $p(D)$ be a real polynomial with K positive real roots, not necessarily distinct. Then the number of sign changes in the sequence p of coefficients of $p(D)$ is at least K.

For a proof we refer the reader to Householder [111]. Now consider a code with a Kth-order spectral null at $\theta = 0$. It follows directly from Descartes Rule of Signs that the minimum squared distance between codewords is at least $2K_d^2$, where d is the minimum distance of the integer alphabet employed (for the bipolar alphabet ± 1, this gives a bound of $8K$. This simple observation is the starting point for the construction of many codes used in magnetic recording applications; more details can be found in Immink and Beenker [115], Karabed and Siegel [121], Eleftheriou and Cideciyan [63], and the survey paper [152]. The objective in all these papers is to separate signals at the output of a partial-response channel by generating codewords at the input with spectral nulls that are matched to those of the channel. The special features of telephone channels and recording channels have also led to new connections between coding theory, dynamical systems, and linear systems theory [151].

M. Partial-Response Channels and Coding with Spectral Constraints

It is natural to try to devise coding schemes that meet both spectral null and minimum distance/coding gain objectives. Starting from an uncoded L-level data sequence (i_k) we want to generate a real-valued sequence (y_k) with nulls at certain prescribed frequencies in such a way that the data (i_k) can be recovered instantly from the sequence (y_k). Fig. 6 shows an input sequence $x(D)$ passing through a partial response channel with impulse response (transfer function) $p(D)$, resulting in an output $y(D) = x(D)p(D)$, which is called a partial-response-coded (PRC) sequence. A white-noise sequence $n(D)$ may be added to $y(D)$ to give a noisy PRC sequence $z(D)$, representing the output of a real channel. The input sequence $x(D)$ can be recovered from the PRC sequence $y(D)$ by passing $y(D)$ through a filter with transfer function $1/p(D)$. (We have to imagine that $x(D)$ "starts" at some finite time for this inverse filtering operation to be well-defined, and we assume the initial values are known.) Thus the sequence

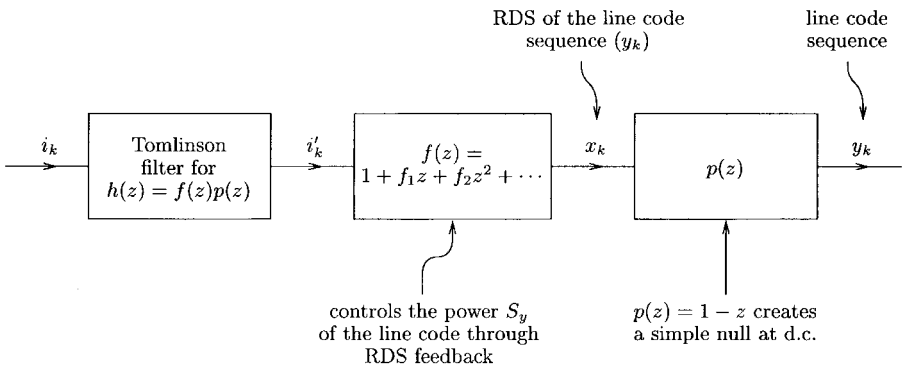

Fig. 6. Diagram for either partial-response signaling or signaling with spectral nulls.

$x(D)$ may be reconstructed as the "running digital sum" (RDS) of the PRC sequence $y(D)$. The spectra of the RDS sequence and PRC sequence are related by the partial-response transfer function expressed in the frequency domain. The order of the spectral null will be the order of the corresponding zero in $p(z)$. This number needs to be doubled to describe the order of the null in the actual power spectrum, which is proportional to $|p(e^{i\theta})|^2$.

We define the *RDS power* S_x as the sample variance of the RDS variables x_k, assuming sufficient stationarity (so that this notion is well-defined), and the *PRC power* S_y as the sample variance of the PRC variables y_k. Neither is necessarily larger than the other. Given $p(z)$, the problem is to choose $f(z)$ so as to minimize S_y subject to the requirement that S_x be held fixed. This will single out a one-parameter family of filters $f(z)$ indexed by the RDS power S_x. It is necessary to constrain $f(z)$, for otherwise the minimizing solution is $f(z) = 1/p(z)$ and the null disappears (the power S_x becomes infinite). Decreasing the width of a spectral null in the line-code spectrum requires a large peak at the appropriate frequency in $f(z)$, and hence large power S_x.

The new information in each symbol x_k is carried by the i.i.d. input i'_k to the filter $f(z)$. The power S of the sequence i'_k is the effective signal power at the output of a minimum mean-squared error (MMSE) predictor for the RDS sequence (x_k). For a single null at dc, Forney and Calderbank [84] show that the filter $f(z) = 1/(1-\beta z)$ gives the best possible tradeoff between the RDS power S_x and the line code power S_y. The optimum tradeoff is shown in Fig. 7 and is given by

$$4\left(\frac{S_x}{S}\right)\left(\frac{S_y}{S} - 1\right) = \left(\frac{S_y}{S}\right)^2.$$

The corresponding PRC spectra are shown in Fig. 8. As S_y approaches S, S_x necessarily increases without bound, and $H_y(\theta)$ becomes flatter with a sharper and sharper null at dc. These power spectra $H_y(\theta)$ are called "first-order power spectra" by Justesen [118], who considers them to be an interesting representative class of simple spectra for sequences with dc nulls, in that they remain small up to some cutoff frequency f_0 and then become approximately constant over the rest of the band. He notes that if f_0 is defined as the frequency at which $H_y(f_0) = S_y/2$ (the "half-power" frequency), then,

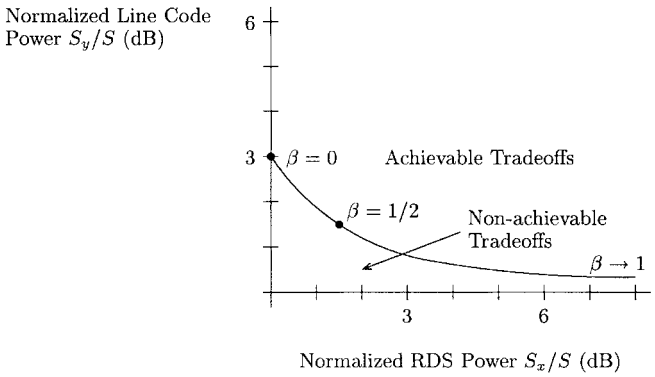

Fig. 7. Optimum tradeoff between S_x/S and S_y/S.

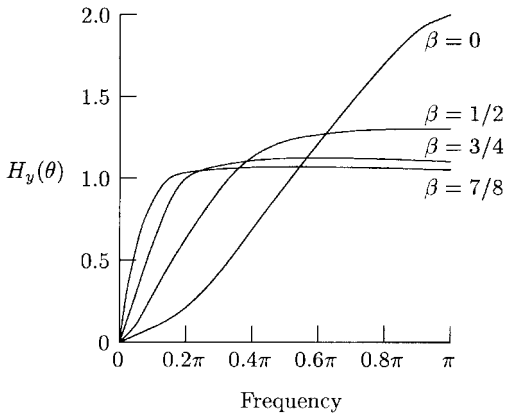

Fig. 8. PRC spectra for first-order autoregressive RDS sequences with parameter β.

for these first-order power spectra

$$\pi f_0 \simeq (S_y/2S_x) f_N$$

(where f_N is the upper Nyquist band-edge frequency), so that $\pi f_0/f_N \simeq 1-\beta$ (or $\theta_0 \simeq 1-\beta$), at least for $\beta \geq 1/2$.

The optimum tradeoff between S_x and S_y for sequences $\{x_k\}$ and $\{y_k\}$ that are related by $y(D) = x(D)p(D)$, where $p(D)$ is a response with arbitrary spectral nulls, was developed in subsequent work by Calderbank and Mazo [30]. Forney and Calderbank have shown that, at least for sequences supporting large numbers of bits per symbol, coset codes can be adapted to achieve effectively the same performance and complexity on

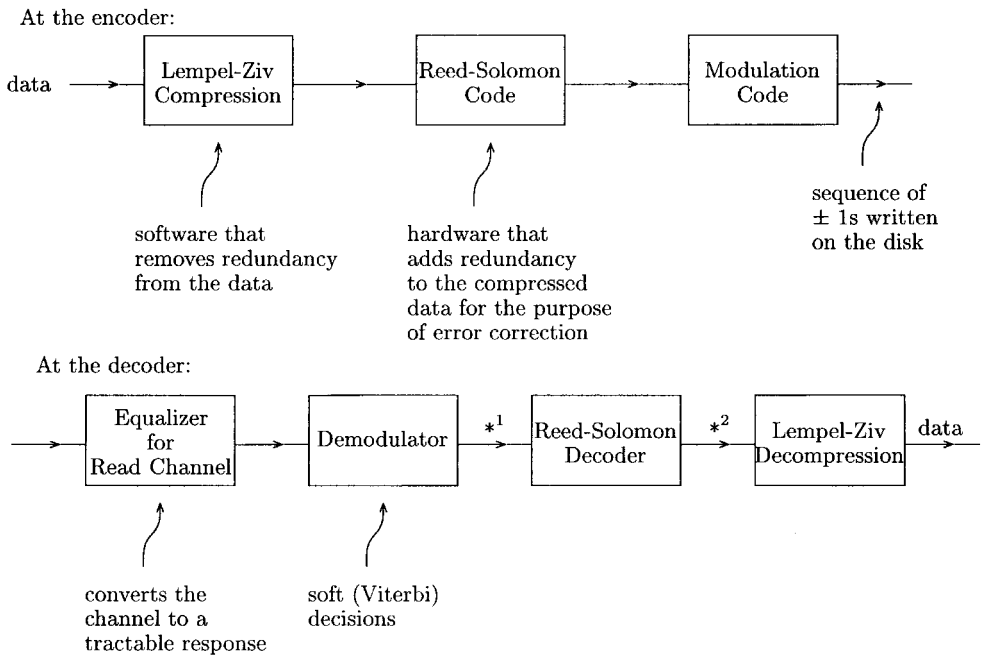

Fig. 9. Concatenation of an inner modulation code with an outer Reed–Solomon code. At $*^1$ the demodular provides a maximum-likelihood estimate of the $\pm$ valued sequence written on the disk, and the bit-error probability might be between 10^{-6} and 10^{-10} depending on the aggressiveness of the modulation strategy. At $*^2$ the bit-error probability needs to be 10^{-18}, that is, essentially error-free.

partial-response channels, or for sequences with spectral nulls, as they do in the ordinary memoryless case. This in addition to the optimum tradeoff between input and output powers.

N. Concatenated Codes

Applications of coding theory (see [45]) from deep-space communication to consumer electronics employ an inner modulation code with an outer algebraic error-correcting code (usually a Reed–Solomon code). Fig. 9 is a representation of a magnetic recording channel. For this application it is likely that in the next five years we will see full integration of demodulation and Reed–Solomon coding (a single-chip solution).

There are opportunities to use soft information calculated by the demodulator in Reed–Solomon decoding. For a small increase in decoder complexity it is possible either to provide reliability information about every demodulated data symbol, or to provide a list of the two or three best estimates of the ± 1-valued sequence written on the disk (see [184] and [106]). For the second alternative, the quantity of interest is the probability that the true write sequence is not among the list of two or three. This quantity may be recast as a decrease in bit-error probability; the old range $[10^{-10}, -10^{-6}]$ becomes $[10^{-14}, 10^{-8}]$, an improvement of about 1.5 dB for the list of three estimates. Both alternatives have the potential to simplify Reed–Solomon decoding, but it is not so easy in practice, and even the declaration of erasures is something of an art. It may in fact be more productive to focus on interpolating reliable symbols as in [203]. Staged decoding can provide additional coding gains of up to 1 dB in concatenated systems. For example, Hagenauer, Offer, and Papke [107] identify Reed–Solomon codewords that are correct with very high probability, and have the inner decoder treat the corresponding information bits as side information in a second round of decoding (*state pinning*). Particularly in magnetic recording, it can be advantageous to reverse the order of modulation and Reed–Solomon encoding (a systematic encoder is required). This reduces error propagation and can result in coding efficiencies (see [17] and [114]).

The theoretical foundations of concatenated coding are found in Forney [74], who showed that for polynomial decoding complexity, the error rate could be made to decrease exponentially with blocklength at any rate less than capacity. The notion of concatenated codes has been pursued with enthusiasm in the Russian literature, and there is a substantial commonality to the generalized cascade codes of Zinoviev [230], and the multilevel codes of Imai and Hirakawa [112]. In algebraic coding theory, Justesen [117] provided an explicit construction of a sequence of codes for which the rate and the normalized distance d/N are both bounded away from zero. For a long time prior to his construction there had been serious doubt as to whether this was possible. Now it is easy to show there exist field elements $\alpha_m \in \mathbb{F}_{2^m}$, so that the binary concatenated codes determined by pairs $(c, \alpha_m c)$, $c \in \mathbb{F}_{2^m}$ meet the Gilbert–Varshamov bound as $m \to \infty$. However, this is not an explicit construction. Justesen's idea was to consider pairs $((c_j), (\alpha_j c_j))$ where the field element α_j depends explicitly on the symbol c_j, but where variation in α_j from symbol to symbol provides the kind of performance attributable to random coding.

IV. Two Important Developments in Algebraic Coding Theory

Even in the 1977 edition of MacWilliams and Sloane there were 1478 references. Since it would be unwise to attempt a

comprehensive array of algebraic coding theory in the space available, we have chosen instead to highlight two developments of particular importance. The first is the geometric-mathematical framework of association schemes presented by Delsarte [48] that provides a common language for coding theory, statistical design, and algebraic combinatorics. The second is grounded in algorithms, and follows developments in cyclic codes through to the creation of algebraic-geometry codes that beat the Gilbert–Varshamov bound.

The theory of association schemes was inspired in part by the MacWilliams Identities, though it is the nonnegativity of the MacWilliams transform that is important, rather than the identity connecting the weight distribution of a linear code to that of the dual code (see [51]). It is these MacWilliams Inequalities that lead to the MRRW linear programming bound on codes, and to lower bounds on combinatorial designs and orthogonal arrays. Many notions of regularity in group theory, combinatorics, and statistics are expressed very naturally in terms of association schemes. For example, the study of distance regular graphs, now a large subject in its own right (see [21]), is the study of association schemes with a particular (P-polynomial) property.

We begin with a section that emphasizes approximation in the theory of combinatorial designs. The notion of approximation is one reason the theoretical computer science community has made extensive use of coding theory in recent years. In particular, codes have been used to design small sample spaces that approximate the behavior of large sample spaces, leading to bounds on the number of random bits used by probabilistic algorithms and the communications complexity of cryptographic protocols. From the perspective of computational complexity it is natural to view random bits as a resource analogous to time and space, and to design algorithms that require as few as possible. For details on this and other applications of coding theory to computational complexity see Feigenbaum [67].

A. Approximation, Combinatorial Designs, and the Johnson Scheme

The concept of approximation is similar but slightly different from that of quantization. The purpose of a design is to capture with a small ensemble the regularity properties of a much larger universe. Designs are concerned with approximating a universe closely, whereas codes are concerned with separating an ensemble widely. Questions in coding theory are packing problems, whereas questions in design theory are covering problems. There is a duality between packing and covering that can be made mathematically precise using the theory of association schemes.

An *association scheme* is a set X together with a partition of the two-element subsets of X into N classes $\Gamma_1, \cdots, \Gamma_N$ satisfying the following conditions:

1) given $x \in X$, the number v_i of points $y \in X$ with $\{x, y\} \in \Gamma_i$ depends only on i;
2) given $x, y \in X$ with $\{x, y\} \in \Gamma_k$, the number of points $z \in X$ with $\{x, z\} \in \Gamma_i$ and $\{y, z\} \in \Gamma_j$ is a constant p_{ij}^k that depends only on i, j, and k.

We may think of an association scheme on X as a coloring of the complete graph on X with colors $c_1, \cdots, c_N$ where an edge has color c_i if it belongs to Γ_i. The first condition asserts that each monochromatic graph (X, Γ_i) is regular. The second condition asserts that the number of triangles with a given coloring on a given base depends only on the coloring and not on the base.

The Johnson scheme $J(v, d)$ offers an algebraic means of quantifying the duality between packing and covering properties of d-subsets of a v-set. The point set Ω of $J(v, d)$ is the set of d-subsets of a v-set, these subsets intersect in $d+1$ possible ways, and the $d+1$ relations

$$R_i = \{(x, y) \mid |x \cap y| = d - i\}, \qquad i = 0, 1, \cdots, d$$

determine an association scheme. Starting from this simple observation, Delsarte [48] used the representation theory of the symmetric group and orthogonal polynomials to derive an algebraic foundation for extremal set theory.

The vector space $\mathbb{R}^\Omega$ consists of all mappings ψ from Ω to $\mathbb{R}$, and is invariant under the natural action of the symmetric group S_v. The irreducible S_v-invariant subspaces under this action are the harmonic spaces $\operatorname{harm}(i)$, $i = 0, 1, \cdots, d$, where $\dim(\operatorname{harm}(i)) = \binom{v}{i} - \binom{v}{i-1}$. The adjacency matrix D_i of the graph (Ω, R_i) is symmetric, and the relations

$$D_i D_j = \sum_k p_{ij}^k D_k$$

imply that the matrices $D_0 = I, D_1, \cdots, D_d$ span a $(d+1)$-dimensional commutative real algebra called the *Bose–Mesner algebra* of the Johnson scheme [20]. The adjacency matrices D_i commute with the natural action of the symmetric group, and Delsarte [48] proved that the $d+1$ eigenspaces common to the matrices D_i are in fact the harmonic spaces. Calderbank, Delsarte, and Sloane [25] constructed an explicit spanning set for each harmonic space $\operatorname{harm}(i)$. For every i-set A, let

$$f_A = \sum_{j=0}^{i} (-1)^j \binom{i}{j}^{-1} \binom{d-j}{i-j} \binom{v-i+1}{j} \sigma_j(A),$$

where $\sigma_j(A)$ is the sum of the characteristic functions of all j-subsets of A. As A ranges over every i-set the vectors f_A span $\operatorname{harm}(i)$.

The harmonic spaces are of combinatorial importance, because if the characteristic function of a family of d-subsets of a v-set is orthogonal to a harmonic space $\operatorname{harm}(j)$, then this family exhibits some regularity with respect to j-sets. To connect this viewpoint with classical design theory, we recall that a $t-(v, d, \lambda)$ design is a collection $\mathfrak{B}$ of subsets of a v-element set such that every member of $\mathfrak{B}$ contains d points and every subset of t points is in λ blocks. Here we are looking at the universe of d-point subsets, and we are approximating the regularity properties of this universe with respect to t-subsets of coordinates.

If ζ is an S_v-invariant subspace of $\mathbb{R}^\Omega$, then we can write

$$\zeta = \sum_{i \in T} \operatorname{harm}(i)$$

for some subset T of $\{0,1,\cdots,d\}$, where $\sum$ denotes orthogonal sum. There are 2^{d+1} such subspaces. Now let $\mathfrak{B}$ be a nonempty family of d-subsets of a v-set. A subspace ζ of $\mathbb{R}^\Omega$ will be said to be $\mathfrak{B}$-*regular* if it satisfies

$$\langle \pi(\mathfrak{B}), \psi\rangle = \frac{|\mathfrak{B}|}{|\Omega|}\langle \pi(\Omega), \psi\rangle, \qquad \text{for all } \psi \in \zeta$$

where $\pi(\,)$ is the characteristic function of a subset of Ω. Here we are thinking about a design as a way of approximating the statistics of the full ensemble Ω of d-subsets of a v-set, using only a proper subset $\mathfrak{B}$. The vector $\pi(\Omega)$ is the all-one function which spans $\mathrm{harm}\,(0)$. Orthogonality implies that the inner product $\langle \pi(\Omega), \psi\rangle$ vanishes for all $\psi \in \mathrm{harm}\,(j)$ with $j \geq 1$. It follows from the definitions that if ζ is S_v-invariant and $\mathfrak{B}$-regular then

$$\langle \pi(\mathfrak{B}), \psi\rangle = 0, \qquad \text{for all } \psi \in \mathrm{harm}\,(j), \text{ with } 0 \neq j \in T.$$

In this case we say $\mathfrak{B}$ is a T-*design* (when $0 \in T$, a T-design is defined to be a T'-design with $T' = T \setminus \{0\}$). The importance of this equation is that it shows the equivalence between the concepts of a T-design in $J(v,d)$ and an S_v-invariant $\mathfrak{B}$-regular subspace of $\mathbb{R}^\Omega$. The following theorem of Delsarte [48] makes the connection with classical design theory.

Theorem: A t-design in $J(v,d)$ is a T-design, where $T = \{1,2,\cdots,t\}$.

Let $\mathfrak{B}$ be a family of d-subsets of a v-set. *The inner distribution* $a = (a_0,\cdots,a_d)$ of $\mathfrak{B}$ is given by

$$a_i = |\mathfrak{B}|^{-1} \sum_{x,y\in\mathfrak{B}} D_i(x,y)$$

which is the average valency of the relation R_i restricted to $\mathfrak{B}$. The information carried by the inner distribution is packing information about the family $\mathfrak{B}$. The $d+1$ numbers in the inner distribution are all that is necessary to calculate the norm of the projection of the characteristic function $\pi(\mathfrak{B})$ on the harmonic spaces $\mathrm{harm}\,(i)$. These norms carry information about how subsets in $\mathfrak{B}$ cover the v points. This is what is meant by quantifying the duality between packing and covering.

Since the Bose–Mesner algebra is semisimple, it has a unique basis of minimal mutually orthogonal idempotent matrices $J_0,\cdots,J_d$. Here $J_0 = J/\binom{v}{d}$, where J is the matrix with every entry 1, and the columns of J_i span the harmonic space $\mathrm{harm}\,(i)$. If

$$D_l = \sum_{i=0}^{d} P_l(i)J_i, \qquad \text{for } l = 0,1,\cdots,d$$

then

$$D_l J_i = P_l(i)J_i$$

so that $P_l(i)$ is the eigenvalue of D_l on the harmonic space $\mathrm{harm}\,(i)$. The $(d+1)\times(d+1)$ matrix P with ilth entry $P_l(i)$ is called the *eigenmatrix* of the Johnson scheme. The eigenvalue $P_l(i) = E_l(i)$, where $E_l(x)$ is the Eberlein polynomial defined by

$$E_l(x) = \sum_{j=0}^{l}(-1)^j \binom{x}{j}\binom{v-x}{l-j}\binom{v-d-x}{l-j}, \qquad l = 0,1,\cdots,d.$$

For a proof, see Delsarte [48]. The matrix $Q = P^{-1}/\binom{v}{d}$, with ilth entry $Q_l(i)$ is called the *dual eigenmatrix*. Note that

$$J_l = \left(\sum_{i=0}^{d} Q_l(i)D_i\right)\Big/\binom{v}{d}.$$

The entry $q_l(i) = H_l(i)$, where $H_l(x)$ is the Hahn polynomial defined by

$$H_l(x) = \left[\binom{v}{l} - \binom{v}{l-1}\right] \times \sum_{i=0}^{l}\left\{(-1)^i\binom{l}{i}\binom{v+1-l}{i}\binom{d}{i}^{-1}\binom{v-d}{i}^{-1}\right\}\binom{x}{i}, \qquad l = 0,1,\cdots,d.$$

Again we refer the reader to Delsarte ([48] or [50]) for a proof.

Given a family $\mathfrak{B}$ of d-subsets of a v-set, the *dual distribution* $b = (b_0,b_1,\cdots,b_d)$ is given by

$$b_i = \frac{|\Omega|}{|\mathfrak{B}|^2}\pi(\mathfrak{B})J_i\pi(\mathfrak{B})^T = \frac{|\Omega|}{|\mathfrak{B}|^2}||\pi_i||^2$$

where π_i is the orthogonal projection of $\pi(\mathfrak{B})$ onto the eigenspace $\mathrm{harm}\,(i)$.

Theorem ([48]): The inner distribution a, and the dual distribution b are related by

$$aQ = |\mathfrak{B}|b$$

where Q is the dual eigenmatrix of the Johnson scheme.

Proof: We have

$$\left(\frac{1}{|\mathfrak{B}|}aQ\right)_l = \frac{1}{|\mathfrak{B}|}\sum_{i=0}^{d} Q_l(i)a_i = \frac{1}{|\mathfrak{B}|^2}\pi(\mathfrak{B})\left(\sum_{i=0}^{d} Q_l(i)D_i\right)\pi(\mathfrak{B})^T = b_l,$$

as required. □

It is also possible to capture the regularity properties of a design $\mathfrak{B}$ through analysis of invariant linear forms. With any t-subset x of V and any integer $i \in [0,t]$ we associate the number $D_i(x)$ that counts the blocks in $\mathfrak{B}$ meeting x in $t-i$ points. Suppose that for all t-subsets x, we have a linear relation

$$f_0D_0(x) + f_1D_1(x) + \cdots + f_tD_t(x) = c$$

where $f_0,f_1,\cdots,f_t$ and c are fixed real numbers. Then we say that the $(t+1)$-tuple $(f_t)_{i=0}^t$ is a t-form for $\mathfrak{B}$. The set of t-forms clearly is a vector space, which will be called the *t-form space* of $\mathfrak{B}$. The dimension of the t-form space measures regularity of $\mathfrak{B}$ with respect to t-subsets, and when $\mathfrak{B}$ is a classical t-design, the t-form space coincides with

$\mathbb{R}^{t+1}$. Calderbank and Delsarte [24] have shown that the t-form space is completely determined by the inner distribution of $\mathfrak{B}$, and that the invariant t-forms can be calculated via a matrix transform that involves a system of dual Hahn polynomials. For example, the inner distribution of octads in the binary Golay code is $(1,0,0,0,280,0,448,0,30)$ and the 7-form space can be generated from the particular 7-form

$$D_0(x) + D_1(x) = 1.$$

It is interesting to note that given any collection of 8-element subsets of a 24-set for which this particular 7-form is invariant, the linear span must be the binary Golay code.

The fundamental question in design theory is usually taken to be: Given v, d, λ does there exist a $t-(v,d,\lambda)$ design? This is certainly a natural question to ask from the perspective of small geometries, but it does not involve the idea of approximation in an essential way. Designs play an important role in applied mathematics and statistics and this author would suggest that questions involving fundamental limits on the quality of approximation are more important than questions involving existence of individual designs.

One of the strengths of the association scheme approach to designs is that it allows arbitrary vectors in $\mathbb{R}^{\Omega}$, not just the characteristic vectors of collections of d-sets, in particular it includes signed designs [177].

We mention briefly an elegant application to extremal set theory that was inspired by Delsarte's thesis. A family $\mathfrak{B}$ of d-element subsets of v-set is called t- *intersecting* if $|B \cap B'| \geq t$ for all $B, B' \in \mathfrak{B}$. The problem of determining the maximum size of t-intersecting families goes back to Erdős, Ko, and Rado [66] who proved the following theorem.

Theorem: Suppose that $\mathfrak{B}$ is a t-intersecting family with $v \geq v_0(d,t)$. Then

$$|\mathfrak{B}| \leq \binom{v-t}{d-t}.$$

The bound is obviously best possible, since we may take $\mathfrak{B}$ to be all d-subsets containing a fixed t-element subset. The best possible value of $v_0(d,t)$ is $(d-t+1)(t+1)$, as was shown by Frankl [89] for $t \geq 15$, and then by Wilson [225] for all t. The eigenvalues of the adjacency matrices D_i are a little difficult to work with, and Wilson used instead the matrices $S(j)$, with rows and columns indexed by d-sets, and where the (A,C) entry counts j-subset B for which $A \cap B = \emptyset$ and $B \subseteq C$. Thus $S(j)$ is a linear combination of $D_d, \cdots, D_{d-j}$ and the eigenvalues turn out to be

$$\lambda(i,j) = (-1)^i \binom{d-i}{j-i}\binom{v-j-i}{d-i}$$

with multiplicity $\binom{v}{i} - \binom{v}{i-1}$. It is interesting to note that when $t=1$, it is easy to prove the Erdős–Ko–Rado theorem using this algebraic framework. An intersecting family determines a principal submatrix of $B(0) = D_d$ that is identically zero, and the size of this submatrix is bounded above by $\binom{v}{d} - \max(l_+, l_-)$, where $l_+(l_-)$ is the number of positive (negative) eigenvalues. We obtain

$$\begin{aligned}|\mathfrak{B}| &\leq \binom{v}{d} \\ &\quad - \left[\binom{v}{d} - \binom{v}{d-1} + \binom{v}{d-2} - \binom{v}{d-3} + \cdots\right] \\ &= \binom{v}{d} - \binom{v-1}{d} \\ &= \binom{v-1}{d-1}\end{aligned}$$

as required. □

We now consider 2-designs in greater detail. If b denotes the number of blocks, and if r denotes the number of blocks containing a given point, then the identities

$$bk = vr \quad \text{and} \quad r(k-1) = (v-1)\lambda$$

restrict the possible parameter sets. These identities are trivial in that they are obtained by elementary counting arguments. It is natural to impose the restriction $k < v$, and in this case we have Fisher's inequality $b \geq v$. Designs with $b = v$ are called symmetric designs. In a symmetric design there is just one intersection number; two distinct blocks always intersect in λ points. Conversely, it is easily shown that a 2-design with one intersection number is a symmetric design. The Bruck–Ryser–Chowla theorem provides a nontrivial restriction on the parameter sets of symmetric designs. Here "nontrivial" means an algebraic condition that is not a consequence of simple counting arguments. The Bruck–Ryser–Chowla theorem also provides a connection between the theory of designs and the algebraic theory of error-correcting codes. The row space of the incidence matrix of a symmetric design determines a self-dual code with respect to some nondegenerate scalar product. The restrictions provided by the theorem are necessary conditions for the existence of these self-dual codes (see Lander [128], Blokhuis and Calderbank [18]).

B. Algebraic Coding Theory and the Hamming Scheme

The Hamming scheme $H(N,q)$ is an association scheme with N classes. The point set X is $\mathbb{F}_q^N$, and a pair of vectors $\{x,y\}$ is in class Γ_i if the Hamming distance $D(x,y) = i$. The adjacency matrices D_i of the graph $(\mathbb{F}_q^N, \Gamma_i)$ generate the Bose–Mesner algebra of the scheme, and there is a second basis $J_0, \cdots, J_N$ of mutually orthogonal idempotent matrices. The two bases are related by

$$D_l = \sum_{i=0}^{N} K_l(i) J_i$$

$$q^N J_l = \sum_{i=0}^{N} K_l(i) D_i,$$

where

$$K_l(z) = \sum_{j=0}^{l} (-1)^j (q-1)^{l-j} \binom{z}{j}\binom{n-z}{l-j}$$

is the lth Krawtchouk polynomial. Recall that $K_l(z)$ is the coefficient of λ^l in

$$(1-\lambda)^x(1+(q-1)\lambda)^{N-x}.$$

In this association scheme, the eigenmatrix P and the dual eigenmatrix Q are identical.

The inner distribution $a = (a_0, \cdots, a_N)$ of a code C is called the *distance distribution*, and the entry a_i is the average number of codewords at distance i from a given codeword. If C is linear then a is simply the weight distribution. The dual distribution $b = (b_0, \cdots, b_N)$ is given by $aQ = |C|b$ which we expand as

$$b_l = \frac{1}{|C|}\sum_{i=0}^{N} a_i K_l(i), \qquad l = 0, 1, \cdots, N.$$

For linear codes C we recognize these equations as the MacWilliams Identities [146], [147] that relate the weight enumerator of a linear code to that of the dual code $C^\perp$. A little rearrangement gives

$$\begin{aligned}|C^\perp| \sum_{x\in C^\perp} \lambda^{\mathrm{wt}\,(x)} &= \sum_i \sum_l a_i K_l(i)\lambda^l \\ &= \sum_i a_i (1-\lambda)^i (1+(q-1)\lambda)^{N-i}\end{aligned}$$

which is the single variable form of the MacWilliams Identities. It is sometimes more convenient to associate to a linear code C a weight enumerator in two variables. Then

$$W_C(x,y) = \sum_{c\in C} x^{N-\mathrm{wt}\,(c)} y^{\mathrm{wt}\,(c)}$$

and the MacWilliams Identities take the form

$$W_{C^\perp}(x,y) = \frac{1}{|C|} W_C(x+(q-1)y, x-y).$$

There are several families of nonlinear codes that have more codewords than any comparable linear code presently known. These are the Nordstrom–Robinson, Kerdock, Preparata, Goethals, and Delsarte–Goethals codes [52], [98], [99], [122], [164], and [174]. Aside from their excellent error-correcting capabilities, these pairs of codes (Kerdock/Preparata and Goethals/Delsarte–Goethals) are remarkable in the sense that although these codes are nonlinear, the weight distribution of one is the MacWilliams transform of the weight distribution of the other code in the pair. Hammons *et al.* [108] provide an algebraic explanation by showing that there is a natural definition of Kerdock and Preparata codes as linear codes over $\mathbb{Z}_4$, and that as $\mathbb{Z}_4$ codes they are duals. The mystery of the weight distributions is resolved by observing that ($\mathbb{Z}_4^N$, Lee distance) and ($\mathbb{F}_2^{2N}$, Hamming distance) are isometric (see Subsection IV-D), and that there is an analog of the standard MacWilliams Identities for codes in the Lee metric. There are in fact a number of different association schemes and MacWilliams Identities that are useful in coding theory. Delsarte and Levenshtein [55] mention five, including the association scheme relative to the split-weight enumerator.

There is a great deal of interesting mathematics associated with self-dual codes. The weight enumerator $W_C(x,y)$ of a binary self-dual code C with all weights divisible by 4 is invariant under the transformations

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

These transformations generate a group containing 192 matrices, and Gleason [97] used a nineteenth century technique called invariant theory to prove that $W_C(x,y)$ is a polynomial in the weight enumerators of the $[8,4,4]$ Hamming and $[24,12,8]$ Golay codes. An immediate corollary is that the blocklength N is divisible by 8. More details and generalizations can be found in a very nice survey by Sloane [201]. There is also a very fruitful connection between self-dual binary codes with all weights divisible by 4 and even unimodular lattices. In fact, there are parallel theorems giving upper and lower bounds on the best codes and lattices, and parallel characterizations of the weight enumerator of the code and the theta series of the lattice (see [44, Ch. 7]).

The most important theorem relating codes and designs is the Assmus–Mattson theorem. The statement of this theorem given below differs from the statement given elsewhere (for example, in Assmus and Mattson [2] or MacWilliams and Sloane [148]) where the conclusion applies only to codewords of sufficiently low weight. This restriction is to exclude designs with repeated blocks. Since we mean to allow t-designs with repeated blocks, we may drop the extra restriction.

Theorem (Assmus–Mattson): Let C be a linear $[v,k,d]$ code over $\mathbb{F}_q$, where the weights of the nonzero codewords are $w_1 = d, w_2, \cdots, w_s$. Let $d', w_2', \cdots, w_{s'}'$ be the nonzero weights in $C^\perp$. Let t be the greatest integer in the range $0 < t < d$, such that there are at most $d-t$ weights w_i' with $0 < w_i' \le v - t$. Then the codewords of any weight w_i in C form a t-design.

The theorem is proved using the MacWilliams Identities. We puncture the code by deleting coordinates $p_1, \cdots, p_t$ to obtain a code $\hat{C}$. The code $\hat{C}^\perp$ is obtained by taking codewords c in $C^\perp$ with $c_{p_1} = c_{p_2} = \cdots = c_{p_t} = 0$ and deleting these coordinates. The MacWilliams Identities allow us to solve for the weight distribution of $\hat{C}$ and the solution is independent of the choice of $p_1, \cdots, p_t$.

Delsarte [49] identified four fundamental properties of a code or more generally, a subset of an association scheme:

- the *minimum distance* d;
- the *degree* s, which is the number of nonzero entries a_i in the inner distribution, not counting $a_0 = 1$;
- the *dual distance* d', which is the index i of the first nonzero entry b_i in the dual distribution, not counting $b_0 = 1$;
- the *dual degree* s' (sometimes called the *external distance*), which is the number of nonzero entries b_i in the dual distribution, not counting $b_0 = 1$.

There is also the (*maximum*) *strength* t which is $d'-1$. In the Johnson scheme, a subset of strength t is a t-design.

The combinatorial significance of the external distance s' is understood through the characteristic polynomial of a code C, which is given by

$$\alpha(\xi) = \frac{q^N}{|C|} \prod_{\substack{1 \le i \le N \\ b_i \neq 0}} \left(1 - \frac{\xi}{i}\right).$$

We expand the shifted polynomials $\xi^m \alpha(\xi)$ in terms of Krawtchouk polynomials

$$\xi^m \alpha(\xi) = \sum_{i=0}^{s'+m} \alpha_i^m K_i(\xi).$$

Now, given a vector $z \in \mathbb{F}_q^N$, let $\beta_i(z)$ be the number of codewords $c \in C$ for which $D(c,z) = i$. Delsarte [48] proved

$$\sum_{i=0}^{s'+m} \alpha_i^m \beta_i(z) = \begin{cases} 1, & \text{if } m = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Taking $m = 0$, we see that the covering radius of C is bounded above by the external distance.

If the minimum distance d is greater than the external distance s', then the coefficients $\alpha_0^0 = \alpha_1^0 = \cdots = \alpha_{d-s'-1}^0 = 1$. For $i < d - s'$, this is proved by choosing $x \in \mathbb{F} - q^N$ at distance i from some $c \in C$. Then by the triangle inequality, every other $c' \in C$ is at distance greater than s' from x. Since $\mathfrak{B}_i(x) = 1$, and $\mathfrak{B}_j(x) = 0$ for $j \le s'$, $j \neq i$ we have $\alpha_i = 1$. This leads to the famous characterization of perfect codes mentioned in Section II.

Theorem: Let C be a code with minimum distance d and external distance s', and let $e = \lfloor (d-1)/2 \rfloor$. Then

$$\sum_{i=0}^{e} \binom{N}{i} (q-1)^i \le \frac{q^N}{|C|} \le \sum_{i=0}^{s'} \binom{N}{i} (q-1)^i.$$

If one of the bounds is attained, then so is the other, the code is perfect, and its characteristic polynomial is

$$\psi_e(z) = \sum_{l=0}^{e} K_l(z).$$

This result is named for Lloyd who obtained the theorem for $q = 2$ by analytic methods prior to the discovery of the MacWilliams Identities. For comparison, the characteristic polynomial of a uniformly packed code is

$$\sum_{l=0}^{e-1} K_l(z) + (1 - \lambda/\mu) K_e(z) + 1/\mu K_{e+1}(z).$$

The problem of finding good upper bounds on the size of a code with minimum distance d can be expressed as a linear program. We treat the entries a_i of the inner distribution as real variables, and we look to maximize the sum $\sum_{i=0}^{N} a_i$ under the linear constraints

$$\begin{aligned} &a_0 = 1,\ a_i = 0, && \text{for } i = 1, \cdots, d-1 \\ &a_i \ge 0, && \text{for } i = d, \cdots, N \\ &\sum_{i=0}^{N} a_i K_l(i) \ge 0, && \text{for } l = 1, \cdots, N. \end{aligned}$$

It has in fact proved more convenient to attack the dual minimization problem. Here we look for a polynomial $F(z)$ of degree at most N, where the coefficient f_k in the Krawtchouk expansion

$$F(z) = \sum_{l=0}^{N} f_l K_l(z)$$

are nonnegative, where $f_0 > 0$, and where $F(i) \le 0$ for $i = d, \cdots, N$. The size of any code with minimum distance d is bounded above by $F(0)/f_0$. The McEliece, Rodemich, Rumsey, Welch (MMRW) bound [156] results from polynomials

$$F(z) = \frac{1}{c-z} (K_t(c) K_{t+1}(z) - K_{t+1}(c) K_t(z))^2$$

where $1 \le t \le \lfloor (N-1)/2 \rfloor$, and c is an appropriately chosen real number. For binary codes the rate R satisfies

$$R \le H_2\left(1/2 - \sqrt{\left[\frac{d}{N}\left(1 - \frac{d}{N}\right)\right]}\right)(1 + o(N)).$$

Strengthening the dual problem by requiring $F(x) \le 0$ for $d \le x \le N$ gives a new problem where the minimizing polynomial can be found explicitly [136], [194]. However, the asymptotics of the solution coincide with the MRRW bound.

A second application of linear programming is to bounding zero-error capacity of a discrete memoryless channel, a concept introduced by Shannon [188] in 1956. Here the input alphabet becomes the vertex set of a graph, and two vertices are joined if the action of noise cannot result in the corresponding symbols being confused at the output of the channel. The problem of determining the zero-error capacity of the pentagon remained unsolved for some 20 years until the linear programming solution by Lovász [144].

The combinatorial significance of the dual distance d' is understood in terms of variation in the inner distribution of translates of C. For example, a code C is said to be *distance-invariant* if the number of codewords at distance i from a given codeword depends only on i and not on the codeword chosen. Linear codes are distance-invariant, as are the binary images of linear codes over $\mathbb{Z}_4$ after applying the Gray map (for example, the Kerdock and Preparata codes). Delsarte [48] proved that a sufficient condition for distance invariance is that the degree s is at most the dual distance d'. The argument rests on degrees of freedom in the MacWilliams transform. If $d' \ge s$ then there is no variance in the distance distribution of translates $C + x$ where $D(x, C) \le d' - s$ is constant (for details see [55] or [38]).

We have seen how the dual degree s' and the minimum distance d can be used to provide upper bounds on the size of codes. We now describe how the degree s and the dual distance d' can be used to provide lower bounds on the size of designs. Given a subset Y of $\mathbb{F}_q^N$, we form the array where the rows are the words in Y. The subset Y is an *orthogonal array of strength t and index λ* if, in each t-tuple of distinct columns of the array, all t-tuples of symbols appear exactly λ

times. Clearly, $|Y| = \lambda q^t$. This what it means to be a design in the Hamming scheme. The two notions of strength coincide, and this is evident when Y is linear.

The counterpart to the characteristic polynomial is the *annihilator polynomial* given by

$$\gamma(\xi) = |Y| \prod_{\substack{1 \le i \le N \\ a_i \neq 0}} \left(1 - \frac{\xi}{i}\right)$$

which we expand in terms of Krawtchouk polynomials

$$\gamma(\xi) = \sum_{l=0}^{s} \gamma_l K_l(\xi).$$

If the maximum strength t is at least the degree s, then the coefficients $\gamma_0 = \cdots = \gamma_{t-s} = 1$. The counterpart of the previous theorem is the following.

Theorem: Let Y be a design with degree s and maximum strength t, and let $f = \lfloor t/2 \rfloor$. Then

$$\sum_{i=0}^{f} \binom{N}{i} (q-1)^i \le |Y| \le \sum_{i=0}^{s} \binom{N}{i} (q-1)^i.$$

If one of the bounds is attained, then so is the other, the design is called *tight* and the annihilator polynomial is

$$\psi_f(z) = \sum_{l=0}^{f} K_l(z).$$

This is the Rao bound [176] for orthogonal arrays of strength t. The corresponding theorem in the Johnson scheme is the Ray–Chaudhuri/Wilson bound for tight designs [178] ($2s$-designs with s different block intersection sizes). For $s > 1$ the only known example is the set of minimum-weight codewords in the perfect binary Golay code.

C. Spherical Codes and Spherical Designs

We begin in real Euclidean space with a mathematical criterion that measures how well a sphere is approximated by a finite point set. Let $\Omega = \{P_1, \cdots, P_M\}$ be a set of M points on the unit sphere

$$S^{N-1} = \{x = (x_1, \cdots, x_N) \in \mathbb{R}^N \mid x \cdot x = 1\}.$$

Then Ω is a *spherical t-design* if the identity

$$\int_{\Omega_d} f(x) d\mu(x) = \frac{1}{M} \sum_{i=1}^{N} f(P_i)$$

(where μ is uniform measure on S^{N-1} normalized to have total measure 1) holds for all polynomials f of degree $\le t$.

For example, a soccer ball is a truncated icosahedron rather than a perfect sphere, and the 60 vertices of the soccer ball form a spherical 5-design. Goethals and Seidel [100] improved upon the standard soccer ball by slightly perturbing the vertices so as to produce a spherical 9-design. This is a very particular spherical design. Seymour and Zaslavsky [186] proved that for any positive integers N and t, and for all sufficiently large M, there exist spherical t-designs of size M in $\mathbb{R}^N$. This result is a remarkable generalization of the mean value theorem and is not constructive.

There are strong structural similarities between the Euclidean sphere S^{N-1} and the Hamming and Johnson schemes. All are *distance-transitive* in the sense that given points x, y, x', y' the distances $D(x,y)$, $D(x',y')$ are equal if and only if there is an isometry g for which $x^g = x'$ and $y^g = y'$. For the Euclidean sphere, isometries are simply orthogonal transformations. Delsarte, Goethals, and Seidel [54] showed that the earlier method of deriving lower bounds on designs remains valid, though the particular orthogonal polynomials are different. Also see [55] for more details.

Delsarte, Goethals, and Seidel [53] also derived upper bounds on the cardinality of sets of lines having prescribed angles both in $\mathbb{R}^N$ and $\mathbb{C}^N$. The inner products between unit vectors in the different lines determine the inner distribution of these spherical codes. Given a spherical code Ω let $A = \{|(a,b)|^2 \mid a \neq b \in \Omega\}$. For $s \in \{0,1\}$ and integers $k \ge 0$, the Jacobi polynomial $Q_{k,s}(x)$ in the real variable x is defined by a three-term recursion that depends on the choice of field.

Theorem [53]: For any $\epsilon > 0$, let $F(x)$ be a polynomial satisfying $\alpha^s F(\alpha) \le 0$ for all $\alpha \in A$, $f_{k,\epsilon} \ge 0$ for all $k \ge 1$, and $f_{0,\epsilon} > 0$, where $f_{k,s}$ is the coefficient of $Q_{k,s}$ in the Jacobi expansion of $F(x)$. Then

$$|\Omega| \le F(1)/f_{0,\epsilon}.$$

This theorem provides upper bounds on the size of families of sequences with favorable correlation properties that are used in spread-spectrum communication. For instance, there is an interesting example involving Kerdock codes. Cameron and Seidel [37] used quadratic forms on $\mathbb{Z}_2^{m+1}$ to construct a family of lines through the origin of $\mathbb{R}^N$, where $N = 2^{m+1}$, such that any two lines are perpendicular or at an angle θ where $\cos\theta = 1/\sqrt{N}$. These line sets are the union of $N/2$ frames corresponding to cosets of the first-order Reed–Muller code in the Kerdock code. König [126] and Levenshtein [135] observed that adding the standard coordinate frame did not increase the set of prescribed angles, and that the augmented system of lines met an upper bound derived from the above theorem. The $\mathbb{Z}_4$-linear Kerdock code determines an extremal system of lines in complex space (see [23]).

D. From Cyclic Codes to Algebraic-Geometry Codes

We take the perspective of frequency-domain techniques particular to finite fields. The notions of time and frequency domain for codes defined over finite fields, and the idea of using constraints in the frequency domain to separate codewords in the time domain are of fundamental importance. This is the foundation for the Reed–Solomon codes that are found everywhere today, from computer disk drives to CD players.

The early theory of cyclic codes was greatly influenced by a series of reports written mostly by Assmus, Mattson, and Turyn in the 1960's and early 1970's. They were much quoted and used extensively by van Lint [139] in his first book on coding theory. These reports were much influenced by the monthly meetings on coding theory held first at Hanscom Field

then at Sylvania involving Assmus, Gleason, Mattson, Pierce, Pless, Prange, Turyn, and the occasional distinguished visitor.

We begin by observing that the binary $[2^m-1, 2^m-m-1, 3]$ Hamming code may be defined as the collection of binary vectors $(a_0, a_1, \cdots, a_{2^m-2})$ that satisfy

$$\sum_{i=0}^{2^m-2} a_i \alpha^i = 0$$

where α is a primitive (2^m-1)th root of unity in the extension field $\mathbb{F}_{2^m}$. (Recall that the Hamming code is the unique binary $[2^m-1, 2^m-m-1, 3]$ code, and the new definition certainly determines a code with these parameters.) We may think of the matrix

$$\left[1, \alpha, \alpha^2, \cdots, \alpha^{2^m-2}\right]$$

as a parity-check matrix for this Hamming code and increase minimum distance by adding a second spectral constraint:

$$\begin{bmatrix} 1, & \alpha, & \alpha^2, & \cdots, & \alpha^{2^m-2} \\ 1, & \alpha^3, & \alpha^6, & \cdots, & \alpha^{3(2^m-2)} \end{bmatrix}.$$

This is the parity-check matrix for the two-error-correcting BCH code. More generally we may define a *BCH code with designed distance* d by means of the parity-check matrix

$$H = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{2^m-2} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(2^m-2)} \\ 1 & \alpha^3 & \alpha^6 & \cdots & \alpha^{3(2^m-2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^{d-2} & \alpha^{2(d-2)} & \cdots & \alpha^{(d-2)(2^m-2)} \end{bmatrix}.$$

Note that the rows of H are not linearly independent: some spectral constraints are inferred by others: for example, $\sum_{i=0}^{2^m-2} a_i \alpha^i = 0$ implies $\sum_{i=0}^{2^m-2} a_i \alpha^{2i} = 0$. The assertion that the minimum distance is at least d amounts to proving that every set of $d-1$ columns is linearly independent. This is a Vandermonde argument.

The Hamming code and the BCH codes with designed distance d are examples of cyclic codes. These codes play an important role in coding practice, and are good in the sense that there are cyclic codes that meet the Gilbert–Varshamov bound. A linear code is *cyclic* if the set of codewords is fixed by a cyclic shift of the coordinates: if $(c_0, \cdots, c_{N-1})$ is a codeword, then so is $(c_{N-1}, c_0, \cdots, c_{N-2})$. To verify that the above codes are indeed cyclic, we apply the identity

$$\alpha^\ell \sum_{i=0}^{2^m-2} a_i \alpha^{\ell i} = \sum_{i=0}^{2^m-2} a_{i+1} \alpha^{\ell i}$$

where subscripts are read modulo 2^m-1. The theory of cyclic codes identifies the sequence $(a_0, a_1, \cdots, a_{N-1})$ with the polynomial $a_0 + a_1 x + \cdots + a_{N-1} x^{N-1}$. Cyclic codes then correspond to ideals in the residue class ring $\mathbb{F}_2[x]/(x^N-1)$, and the structure theory of principal ideal rings can be brought to bear. It is also possible to approach cyclic codes through a discrete analog of the Fourier transform called the *Mattson–Solomon polynomial* [154]. The vector

$$a(x) = a_0 + a_1 x + \cdots + a_{N-1} x^{N-1}$$

is represented by the polynomial $A(x) = \sum_{j=1}^{N} A_j X^N$ where

$$A_j = a(\alpha^j) = \sum_{i=0}^{N-1} a_i \alpha^{ij}.$$

The BCH code with designed distance $d = 2\delta$ is then the set of all vectors a for which $A_1 = A_2 = \cdots = A_{\delta-1} = 0$. VLSI implementation of Reed–Solomon decoding has inspired a great deal of creativity regarding effective computation in finite fields, for example Berlekamp's bit-serial multiplication circuits. For an introduction to this area see McEliece [155], and note the dedication to Solomon.

1) Cyclic Codes Obtained by Hensel Lifting: A binary cyclic code C is generated by a divisor $g_2(x)$ of x^N-1 in $\mathbb{F}_2[x]$. Hensel's Lemma allows us to refine a factorization $x^N - 1 = g_2(x)h_x(x)$ modulo 2, to a factorization $x^N - 1 = g_{2^a}(x)h_{2^a}(x)$ modulo 2^a, and to a factorization $x^N - 1 = g(x)h(x)$ over the 2-adic integers. The polynomial $g_{2^a}(x)$ generates a cyclic code C_{2^a} over the ring of integers $\mathbb{Z}_{2^a}$, and the polynomial $g(x)$ generates a cyclic code C_∞ over the 2-adic integers. The codes over $\mathbb{Z}_{2^a}$ can also be described in terms of parity checks involving Galois rings, and this is completely analogous to the construction of binary cyclic codes through parity checks involving finite fields.

A very striking theorem of McEliece (generalized to Abelian codes in [56]) characterizes the possible Hamming weights that can appear in a binary cyclic code C in terms of l, the smallest number such that l nonzeros of C (roots of $h_2(x)$) have product 1. The characterization is that all Hamming weights are divisible by 2^{l-1}, and there is a weight not divisible by 2^l. Though this theorem has been generalized to cyclic codes obtained by Hensel lifting [28] there remains the possibility of using the codes C_{2^a}, C_∞ to infer additional properties of C. We might, for example, hope to resolve the deceptively innocent question of given two m-sequences, whether or not -1 must appear as a crosscorrelation value.

A special case of particular interest is cyclic codes over $\mathbb{Z}_4$ that are obtained from binary cyclic codes by means of a single Hensel lift. It will be of interest to characterize the possible Lee weights that can appear in this cyclic code. Recall the the *Lee weights* of the elements $0, 1, 2, 3$ of $\mathbb{Z}_4$ are, respectively, $0, 1, 2, 1$ and that the Lee weight of a vector in $\mathbb{Z}_4^N$ is just the rational sum of the Lee weights of its components. This weight function defines the *Lee metric* on $\mathbb{Z}_4^N$. If we imagine $0, 1, 2, 3$ as labeling (clockwise) four equally spaced points on a circle, then Lee distance is distance around this circle. The Lee metric is important because there is a natural isometry from ($\mathbb{Z}_4^N$, Lee Metric) to ($\mathbb{F}_2^{2N}$, Hamming Metric) called the Gray map. This map ϕ is defined from $\mathbb{Z}_4$ to $\mathbb{Z}_2$ by

$$\phi(0) = (00) \quad \phi(1) = (01) \quad \phi(2) = (11) \quad \phi(3) = (10)$$

and is extended in the obvious way to a map ϕ from $\mathbb{Z}_4^N$ to $\mathbb{F}_2^{2N}$. It is evidently distance preserving. Hammons *et al.* [108] proved that the Gray image of the Hensel lift of the first-order Reed–Muller code RM $(1, m)$ is the Kerdock code [122]. The Gray image of the Hensel lift of the extended Hamming code differs slightly from the standard Preparata code [174], but

shares the same distance structure. The Kerdock, Preparata, and Delsarte–Goethals codes are nonlinear binary codes, defined via quadratic forms, that contain more codewords than any linear code presently known. What remains mysterious is how to construct efficient linear codes over $\mathbb{Z}_4$ that correct more than three errors by specifying parity checks involving Galois rings. We do not have any counterpart to the BCH, Hartmann–Tzeng, and Roos bounds for classical cyclic codes (for a unified approach to these bounds see [140]).

2) Algebraic-Geometry Codes: The last 20 years have seen the construction of algebraic-geometry codes that can be encoded and decoded in time polynomial in the blocklength N, and with performance that matches or exceeds the Gilbert–Varshamov bound. This was proved by Tsfasman, Vlăduţ, and Zink [211] for finite fields $\mathbb{F}_q$, where q is a square and $q \geq 49$, but this numerical restriction on q may not be essential. It was and is a spectacular result, so spectacular that it motivated many mathematicians to learn some coding theory, and many engineers to learn some algebraic geometry. The consequence has been a fascinating combination of abstract geometry and efficient computational methods that has been described in a number of excellent surveys and introductory articles, for example, [110], [204], and [16].

We begin by describing the codes proposed by Reed and Solomon [179], that are now found everywhere from computer disk drives to CD players. Even these codes did not go into use immediately because fast digital electronics did not exist in 1960. Consider the vector space

$$L_r = \{f(z) \in \mathbb{F}_q[z] \mid \deg f \leq r\}$$

of polynomials with coefficients in the field $\mathbb{F}_q$ and degree at most r. Let $\alpha_1, \cdots, \alpha_N$ be distinct elements of $\mathbb{F}_q$, and define the evaluation map

$$\mathrm{ev}\,(f) = (f(\alpha_1), \cdots, f(\alpha_N)).$$

The evaluation map is linear, and if $r < N$ it is $1-1$. The image of L_r is a Reed–Solomon code with dimension $r+1$ and minimum distance $d = N-r$. Reed–Solomon codes are optimal in the sense that they meet the Singleton bound $d \leq n - k + 1$. The only drawback is that the length N is constrained by the size of the field $\mathbb{F}_q$, though this constraint can be removed by passing to general BCH codes.

The construction of Reed–Solomon codes can be generalized by allowing polynomials $f(z_1, \cdots, z_m)$ in several variables, and by evaluating these polynomials on a subset of the affine space $\mathbb{A}^m$. In general, the result will be a code with a poor tradeoff between k/N and d/N. However, the Russian mathematician Goppa [103] made the inspired suggestion of choosing the subset of $\mathbb{A}^m$ to be points on a curve. Tsfasman, Vlăduţ and Zink recognized that existence of asymptotically good codes required curves over finite fields with many rational points, hence the entrance of modular curves. Table IV juxtaposes developments in algebraic geometry codes with the corresponding theory for BCH codes.

V. The Next Fifty Years

We have chosen to highlight two very different challenges, the creation of a quantum information theory, and the development of coding techniques for data networks in general, and wireless networks in particular.

In 1948 the main thread connecting information theory and physics was understanding the new perspective on entropy and its relation to the laws of thermodynamics. Today the main thread is quantum mechanics, as methods in information theory and computing have been extended to treat the transmission and processing of intact quantum states, and the interaction of such "quantum information" with classical information. According to Bennett and Shor [10]

> It has become clear that an information theory based on quantum principles extends and completes classical information theory, somewhat as complex numbers extend and complete the reals. The new theory includes quantum generalizations of classical notions such as sources, channels and codes, and two complementary, quantifiable kinds of information—classical information and quantum entanglement.

In this perspective we focus on the development of quantum error-correcting codes.

We then turn to 21st century communication. Fifty years from now it will be disappointing if the focus of coding theory is point-to-point communication in the presence of noise. Telecommunications will likely be dominated by packet data/voice transmitted over wide-area networks like the Internet where network management is distributed. The reliability and even the nature of individual links will be of secondary importance, and the challenge will be to understand the network as a whole and to guarantee end-to-end quality of service.

A. Quantum Error-Correcting Codes

Classical bits take the values 0 or 1 at all times, but quantum bits or qubits occupy a superposition of the 0 and 1 states. This is not to say that the qubit has some intermediate value between 0 and 1. Rather, the qubit is in both the 0 state and the 1 state at the same time to varying extents. Mathematically, a qubit is a two-dimensional Hilbert space, and a quantum state is a vector

$$\alpha|0\rangle + \beta|1\rangle, \qquad \text{where } |\alpha|^2 + |\beta|^2 = 1.$$

A collection of N different two-state memory cells is then expressed as the tensor product of the individual two-dimensional Hilbert spaces, so we are led to vectors

$$\sum_{v \in V} \alpha_v |v\rangle, \qquad \text{where } V = \mathbb{Z}_2^N \text{ and } \sum_{v \in V} |\alpha_v|^2 = 1.$$

When the qubit $\alpha|0\rangle + \beta|1\rangle$ is measured with respect to the basis $|0\rangle$, $|1\rangle$ the probability that the qubit is found in a particular state is the square of the absolute value of the corresponding amplitude. The evolution of an isolated quantum system conserves superposition and distinguishability, and is described by a unitary transformation that is linear and preserves inner products. This is the analog in Hilbert space of rigid rotation in Euclidean space.

TABLE IV
PARALLEL DEVELOPMENTS IN THE THEORY OF BCH AND ALGEBRAIC-GEOMETRY CODES

BCH Codes	Algebraic Geometry Codes
Flexibility in choice of code (block length N not constrained by field size as with RS codes)	Flexibility regarding block length requires curves with many rational points
	Block length of codes from plane curves (e.g. Hermitian curves) bounded by q^2+q+1
[137] Performance degrades with large block length (cannot bound both d/N and k/N away from zero asymptotically)	[211] Existence of codes from modular curves with polynomial complexity that exceed the GV bound for alphabets of size $q \geq 49$
	[93] Explicit curves (from here to codes is still a lot of work)
Gorenstein-Peterson-Zierler Algorithm: Complexity $O(N^3)$	Basic Algorithm [119]: Complexity $O(N^3)$, restricted to plane curves
	Modified Algorithm [195]: Arbitrary curves
	Porter's Algorithm [171], [172]: generalization of Euclidean algorithm for decoding classical Goppa codes equivalent to modified algorithm
[60], [169] Error locating pairs of vector spaces common framework for decoding cyclic codes up to and beyond the BCH bound, and the Basic Algorithm	
[70] Decoding beyond the BCH designed distance	[69], [59] Decoding up to the Goppa designed distance, using majority voting to find additional syndromes for the error vector
	[61] gives a different solution
	[123], [229] Decoding Hermitian codes up to the actual distance using special properties of the affine ring of the curve
1-dimensional Berlekamp-Massey Algorithm: Complexity $O(N^2)$	[180], [181] Multidimensional generalization of the Berlekamp-Massey Algorithm with complexity $O(N^{3-2/m+1})$ for curves in $\mathbb{A}^m$
	[182] Incorporates majority voting
[205] List decoding of Reed-Solomon codes	[189] List decoding of algebraic geometry codes

The first work connecting information theory and quantum mechanics was that of Landauer and Bennett who were looking to understand the implications of Moore's Law; every two years for the past 50 years, computers have become twice as fast while components have become twice as small. As the components of computer circuits become very small, their description must be given by quantum mechanics. Over time there developed a curiosity about the power of quantum computation, until in 1994 Shor [190] found a way of exploiting quantum superposition to provide a polynomial time algorithm for factoring integers. This was the first example of an important problem that a quantum computer could solve more efficiently than a classical computer. The design of quantum algorithms for different classes of problem, for instance finding short vectors in lattices, is currently an active area of research.

The effectiveness of quantum computing is founded on coherent quantum superposition or entanglement, which allows exponentially many instances to be processed simultaneously. However, no quantum system can be perfectly isolated from the rest of the world and this interaction with the environment causes decoherence. This error process is expressed mathematically in terms of Pauli matrices. A bit error in an individual qubit corresponds to applying the Pauli matrix $\sigma_x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ to that qubit, and a phase error to the Pauli matrix $\sigma_z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. The third Pauli matrix, $\sigma_y = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right) = i\sigma_x\sigma_z$, corresponds to a combination of bit and phase errors. The group E of tensor products $\pm w_1 \otimes \cdots \otimes w_N$ and $\pm i w_1 \otimes \cdots \otimes w_N$, where each w_j is one of I, σ_x, σ_y, σ_z, describes the possible errors in N qubits. The *Error Group* E is a subgroup of the unitary group $U(2^N)$. In general, there is a continuum of possible errors in qubits, and there are errors in sets of qubits which cannot

be described by a product of errors in individual qubits. For the purposes of quantum error correction, however, we need consider only the three types of errors σ_x, σ_y, and σ_z, since any error-correcting code which connects t of these errors will be able to correct arbitrary errors in t qubits [62], [9]. We do not go into the details of this result, but essentially it follows from the fact that the matrices I, σ_x, σ_y, and σ_z form a basis for the space of all 2×2 matrices, and so the tensor products of t of these errors form a basis for the space of $2^t \times 2^t$ matrices.

In classical computing one can assemble computers that are much more reliable than any of their individual components by exploiting error-correcting codes. In quantum computing this was initially thought to be precluded by the Heisenberg Uncertainty Principle (HUP) which states that observations of a quantum system, no matter how delicately performed, cannot yield complete information on the system's state before observation. For example, we cannot learn more about a single photon's polarization by amplifying it into a clone of many photons—the HUP introduces just enough randomness into the polarizations of the daughter photons to nullify any advantage gained by having more photons to measure. At first, error correction was thought to be impossible in the quantum world because the HUP prevents duplication of quantum states. This is not so—only repetition codes are eliminated. The trick is to take quantum superposition + decoherence, to measure the decoherence in a way that gives no information about the original superposition, and then to correct the measured decoherence. The first codes were discovered quite recently [191], [203], [8], [35] but there is now a beautiful group-theoretic framework for code construction [32], [105], [33].

Commutative subgroups of the error group E play a special role. The quantum error-correcting code is the subspace fixed by the commutative subgroup—hence the name *stabilizer codes*. Errors move the fixed subspace to a different eigenspace of the original commutative subgroup. This eigenspace is identified by a process similar to that of calculating a syndrome in the classical world. Note that syndrome decoding identifies the coset of a linear code containing the received vector, and not an error pattern. However, given the coset, there is a coset leader that gives the most probable error pattern. Likewise, in the quantum world there is an error that is most probable given the eigenspace that has been identified.

The error group in classical theory is the subgroup B of bit errors. It is possible to describe classical linear codes as the fixed spaces of commutative subgroups of B, so the new framework is a graceful extension of the classical theory. Recent developments in quantum coding theory include a quantum analog of the MacWilliams Identities in classical coding theory [192].

B. The Changing Nature of Data Network Traffic

Today we lack fundamental understanding of network data traffic, and we need to replace network engineering methods developed for voice traffic. Information theory and coding may have an important role to play, but the first step must be to develop channel models through active and passive network measurement, that capture the interaction of applications, protocols, and end-to-end congestion control mechanisms.

A. K. Erlang (1878–1929) was the first person to study call blocking in telephone networks. By taking measurements in a small village telephone exchange he worked out a formula, now known as Erlang's formula, that expresses the fraction of callers that must wait because all lines are in use. Ever since Erlang, the nature of voice telephone traffic—exponentially distributed interarrival and holding times—has remained unchanged. However, Erlang did not anticipate fax, nor could he imagine the emergence of data networks where computers talk rather than humans. For voice networks the only statistic that matters is the mean traffic rate. By contrast, data traffic is extremely bursty and looks the same when viewed over a range of different time scales. More precisely, aggregate packet-level network traffic exhibits fractal-like scaling behavior over time scales on the order of a few hundred milliseconds and larger, if and only if the durations (in second) or sizes (in bytes) of the individual sessions or connections that generate the aggregate traffic have a heavy-tailed distribution with infinite variance. The self-similar nature of data network traffic was an empirical discovery made by Leland, Taqqu, Willinger, and Wilson [133] from extensive measurements on different local-area networks. The fact that heavy tails are found everywhere from sizes of files in a file system to bursts and idle periods of individual Ethernet connections, leads to self-similarity at the packet level across local- and wide-area networks (see [224] or [223] for a popular article). Above a certain time scale there are no surprises in voice traffic since everything reduces to the long-term arrival rate. For data traffic, significant variation on quite coarse time scales means that routers require large buffers, that safe operating points have to be set very conservatively, and that overall network performance is no longer a guarantee of individual quality of service. Absent new insights from coding and information theory, these variations are likely to be magnified on wireless channels by the rapidly changing nature of fading and interference.

The flow of packets at the different layers in the TCP/IP hierarchy is determined by Internet protocols and end-to-end congestion control mechanisms. The impact of the network on traffic shows up on small time scales, from a few hundred milliseconds and downwards. Feldmann, Gilbert, and Willinger [68] have proposed cascades (or multiplicative processes) as an explanation for the more complex (multifractal rather than monofractal) scaling behavior exhibited by measured TCP/IP and ATM wide-area network traffic. The thought is that cascades allow refinement of self-similarity (monofractal scaling) to account for local irregularities in WAN traffic that might be associated with networking mechanisms such as TCP flow control that operate on small time scales. Fig. 10 is taken from [68] and it compares local scaling behavior of exactly self-similar traffic with that of measured WAN traffic. This author would suggest that particularly on wireless channels, we need to change the metrics we use to evaluate systems, de-emphasizing long-term average packet loss statistics, and augmenting throughput with appropriate measures of delay.

C. It is Dangerous to Put Limits on Wireless

The heading is a quotation of Marconi from 1932. Fig. 11 superimposes research issues in wireless communication on

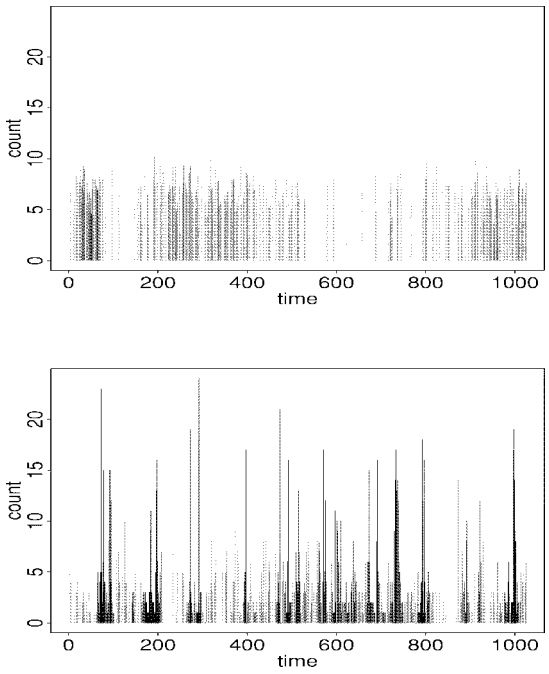

Fig. 10. Local scaling analysis of packet-level data traffic; different shades of gray indicate different magnitudes of the local scaling exponents at the different point in the traffic trace (black for small scaling exponents or "bursty" instants, light for large scaling exponents or "lull" periods). From top to bottom: (exactly) self-similar traffic, and WAN trace at the 1-ms time scale.The latter trace was gathered from an FDDI ring (with typical utilization levels of 5–10%) that connects about 420 modems to the Internet. It was collected between 22:00 and 23:00, July 22, 1997 and contains modem user as well as nonmodem user traffic totalling 8 910 014 packets.

a plot that displays the increasing size of the U.S. cellular market. Unlike the Gaussian channel, the wireless channel suffers from attenuation due to destructive addition of multipaths in the propagation media and due to interference from other users. Severe attenuation makes it impossible to determine the transmitted signal unless some less-attenuated replica of the transmitted signal is provided to the receiver. This resource is called *diversity* and it is the single most important contributor to reliable wireless communications. Examples of diversity techniques are (but are not restricted to) as follows.

- *Temporal Diversity:* Channel coding in connection with time interleaving is used. Thus replicas of the transmitted signal are provided to the receiver in the form of redundancy in temporal domain.
- *Frequency Diversity:* The fact that waves transmitted on different frequencies induce different multipath structure in the propagation media is exploited. Thus replicas of the transmitted signal are provided to the receiver in the form of redundancy in the frequency domain.
- *Antenna Diversity:* Spatially separated or differently polarized antennas are used. Replicas of the transmitted signal are provided to the receiver in the form of redundancy in spatial domain. This can be provided with no penalty in bandwidth efficiency.

When possible, cellular systems should be designed to encompass all forms of diversity to ensure adequate performance. For instance, cellular systems typically use channel coding in combination with time interleaving to obtain some form of temporal diversity [206]. In TDMA systems, frequency diversity is obtained using a nonlinear equalizer [4] when multipath delays are a significant fraction of symbol interval. In DS-CDMA, RAKE receivers are used to obtain frequency diversity, and more general two-dimensional RAKE structures have been proposed [159] that exploit temporal and spatial structure in the received multipath signal. Antenna diversity is typically used in the uplink (mobile-to-base) direction to provide link margin and cochannel interference suppression. This is necessary to compensate for the low-power transmission from mobiles [96]. The focus here will be narrowband 30-kHz TDMA (IS-136) channels, specifically the design of channel codes for improving the data rate and/or the reliability of communications over fading channels using multiple transmit and receive antennas. Information-theoretic aspects of transmit diversity were addressed by Telatar [210] and by Foschini and Gans [88]. They derived the outage capacity curves shown in Fig. 12 under the assumption that fading is quasistatic, that is constant over a long period of time and then changing in an independent manner. Recall that 10% outage capacity is the transmission rate that can be achieved 90% of the time. With only two antennas at both the base station and the mobile there is the potential to increase the achievable data rate by a factor of 6.

Transmit diversity schemes use linear processing at the transmitter to spread the information across the antennas. At the receiver, the demodulator computes a decision statistic based on the received signals arriving at each receive antenna $1 \le j \le m$. The signal d_t^j received by antenna j at time t is given by

$$d_t^j = \sum_{i=1}^{n} \alpha_{i,j} c_t^i \sqrt{E_s} + \eta_t^j$$

where the noise η_t^j at time t is modeled as independent samples of a zero-mean complex Gaussian random variable with variance $N_0/2$ per dimension. The coefficient $\alpha_{i,j}$ is the path gain from transmit antenna i to receive antenna j. It is assumed that these path gains are constant during a frame and vary from one frame to another (quasistatic flat fading). Feedforward information (the path gains α_{ij}) is required to estimate the channel from the transmitter to the receiver. The first scheme of this type was proposed by Wittneben [226] and it includes the delay diversity schemes of Seshadri and Winters [185] as a special case. In delay diversity there are two transmit antennas, and a signal is transmitted from the first antenna, then delayed one time slot, and transmitted from the second antenna ($c_t^2 = c_{t-1}^1$). It has been shown by Wittneben that delay diversity schemes are optimal in providing diversity in the sense that the diversity advantage experienced by an optimal receiver is equal to the number of transmit antennas. There is, however, no "coding gain." For wireless systems employing small numbers of antennas, the space–time codes constructed by Tarokh, Seshadri, and Calderbank [209] provide both coding gain and diversity, and using only a 64-state decoder come within 2–3 dB of outage capacity. The general problem of combined coding and

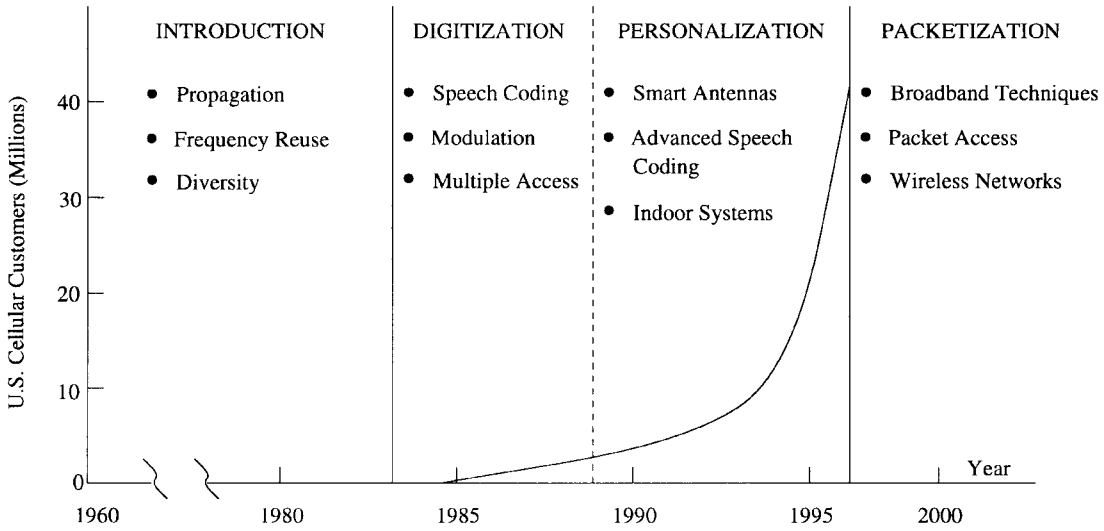

Fig. 11. Progress in wireless communications.

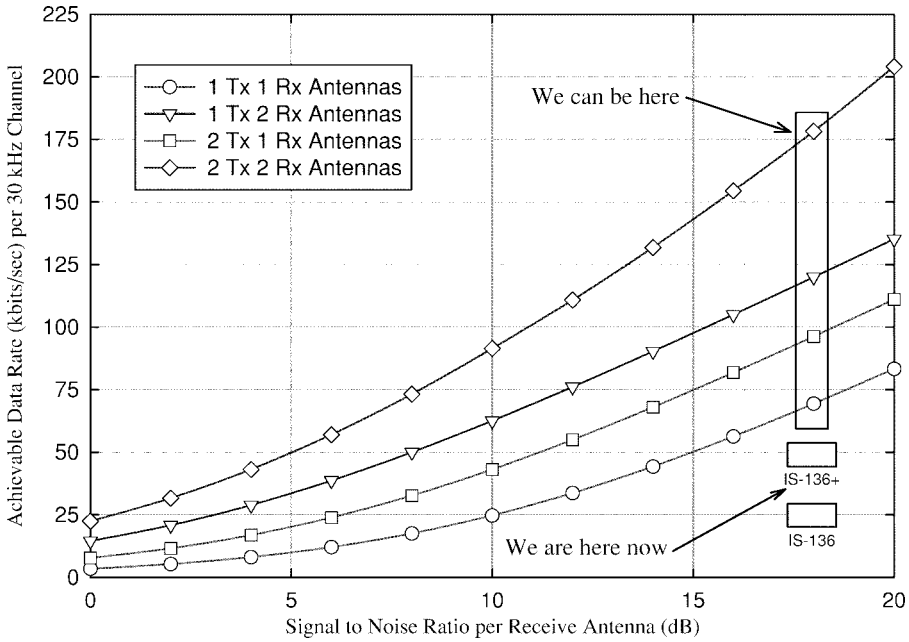

Fig. 12. Achievable data rates with multiple antennas at 10% outage capacity.

modulation for multi-input (multiple transmit antennas) multi-output (multiple receive antennas) fading channels is a new research area with great potential.

D. Interference Suppression

The challenge in designing cellular radio networks is to satisfy large demand with limited bandwidth. Limits on the available radio spectrum means that cochannel interference is inevitable when a cellular radio network is operating near capacity. The standard solution is to treat cochannel interference as Gaussian noise, and to employ powerful channel codes to mitigate its effect. This solution is far from optimal, since the decoder is using a mismatched metric. Interference is often due to a few dominant cochannel users, and this cannot be described as additive white Gaussian noise.

A second method of providing interference suppression is adaptive antenna array processing at the receiver. Here a substantial body of work by Winters and colleagues (see [96]) has shown that a receiver using N-branch spatial diversity can completely eliminate $N-1$ interferers using optimal linear combining.

The challenge for coding theory is to provide immunity to multiple channel impairments, in this case fading and cochannel interference. This author advocates a divide-and-conquer strategy, specifically the development of concatenated coding schemes where an inner component code might enable interference suppression, and an appropriate outer code might provide additional immunity to fading. For narrowband 30-kHz TDMA channels it is possible to design very simple space–time block codes that provide diversity gain using only

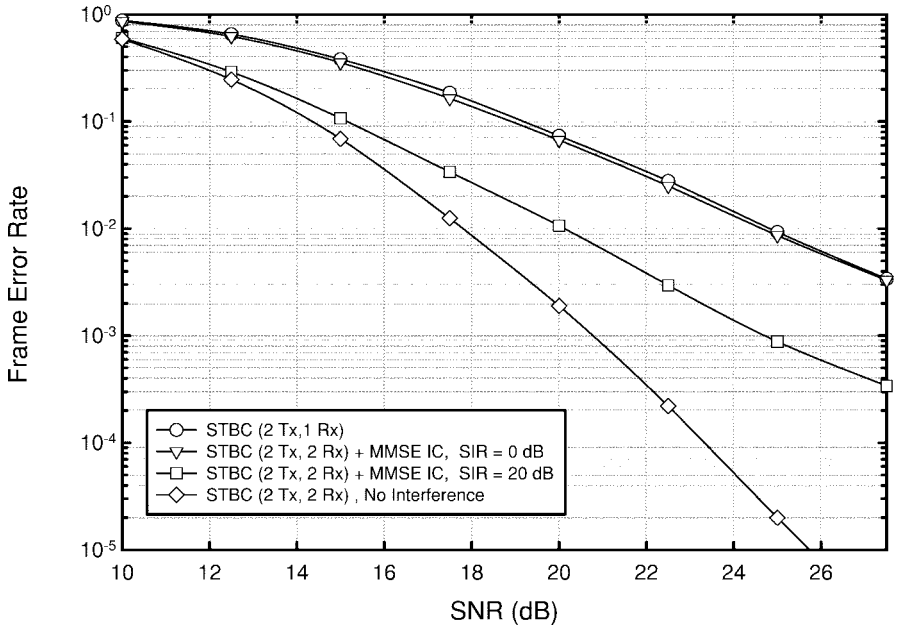

Fig. 13. Frame error rate performance of 8-PSK modulation with a space–time block code and interference suppression.

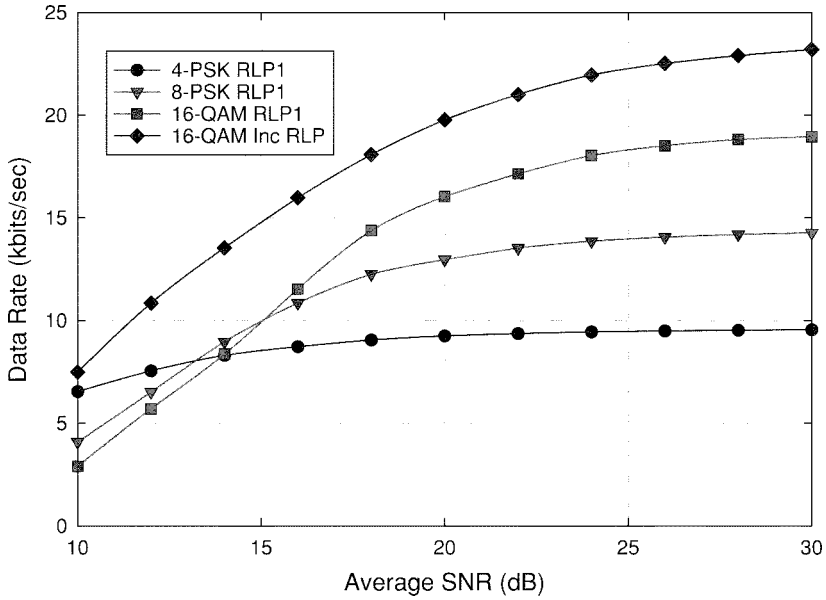

Fig. 14. Throughput of an incremental redundancy radio link protocol on a narrowband 30-kHz IS-136 channel.

transmit antennas. For example, Alamouti [1] presents the code

$$[c_1, c_2] \rightarrow \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}$$

where the signals r_1, r_2 received over two consecutive symbol periods are given by

$$\begin{pmatrix} r_1 \\ r_2^* \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}.$$

Assuming that channel state information is known to the receiver, we may form

$$\begin{pmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2^* \end{pmatrix} = (|h_1|^2 + |h_2|^2) \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} n_1' \\ n_2' \end{pmatrix}$$

where the noise vector (n_1', n_2') has zero mean and covariance $(|h_1|^2 + |h_2|^2)I_2$, and take the vector that results to a slicer. This code provides diversity gain (but no coding gain) and decoding is remarkably simple. The 2×2 matrix that describes transmission is a particularly simple example of an orthogonal design [94] and this rather arcane mathematical theory provides generalizations to more antennas.

If two antennas are available at the receiver, then Naguib and Seshadri [160] have shown that it is possible to suppress interference from a second space–time user by exploiting the special structure of the inner space–time block code. Fig. 13 shows the performance of their scheme with 8-PSK modulation. When the signal power of the interferer is equal to that of the desired signal, performance is the same as that

of a system employing two transmit and one receive antenna. When there is no interference, the second antenna provides additional immunity to fading. The decoder does not require any information about the interference, and simply adapts automatically.

E. Radio Link Protocols

In wireless communication, coding theory is associated with the physical layer which lies at the bottom of the protocol stack. The next layer is radio link protocols which are designed to deliver error-free packets to the higher networking layers. The gains that come from joint optimization of the physical and radio link layers are substantial, and may well be essential to the engineering of attractive wireless data services.

A very interesting idea with great potential is that of incremental redundancy. Packets received in error are stored at the receiver, and additional parity packets are transmitted until the original packet is decoded successfully. The type of hybrid radio link protocol is extremely flexible and can be tuned to different delay/throughput characteristics by adjusting the coding strategy and packet size (see [163]). Fig. 14 shows the throughput that can be achieved on narrowband 30-kHz channels. An alternative method of increasing throughput is to measure the signal-to-noise ratio (SNR) at the receiver and adapt coding and modulation to the measured SNR. It is difficult to do this accurately (within 1 dB) in the presence of rapid fading, and changing interference. Furthermore, the SNR values that trigger changes in coding and modulation vary with mobile speed so that collection of second-order statistics is necessary. The incremental redundancy radio link protocol implicitly adapts to SNR and provides superior performance. The rise of the Internet shows the power of distributed control in communications systems, and lightweight engineering is another reason to prefer implicit adaptation to SNR over explicit measurement and adaptation to SNR. The radio link protocol described by van Nobelen [163] has been accepted as a standard for the IS-136 high-speed packet state mode, and has similar potential to improve the proposed GSM EDGE standard.

ACKNOWLEDGMENT

The author is grateful to Peter Cameron for information on Fisher and the Hamming codes, to Jim Reeds for historical detail on telegraph codebooks, to Alexander Vardy for the statistics in Table I on soft-decision decoding, and to Walter Willinger for education on data network traffic. He would also like to thank Alexander Barg, Ian Blake, Iwan Duursma, G. David Forney Jr., Jon Hall, Ayman Naguib, Colin Mallows, Nambi Seshadri, Emira Soljanin, and Vahid Tarokh for their advice on earlier drafts of the manuscript. The author takes full responsibility for the faults that remain.

REFERENCES

[1] S. Alamouti, "Space block coding: A simple transmitter diversity scheme for wireless communications," submitted to *IEEE J. Select. Areas Commun.*, 1997.

[2] E. F. Assmus Jr. and H. F. Mattson Jr., "New 5-designs," *J. Combin. Theory*, vol. 6, pp. 122–151, 1969.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, 1974.

[4] N. Balaban and J. Salz, "Dual diversity combining and equalization in digital cellular mobile radio," *IEEE Trans. Veh. Tech.nol*, vol. 40, pp. 342–354, 1991.

[5] A. Barg, "At the dawn of the theory of codes," *Math. Intelligencer*, vol. 15, pp. 20–26, 1993.

[6] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite-state Markov chains," *Ann. Math. Statist.*, vol. 37, pp. 1554–1563, 1966.

[7] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 409–428, 1996.

[8] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, "Purification of noisy entanglement and faithful teleportation via noisy channels," *Phys. Rev. Lett.*, vol. 76, pp. 722–725, 1996; also LANL eprint quant-ph/9511027.

[9] C. H. Bennett, D. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed state entanglement and quantum error correction," *Phys. Rev. A*, vol. 54, pp. 3824–3851, 1996; also LANL eprint quant-ph/9604024.

[10] C. H. Bennett and P. W. Shor, "Quantum information theory," this issue, pp. 2724–2742.

[11] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[12] C. Berrou, A. Glavieux, and P. Thitmajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," in *Proc. Int. Conf. Communications*, 1993, pp. 1063–1070.

[13] W. Betts, A. R. Calderbank, and R. Laroia, "Performance of nonuniform constellations on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1633–1638, 1994.

[14] R. E. Blahut, *Algebraic Methods for Signal Processing and Communications Codes*. New York: Springer-Verlag, 1991.

[15] I. F. Blake, *Algebraic Coding Theory; History and Development*. Stroudsburg, PA: Dowden, Hutchinson and Ross, 1973.

[16] I. Blake, C. Heegard, T. Høholdt, and V. K. Wei, "Algebraic-geometry codes," this issue, pp. 2596–2618.

[17] W. G. Bliss, "Circuitry for performing error correction calculations on baseband encoded data to eliminate error propagation," *IBM Tech. Discl. Bull.*, vol. 23, pp. 4633–4634, 1981.

[18] A. Blokhuis and A. R. Calderbank, "Quasisymmetric designs and the smith normal form," *Des., Codes Cryptogr.*, vol. 2, pp. 189–206, 1992.

[19] A. Bonnecaze, A. R. Calderbank, and P. Solé, "Quaternary quadratic residue codes and unimodular lattices," *IEEE Trans. Inform. Theory*, vol. 41, pp. 366–377, 1995.

[20] R. C. Bose and D. M. Mesner, "On linear associative algebras corresponding to association schemes of partially balanced designs," *Ann. Math. Statist.*, vol. 30, pp. 21–38, 1959.

[21] A. E. Brouwer, A. M. Cohen, and A. Neumaier, *Distance Regular Graphs*. Berlin, Germany: Springer-Verlag, 1989.

[22] A. R. Calderbank, "Multilevel codes and multistage decoding," *IEEE Trans Commun.*, vol. 37, pp. 222–229, 1989.

[23] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel, "$\mathbb{Z}_4$-Kerdock codes, orthogonal spreads and extremal euclidean line sets," in *Proc. London Math. Soc.*, vol. 75, pp. 436–480, 1997.

[24] A. R. Calderbank and P. Delsarte, "Extending the t-design concept," *Trans. Amer. Math. Soc.*, vol. 338, pp. 941–962, 1993.

[25] A. R. Calderbank, P. Delsarte, and N. J. A. Sloane, "Strengthening of the Assmus–Mattson theorem," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1261–1268, 1991.

[26] A. R. Calderbank, G. D. Forney, Jr., and A. Vardy, "Minimal tail-biting representations of the golay code and others," *IEEE Trans. Inform. Theory*, to be published.

[27] A. R. Calderbank, R. Laroia, and S. W. McLaughlin, "Partial response codes for electron trapping optical memories," *IEEE Trans. Commun.*, vol. 46, pp. 1011–1019, 1998.

[28] A. R. Calderbank, W.-C. W. Li, and B. Poonen, "A 2-adic approach to the analysis of cyclic codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 977–986, 1997.

[29] A. R. Calderbank, G. McGuire, P. V. Kumar, and T. Helleseth, "Cyclic codes over $\mathbb{Z}_4$, locator polynomials and Newton's identities," *IEEE Trans. Inform. Theory*, vol. 43, pp. 217–226, 1996.

[30] A. R. Calderbank and J. E. Mazo, "Baseband line codes via spectral factorization," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 914–928, 1989.

[31] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 33, pp. 726–740, 1990.

[32] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, "Quantum error correction and orthogonal geometry," *Phys. Rev. Lett.*, vol. 78, pp. 405–409, 1997; also LANL eprint quant-ph/9605005.

[33] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, "Quantum error correction via codes over GF (4)," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1369–1387, July 1998.

[34] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1234–1248, 1993.

[35] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Phys. Rev. A*, vol. 54, pp. 1098–1105, 1996; also LANL eprint quant-ph/9512032.

[36] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. 33, pp. 177–195, 1987.

[37] P. J. Cameron and J. J. Seidel, "Quadratic forms over $\mathrm{GF}(2)$," *Indag. Math.*, vol. 35, pp. 1–8, 1973.

[38] P. Camion, "Codes and association schemes," in *Handbook of Coding Theory*, R. A. Brualdi, W. C. Huffman, and V. Pless, Eds. Amsterdam, The Netherlands: Elsevier, to be published.

[39] H. Chen, "On minimum Lee weights of Hensel lifts of some binary BCH codes," *IEEE Trans. Inform. Theory*, submitted for publication.

[40] R. T. Chien, "Cyclic decoding procedure for the Bose–Chaudhuri–Hocquengham codes," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 357–363, 1964.

[41] O. M. Collins, "The subtleties and intricacies of building a constraint length 15 convolutional decoder," *IEEE Trans. Commun.*, vol. 40, pp. 1810–1819, 1992.

[42] J. H. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 820–824, 1983.

[43] ———, "Soft decoding techniques for codes and lattices including the Golay code and the Leech lattice," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 41–50, 1986.

[44] ———, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.

[45] D. J. Costello, Jr., J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error control coding," this issue, pp. 2531–2560.

[46] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, 1972.

[47] R. V. Cox, J. Hagenauer, N. Seshadri, and C.-E. Sundberg, "Variable rate sub-band speech coding and matched convolutional channel coding for mobile radio channels," *IEEE Trans. Signal Processing*, vol. 39, pp. 1717–1731, 1991.

[48] P. Delsarte, "An algebraic approach to the association schemes of coding theory," *Philips Res. Rep. Suppl.*, no. 10, 1973.

[49] ———, "Four fundamental parameters of a code and their combinatorial significance," *Inform. Contr.*, vol. 23, pp. 407–438, 1973.

[50] ———, "Hahn polynomials, discrete harmonics and t-designs," *SIAM J. Appl. Math.*, vol. 34, pp. 157–166, 1978.

[51] ———, "Application and generalization of the MacWilliams transform in coding theory," in *Proc. 15th Symp. Information Theory in the Benelux* (Louvain-la-Neuve, Belgium, 1989), pp. 9–44.

[52] P. Delsarte and J.-M. Goethals, "Alternating bilinear forms over $\mathrm{GF}(q)$," *J. Comb. Theory (A)*, vol. 19, pp. 26–50, 1975.

[53] P. Delsarte, J.-M. Goethals, and J. J. Seidel, "Bounds for systems of lines and Jacobi polynomials," *Philips Res. Rep.*, vol. 30, pp. 91–105, 1975.

[54] ———, "Spherical codes and designs," *Geom. Dedicata*, vol. 6, pp. 363–388, 1977.

[55] P. Delsarte and V. I. Levenshtein, "Association schemes and coding theory," this issue, pp. 2477–2504.

[56] P. Delsarte and R. J. McEliece, "Zeros of functions in finite abelian group algebras," *Amer. J. Math.*, vol. 98, pp. 197–224, 1976.

[57] P. Delsarte and J. J. Seidel, "Fisher type inequalities for euclidean t-designs," *Lin. Alg. Appl.*, vols. 114/115, pp. 213–230, 1989.

[58] M. R. de Prony, "Essai expérimentalle et analytique," *J. École Polytech. Paris*, vol. 1, pp. 24–76, 1795.

[59] I. Duursma, "Majority coset decoding," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1067–1071, 1993.

[60] I. Duursma and R. Kötter, "Error-locating pairs for cyclic codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1108–1121, 1994.

[61] D. Ehrhard, "Achieving the designed error capacity in decoding algebraic-geometric codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 743–751, 1993.

[62] A. Ekert and C. Macchiavello, "Error correction in quantum communication," *Phys. Rev. Lett.*, vol. 77, pp. 2585–2588, 1996; also LANL eprint quant-phi/9602022.

[63] E. Eleftheriou and R. Cideciyan, "On codes satisfying Mth order running digital sum constraints," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1294–1313, 1991.

[64] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, 1982.

[65] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, 1991.

[66] P. Erdős, C. Ko, and R. Rado, "Intersection theorems for systems of finite sets," *Quart. J. Math. Oxford*, vol. 12, pp. 313–320, 1961.

[67] J. Feigenbaum, "The use of coding theory in computational complexity," in *Different Aspects of Coding Theory, Proc. Symp. in Appl. Math.*, vol. 50. Providence, RI: Amer. Math. Soci., 1995, pp. 207–233.

[68] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of internet WAN traffic," to be published in *Proc. ACM/SIGCOMM'98*.

[69] G.-L. Feng and T. R. N. Rao, "Decoding of algebraic geometric codes up to the designed minimum distance," *IEEE Trans. Inform. Theory*, vol. 39, pp. 37–45, 1993.

[70] G.-L. Feng and K. K. Tzeng, "A new procedure for decoding cyclic and BCH codes up to the actual minimum distance," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1364–1374, 1994.

[71] R. A. Fisher, "The theory of confounding in factorial experiments in relation to the theory of groups," *Ann. Eugenics*, vol. 11, pp. 341–353, 1942.

[72] R. A. Fisher, "A system of confounding for factors with more than two alternatives, giving completely orthogonal cubes and higher powers," *Ann. Eugenics*, vol. 12, pp. 2283–2290, 1945.

[73] G. D. Forney, Jr., "On decoding BCH codes," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 549–557, 1965.

[74] ———, *Concatenated Codes*. Cambridge, MA: MIT Press, 1966.

[75] ———, "Convolutional codes I: Algebraic structure," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 720–738, 1970; correction in vol. IT-17, p. 360, 1971.

[76] ———, "Maximum likelihood sequence estimation in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, 1972.

[77] ———, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 267–278, 1973.

[78] ———, "Minimal bases of rational vector spaces wtih applications to multivariable linear systems," *SIAM J. Contr.*, vol. 13, pp. 439–520, 1975.

[79] ———, "Coset codes—Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, 1988.

[80] ———, "Coset codes—Part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, 1988.

[81] ———, "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, 1991.

[82] ———, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, pp. 281–300, 1992.

[83] G. D. Forney, Jr., L. Brown, M. V. Eyuboglu, and J. L. Moran III, "The V.34 high-speed modem standard," *IEEE Commun. Mag.*, vol. 34, pp. 28–33, Dec. 1996.

[84] G. D. Forney, Jr. and A. R. Calderbank, "Coset codes for partial response channels; or coset codes with spectral nulls," *IEEE Trans. Inform. Theory*, vol. 35, pp. 925–943, 1989.

[85] G. D. Forney, Jr. and M. Trott, "The dynamics of group codes: State spaces, trellis diagrams and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, 1993.

[86] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," this issue, pp. 2384–2415.

[87] G. D. Forney, Jr. and A. Vardy, "Generalized minimum distance decoding of euclidean-space codes and lattices," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1992–2026, 1996.

[88] G. J. Foschini, Jr. and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, to be published.

[89] P. Frankl, "The Erdős-Ko-Rado theorem is true for $n = ckt$," in *Combinatorics, Proc. 5th Hungarian Colloq. Combinatorics, Keszthely*. Amsterdam, The Netherlands: North-Holland, 1976, pp. 365–375.

[90] B. J. Frey and F. R. Kschischang, "Probability propagation and iterative decoding," in *Proc. 34th Allerton Conf. Communication, Control, and Computing*, 1996, pp. 482–493.

[91] R. G. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. IT-8, pp. 21–28, 1962.

[92] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge MA: MIT Press, 1963.

[93] A. Garcia and H. Stichtenoth, "A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vlăduţ bound," *Invent. Math.*, vol. 121, pp. 211–222, 1995.

[94] A. V. Geramita and J. Seberry, "Orthogonal designs, quadratic forms and Hadamard matrices," in *Lecture Notes in Pure and Applied Mathematics*,

vol. 43. New York and Basel: Marcel Decker, 1979.
[95] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Syst. Tech. J.*, vol. 31, pp. 504–522, 1952.
[96] R. D. Gitlin, J. Salz, and J. H. Winters, "The capacity of wireless communications systems can be substantially increased by the use of antenna diversity," *IEEE J. Select. Areas Commun.*, to be published.
[97] A. M. Gleason, "Weight polynomials of self-dual codes and the MacWilliams identities," in *Actes Congrés Int. de Math.*, vol. 3, pp. 211–215, 1970.
[98] J.-M. Goethals, "Two dual families of nonlinear binary codes," *Electron. Lett.*, vol. 10, pp. 471–472, 1974.
[99] ——, "Nonlinear codes defined by quadratic forms over $\mathrm{GF}(2)$," *Inform. Contr.*, vol. 31, pp. 43–74, 1976.
[100] J.-M. Goethals and J. J. Seidel, "The football," *Nieuw. Arch. Wiskunde*, vol. 29, pp. 50–58, 1981.
[101] J.-M. Goethals and H. C. A. van Tilborg, "Uniformly packed codes," *Philips Res. Rep.*, vol. 30, pp. 9–36, 1975.
[102] M. J. E. Golay, "Notes on digital coding," *Proc. IEEE*, vol. 37, p. 657, 1949.
[103] V. D. Goppa, "Codes on algebraic curves," *Sov. Math.–Dokl.*, vol. 24, pp. 170–172, 1981. Translation from *Dokl. Akad. Nauk S.S.S.R.*, vol. 259, pp. 1289–1290, 1981.
[104] D. C. Gorenstein and N. Zierler, "A class of error correcting codes in p^m symbols," *SIAM J.*, vol. 9, pp. 207–214, 1971.
[105] D. Gottesman, "A class of quantum error-correcting codes saturating the quantum Hamming bound," *Phys. Rev. A*, vol. 54, pp. 1862–1868, 1996; also LANL eprint quant-ph/9604038.
[106] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *IEEE Globecom'89*, 1989, pp. 1680–1685.
[107] J. Hagenauer, E. Offer, and L. Papke, "Matching Viterbi decoders and Reed–Solomon decoders in concatenated systems," in *Reed–Solomon Codes and Their Applications*, S. B. Wicker and V. K. Bhargava, Eds. Piscataway, NJ: IEEE Press, 1994, pp. 242–271.
[108] A. R. Hammons, Jr., P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, and P. Solé, "The $\mathbb{Z}_4$-linearity of Kerdock, Preparata, Goethals and related codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 301–319, 1994.
[109] R. H. Hardin and N. J. A. Sloane, "Codes (spherical) and designs (experimental)," in *Different Aspects of Coding Theory, Proc. Symp. Applied Mathematics*, vol. 50. Providence RI: Amer. Math. Soc., 1995, pp. 179–206.
[110] T. Høholdt and R. Pellikaan, "On the decoding of algebraic-geometric codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1589–1614, 1995.
[111] A. S. Householder, *Principles of Numerical Analysis*. New York: McGraw-Hill, 1953.
[112] H. Imai and S. Hirakawa, "A new multi-level coding method using error correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, 1977.
[113] K. A. S. Immink, "RS codes and the compact disk," in *Reed–Solomon Codes and Their Applications*, S. B. Wicker and V. K. Bhargava, Eds. Piscataway, NJ: IEEE Press, 1994, pp. 41–59.
[114] ——, "A practical method for approaching the channel capacity of constrained channels," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1389–1399, 1997.
[115] K. A. S. Immink and G. Beenker, "Binary transmission codes with higher order spectral nulls at zero frequency," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 452–454, 1987.
[116] International Telecommunication Union, ITU-T Recommendation V.34, "A modem operating at data signaling rates of up to 28,800 bits/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits," Sept. 1994.
[117] J. Justesen, "A class of constructive asymptotically good algebraic codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 652–656, 1972.
[118] ——, "Information rates and power spectra of digital codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 457–472, 1982.
[119] J. Justesen, K. J. Larsen, H. E. Jensen, A. Havemose, and T. Høholdt, "Construction and decoding of a class of algebraic geometry codes," *IEEE Trans. Inform. Theory*, vol. 35, pp. 811–821, 1989.
[120] D. Kahn, *The Codebreakers, The Story of Secret Writing*. New York: MacMillan, 1967.
[121] R. Karabed and P. H. Siegel, "Matched spectral null codes for partial response channels," *IEEE Trans. Inform. Theory*, vol. 37, pp. 818–855, 1991.
[122] A. M. Kerdock, "A class of low-rate nonlinear binary codes," *Inform. Contr.*, vol. 20, pp. 182–187, 1972.
[123] C. Kirfel and R. Pellikaan, "The minimum distance of codes in an array coming from telescopic subgroups," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1720–1731, 1995.
[124] H. Kobayashi, "Correlative level coding and maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 586–594, 1971.
[125] H. Kobayashi and D. T. Tang, "Applications of partial response channel coding to magnetic recording systems," *IBM J. Res. Develop.*, vol. 14, pp. 368–375, 1970.
[126] H. König, *Isometric Embeddings of Euclidean Spaces into Finite-Dimensional lp-Spaces* (Banach Center Publications, vol. 34.). Warsaw, Poland: PWN, 1995, pp. 79–87.
[127] V. Koshélev, "Estimation of mean error for a discrete successive approximation scheme," *Probl. Inform. Transm.*, vol. 17, pp. 20–33, 1981.
[128] E. S. Lander, *Symmetric Designs: An Algebraic Approach* (London Math. Soc. Lecture Notes, vol. 74). London/New York: Cambridge Univ. Press, 1983.
[129] R. Laroia, "Coding for intersymbol interference channels—Combined coding and precoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1053–1061, 1996.
[130] R. Laroia, N. Farvardin, and S. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1044–1056, 1994.
[131] J. Leech, "Notes on sphere packings," *Canadian J. Math.*, vol. 19, pp. 251–267, 1967.
[132] J. Leech and N. J. A. Sloane, "Sphere packings and error-correcting codes," *Canadian J. Math.*, vol. 23, pp. 718–745, 1971.
[133] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, 1994.
[134] V. I. Levenshtein, "On bounds for packings in n-dimensional euclidean space," *Sov. Math.–Dokl.*, vol. 20, no. 2, pp. 417–421, 1979.
[135] ——, "Bounds on the maximum cardinality of a code with bounded modulus of the inner product," *Sov. Math.–Dokl.*, vol. 25, pp. 526–531, 1982.
[136] ——, "Designs as maximum codes in polynomial metric spaces," *Acta Applic. Math.*, vol. 29, pp. 1–82, 1992.
[137] S. Lin and E. J. Weldon, Jr., "Long BCH codes are bad," *Inform. Contr.*, vol. 11, pp. 452–495, 1967.
[138] J. H. van Lint, "Nonexistence theorems for perfect error correcting codes," in *Computers in Algebraic Number Theory, SIAM-AMS Proc.*, vol. IV, 1971.
[139] ——, "Coding theory," in *Springer Lecture Notes*, vol. 201. Berlin-Heidelberg-New York: Springer, 1971.
[140] J. H. van Lint and R. M. Wilson, "On the minimum distance of cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 23–40, 1986.
[141] S. P. Lloyd, "Binary block coding," *Bell Syst. Tech. J.*, vol. 36, pp. 517–535, 1957.
[142] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1767–1773, 1997.
[143] T. D. Lookabaugh and R. M. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1020–1033, 1989.
[144] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 1–7, 1979.
[145] D. J. C. MacKay and R. M. Neil, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 32, pp. 1645–1646, 1996 (reprinted vol. 33, pp. 457–458, 1997).
[146] F. J. MacWilliams, "Combinatorial properties of elementary abelian groups," Ph.D. dissertation, Harvard Univ., Cambridge, MA, 1962.
[147] ——, "A theorem on the distribution of weights in a systematic code," *Bell Syst. Tech. J.*, vol. 42, pp. 79–94, 1963.
[148] F. J. MacWilliams and N. J. A. Sloane," *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1977.
[149] F. J. MacWilliams, N. J. A. Sloane, and J. G. Thompson, "Good self-dual codes exist," *Discr. Math.*, vol. 3, pp. 153–162, 1972.
[150] M. W. Marcellin and T. R. Fischer, "Trellis-coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, pp. 82–93, 1993.
[151] B. Marcus, "Symbolic dynamics and connections to coding theory automata theory and system theory," in *Different Aspects of Coding Theory, Proc. Symp. Applied Mathematics*, vol. 50. Providence, RI: Amer. Math. Soc., 1995, pp. 95–108.
[152] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 5–37, 1992.
[153] J. L. Massey, "Shift register synthesis and BCH decoding," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 122–127, 1969.
[154] H. F. Mattson, Jr. and G. Solomon, "A new treatment of Bose-Chaudhuri codes," *SIAM J.*, vol. 9, pp. 654–669, 1961.
[155] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*. Boston-Dordrecht-Lancaster: Kluwer, 1987.
[156] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams

inequalities," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 157–166, 1977.
[157] H. Miyakawa and H. Harashima, "Capacity of channels with matched transmission technique for peak transmitting power limitation," in *Nat. Conv. Rec. IECE Japan*, 1969, pp. 1268–1264.
[158] D. J. Muder, "Minimal trellises for block codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1049–1053, 1988.
[159] A. Naguib, "Adaptive antennas for CDMA wireless networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1996.
[160] A. F. Naguib and N. Seshadri, "Combined interference cancellation and maximum likelihood decoding of space-time block codes," preprint, 1998.
[161] K. Nakamura, "A class of error correcting codes for DPSK channels," in *Proc. Int. Conf. Communication*, 1979, pp. 45.4.1–45.4.5.
[162] Y. Nakamura, Y. Saito, and S. Aikawa, "256-QAM modem for multi-carrier 400 Mb/s digital radio," *IEEE J. Select. Areas. Commun.*, vol. JSAC-5, pp. 329–335, 1987.
[163] R. van Nobelen, "Toward higher data rates for IS-36," in *Proc. IEEE Vehicular Technology Conf.*, 1998, pp. 2403–2407.
[164] A. W. Nordstrom and J. P. Robinson, "An optimum nonlinear code," *Inform. Contr.*, vol. 11, pp. 613–616, 1967.
[165] A. M. Odlyzko and N. J. A. Sloane, "New bounds on the number of spheres that can touch a unit sphere in n-dimensions," *J. Combin. Theory*, vol. (A)26, pp. 210–214, 1979.
[166] L. H. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, 1980.
[167] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan-Kaufmann, 1988.
[168] J. B. H. Peek, "Communications aspects of the compact disk digital audio system," *IEEE Commun. Mag.*, vol. 23, pp. 7–15, Feb. 1985.
[169] R. Pellikaan, "On the decoding by error location and the number of dependent error positions," *Discr. Math.*, vols. 106/107, pp. 369–381, 1992.
[170] W. W. Peterson, "Encoding and error correction procedures for the Bose-Chaudhuri codes," *IEEE Trans. Inform. Theory*, vol. 6, pp. 459–470, 1960.
[171] S. C. Porter, "Decoding codes arising from Goppa's construction on algebraic curves," Ph.D. dissertation, Yale Univ., New Haven, CT, 1988.
[172] S. C. Porter, B.-Z. Shen, and R. Pellikaan, "On decoding geometric Goppa codes using an extra place," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1663–1676, 1992.
[173] G. J. Pottie and D. P. Taylor, "Multilevel codes based on partitioning," *IEEE Trans. Inform. Theory*, vol. 35, pp. 87–98, 1989.
[174] F. P. Preparata, "A class of optimum nonlinear double error correcting codes," *Inform. Contr.*, vol. 13, pp. 378–400, 1968.
[175] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 6–23, 1993.
[176] C. R. Rao, "Factorial experiments derivable from combinatorial arrangements of arrays," *J. Roy. Statist. Soc.*, vol. 9, pp. 128–139, 1947.
[177] D. K. Ray-Chaudhuri and N. M. Singhi, "On existence of t-designs with large v and λ," *SIAM J. Discr. Math.*, vol. 1, pp. 98–104, 1988.
[178] D. K. Ray-Chaudhuri and R. M. Wilson, "On t-designs," *Osaka J. Math.*, vol. 12, pp. 737–744, 1975.
[179] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *SIAM J.*, vol. 8, pp. 300–304, 1960.
[180] S. Sakata, "Finding a minimal set of linear recurring relations capable of generating a given finite two-dimensional array," *J. Symbolic Comput.*, vol. 5, pp. 321–337, 1988.
[181] S. Sakata, "Extension of the Berlekamp-Massey algorithm to N dimensions," *Inform. Comput.*, vol. 84, pp. 207–239, 1990.
[182] S. Sakata, J. Justesen, Y. Madelung, H. E. Jensen, and T. Høholdt, "Fast decoding of algebraic geometric codes up to the designed minimum distance," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1672–1677, 1995.
[183] N. V. Semakov, V. A. Zinoviev, and G. V. Zaitzev, "Uniformly packed codes," *Probl. Pered. Inform.*, vol. 7, pp. 38–50, 1971 (in Russian).
[184] N. Seshadri and C.-E. Sundberg, "Generalized Viterbi detection with convolutional codes," in *IEEE Globecom'89*, 1989, pp. 1534–1538.
[185] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, pp. 49–60, 1994.
[186] P. D. Seymour and T. Zaslavsky, "Averaging sets," *Adv. Math.*, vol. 52, pp. 213–240, 1984.
[187] C. E. Shannon, "A mathematical theory of communication I, II," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423; 623–656, 1948; Reprinted in C. E. Shannon and W. Weaver, *A Mathematical Theory of Communication*. Urbana, IL: Univ. Illinois Press, 1949.
[188] ———, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.
[189] M. A. Shokorollahi and H. Wassermann, "Decoding algebraic-geometric codes beyond the error-correction bound," preprint, 1997.
[190] P. W. Shor, "Algorithms for quantum computation: Discrete logarithm and factoring," in *Proc. 35th Annu. Symp. Foundations of Computer Science*, 1994, pp. 124–134.
[191] ———, "Scheme for reducing deroherence in quantum computer memory," *Phys. Rev. A.*, vol. 52, pp. 2493–2496, 1995.
[192] P. W. Shor and R. Laflamme, "Quantum analog of the MacWilliams identities in classical coding theory," *Phys. Rev. Lett.*, vol. 78, pp. 1600–1602, 1997; also LANL eprint quant-ph/9610040.
[193] V. M. Sidel'nikov, "Extremal polynomials used in bounds of code volume," *Probl. Pered. Inform.*, vol. 16, pp. 17–39, 1980 (in Russian). English translation in *Probl. Inform. Transm.*, vol. 16, pp. 174–186, 1980.
[194] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1710–1722, 1996.
[195] A. N. Skorobogatov and S. G. Vlăduţ, "On the decoding of algebraic geometric codes," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1461–1463, 1990.
[196] D. Slepian, "A class of binary signaling alphabets," *Bell Syst. Tech. J.*, vol. 35, pp. 203–234, 1956.
[197] ———, "A note on two binary signaling alphabets," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 84–86, 1956.
[198] ———, "Some further theory of group codes," *Bell Syst. Tech. J.*, vol. 39, pp. 1219–1252, 1960.
[199] ———, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, 1968.
[200] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.
[201] N. J. A. Sloane, "Error-correcting codes and invariant theory: New applications of a nineteenth century technique," *Amer. Math. Monthly*, vol. 84, pp. 82–107, 1977.
[202] G. Solomon and H. C. A. van Tilborg, "A connection between block and convolutional codes," *SIAM J. Appl. Math.*, vol. 37, pp. 358–369, 1979.
[203] A. M. Steane, "Error correcting codes in quantum theory," *Phys. Rev. Lett.*, vol. 77, pp. 793–797, 1996.
[204] H. Stichtenoth, "Algebraic geometric codes," in *Different Aspects of Coding Theory, Proc. Symp. App. Math.*, vol. 50. Providence, RI: Amer. Math. Soc., 1995,
[205] M. Sudan, "Decoding of Reed–Solomon codes beyond the error-correction bound," *J. Complexity*, vol. 13, pp. 180–193, 1997.
[206] C.-E. Sundberg and N. Seshadri, "Digital cellular systems for North America," in *IEEE Globecom'90*, 1990, pp. 533–537.
[207] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 533–547, 1981.
[208] V. Tarokh and I. F. Blake, "Trellis complexity versus the coding gain of lattices, parts I and II," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1796–1807 and 1808–1816, 1996.
[209] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, 1998.
[210] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs. Internal Tech. Memo, June 1995.
[211] M. A. Tsfasman, S. G. Vlăduţ, and T. Zink, "On Goppa codes which are better than the Varshamov-Gilbert bound," *Math. Nach.*, vol. 109, pp. 21–28, 1982.
[212] A. Tietäväinen, "On the nonexistence of perfect codes over finite fields," *SIAM J. Appl. Math.*, vol. 24, pp. 88–96, 1973.
[213] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, 1971.
[214] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, 1993.
[215] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, 1982.
[216] A. Vardy and Y. Be'ery, "More efficient soft decoding of the Golay codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 667–672, 1991.
[217] ———, "Maximum-likelihood decoding of the Leech lattice," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1435–1444, 1993.
[218] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," *Dokl. Akad. Nauk. SSSR*, vol. 117, pp. 739–741, 1957.
[219] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, 1967.
[220] U. Wachsmann and J. Huber, "Power and bandwidth efficient digital communication using turbo codes in multilevel codes," *Euro. Trans. Telecommun.*, vol. 6, pp. 557–567, 1995.

[221] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Linköping Univ., Linköping, Sweden, 1996.
[222] N. Wiberg, H.-A. Loeliger, and R. Kötter, "Codes and iterative decoding on general graphs," *Euro. Trans. Telecommun.*, vol. 6, pp. 513–525, 1995.
[223] W. Willinger and V. Paxson, "Where mathematics meets the internet," to be published in the *Notices Amer. Math. Soc.*
[224] W. Willinger, M. S. Taqqu, and A. Erramilli, "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks," in *Stochastic Networks: Theory and Applications, Royal Statistical Lecture Note Series*, vol. 4, F. P. Kelly, S. Zachary, and I. Ziedins, Eds. Oxford, U.K.: Clarendon Press, 1996, pp. 339–366.
[225] R. M. Wilson, "The exact bound in the Erdős-Ko-Rado theorem," *Combinatorica*, vol. 4, pp. 247–257, 1984.
[226] A. Wittneben, "Base station modulation diversity for digital SIMULCAST," in *Proc. IEEE Vehicular Technology Conf.*, 1993, pp. 505–511.
[227] J. K. Wolf, "Decoding of Bose–Chaudhuri–Hocquenghem codes and Prony's method for curve fitting," *IEEE Trans. Inform. Theory*, vol. IT-3, p. 608, 1967.
[228] J. A. Wood, "Extension theorems for linear codes over finite rings," preprint 1997.
[229] K. Yang and P. V. Kumar, "On the true minimum distance of hermitian codes," in *Coding Theory and Algebraic Geometry* (Lecture Notes in Mathematics, vol. 1518). Berlin, Germany: Springer, 1992, pp. 99–107.
[230] V. A. Zinoviev, "Generalized cascade codes," *Probl. Pered. Inform.*, vol. 12, pp. 2–9, 1976 (in Russian).
[231] V. A. Zinoviev and V. K. Leontiev, "The nonexistence of perfect codes over Galois fields," *Probl. Contr. Inform. Theory*, vol. 2, pp. 123–132, 1973.
[232] V. V. Zyablov and M. S. Pinsker, "Estimates of the error-correction complexity of Gallager's low density codes," *Probl. Inform. Transm.*, vol. 11, pp. 18–28, 1976.