Abstract—Shannon’s determination of the capacity of the linear Gaussian channel has posed a magnificent challenge to succeeding generations of researchers. This paper surveys how this challenge has been met during the past half century. Orthogonal minimum-bandwidth modulation techniques and channel capacity are discussed. Binary coding techniques for low-signal-to-noise ratio (SNR) channels and nonbinary coding techniques for high-SNR channels are reviewed. Recent developments, which now allow capacity to be approached on any linear Gaussian channel, are surveyed. These new capacity-approaching techniques include turbo coding and decoding, multilevel coding, and combined coding/precoding for intersymbol-interference channels.

Index Terms—Binary coding, block codes, canonical channel, capacity, coding gain, convolutional codes, lattice codes, multicarrier modulation, nonbinary coding, orthogonal modulation, precoding, shaping gain, single-carrier modulation, trellis codes.

I. INTRODUCTION

The problem of efficient data communication over linear Gaussian channels has driven much of communication and coding research in the half century since Shannon’s original work [93], [94], where this problem was central. Shannon’s most celebrated result was the explicit computation of the capacity of the additive Gaussian noise channel. His result has posed a magnificent challenge to succeeding generations of researchers. Only in the past decade can we say that methods of approaching capacity have been found for practically all linear Gaussian channels. This paper focuses on the modulation, coding, and equalization techniques that have proved to be most effective in meeting Shannon’s challenge.

Approaching channel capacity takes different forms in different domains. There is a fundamental difference between the power-limited regime with low signal-to-noise ratio (SNR) and the high-SNR bandwidth-limited regime.

Early work focused on the power-limited regime. In this domain, binary codes suffice, and intersymbol interference (ISI) due to nonflat channel responses is rarely a serious problem. As early as the 1960’s, sequential decoding [119] of binary convolutional codes [31] was shown to be an implementable method for achieving the cutoff rate $R_0$, which at low SNR is 3 dB away from the Shannon limit. Many leading communications engineers took $R_0$ [78], [118] to be the “practical capacity,” and concluded that the problem of reaching capacity had effectively been solved. Nonetheless, only in the 1990’s have coding schemes that reach or exceed $R_0$ actually been implemented, even for such important and effectively ideal additive white Gaussian noise (AWGN) channels as deep-space communication channels.

Meanwhile, in the bandwidth-limited regime there was essentially no practical progress beyond uncoded multilevel modulation until the invention of trellis-coded modulation in the mid-1970’s and its widespread implementation in the 1980’s [104]. An alternative route to capacity in this regime is via multilevel coding, which was also introduced during the 1970’s [60], [75]. Trellis-coded modulation and multilevel coding are both based on the concept of set partitioning. In the late 1980’s, constellation shaping was recognized as a separable part of the high-SNR coding problem, which yields a small but essential contribution to approaching capacity [48].

For channels with strongly frequency-dependent attenuation or noise characteristics, it had long been known that multicarrier modulation could in principle be used to achieve the power and rate allocations prescribed by “water pouring” (see, e.g., [51]). However, practical realizations of multicarrier modulation in combination with powerful codes have been achieved only in recent years [90]. Multicarrier techniques are particularly attractive for channels for which the capacity-achieving band consists of multiple disconnected frequency intervals, as may happen with narrowband interference.

When the capacity-achieving band is a single interval, methods for approaching capacity using serial transmission and nonlinear transmitter pre-equalization (“precoding”) have been developed in the past decade [34]. With these methods, the received signal is apparently ISI-free, so ideal-channel decoding can be performed at the channel output. In addition, signal redundancy at the channel output is exploited for constellation shaping at the channel input. The performance achieved is equivalent to the performance that would be obtained if ISI could be perfectly canceled in the receiver with decision-feedback equalization (DFE). At high SNR, the effective SNR is approximately equal to the optimal effective SNR achieved by water pouring. The capacity of an ISI channel can therefore be approached as closely as the capacity of an ideal ISI-free channel.

Currently, the invention of turbo codes [13] and the rediscovery of low-density parity-check (LDPC) codes [50]
have created tremendous excitement. These schemes operate successfully at rates well beyond the cutoff rate $R_0$, within tenths of a decibel of the Shannon limit, in both the low-SNR and high-SNR regimes. Whereas coding for rates below $R_0$ is a rather mature field, coding in the beyond-$R_0$ regime is in its early stages, and theoretical understanding is still weak. At $R_0$, it seems that a “phase transition” occurs from a static regime of regular code structures to a statistical regime of quasirandom codes, comparable to a phase transition between a solid and a liquid or gas.

In Section II, we define the general linear Gaussian channel and the ideal band-limited additive white Gaussian noise (AWGN) channel. We review how orthogonal modulation techniques convert a waveform channel to an equivalent ideal discrete-time AWGN channel. We discuss the Shannon limit for such channels, and emphasize the difference between the low-SNR and high-SNR regimes. We give baseline performance curves for uncoded $M$-ary PAM modulation, and evaluate the error probability as a function of the gap to capacity. We review union-bound performance analysis, which is a useful tool below $R_0$, but useless beyond $R_0$.

In Section III, we discuss the performance and complexity of binary block codes and convolutional codes for the low-SNR regime at rates below $R_0$. We also examine more powerful binary coding and decoding techniques such as sequential decoding, code concatenation with outer Reed–Solomon codes, turbo codes, and low-density parity-check codes.

In Section IV, we do the same for lattice and nonbinary trellis codes, which are the high-SNR analogs of binary block and convolutional codes, respectively. We also discuss more powerful schemes, such as multilevel coding with binary turbo codes [109].

Finally, in Section V, we consider the general linear Gaussian channel. We review water pouring and discuss approaching capacity with multicarrier modulation. For serial single-carrier transmission, we explain how to reduce the channel to an equivalent discrete-time channel without loss of optimality. We show that capacity can be approached if the intersymbol interference in this channel can be eliminated, and discuss various transmitter precoding techniques that have been developed to achieve this objective.

Appendix I discusses various forms of multicarrier modulation. In Appendix II, several uniformity properties that have proved to be helpful in the design and analysis of Euclidean-space codes are reviewed. Appendix III addresses discrete-time spectral factorization.

II. MODULATION AND CODING FOR THE IDEAL AWGN CHANNEL

In this section, we first define the general linear Gaussian channel and the ideal band-limited AWGN channel. We then review minimum-bandwidth orthogonal pulse amplitude modulation (PAM) for serial transmission of real and complex symbol sequences over the ideal AWGN channel. As is well known, orthogonal transmission at a rate of $1/T$ real (respectively, complex) symbols per second requires a minimum one-sided bandwidth of $1/2T$ Hz (respectively, $1/T$ Hz).

This can be accomplished in a variety of ways. Any such modulation technique converts the continuous-time AWGN channel without loss of optimality to an ideal discrete-time AWGN channel.

We then review the channel capacity of the ideal AWGN channel, and give capacity curves for equiprobable $M$-ary PAM ($M$-PAM) inputs. We emphasize the significant differences between the low-SNR and high-SNR regimes. As a fundamental figure of merit for coding, we use the normalized signal-to-noise ratio $SNR_{\text{norm}}$ defined in Section II-E, and compare it with the traditional figure of merit $E_b/N_0$. We propose that $E_b/N_0$ should be used only in the low-SNR regime. We give baseline performance curves for uncoded 2-PAM modulation as a function of $SNR_{\text{norm}}$ and $E_b/N_0$, and for $M$-PAM as a function of $SNR_{\text{norm}}$. Finally, we discuss union-bound performance analysis.

A. General Linear Gaussian Channel and Ideal AWGN Channel

The general linear Gaussian channel is a real waveform channel with input signal $s(t)$, a channel impulse response $g(t)$, and additive Gaussian noise $n(t)$. The received signal is

$$ r(t) = s(t) * g(t) + n(t) $$

where “$*$” denotes convolution. The Fourier transform (spectral response) of $g(t)$ will be denoted by $G(f)$. The input signal is subject to a power constraint $P$, and the noise has one-sided power spectral density (p.s.d.) $N(f)$.

Since all waveforms are real and thus have Hermitian-symmetric spectra about $f = 0$, only positive frequencies need to be considered. We will follow this long-established tradition, although for analytic purposes it is often preferable to consider both positive and negative frequencies. All bandwidths and power spectral densities in this paper will therefore be “one-sided.”

The ideal band-limited additive white Gaussian noise (AWGN) channel (for short, the ideal AWGN channel) is a linear Gaussian channel with flat (constant) $G(f)$ and $N(f)$ over a frequency band $B$ of one-sided bandwidth $W = \int_B df$. Signals can be transmitted only in the band $B$, which is not necessarily one continuous interval. The restriction to $B$ may be due either to the channel ($G(f) = 0$ for $f \not\in B$) or to system constraints.

Without loss of generality, we may normalize $G(f)$ to 1 within $B$, so that

$$ r(t) = s(t) + n(t) $$

where $n(t)$ is AWGN with one-sided p.s.d. $N(f) = N_0$. The signal-to-noise ratio is then

$$ SNR = P/N_0W. $$

B. Real and Complex Pulse Amplitude Modulation

Let $\{a_k\}$ be a sequence of real modulation symbols representing digital data. Assume that $\{\phi_k(t)\}$ represents a set of real orthonormal waveforms; i.e., $\int_0^T \phi_{k}(t)\phi_{k'}(t)\,dt = \delta_{k,k'}$. 


where the integral is over all time, and \( \delta_\ell \) denotes the Kronecker delta function: \( \delta_\ell = 1 \) if \( \ell = 0 \), else 0. A PAM modulator transmits the continuous-time signal
\[
s(t) = \sum_i a_i \phi(t)
\] over an ideal AWGN channel. An optimal demodulator recovers a sequence \( \{z_i\} \) of noisy estimates of the transmitted symbols \( \{a_i\} \) by correlating the received signal \( r(t) = s(t) + n(t) \) with the waveforms \( \{\phi_i(t)\} \) (matched filtering)
\[
z_i = \int r(t) \phi_i(t) dt = a_i + w_i.
\]
By optimum detection theory [118], the sequence \( \{z_i\} \) is a set of sufficient statistics about the sequence \( \{a_i\} \); i.e., all information about \( \{a_i\} \) that is contained in the continuous-time received signal \( r(t) \) is condensed into the discrete-time sequence \( \{z_i\} \). Moreover, the orthonormality of the \( \{\phi_i(t)\} \) ensures that there is no intersymbol interference (ISI), and that the sequence \( \{w_i\} \) is a set of independent and identically distributed (i.i.d.) Gaussian random noise variables with mean zero and variance \( \sigma^2_w = N_0/2 \). The waveform channel is thus reduced to an equivalent discrete-time ideal AWGN channel.

For serial PAM transmission, the orthogonal waveforms are chosen to be time shifts of a pulse \( p(t) \) by integer multiples of the modulation interval \( T \); i.e., \( \{\phi_i(t)\} = \{p(t - \ell T)\} \), so that the transmitted signal becomes
\[
s(t) = \sum_i a_i p(t - \ell T).
\]

The orthonormality condition requires that
\[
\int p(t - \ell T) p(t - \ell' T) dt = \delta_{\ell, \ell'}
\]
which is equivalent to
\[
p(t) * p(-t) = h(t) = \delta_{\ell, \ell'}.
\]
The frequency-domain equivalent to (2.7) is known as the Nyquist criterion for zero ISI, namely,
\[
\tilde{H}(f) = \frac{1}{T} \sum_{m \in \mathbb{Z}} H(f + m/T) = \frac{1}{T} \sum_{m \in \mathbb{Z}} \left| P(f + m/T) \right|^2 = 1, \quad \text{for all } f
\]
where \( H(f) = \left| P(f) \right|^2 \) is the Fourier transform of \( h(t) = p(t) * p(-t) \) and \( \tilde{H}(f) \) is the \((1/T)\)-aliased spectrum of \( H(f) \).
The Nyquist criterion imposes no restriction on the phase of \( P(f) \). Because \( \tilde{H}(f) \) is periodic with period \( 1/T \), and for real signals \( H(f) \) and thus \( \tilde{H}(f) \) are symmetric around \( f = 0 \), it suffices that \( \tilde{H}(f) = 1 \) for the frequency interval \( B_N = \{f: 0 \leq f \leq 1/2T\} \). This implies that for each \( f \in B_N, \ H(f + m/T) \) > 0 for at least one \( m \in \mathbb{Z} \). It follows that serial ISI-free transmission of \( 1/T \) real symbols per second over a real channel requires a minimum one-sided bandwidth of \( W = 1/2T \) Hz, or 1/2 Hz per symbol dimension transmitted per second (1/2 Hz/dim/s).

For serial baseband transmission, the Nyquist criterion is met with minimum bandwidth by the brick-wall spectrum
\[
\left| P(f) \right|^2 = H_b(f) = \begin{cases} T, & |f| \leq 1/2T \\ 0, & \text{elsewhere} \end{cases}
\]
\[
\iff p_b(t) * p_b(-t) = h_b(t) = \sin(\pi f/T)/\pi f/T.
\]

Solutions requiring more bandwidth include spectra with symmetric finite rolloffs at the band-edge frequencies \( \pm 1/2T \), e.g., the well-known family of raised-cosine spectra [74]. Of course, the Nyquist criterion is also satisfied by the spectrum \( \left| P(f) \right|^2 = T \sin(\pi f T)/(\pi f T) \), which corresponds to “unfiltered” modulation with the rectangular pulse \( p(t) = 1/\sqrt{T} \) for \( 0 \leq t \leq T \) (0 elsewhere).

Serial passband transmission of real modulation symbols can be achieved with minimum bandwidth by \( |P(f)|^2 = T \) for \( |f|/2T \leq |f| \leq (\ell + 1)/2T, \ell \geq 1 \), and zero elsewhere. This type of modulation may be referred to as carrierless single-sideband (CSSB) modulation; it is not often used. Serial passband transmission of complex modulation symbols \( a_i = a_i^R + j a_i^I \) by carrierless amplitude-phase (CAP) modulation or quadrature amplitude modulation (QAM) is usually preferred.

To transmit a complex signal over a real passband channel, one may use a complex signal with only positive-frequency components. Sending the real part thereof creates a negative-frequency image spectrum. In the receiver, the complex signal is recovered by suppressing the negative-frequency spectrum. Let \( p_c(t) \) be as defined in (2.9), and let
\[
p_c(t) = p_c(t) \exp(j 2\pi f_c t) = p_c^R(t) + j p_c^I(t)
\]
where \( f_c > T/2 \) so that the Fourier transform of \( p_c(t) \) is analytic, i.e., \( \hat{p}_c(f) = 0 \) for \( f < 0 \). Then \( p_C(t) \) is the Hilbert transform of \( p_c(t) \), and the two pulses are said to form a Hilbert pair. The two pulses are orthogonal with energy \( \sigma^2_w = N_0/2 \). The waveform channel is thus reduced to an equivalent discrete-time ideal AWGN channel.

The transmit signals for minimum-bandwidth CAP and QAM are, respectively,
\[
s_{\text{CAP}}(t) = \text{Re}\left\{ \sum_i a_i p_c(t - i T) \right\}
\]
\[
= \sum_i a_i^R p_c^R(t - i T) - a_i^I p_c^I(t - i T) \quad (2.10)
\]
and
\[
s_{\text{QAM}}(t) = \text{Re}\left\{ \sum_i a_i p_c(t - i T) \exp(j 2\pi f_c t) \right\}
\]
\[
= \sum_i b_i^R p_c^R(t - i T) - b_i^I p_c^I(t - i T),
\]
where \( b_i = a_i \exp(j 2\pi f_c t) \). Notice that QAM is equivalent to CAP with the symbols \( a_i \) replaced by the rotated symbols \( b_i \). If \( f_c \) is an integer multiple of \( 1/T \), then there is no difference between CAP and QAM.

An advantage of CAP over CSSB is that the CAP spectrum does not need to be aligned with integer multiples of \( 1/2T \); as is the case for CSSB. CAP may be preferred over QAM when \( f_c \) does not substantially exceed \( 1/2T \). On the other hand, QAM is the modulation of choice when \( f_c \gg 1/2T \). For ISI-free transmission of \( 1/T \) complex (two-dimensional) signals.
CAP and QAM require a minimum one-sided bandwidth of \( W = 1/T \) Hz, which again is 1/2 Hz/dim/s.

There exists literally an infinite variety of orthonormal modulation schemes with the same bandwidth efficiency. Consider general serial/parallel PAM modulation in which real \( N \)-vectors \( \mathbf{a}_i = [a_i^0, a_i^1, \ldots, a_i^{N-1}] \) are transmitted at a rate of \( 1/T \) vectors per second over an ideal AWGN channel. Let \( \mathbf{p}(t) = [p^0(t), p^1(t), \ldots, p^{N-1}(t)] \) be an \( N \)-vector of real pulses. A general PAM modulator transmits

\[
\mathbf{s}(t) = \sum_i \langle \mathbf{a}_i, \mathbf{p}(t-iT) \rangle \tag{2.12}
\]

where \( \langle \cdot, \cdot \rangle \) denotes the inner product. If \( \{p^n(t-iT)\} \) is a set of orthonormal functions, an optimum demodulator may recover noisy estimates of \( \mathbf{z}_i \) of \( \mathbf{a}_i \) from \( r(t) \) by matched filtering

\[
\mathbf{z}_i = \int r(t) \mathbf{p}(t-iT) dt = \mathbf{a}_i + \mathbf{w}_i. \tag{2.13}
\]

The sequence \( \{\mathbf{w}_i\} \) is a sequence of i.i.d. Gaussian \( N \)-vectors, whose elements have variance \( \sigma_w^2 \). Thus the AWGN waveform channel is converted to a discrete-time vector AWGN channel, which is equivalent to \( N \) real discrete-time AWGN channels operating in parallel.

Let \( \mathbf{P}(f) = [P^0(f), P^1(f), \ldots, P^{N-1}(f)] \) be the vector of the Fourier transforms of the elements of \( \mathbf{p}(t) \). The frequency-domain condition for orthonormality is the generalized Nyquist criterion [74]

\[
\hat{\mathbf{H}}(f) = \frac{1}{T} \sum_{m \in \mathbb{Z}} \mathbf{H}(f + m/T) = \frac{1}{T} \sum_{m \in \mathbb{Z}} \mathbf{P}^*(f + m/T) \mathbf{P}(f + m/T) = I_N, \quad \text{for all} \; f, \tag{2.14}
\]

In (2.14), \( \mathbf{P}^*(f) \) denotes the conjugate transpose of \( \mathbf{P}(f) \). \( \mathbf{H}(f) \) is the \( N \times N \) matrix \( \mathbf{H}(f) = \{[\mathbf{p}^m(f)]^* \mathbf{p}^n(f), 0 \leq n, m \leq N-1\} \). \( \hat{\mathbf{H}}(f) \) denotes the \((1/T)\)-aliased spectrum of \( \mathbf{H}(f) \), and \( I_N \) is the \( N \times N \) identity matrix. As with (2.8), it suffices that \( \hat{\mathbf{H}}(f) = I_N \) for the frequency interval \( B_N = \{f: 0 \leq f \leq 1/2T\} \). Now consider the \( \infty \times N \) matrix

\[
\mathbf{U}(f) = T^{-1/2} \times \{\mathbf{p}^m(f + m/T), m \in \mathbb{Z}, 0 \leq n \leq N-1\}. \tag{2.15}
\]

Then (2.14) can be written as \( \hat{\mathbf{H}}(f) = \mathbf{U}^*(f) \mathbf{U}(f) = I_N \), which implies that \( \mathbf{U}(f) \) must have rank \( N \) for all \( f \). It follows that the total one-sided bandwidth for which some \( \mathbf{P}^m(f) \), \( 0 \leq n \leq N-1, f \geq 0 \) is nonzero must be at least \( N/2T \) Hz. Because \( N/T \) symbol dimensions are transmitted per second, again a minimum one-sided bandwidth of \( 1/2 \) Hz/dim/s is needed.

If \( \mathbf{U}(f) \) meets the minimum-bandwidth condition, then the \( N \) nonzero rows of \( \mathbf{U}(f) \) must form an \( N \times N \) unitary matrix in order that \( \mathbf{U}^*(f) \mathbf{U}(f) = I_N \). Conversely, any set of unitary matrices \( \{\mathbf{U}(f), f \in B_N\} \) defines a minimum-bandwidth orthonormal transmission scheme. The number of such schemes is therefore uncountably infinite.

Appendix I describes two currently popular multicarrier modulation schemes, namely, discrete multitone (DMT) and discrete wavelet multitone (DWMT) modulation.

The optimum design of transmitter and receiver filters for serial/parallel transmission over general noisy linear channels has been discussed in [28].

### C. Capacity of the Ideal AWGN Channel

We have seen that all orthonormal modulation schemes with optimum matched-filter detection are equivalent to an ideal discrete-time AWGN channel with real input symbols \( a_i \) and real output signals \( z_i \), where \( z_i = a_i + w_i \). The i.i.d. Gaussian noise variables \( w_i \) have zero mean and variance \( \sigma_w^2 = N_0/2 \).

Let the average input-symbol energy per dimension be \( E_s = E[a_i^2] \), and let the average energy per information bit be \( E_b = E_s/R \), where \( R \) is the code rate in bits per dimension (b/dim). Because \( P = E_s/T \) and \( W = 1/2T \), the signal-to-noise ratio (SNR) of the AWGN waveform channel and the discrete-time AWGN channel are identical:

\[
\text{SNR} = \frac{P}{N_0 W} = \frac{E_s}{\sigma_w^2} = 2R \frac{E_b}{N_0}. \tag{2.16}
\]

The channel capacity is the maximum mutual information between channel input and output, which is obtained with a Gaussian distribution over the symbols \( a_i \) with average power \( E_s \). Shannon’s most famous result states the capacity in bits per second as

\[
C_{[b/s]} = W \log_2 (1 + \text{SNR}), \tag{2.17}
\]

Equivalently, the capacity in bits per dimension is given by

\[
C = \frac{1}{2} \log_2 (1 + \text{SNR}) \; [\text{b/dim}]. \tag{2.18}
\]

Shannon [93], [94] showed that reliable transmission is possible for any rate \( R < C \), and impossible if \( R > C \). (Unless stated otherwise, in this paper capacity \( C \) and code rates \( R \) are given in bits per dimension.)

The capacity can be upper-bounded and approximated for low SNR by a linear function in SNR, and lower-bounded and approximated for high SNR by a logarithmic function in SNR, as follows:

\[
\text{SNR} \ll 1: \quad C \approx \frac{1}{2} \frac{1}{\text{SNR}} \log_2 \left( \frac{e}{\text{SNR}} \right), \tag{2.19}
\]

\[
\text{SNR} \gg 1: \quad C \approx \frac{1}{2} \log_2 (\text{SNR}). \tag{2.20}
\]

A finite set \( A \) from which modulation symbols \( a_i \) can be chosen is called a signal alphabet or signal constellation. An \( M \)-PAM constellation contains \( M \geq 2 \) equidistant real symbols centered on the origin; i.e., \( A = \{d_0/2\{-M + 1, -M + 3, \ldots, M - 1\}\} \), where \( d_0 \) is the minimum distance between symbols. For example, \( A = \{-3, -1, +1, +3\} \) is a 4-PAM constellation with \( d_0 = 2 \). If the symbols are equiprobable, then the average symbol energy is

\[
E_s = (M^2 - 1)d_0/12. \tag{2.21}
\]
Fig. 1 shows the capacity and the mutual information achieved with equiprobable $M$-PAM ("equiprobable $M$-PAM capacity"), for $M = 2, 4, \cdots, 64$, as a function of SNR. The $M$-PAM curves saturate because information cannot be sent at a rate higher than $R = \log_2 M$.

D. Low-SNR and High-SNR Regimes

We see from Fig. 1 that in the low-SNR regime an equiprobable binary alphabet is nearly optimal. For SNR < 1 (0 dB), the reduction in capacity is negligible.

In the high-SNR regime, the capacity of equiprobable $M$-PAM constellations asymptotically approaches a straight line parallel to the capacity of the AWGN channel. The asymptotic loss of $\pi e/6$ (1.53 dB) is due to using a uniform rather than a Gaussian distribution over the signal set. To achieve capacity, the use of powerful coding with equiprobable $M$-PAM signals is not enough. To obtain the remaining 1.53 dB, constellation-shaping techniques that produce a Gaussian-like distribution over an $M$-PAM constellation are required (see Section IV-B).

Thus coding techniques for the low-SNR and high-SNR regimes are quite different. In the low-SNR regime, binary codes are nearly optimal, and no constellation shaping is required. Good binary coding and decoding techniques have been known since the 1960’s.

On the other hand, in the high-SNR regime, nonbinary signal constellations must be used. Very little progress in developing practical codes for this regime was made until the advent of trellis-coded modulation in the late 1970’s and 1980’s. To approach capacity, coding techniques must be supplemented with constellation-shaping techniques. Moreover, on bandwidth-limited channels, ISI is often a dominant impairment, and practical techniques for combined coding, shaping, and equalization are required to approach capacity. Most of the progress in these areas has been made only in the past decade.

For these reasons, we discuss coding for the low-SNR and high-SNR regimes separately in Sections III and IV, and coding for the general linear Gaussian channel in Section V.

E. Baseline Performance of Uncoded $M$-PAM and Normalized SNR

In an uncoded $M$-PAM system, $R = \log_2 M$ information bits are independently encoded into each $M$-PAM symbol transmitted. In the receiver, the optimum decoding rule is then to make independent symbol-by-symbol decisions. The probability that a Gaussian noise variable $w_i$ exceeds half of the distance $d_0$ between adjacent $M$-PAM symbols is $Q(d_0/2\sigma_w)$, where \[ Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-y^2/2)\,dy < \exp(-x^2/2) \] (2.22)

The error probability per symbol is given by $P_s(E_{\text{outer}}) = Q(d_0/2\sigma_w)$ for the two outer points, and by $P_s(E_{\text{inner}}) = 2Q(d_0/2\sigma_w)$ for the $M - 2$ inner points, so the average error probability per symbol is

\[ P_s(E) = \frac{2(M - 1)}{M} Q(d_0/2\sigma_w) = \frac{2(M - 1)}{M} Q\left(\sqrt{\frac{3\text{SNR}}{M^2 - 1}}\right) \] (2.23)

where $(d_0/2\sigma_w)^2 = 3\text{SNR}/(M^2 - 1)$ follows from (2.16) and (2.21). Thus $P_s(E)$ is a function only of $M$ and SNR. The circles in Fig. 1 indicate the values of SNR for which $P_s(E) = 10^{-6}$ is achieved.

The capacity formula (2.18) can be rewritten as $\text{SNR}/(2^R - 1) = 1$. This suggests defining the normalized SNR

\[ \text{SNR}_{\text{norm}} = \frac{\text{SNR}}{2^R - 1} \] (2.24)
where $R$ is the actual data rate of a given modulation and coding scheme. For a capacity-achieving scheme, $R$ equals the channel capacity $C$ and $\text{SNR}_{\text{norm}} = 1$ (0 dB). If $R < C$, as will always be the case in practice, then $\text{SNR}_{\text{norm}} > 1$. The value of $\text{SNR}_{\text{norm}}$ thus signifies how far a system is operating from the Shannon limit (the “gap to capacity”).

For uncoded $M$-PAM, from $R = \log_2 M$ and (2.24) one obtains

$$\text{SNR} / (M^2 - 1) = \text{SNR} / (2^R - 1) = \text{SNR}_{\text{norm}}.$$ 

Therefore, the average error probability per symbol of uncoded $M$-PAM can be written as

$$P_b(\mathcal{E}) = \frac{2(M - 1)}{M} Q(\sqrt{3\text{SNR}_{\text{norm}}}) \approx 2Q(\sqrt{3\text{SNR}_{\text{norm}}}) (M \text{ large}). \tag{2.25}$$

Note that the baseline $M$-PAM performance curve of $P_b(\mathcal{E})$ versus $\text{SNR}_{\text{norm}}$ is nearly independent of $M$, if $M$ is large. This shows that $\text{SNR}_{\text{norm}}$ is appropriately normalized for rate in the high-SNR regime.

Because $E_b/N_0 = \text{SNR}/2R$ by (2.16), the general relation between $\text{SNR}_{\text{norm}}$ and $E_b/N_0$ at a given rate $R$ (in bits per dimension) is given by

$$\frac{E_b}{N_0} = \frac{2^R - 1}{2R} \text{SNR}_{\text{norm}}. \tag{2.26}$$

If $R \ll 1$, then $E_b/N_0 \approx (\ln 2) \text{SNR}_{\text{norm}}$, so the two figures of merit are equivalent. If $R = 1/2$, then $E_b/N_0 = \text{SNR}_{\text{norm}}$; if $R = 1$, then $E_b/N_0 = (3/2) \text{SNR}_{\text{norm}}$.

In the low-SNR regime, if bandwidth is truly unconstrained, then as bandwidth is increased to permit usage of powerful low-rate binary codes, both SNR and $R$ tend toward zero. In this power-limited regime, it is appropriate to use as a figure of merit the traditional ratio $E_b/N_0$.

From the general Shannon limit $\text{SNR}_{\text{norm}} > 1$, we obtain the Shannon limit on $E_b/N_0$ for a given rate $R$

$$E_b/N_0 > (2^R - 1)/2R. \tag{2.27}$$

This lower bound decreases monotonically with $R$, and as $R$ approaches 0, it approaches the ultimate Shannon limit

$$E_b/N_0 > \ln 2 (-1.59 \text{ dB}). \tag{2.28}$$

However, if bandwidth is limited, then the Shannon limit on $E_b/N_0$ is higher. For example, if $R = 1/2$, then $E_b/N_0 > 1$ (0 dB).

Because $E_b/N_0 = (3/2) \text{SNR}_{\text{norm}}$ at $R = 1$, the error probability per symbol (or per bit) of uncoded 2-PAM may be expressed in two equivalent ways

$$P_b(\mathcal{E}) = Q(\sqrt{3\text{SNR}_{\text{norm}}}) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \tag{2.29}$$

In this paper we will use $E_b/N_0$ in the low-SNR regime and $\text{SNR}_{\text{norm}}$ in the high-SNR regime as the fundamental figures of merit of uncoded and coded modulation schemes. This practice appears to be on its way to general adoption.

The “effective coding gain” of a coded modulation scheme is measured by the reduction in required $E_b/N_0$ or $\text{SNR}_{\text{norm}}$ to achieve a certain target error probability relative to a baseline uncoded scheme. In the low-SNR regime, the baseline will be taken as 2-PAM; in the high-SNR regime, the baseline will be taken as $M$-PAM ($M$ large).

Fig. 2 gives the probability of bit error $P_b(\mathcal{E})$ for uncoded 2-PAM as a function of both $\text{SNR}_{\text{norm}}$ and $E_b/N_0$. Note that the axis for $E_b/N_0$ is shifted by a factor of $3/2$ (1.76 dB) relative to the axis for $\text{SNR}_{\text{norm}}$. At $P_b(\mathcal{E}) = 10^{-6}$, the baseline
unencoded binary modulation scheme operates about 12.5 dB away from the Shannon limit. Therefore, a coding gain of up to 12.5 dB in $E_b/N_0$ is in principle possible at this $H_b(\mathcal{E})$, provided that bandwidth can be expanded sufficiently to permit the use of powerful very-low-rate binary codes ($R \ll 1$). If bandwidth can be expanded by a factor of only 2, then with binary codes of rate $R = 1/2$ a coding gain of up to about 10.8 dB can be achieved.

Fig. 2 also depicts the probability of symbol error $P_s(\mathcal{E})$ for unencoded $M$-PAM as a function of $SNR_{\text{XMM}}$ for large $M$ (in this case, the $E_b/N_0$ axis should be ignored). At $P_s(\mathcal{E}) = 10^{-6}$, a baseline unencoded $M$-PAM modulation scheme operates about 9 dB away from the Shannon limit in the bandwidth-limited regime. Thus if bandwidth is a fixed, nonexpandable resource, a coding gain of up to about 9 dB in $SNR_{\text{XMM}}$ is in principle possible at $P_s(\mathcal{E}) = 10^{-6}$. This conclusion holds even at $R = 1$.

### F. The Union Bound

The union bound is a useful tool for evaluating the performance of moderately powerful codes, although it breaks down for rates beyond the cutoff rate $R_0$.

The union bound is based on evaluating pairwise error probabilities between pairs of coded sequences. On an ideal real discrete-time AWGN channel, the received sequence $z = a + w$ is the sum of the transmitted coded sequence $a$ and an i.i.d. Gaussian sequence $w$ with mean 0 and variance $\sigma_w^2$ per symbol. Because $p(z|a) \propto \exp\left(-\frac{|z-a|^2}{2\sigma_w^2}\right)$, maximum-likelihood (ML) decoding is equivalent to minimum-distance decoding.

Given two coded sequences $a$ and $a'$ that differ by the Euclidean distance $d(a, a')$, the probability that the received sequence $z = a + w$ will be closer to $a'$ than to $a$ is given by

$$P(\mathcal{E}|a) = Q(d(a, a')/2\sigma_w). \quad (2.30)$$

The probability that $z$ will be closer to $a'$ than to $a$ for any $a' \neq a$, which is precisely the probability of error with ML decoding, is thus upperbounded by

$$P(\mathcal{E}|a) \leq \sum_{a' \neq a} Q(d(a', a)/2\sigma_w). \quad (2.31)$$

The average probability of error over all coded sequences $a$ is upperbounded by the union bound

$$P(\mathcal{E}) \leq \frac{K_d}{d}Q(d/2\sigma_w) \quad (2.32)$$

where $K_d$ is the average number of coded sequences $a' \neq a$ at distance $d$ from $a$.

The Gaussian error probability function $Q(x)$ decays exponentially as $\exp(-x^2/2)$. Therefore, if $K_d$ does not rise too rapidly with $d$, the union bound is dominated by its first term, which is called the union bound estimate (UBE)

$$P(\mathcal{E}) \approx K_{\min}Q(d_{\min}/2\sigma_w) \quad (2.33)$$

where $d_{\min}$ is the minimum distance between coded sequences, and $K_{\min}$ is the average number of sequences at minimum distance $d_{\min}$ from a given coded sequence. The squared argument of the $Q(\bullet)$ function in the union bound estimate yields a first-order estimate of performance, namely,

$$\frac{P^2_{\text{min}}}{4\sigma_w^2} = \frac{d_{\min}^2}{2\sigma_w^2} R 2E_b \quad (2.34)$$

where $d_0^2$ is the squared baseline distance, $R_0 = \log_2 M$ is the baseline rate, and $R \leq R_0$ is the actual rate.

In comparison to the baseline squared distance, there is a distance gain of a factor of $d_{\min}^2/d_0^2$. However, in comparison to the baseline rate $R_0$ there is also a redundancy loss. In the high-SNR regime, the redundancy loss is approximately $2^{2R} = 2^{2(R-R_0)}$, where $r = R_0 - R$ is the redundancy of the code in b/dim. In the low-SNR regime, with 2-PAM signaling ($R_0 = 1$), the redundancy loss is simply a factor of $R/R_0 = R$. The product of these factors gives the nominal (or asymptotic) coding gain, which is based solely on the argument of the $Q(\bullet)$ function in the UBE.

The effective coding gain is the difference between the signal-to-noise ratios required to achieve a certain target error rate with a coded modulation scheme and to obtain the same error rate with an unencoded scheme of the same rate $R$. The effective coding gain is typically less than the nominal coding gain, because the error coefficient $K_{\min}$ is typically larger than the baseline error coefficient. Note that $K_{\min}$ depends on exactly what is being estimated, e.g., bit-error probability, block-error probability, etc., over which interval, per bit, per dimension, per block, etc. A rule of thumb based on the slope of the $Q(\bullet)$ curve in the $10^{-5}$–$10^{-6}$ region is that every increase of a factor of two in the error coefficient $K_{\min}$ costs about 0.2 dB in effective coding gain, provided that $K_{\min}$ is not too large.

The union bound blows up even for optimal codes at the cutoff rate $R_0$, i.e., the corresponding error exponent goes to zero [51]. For low-SNR channels, the cutoff rate is 3 dB away from the Shannon limit; i.e., the cutoff rate limit is $E_b/N_0 = 2 \ln 2$ (1.42 dB). For high-SNR channels, the cutoff rate corresponds to an $SNR_{\text{norm}}$ which is a factor of $4/e$ away from the Shannon limit [95]; i.e., the cutoff rate limit is $SNR_{\text{norm}} = 4/e$ (1.68 dB).

Beyond $R_0$, codes cannot be analyzed by using the union bound estimate. Any apparent agreement between a union bound estimate and actual performance in the region between $R_0$ and $C$ must be regarded as fortuitous. Because the operational significance of minimum distance or nominal coding gain is that they give a good estimate of performance via the union bound, their significance in the beyond-$R_0$ regime is questionable.

### III. Binary Codes for Power-Limited Channels

In this section we discuss coding techniques for power-limited (low-SNR) ideal AWGN channels. As we have seen, low-rate binary codes are near-optimal in this regime, as long as soft decisions are used at the output. We develop a union bound estimate of probability of error per bit with maximum-likelihood (ML) decoding. We also show that the nominal
coding gain of a rate-$k/n$ binary code $C$ with minimum Hamming distance $d$ over baseline 2-PAM is simply $\gamma_c(C) = (k/n)d$.

Then we give the effective coding gains of known moderate-complexity block and convolutional codes versus the branch complexity of their minimal trellises, assuming ML decoding at rates below $R_0$. Convolutional codes are clearly superior in terms of coding gain versus trellis complexity.

Finally, we briefly discuss higher-performance codes, including convolutional codes with sequential decoding, concatenated codes with outer Reed–Solomon codes, and turbo codes and other capacity-approaching codes.

A. Optimality of Low-Rate Binary Linear Codes with Soft Decisions

We have seen in Fig. 1 that on an AWGN channel with $\text{SNR} < 1$, the capacity with equiprobable binary signaling is negligibly less than the true capacity. Moreover, on symmetric channels such as the ideal AWGN channel, it has long been known that there is no reduction in capacity if one restricts one’s attention to linear binary codes [31], [51].

As shown by (2.26), if the code rate $R = k/n$ is greater than zero, the figure of merit $E_b/N_0$ is lower-bounded by $(2^{2R} - 1)/2R$, which exceeds the ultimate Shannon limit of $\ln 2 (\sim 1.59 \text{ dB})$. If $R$ is small, then

$$(2^{2R} - 1)/2R \approx (\ln 2)(1 + R \ln 2) \approx (\ln 2)2^R$$

so the lower bound is approximately

$$E_b/N_0 > (\ln 2)2^R (-1.59 + 3.01R \text{ decibels}).$$

For example, the penalty to operate at rate $R = 1/4$ is about 0.77 dB, which is not insignificant. However, there may be a limit on bandwidth, which scales as $1/R$. Moreover, even if bandwidth is unlimited, there is usually a technology-dependent lower limit on the energy $E_b = R\sigma_w^2$ per transmitted symbol below which the performance of other system elements (e.g., tracking loops) degrades significantly.

On the other hand, whereas two-level quantization of the channel input costs little, two-level quantization of the channel output (hard decisions) costs of the order of 2–3 dB [118]. Optimized three-level quantization (hard decisions with erasures) costs only about half as much (1–1.5 dB) [39]. However, optimized uniform eight-level (3-bit) quantization (quantized soft decisions) costs only about 0.2 dB, and has often been used in practice [55].

B. The Union Bound and Nominal Coding Gain

An $(n, k, d)$ binary linear block code $C$ is a $k$-dimensional subspace of the $n$-dimensional binary vector space $\mathbb{F}_2^n$ such that the minimum Hamming distance between any two code $n$-tuples is $d$. Its size is $|C| = 2^k$.

For use on the Gaussian channel, code bits are usually mapped to real numbers via the standard 2-PAM map $m : \{0, 1\} \rightarrow \{\pm d_0/2\}$. The Euclidean image $m(C)$ is then a subset of $2^k$ vertices of the $2^n$ vertices $m(\mathbb{F}_2^n) \subseteq \mathbb{R}^n$ of an $n$-cube of side $d_0$ centered on the origin. If two code $n$-tuples $\mathbf{c}, \mathbf{d} \in C$ differ in $d_H$ places, then the squared distance between their Euclidean images $m(\mathbf{c})$ and $m(\mathbf{d})$ is $d_H^2d_0^2$. Consequently, the minimum squared Euclidean distance between any two vectors in $m(C)$ is

$$d_{\text{min}}^2(C) = d_H^2d_0^2.$$  \hspace{1cm} (3.3)

Moreover, if $C$ has $K_d$ words of minimum weight $d$, then by linearity there are $K_d$ vectors $m(\mathbf{c})$ at minimum squared distance $d_{\text{min}}^2(C)$ from every vector $m(\mathbf{a})$ in $m(C)$.

The union bound estimate (UBE) of the probability of a block decoding error $P(\mathcal{E})$ is then

$$P(\mathcal{E}) \approx K_dQ(d_{\text{min}}^2(C)/2\sigma_w^2) = K_dQ(\sqrt{d_H^2d_0^2}/2\sigma_w^2).$$

Because $E_b = (n/k)d_0^2/4$ and $\sigma_w^2 = N_0/2$, we may write the UBE as

$$P(\mathcal{E}) \approx K_dQ(\sqrt{\gamma_c(C)}E_b/N_0)$$

where the nominal coding gain of an $(n, k, d)$ binary linear code $C$ is defined as

$$\gamma_c(C) = (k/n)d.$$  \hspace{1cm} (3.6)

This is the product of a distance gain of a factor of $d$ with a redundancy loss of a factor of $R = k/n$. The uncoded 2-PAM baseline corresponds to a (1, 1, 1) code, for which $\gamma_c(C) = 1$ and $K_d = 1$.

For comparability with the uncoded baseline, it is appropriate to normalize $P(\mathcal{E})$ by $k$ to obtain the probability of block decoding error per information bit (not the bit-error probability)

$$P_b(\mathcal{E}) \approx K(C)Q(\sqrt{\gamma_c(C)}E_b/N_0)$$

in which the normalized error coefficient is $K(C) = K_d/k$.

Graphically, a curve of the form of (3.7) may be obtained simply by moving the baseline curve $P_b(\mathcal{E}) = Q(\sqrt{2E_b/N_0})$ of Fig. 2 to the left by $\gamma_c(C)$ (in decibels), and upward by a factor of $K(C)$.

C. Biorthogonal Codes

Biorthogonal codes are asymptotically optimal codes for the ideal AWGN channel, provided that bandwidth is unlimited. They illustrate both the usefulness and the limitations of union bound estimates.

For any integer $k > 1$ there exists a $(2^{k-1}, k, 2^{k-2})$ binary linear code $C$, e.g., a first-order Reed–Muller code, such that its Euclidean image $m(C)$ is a biorthogonal signal set in $2^{k-1}$-space (i.e., a set of $2^{k-1}$ orthogonal vectors and their negatives). Its nominal coding gain is

$$\gamma_c(C) = k/2$$

which goes to infinity as $k$ goes to infinity.

Such a code consists of the all-zero word, the all-one word, and $K_d = 2^{k-2}$ codewords of weight $d = 2^{k-2}$. The union bound on block error probability thus becomes

$$P(\mathcal{E}) \leq (2^k - 2)Q(\sqrt{kE_b/N_0}) + Q(\sqrt{2kE_b/N_0})$$

$$< 2^kQ(\sqrt{E_b/N_0})$$

$$< \exp(-k(E_b/2N_0 - \ln 2))$$

(3.9)
where we have used the upper bound $Q(x) < \exp(-x^2/2)$. Thus if $E_b/N_0 > 2 \ln 2$ (1.42 dB), or if $\text{SNR}_{\text{norm}} > 2$, then the union bound approaches zero exponentially with $k$. Thus the union bound shows that with biorthogonal codes it is possible to approach within 3 dB of the Shannon limit, which corresponds to the cutoff rate $R_0$ of the low-SNR AWGN channel.

A more refined upper bound [118] shows that for

$$P(\mathcal{E}) < \exp(-k(\sqrt{E_b/N_0} - \sqrt{\ln 2})^2)$$

which approaches zero exponentially with $k$ if $E_b/N_0 > \ln 2$ (-1.59 dB), or if $\text{SNR}_{\text{norm}} > 1$. In other words, this bound shows that biorthogonal codes can achieve an arbitrarily small error probability for any $E_b/N_0$ above the ultimate Shannon limit. It also illustrates the limitations of the union bound, which blows up 3 dB away.

Of course, the code rate $R = k2^{-k+1}$ approaches zero rapidly as $k$ becomes large, so long biorthogonal codes can be used only if bandwidth is effectively unlimited. They can be decoded efficiently by fast Hadamard transform techniques [43], but even so the decoding complexity is proportional to $k2^k$, which increases exponentially with $k$.

A $(32, 6, 16)$ biorthogonal code decoded by a fast Hadamard transform (“Green machine”) was used for an early deep-space mission (Mariner, 1969) [79].

**D. Effective Coding Gains of Known Codes**

For moderate codelengths, Bose–Chaudhuri–Hocquenghem (BCH) codes form a large and well-known class of binary linear block codes that often are the best known codes in terms of their parameters $(n, k, d)$. Unfortunately, not much is known about soft-decision ML decoding algorithms for such codes. The most efficient general methods are trellis-based methods—i.e., the code is represented by a trellis, and the Viterbi algorithm (VA) is used for ML decoding.

There has been considerable recent work on efficient trellis representations of binary linear block codes, including BCH codes [107]. Reed–Muller (RM) codes are as good as BCH codes for $n \leq 32$, and almost as good for $n \leq 128$ in terms of the parameters $(n, k, d)$. Efficient trellis representations are known for all RM codes [43]. In fact, in terms of nominal coding gain $(k/n)d$ or effective coding gain versus trellis complexity, RM codes are often better than BCH codes [107]. Moreover, the number $K_d$ of minimum-weight codewords is known for all RM codes, but not for all BCH codes [77].

In Table I we give the parameters $n, k, d, N_d, \gamma(C)$, and $\gamma_{\text{eff}}(C)$ for all RM codes with lengths $n$ up to 256 and code rates $R = k/n$ up to 1/2, including biorthogonal codes, and also for the $(24, 12, 8)$ Golay code. We also give a parameter $s$ that measures trellis complexity (the log branch complexity). To estimate from $\gamma(C)$ the effective coding gain $\gamma_{\text{eff}}(C)$, we use the normalized error coefficient $K(C) = K_d/k$, and apply the “0.2-dB loss per factor of 2” rule, which is not very accurate for large $K_d$. We remind the reader that these estimated effective coding gains assume soft decisions, ML decoding, and the accuracy of the union bound estimate, and that the complexity parameter assumes trellis-based decoding. Moreover, complexity comparisons are always technology-dependent.

There also exist algebraic decoding algorithms for all of these codes, as well as for BCH codes. Algebraic error-correcting decoders that are based on hard decisions are not appropriate for the AWGN channel, since hard decisions cost 2 to 3 dB. For BCH and other algebraic block codes, there do exist efficient soft-decision decoding algorithms such as generalized minimum distance (GMD) decoding [39] and the Chase algorithms [19] that are capable of approaching ML performance; furthermore, the complexity of “one-pass” GMD decoding [12] is not much greater than that of error-correction only. However, we know of no published results showing that

<table>
<thead>
<tr>
<th>$(n, k, d)$</th>
<th>$\gamma(C)$</th>
<th>$(dB)$</th>
<th>$N_d$</th>
<th>$K(C)$</th>
<th>$\gamma_{\text{eff}}(C)$ $(dB)$</th>
<th>$s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(8, 4, 4)$</td>
<td>2</td>
<td>3.01</td>
<td>14</td>
<td>3.5</td>
<td>2.6</td>
<td>3</td>
</tr>
<tr>
<td>$(16, 5, 8)$</td>
<td>2.5</td>
<td>3.98</td>
<td>30</td>
<td>6</td>
<td>3.5</td>
<td>4</td>
</tr>
<tr>
<td>$(24, 12, 8)$</td>
<td>4</td>
<td>6.02</td>
<td>759</td>
<td>63.2</td>
<td>4.8</td>
<td>9</td>
</tr>
<tr>
<td>$(32, 6, 16)$</td>
<td>3</td>
<td>4.77</td>
<td>62</td>
<td>10.3</td>
<td>4.1</td>
<td>5</td>
</tr>
<tr>
<td>$(32, 16, 8)$</td>
<td>4</td>
<td>6.02</td>
<td>620</td>
<td>38.8</td>
<td>5.0</td>
<td>9</td>
</tr>
<tr>
<td>$(64, 7, 32)$</td>
<td>3.5</td>
<td>5.44</td>
<td>126</td>
<td>18</td>
<td>4.6</td>
<td>6</td>
</tr>
<tr>
<td>$(64, 22, 16)$</td>
<td>5.5</td>
<td>7.40</td>
<td>2604</td>
<td>118.4</td>
<td>6.0</td>
<td>14</td>
</tr>
<tr>
<td>$(128, 8, 64)$</td>
<td>4</td>
<td>6.02</td>
<td>254</td>
<td>31.8</td>
<td>5.0</td>
<td>7</td>
</tr>
<tr>
<td>$(128, 29, 32)$</td>
<td>7.25</td>
<td>8.60</td>
<td>10668</td>
<td>367.9</td>
<td>6.9</td>
<td>20</td>
</tr>
<tr>
<td>$(128, 64, 16)$</td>
<td>8</td>
<td>9.03</td>
<td>94488</td>
<td>1476.4</td>
<td>6.9</td>
<td>30</td>
</tr>
<tr>
<td>$(256, 9, 128)$</td>
<td>4.5</td>
<td>6.53</td>
<td>510</td>
<td>56.7</td>
<td>5.4</td>
<td>8</td>
</tr>
<tr>
<td>$(256, 37, 64)$</td>
<td>9.25</td>
<td>9.66</td>
<td>43180</td>
<td>1167.0</td>
<td>7.6</td>
<td>27</td>
</tr>
<tr>
<td>$(256, 93, 32)$</td>
<td>11.625</td>
<td>10.65</td>
<td>777240</td>
<td>8357.4</td>
<td>8.0</td>
<td>50</td>
</tr>
</tbody>
</table>
the performance–complexity tradeoff for moderate-complexity binary block codes with soft-decision algebraic decoding can be better than that of moderate-complexity binary convolutional codes with trellis-based decoding on AWGN channels.

From this table it appears that the best tradeoff between effective coding gain and trellis complexity is obtained with low-rate biorthogonal codes, if bandwidth is not an issue. An effective coding gain of about 5 dB can be obtained with the (2, 1, 3) code, whose code rate is \( k = 0 \), with a trellis branch complexity of only \( s = 2 \). The (2, 1, 5) and (2, 1, 7) codes, which are close cousins, can also obtain an effective coding gain of about 5 dB, with a somewhat greater branch complexity of \( s = 4 \), but with a more comfortable code rate of \( k = 0 \). Beyond these codes, complexity increases rapidly.

A similar analysis can be performed for binary linear convolutional codes. For a rate-\( k/n \) binary linear convolutional code \( C \) with free distance \( d \), the nominal coding gain is again \( \gamma_c(C) = (k/n)d \). If \( K_d \) is the number of weight-\( d \) code sequences that start in a given \( k \)-input, \( n \)-output block, then the appropriate error coefficient to measure the probability of error event per information bit is again \( K_d/k \).

Table II gives the parameters of the best known rate-\( 1/n \) time-invariant binary linear convolutional codes for \( n = 2, 3, 4 \) with constraint lengths \( \nu \leq 7 \) [18]. The log branch complexity parameter \( s \) equals \( \nu + 1 \). (For some combinations of parameters \( (n, k, \nu) \), superior time-varying codes are known [70].)

It is apparent from a comparison of Tables I and II that convolutional codes offer a much better performance/complexity tradeoff than block codes. It is possible to obtain more than 6 dB of effective coding gain with a 32-state or 64-state rate-1/2 or rate-1/3 convolutional code with a simple, regular trellis (VA) decoder. No block code approaches such performance with such modest complexity.

Consequently, for power-limited channels such as the deep-space channel, convolutional codes rather than block codes have been used almost from the earliest days [26]. Although this is due in part to the historical fact that efficient maximum-likelihood (ML) or near-ML soft-decision decoding algorithms were invented much earlier for convolutional codes, Tables I and II show that convolutional codes have an inherent performance/complexity advantage.

### E. Sequential Decoding

Sequential decoding of convolutional codes was perhaps the earliest near-ML decoding algorithm, and is still one of the few that can operate at the cutoff rate \( R_0 \). For this reason,
sequential decoding was the first coding scheme used for the deep-space application (Pioneer, 1968) [26].

Although there exist many variants of sequential decoding, they have in common a sequential search through a code tree (not a trellis) for the ML code sequence. The code constraint length $\nu$ may be infinite in principle, although for synchronization purposes (whether via zero-tail termination, tail-biting, or self-resynchronization) it is usually chosen to be not too large (of the order of 15–30).

The amount of computation to decode a given number of information bits is (highly) variable in sequential decoding, but is more or less independent of $\nu$. The distribution of computation $N$ follows a Pareto distribution, $\Pr(N \geq x) \propto x^{-\alpha}$, where $\alpha$ is the Pareto exponent [63]. The Pareto exponent is equal to 1 at the cutoff rate $R_0$, which is usually considered to be the practical limit of sequential decoding (because a Pareto distribution has a finite mean only if $\alpha > 1$). Thus on the low-SNR ideal AWGN channel, sequential decoding of low-rate binary linear convolutional codes can approach the Shannon limit within a factor of about 3 dB.

If $\nu$ is large enough, then sequential decoders rarely make errors; rather, they fail to decode due to excessive computation. This error-detection property is useful in many applications, such as deep-space telemetry.

Moreover, experiments have shown that bidirectional sequential decoding of moderate-length blocks can approximately double the Pareto exponent and thus significantly reduce the gap to capacity, even after giving effect to the rate loss due to zero-tail termination [65].

In view of these desirable properties, it is something of a mystery why sequential decoding has received so little practical attention during the past 30 years.

**F. Concatenated Codes and RS Codes**

Like the Viterbi algorithm and the Lempel–Ziv algorithm, concatenated codes were originally introduced to solve a theoretical problem [39], but have turned out to be useful for a variety of practical applications.

The basic idea is to use a moderate-strength “inner code” with an ML or near-ML decoding algorithm to achieve a moderate error rate like $10^{-2}$–$10^{-3}$ at a code rate as close to capacity as possible. Then a powerful algebraic “outer code” capable of correcting many errors with low redundancy is used to drive the error rate down to as low an error rate as may be desired. It was shown in [39] that the error rate could be made to decrease exponentially with blocklength at any rate less than capacity, while decoding complexity increases only polynomially.

Reed–Solomon (RS) codes are ideal for use as outer codes. (They are not suitable for use as inner codes because they are nonbinary.) RS codes are defined over finite fields $\mathbb{F}_q$ whose order $q$ is typically large (e.g., $q = 2^m$). An $(n, k, n-k+1)$ RS code over $\mathbb{F}_q$ exists for every $n \leq q + 1$ and $1 \leq k \leq n$ [77]. The minimum distance $n-k+1$ is as large as possible, in view of the Singleton bound [77].

Very efficient polynomial-time algebraic decoding algorithms are known, not just for error correction but also for erasure-and-error correction and even for near-ML soft-decision decoding (i.e., generalized minimum-distance decoding [39]). RS error-correction algorithms are standard in VLSI libraries, and today are typically capable of decoding dozens of errors per block over $\mathbb{F}_{256}$ at data rates of tens of megabits per second (Mb/s) [98].

When used with interleaving, RS codes are also ideal burst-error correctors in the sense of requiring the least possible guard space [40].

For these reasons RS codes are used in a large variety of applications, including the NASA deep-space standard adopted in the 1970’s [24]. RS codes are clearly the outstanding practical success story of the field of algebraic block coding.

**G. Turbo Codes and Other Capacity-Approaching Codes**

The invention of “turbo codes” [13] put a decisive end to the long-standing conjecture that the cutoff rate $R_0$ might represent the “practical capacity.” Performance within tenths of a decibel of the Shannon limit is now routinely demonstrated with reasonable decoding complexity, albeit with large delay [61].

The original turbo code operates as follows. An information bit sequence is encoded in a simple (e.g., 16-state) recursive systematic rate-1/2 convolutional encoder to produce one check bit sequence. The same information bit sequence is permuted in a very long (e.g., $10^4$–$10^5$ bits) interleaver and then encoded in a second recursive systematic rate-1/2 convolutional encoder to produce a second check bit sequence. The information bit sequence and both check bit sequences are transmitted, so the code rate is 1/3. (Puncturing can be used to raise the rate.)

Decoding is performed in an iterative fashion as follows. The received sequences corresponding to the information bit sequence and the first check bit sequence are decoded by a soft-decision decoder for the first convolutional code. The outputs of this decoder are a sequence of soft decisions for each bit of the information sequence. (This is done by some version of the forward–backward algorithm [5], [6].) These soft decisions are then used by a similar decoder for the second convolutional code, which hopefully produces still better soft decisions that can then be used by the first decoder for a new iteration. Decoding iterates in this way for 10 to 20 cycles, before hard decisions are finally made on the information bits.

Empirically, operation within 0.3–0.7 dB of the Shannon limit can be achieved at moderate error rates. Theoretical understanding of turbo codes is still weak. At low $E_b/N_0$, turbo codes appear to behave like random codes whose blocklength is comparable to the interleaver length [84]. At higher $E_b/N_0$, the performance of a turbo decoder is dominated by certain low-weight codewords, whose multiplicity is inversely proportional to the interleaver length (the so-called “error floor” effect) [11].

The turbo decoding algorithm is now understood to be the general sum-product (APP) decoding algorithm, with a particular update schedule that is well matched to the turbo-code structure [116]. However, little is known about the theoretical performance and convergence properties of this
algorithm applied to turbo codes, or more generally to codes defined on graphs with cycles.

Many variants of the turbo coding idea have been studied. Different arrangements of compound codes have been investigated, including “serially concatenated” codes like those of [39], which mitigate the error floor effect [9]. The decoding algorithm has been refined in various ways. Overall, however, improvements have been minor.

A turbo coding scheme is now being standardized for future deep-space missions [29].

Even more recently, codes and decoding algorithms that approach turbo coding performance have been devised according to quite different principles. With these methods turbo-code performance has been approached and even exceeded. Notable among these are low-density parity-check (LDPC) codes, originally proposed by Gallager in the early 1960’s [50], subsequently almost forgotten, and recently rediscovered by various authors [76], [96], [116]. Like turbo codes, these codes may be viewed as “codes defined on graphs” and may be decoded by similar iterative APP decoding algorithms [116]. Long LDPC codes can operate well beyond $R_0$; furthermore, they have no error floor and rarely make decoding errors, but rather simply fail to decode [76]. Theoretically, it has been shown that long LDPC codes can achieve rates up to capacity with ML decoding [50], [76]. Very recently, nonbinary LDPC codes have been devised that outperform turbo codes [27]. However, the parallel (“flooding”) version of the iterative decoding algorithm that is usually used with these codes is computationally expensive, and their lengths are comparable to those of turbo codes.

IV. NONBINARY CODES FOR BANDWIDTH-LIMITED CHANNELS

In this section we discuss coding techniques for high-SNR ideal bandwidth-limited AWGN channels. In this regime nonbinary signal alphabets such as $M$-PAM must be used to approach capacity. Using large-alphabet approximations, we show that the total coding gain of a coded-modulation scheme for the high-SNR ideal AWGN channel is the sum of a coding gain and a shaping gain. At high SNR’s, the coding and shaping problems are separable.

Shaping gains are obtained by using signal constellations in high-dimensional spaces that are bounded by a quasispherical region, rather than the cubical region that results from independent $M$-PAM signaling; or, alternatively, by using points in a low-dimensional constellation with a Gaussian-like probability distribution rather than equiprobably. The maximum possible shaping gain is a factor of $\pi e/6$ (1.53 dB). We briefly mention several shaping methods that can easily obtain about 1 dB of shaping gain.

In the high-SNR regime, the SNR gap between uncoded baseline performance at $P_b(x) \simeq 10^{-6}$ and the cutoff limit without shaping is about 5.8 dB. This gap arises as follows: the gap to the Shannon limit is 9 dB; with shaping the cutoff limit is 1.7 dB below the Shannon limit; and without shaping the cutoff limit is 1.5 dB lower than with shaping. Coding gains of the order of 3–5 dB at error probabilities of $10^{-5}$–$10^{-6}$ can be obtained with moderate complexity.

The two principal classes of high-SNR codes are lattices and trellis codes, which are analogous to binary block and convolutional codes, respectively. By now the principles of construction of such codes are well understood, and it seems likely that the best codes have been found. We plot the effective coding gains of these known moderate-complexity lattices and trellis codes versus the branch complexity of their minimal trellises, assuming maximum-likelihood decoding. Trellis codes are somewhat superior, due mainly to their lower error coefficients.

We briefly discuss coding schemes that can achieve coding gains beyond $R_0$, of the order of 5 to 7 dB, at error probabilities of $10^{-6}$–$10^{-5}$. Multilevel schemes with multistage decoding allow the use of high-performance binary codes such as those described in the previous section to be used to approach the Shannon limit of high-SNR channels. Such high-performance schemes naturally involve greater decoding complexity and large delays.

We particularly note techniques that have been used in telephone-line modem standards, which have generally reflected the state of the art for high-SNR channels.

A. Lattice Constellations

It is clear from the proof of Shannon’s capacity theorem that an optimal block code for a high-SNR ideal band-limited AWGN channel consists of a dense packing of signal points within a sphere in a high-dimensional Euclidean space. Most of the known densest packings are lattices [25]. In this section we briefly describe lattice constellations, and analyze their performance using the union bound estimate and large-constellation approximations.

An $N$-dimensional lattice $\Lambda$ is a discrete subgroup of $N$-space $\mathbb{R}^N$, which without essential loss of generality may be assumed to span $\mathbb{R}^N$. The points of the lattice form a uniform infinite packing of $\mathbb{R}^N$. By the group property, each point of the lattice has the same number of neighbors at each distance, and all decision regions of a minimum-distance decoder (Voronoi regions) are congruent and tessellate $\mathbb{R}^N$. These properties hold for any lattice translate $\Lambda + t$.

The key geometrical parameters of a lattice are the minimum squared distance $d_{\min}^2(\Lambda)$ between lattice points, the kissing number $K_{\min}(\Lambda)$ of nearest neighbors to any lattice point, and the volume $V(\Lambda)$ of $N$-space per lattice point, which is equal to the volume of any Voronoi region [25]. The Hermite parameter is the normalized parameter $\gamma(\Lambda) = d_{\min}^2(\Lambda)/V(\Lambda)^{2/N}$, which we will soon identify as the nominal coding gain of $\Lambda$.

A lattice constellation $C(\Lambda, R) = (\Lambda + t) \cap R$ is the finite set of points in a lattice translate $\Lambda + t$ that lie within a compact bounding region $R$ of $N$-space. The key geometric properties of the region $R$ are its volume $V(R)$ and the average energy $P(R)$ per dimension of a uniform probability density function over $R$:

$$P(R) = \int_R \langle|\mathbf{x}|^2/N\rangle \, d\mathbf{x}/V(R). \tag{4.1}$$

The normalized second moment of $R$ is defined as $G(R) = P(R)/V(R)^{2/N}$ [25].
For example, an $M$-PAM constellation with $M$ even
\[(d_0/2)\{\pm 1, \pm 3, \cdots, \pm (M - 1)\}\]
is a one-dimensional lattice constellation $C(Z, R)$ with $\Lambda + \ell =
d_0(Z + 1/2)$ and $R = (d_0/2)[-M, M]$. The key geometrical parameters of $\Lambda = d_0 Z$ are
\[
\begin{align*}
d_{\text{min}}^2(d_0 Z) &= d_0^2 \\
K_{\text{min}}(d_0 Z) &= 2 \\
V(d_0 Z) &= d_0 \\
\gamma_\text{c}(d_0 Z) &= 1.
\end{align*}
\] (4.2)
The key parameters of $R = (d_0/2)[-M, M]$ are
\[
\begin{align*}
V(R) &= d_0 M \\
P(R) &= (d_0 M)^2/12 \\
G(R) &= 1/12.
\end{align*}
\] (4.3)

Both $\gamma_\text{c}(\Lambda)$ and $G(R)$ are invariant under scaling, orthogonal transformations, and Cartesian products; i.e., $\gamma_\text{c}(\alpha U^N) =
\gamma_\text{c}(\Lambda)$ and $G(\alpha U^N) = G(R)$, where $\alpha > 0$ is any scale factor, $U$ is any orthogonal matrix, and $N \geq 1$ is any positive integer. In particular, this implies that $\gamma_\text{c}(\alpha U^N) = 1$ for any version of an integer lattice $Z^N$, and that the normalized second moment of any $N$-cube centered at the origin is $1/12$.

For large lattice constellations, one may use the following approximations, the first two of which are collectively known as the continuous approximation [48]:
- the size of the constellation is $|C(\Lambda, R)| \approx V(R)/V(\Lambda)$;
- the average energy per dimension of an equiprobable distribution over $C(\Lambda, R)$ is $P(C(\Lambda, R)) \approx P(R)$;
- for large rates $R = (1/N) \log_2|C(\Lambda, R)|$, we have $2^{2R} - 1 \approx 2^{2R}$;
- the average number of nearest neighbors to any point in $C(\Lambda, R)$ is approximately $K_{\text{min}}(\Lambda)$.

The union bound estimate on probability of block decoding error is
\[
P(\mathcal{E}) \approx K_{\text{min}}(\Lambda)Q(\gamma_{\text{c}}(\Lambda)/2\sigma_v),
\] (4.4)
Because
\[
R = (1/N) \log_2|C(\Lambda, R)| \approx (1/N) \log_2V(R)/V(\Lambda),
\] (4.5)
\[
\text{SNR} = P(C(\Lambda, R))/\sigma_v^2 \approx P(R)/\sigma_v^2,
\] (4.6)
\[
\text{SNR}_{\text{norm}} \approx \text{SNR}/2^{2R} = (V(\Lambda)^{2/N}/V(R)^{2/N})(P(R)/\sigma_v^2),
\] (4.7)
we may write the union bound estimate as
\[
P(\mathcal{E}) \approx K_{\text{min}}(\Lambda)Q\left(\sqrt{\gamma_\text{c}(\Lambda)\gamma_\text{s}(R)/3}\text{SNR}_{\text{norm}}\right),
\] (4.8)
where the nominal coding gain of $\Lambda$ and the shaping gain of $R$ are defined, respectively, as
\[
\gamma_\text{c}(\Lambda) = d_{\text{min}}^2(\Lambda)/V(\Lambda)^{2/N},
\] (4.9)
\[
\gamma_\text{s}(R) = V(R)^{2/N}/(12P(R)) = (1/12)/G(R),
\] (4.10)
For a baseline $M$-PAM constellation, we have $\gamma_\text{c}(\Lambda) =
\gamma_\text{s}(R) = 1$, and the UBE reduces to
\[
P(\mathcal{E}) \approx 2Q(\sqrt{3}\text{SNR}_{\text{norm}}).
\] (4.11)

The nominal coding gain $\gamma_\text{c}(\Lambda)$ measures the increase in density of $\Lambda$ over a baseline lattice, $Z$ or $Z^N$. The shaping gain $\gamma_\text{s}(R)$ measures the decrease in average energy of $R$ relative to a baseline region, namely, an interval $[-d_0/2, d_0/2]$ or an $N$-cube $[-d_0/2, d_0/2]^N$. Both contribute a multiplicative factor of gain to the argument of the $Q(\sqrt{3})$ function.

As before, the effective coding gain is reduced by the error coefficient $K_{\text{min}}(\Lambda)$. For comparability with the $M$-PAM uncoded baseline, it is appropriate to normalize $P(\mathcal{E})$ by the dimension $N$ to obtain the probability of block decoding error per symbol
\[
P_s(\mathcal{E}) = P(\mathcal{E})/N \approx K(\Lambda)Q\left(\sqrt{\gamma_\text{c}(\Lambda)\gamma_\text{s}(R)/3}\text{SNR}_{\text{norm}}\right),
\] (4.11)
in which the normalized error coefficient is $K(\Lambda) =
K_{\text{min}}(\Lambda)/N$.

Graphically, a curve of the form of (4.11) may be obtained simply by moving the baseline curve $P_s(\mathcal{E}) = 2Q(\sqrt{3}\text{SNR}_{\text{norm}})$ of Fig. 2 to the left by $\gamma_\text{c}(\Lambda)$ and $\gamma_\text{s}(R)$ (in decibels), and upward by a factor of $K(\Lambda)/2$.

B. Shaping Gain and Shaping Techniques

Although shaping is a newer and less important topic than coding, we discuss it first because its story is quite simple.

The optimum $N$-dimensional shaping region is an $N$-sphere. The key geometrical parameters of an $N$-sphere $(= \odot)$ of radius $r$ for $N$ even are [25]
\[
V(\odot) = \left(\frac{\pi^\frac{N}{2}}{N/2}!\right),
\]
\[
P(\odot) = \frac{r^2}{N + 2},
\]
\[
G(\odot) = P(\odot)V(\odot)^{2/N} = \frac{((N/2)^{2/N}}{\pi(N + 2)}.
\] (4.12)
By Stirling’s approximation, namely $m! \approx (m/e)^m$ as $m$ goes to infinity [51], we have
\[
G(\odot) \xrightarrow{N \to \infty} 1/2\pi e,
\] (4.13)

The shaping gain of an $N$-sphere is plotted for dimensions $N = 1, 2, 4, 8, \cdots, 512$ in Fig. 3. Note that the shaping gain of a 16-sphere is approximately 1 dB. However, for larger dimensions, we see from Fig. 3 that the shaping gain approaches the ultimate shaping gain $\pi e/6$ (1.53 dB) rather slowly.

The projection of a uniform probability distribution over an $N$-sphere onto one or two dimensions is a nonuniform probability distribution that approaches a Gaussian density as $N \to \infty$. The ultimate shaping gain of $\pi e/6$ (1.53 dB) may be derived alternatively as the difference between the average power of a uniform density over an interval and that of a Gaussian density with the same differential entropy [48].

Shaping therefore induces a Gaussian-like probability distribution on a one-dimensional PAM (or two-dimensional QAM) constellation, rather than an equiprobable distribution.
In principle, with spherical shaping, the lower dimensional constellation will become arbitrarily large, even with fixed average power. In practice, the lower dimensional “shaping constellation expansion” is constrained by design to a permitted peak amplitude. If the $N$-dimensional shape approximates spherical shaping subject to this constraint, then the lower dimensional probability distribution approaches a truncated Gaussian distribution [48].

A nonuniform distribution over a low-dimensional constellation may alternatively be obtained by codes over bits that specify regions of the constellation [16].

With large constellations, shaping can be implemented more or less independently of coding by operations on the “most significant bits” of PAM or QAM constellation labels, which affect the large-scale shape of the $N$-dimensional constellation. In contrast, coding affects the “least significant bits” and determines small-scale structure.

Two practical schemes that can easily obtain shaping gains of 1 dB or more while limiting two-dimensional (2-D) shaping constellation expansion to a factor of $1.5$ or less are “trellis shaping” and “shell mapping.”

Trellis shaping [45] is a kind of dual to trellis coding. Using a trellis, one defines a set of equivalence classes of coded sequences. The shaping operation consists of determining among equivalent sequences in the trellis the minimum-energy sequence. Shaping gains of the order of 1 dB are easily obtained.

In shell mapping [49], [66], [68], [69], [72] the signals in an $N$-dimensional lattice constellation are in principle labeled in order of increasing signal energy, and a one-to-one map is defined between possible input data sequences and an equal number of least energy constellation points. In practice, generating-function methods are used to label points in Cartesian-product constellations in approximate increasing order of energy based on shell-mapping labelings of lower dimensional constituent constellations. The V.34 modem uses 16-dimensional shell mapping, and obtains shaping gains of the order of 0.8 dB with 2-D shaping constellation expansion limited to 25% [47]. These shaping methods are also useful for accommodating noninteger numbers of bits per symbol. The latter objective is also achieved by a mapping technique known as “modulus conversion” [3], which is an enumerative encoding technique for mapping large integers into blocks (“mapping frames”) of signals in lexicographical order, where the sizes of the signal constellations do not need to be powers of $2$.

### C. Coding Gains of Dense Lattices

Finding the densest lattice packings in a given number of dimensions is a mathematical problem of long standing. A summary of the densest known packings is given in [25]. The nominal coding gains of these lattices in up to 24 dimensions is plotted in Fig. 4. Notable lattices include the eight-dimensional Gosset lattice $E_8$, whose nominal coding gain is 2 (3 dB), and the 24-dimensional Leech lattice $L_{24}$, whose nominal coding gain is 4 (6 dB).

In contrast to shaping gain, the nominal coding gains of dense $N$-dimensional lattices become infinite as $N \to \infty$. For example, for any integer $m \geq 0$, there exists a $2^{m+1}$-dimensional Barnes–Wall lattice whose nominal coding gain is $2^{m/2}$ (which is by no means the densest possible for large $m$) [25].

However, effective coding gains cannot become infinite. Indeed, the Shannon limit shows that no lattice can have a combined effective coding gain and shaping gain greater than 9 dB at $P(E) \approx 10^{-6}$. This limits the maximum possible effective coding gain to 7.5 dB, because shaping gain can contribute up to 1.53 dB.

What limits effective coding gain is the number of near neighbors, which becomes very large for high-dimensional dense lattices. For example, the kissing number of the $2^{m+1}$-dimensional Barnes–Wall (BW) lattice is [25]

$$K_{\min}(BW_{2^{m+1}}) = \prod_{1 \leq j \leq m} (2^j + 2)$$

which yields, for example, $K_{\min}(BW_{32}) = 146880$ and $K_{\min}(BW_{128}) = 1200230400$. 

---

**Fig. 3.** Shaping gains of $N$-spheres over $N$-cubes for $N = 1, 2, 4, \ldots, 512$. 

---
D. Trellis Codes

Trellis codes are dense packings in infinite-dimensional Euclidean sequence space. Trellis codes are to lattices as binary convolutional codes are to block codes. We will see that trellis codes give a better performance/complexity tradeoff than lattices in the high-SNR regime, just as convolutional codes give a better performance/complexity tradeoff than block codes in the low-SNR regime, although the difference is not as dramatic.

The key ideas in the invention of trellis codes [104] were:

- use of minimum squared Euclidean distance as the design criterion;
- set partitioning of a constellation $A$ into $2^n$ subsets;
- maximizing intrasubset minimum squared distances;
- rate-$k/n$ convolutional coding over the subsets;
- selection of signals within subsets by further uncoded bits (if necessary).

Most practical trellis codes have used either $N$-dimensional lattice constellations based on versions of $\mathbb{Z}^N$ with $N = 1, 2, 4, 8$ or $2K$-dimensional $K \times M$-PSK constellations with $K = 1, 2, 3, 4$ and $M = 8$. We will focus on lattice-type trellis codes in this section.

Set partitioning of an $N$-dimensional constellation $A$ into $2^n$ subsets is usually performed by $n$ partitioning steps, with each two-way selection being identified by a label bit $c^j \in \{0, 1\}$, $1 \leq j \leq n$ [104], [105]

$$A \leadsto B(c_1) \leadsto C(c_2, c_1) \leadsto \cdots \leadsto S(c^n).$$

(4.15)

The quantities $\Delta_{0}^{2}, j = 0, \ldots, n$, are the intrasubset minimum squared distances of the $j$th-level (sub)sets $A, B(c^2), C(c^2, c^1), \ldots$. The objective of set partitioning is to increase these distances as much as possible at every subset level. Set partitioning is continued until $\Delta_{0}^{2}$ is at least as large as the desired minimum squared distance $d^{2}_{\text{min}}(C)$ between trellis code sequences. The binary $n$-tuples $c = [c^n \cdots c^1]$ are called the subset labels of the final subsets $S(c)$.

Partitioning one- or two-dimensional constellations in this way is trivial [104]. Partitioning of higher dimensional constellations was first performed starting from lower dimensional partitions [113]. Subsequently, partitioning of lattice constellations was introduced using lattice partitions [17], [42], [43] (or, more generally, group-theoretic partitions [44]), as we will discuss below.

In almost all cases of practical interest, the two first-level subsets $B(0)$ and $B(1)$ will be geometrically congruent to each other (see Appendix II-B); i.e., they will differ only by translation, rotation, and/or reflection.

An encoder for a trellis code $C$ then operates as shown in Fig. 5 [105]. Given $m$ input bits per $N$-dimensional symbol, $k$ input bits are encoded by a rate-$k/n$ binary convolutional encoder with $2^k$ states into a coded sequence of subset labels $c_i$. At time $i$, the label $c_i$ is used to select subset $S(c_i)$. The remaining $m - k$ input bits are used to choose one signal $a_i$ from $2^{m-k}$ signals in the selected subset $S(c_i)$. The size of $A$ must therefore be $2^{m/(n-k)}$, or a factor of $2^{n-k}$ (the “coding constellation expansion” factor per $N$ dimensions) larger than needed to send $m$ uncoded bits per symbol. (If there is any shaping, it is performed on the uncoded bits and results in further “shaping constellation expansion.”)
A trellis code may be maximum-likelihood decoded by a Viterbi algorithm (VA) decoder as follows. Given a received point \( z_i \in \mathbb{R}^N \), the receiver first finds the closest signal point \( \hat{s}(c_i) \) in each subset \( S(c_i) \). This is called subset decoding. A VA decoder then finds the code sequence \( \{c_i\} \) for which the signals chosen in the subsets are closest to the entire received sequence \( \{z_i\} \). The decoding complexity is dominated by the complexity of the VA decoder, which is approximately given by the branch complexity of the convolutional code, normalized by the dimension \( N \).

In practice, the redundancy of the convolutional code is almost always one bit per \( N \)-dimensional symbol \((k = n - 1)\), so that the coding constellation expansion factor is 2 per \( N \) dimensions. The lowest level (“least significant”) label bit \( c^0 \) is then the sole parity-check bit. If the convolutional code is linear, then in \( D \)-transform notation the check sequence \( c^0(D) \) is determined from the information bit sequences \( c^1(D), \ldots, c^m(D) \) by a linear parity-check equation of the form

\[
h^n(D)c^0(D) + \cdots + h^2(D)c^2(D) + h^1(D)c^1(D) = 0 \tag{4.16}
\]

where

\[
\{h^j(D) = h^j_1 D^j + \cdots + h^j_n D^n \mid 1 \leq j \leq n\}
\]

is a set of \( n \) relatively-prime parity-check polynomials. If the greatest degree of any of these polynomials is \( \tau \), then a minimal encoder for the code has \( 2^\tau \) states.

The convolutional code is chosen primarily to maximize the minimum squared distance \( d^2_{\text{min}}(C) \) between trellis-code sequences. If the redundancy is one bit per symbol, then a good first step is to ensure that in any encoder state \( s_i \) the set of possible next outputs is either \( B(0) \) or \( B(1) \), so that a squared distance contribution of \( \Delta^2 \) is immediately obtained when code sequences diverge from a common state.

A linear convolutional code, this is achieved by choosing the low-order coefficients \( h^0_j \) of the parity-check polynomials \( h^j(D) \) such that \( h^0_1 = 1 \) and \( h^0_j = 0 \), \( 2 \leq j \leq n \). The same argument applies in the reverse time direction to the high-order coefficients \( h^-_j \). This design rule guarantees that \( d^2_{\text{min}}(C) \geq 2\Delta^2 \). [104]

An important consequence of such a choice is that at time \( i \) the lowest level label bit \( c^1_i \) is determined from the encoder state \( s_i \). This observation leads to the idea of feedback trellis encoding [71], which was proposed during the development of the V.34 modem standard to permit an improved type of precoding (see Section V). With this technique, the sequence of encoding operations at time \( i \) is as follows:

a) the check bit \( c^1_i \) is determined from the encoder state \( s_i \);
b) the \( m \) input bits at time \( i \) select a signal \( s_i \) from \( B(c^1_i) \);
c) the remaining label bits \( c^2_i, \ldots, c^m_i \) are determined from \( a_i \);
d) the next encoder state \( s_{i+1} \) is determined from \( s_i \) and \( c_i \).

E. Lattice Constellation Partitioning

Partitioning of lattice constellations by means of lattice partitions is done as follows [17], [43]. One starts with an \( N \)-dimensional lattice constellation \( C(\Lambda, \mathbb{R}) \), where \( \Lambda \) is almost always a version of an \( N \)-dimensional integer lattice \( \mathbb{Z}^N \). The constellation \( C(\Lambda, \mathbb{R}) \) is partitioned into \( 2^n \) subsets of equal size that are congruent to a sublattice \( \Lambda' \) of index \( |\Lambda'/\Lambda| = 2^n \) in \( \Lambda \), so that \( \Lambda' \) is the union of \( 2^n \) cosets (translates) of \( \Lambda' \). The \( 2^n \) subsets are then the points of \( C(\Lambda, \mathbb{R}) \) that lie in each such subset, which form sublattice constellations of the form \( C(\Lambda', \mathbb{Z}) \). The region \( \mathbb{R} \) must be chosen so that there are an equal number of points in each subset. The sublattice \( \Lambda' \) is usually chosen to be as dense as possible.

Codes based on lattice partitions \( \Lambda'/\Lambda \) are called coset codes. The nominal coding gain of a coset code based on a lattice \( \Lambda = \mathbb{Z}^N \) is [42]

\[
\gamma_{\text{C}}(C) = \frac{d^2_{\text{min}}(C)}{2^{2\rho(C)}} \tag{4.17}
\]

where \( \rho(C) = 1/N \) is the redundancy of the convolutional encoder in bits per dimension. Again, this coding gain is the product of a distance gain of \( d^2_{\text{min}}(C) \) over \( d^2_{\text{min}}(\mathbb{Z}^N) = 1 \), and a redundancy loss \( 2^{-2\rho(C)} \). The effective coding gain is reduced by the amount that the error coefficient \( K_{\text{min}}(C)/N \) per dimension exceeds the baseline \( M \)-PAM error coefficient of 2 per dimension. Again, the rule of thumb that an increase of a factor of two costs 0.2 dB may be used.

The encoder redundancy \( \rho(C) \) also leads to a “coding constellation expansion ratio” of a factor of \( 2^{2\rho(C)} \) per two dimensions [42]—i.e., a factor of \( 4, 2, \sqrt{2} \cdots \) for 1D, 2D, 4D, \cdots, codes, respectively. Minimization of coding constellation expansion has motivated the use of higher dimensional trellis codes.

F. Coding Gains of Known Trellis Codes

Fig. 6 shows the effective coding gains of important families of trellis codes for lattice-type signal constellations. The effective coding gains are plotted versus their VA decoding
complexity, measured by a detailed operation count. The codes considered are as follows:

1) The original 1D (PAM) trellis codes with \(\rho(C) = 1\) of Ungerboeck [104], based on rate-1/2 convolutional codes with \(2 \leq \nu \leq 9\) and the four-way partition \(\mathbb{Z}/4\mathbb{Z}\).

2) The 2D (QAM) trellis codes with \(\rho(C) = 1/2\) of Ungerboeck [104], based on (except for the simplest code with \(\nu = 2\)) rate-2/3 convolutional codes with \(3 \leq \nu \leq 9\) and the eight-way partition \(\mathbb{Z}^2/2R\mathbb{Z}^2\) (where \(R\) is a \(2 \times 2\) Hadamard matrix).

3) The 4D trellis codes with \(\rho(C) = 1/4\) of Wei [113], based on
   a) rate-2/3 8- and 16-state convolutional codes and the eight-way partition \(\mathbb{Z}^4/RD_4\);
   b) a rate-3/4 32-state convolutional code and the 16-way partition \(\mathbb{Z}^4/2\mathbb{Z}^4\);
   c) a rate-4/5 64-state convolutional code and the 32-way partition \(\mathbb{Z}^4/2D_4\).

4) Two families of 8D trellis codes of Wei [113].

The V.32 modem (1984) uses an eight-state 2D trellis code, also due to Wei [112]. The performance/complexity tradeoff is the same as that of the original eight-state 2D Ungerboeck code. However, the Wei code uses a nonlinear convolutional encoder to achieve 90° rotational invariance. This code has an effective coding gain of about 3.6 dB, a branch complexity of \(2^5\) per two dimensions, and a coding constellation expansion ratio of 2.

The V.34 modem (1994) specifies three 4D trellis codes, with performance and complexity equivalent to the 4D Wei codes circled on Fig. 6 [47]. All have a coding constellation expansion ratio of \(\sqrt{2}\). The 16-state code, which is the only code implemented by most manufacturers, is the original 16-state 4D Wei code, which has an effective coding gain of about 4.2 dB and a branch complexity of \(2^6\) per four dimensions. The 32-state code is due to Williams [117] and is based on a 16-way partition of \(\mathbb{Z}^4\) into 16 subsets congruent to \(H\mathbb{Z}^4\), where \(H\) is a \(4 \times 4\) Hadamard matrix, to ensure that there are no minimum-distance error events whose length is only two dimensions; it has an effective coding gain of about 4.5 dB and a branch complexity of \(2^8\) per four dimensions. The 64-state code is a modification of the original 4D Wei code, modified to prevent quasicatastrophic error propagation; it has an effective coding gain of about 4.7 dB and a branch complexity of \(2^{10}\) per four dimensions.

It is noteworthy that no one has improved on the performance/complexity tradeoff of the original 1D and 2D trellis codes of Ungerboeck or the subsequent multidimensional codes of Wei. By this time it seems safe to predict that no one will ever do so. There have, however, been new trellis codes that feature other properties and have about the same performance and complexity, as described in the previous two paragraphs, and there may still be room for further improvements of this kind.

Finally, we see that trellis codes have a performance/complexity advantage over lattice codes, when used with maximum-likelihood decoding. Effective coding gains of 4.2–4.7 dB, better than that of the Leech lattice or of BW\(\mathbb{Z}_2\), are attainable with less complexity and much less constellation expansion. With the 512-state 1D or 2D trellis codes, effective coding gains of the order of 5.5 dB can be achieved. These gains are larger than the gains that can be obtained with lattice codes of far greater complexity.

On the other hand, it seems very difficult to obtain effective coding gains approaching 6 dB. This is not surprising, because at \(P(\mathcal{E}) \approx 10^{-6}\) the effective coding gain at the Shannon limit...
is about 7.5 dB, and at the cutoff rate limit it is about 5.8 dB. To approach the Shannon limit, codes and decoding methods of higher complexity are necessary.

So far in this section we have only discussed trellis codes for lattice-type signals. However, the principle of set partitioning and the general encoder structure of Fig. 5 are also valid for trellis codes with signals from constellations which are not of lattice type, such as 8-PSK and 16-PSK constellations. Code tables for 8-PSK and 16-PSK trellis codes are given in [104] and [105]. A number of multidimensional linear PSK-type trellis codes are described in [114], all fully rotationally invariant. In [87], principles of set partitioning of multidimensional $K \times M$-PSK constellations are given, based on concepts of multilevel block coding; tables of linear $K \times M$-PSK trellis codes are presented for $M = 4, 8, 16$ and $K = 1, 2, 3, 4$, with the largest degree of rotational invariance given for each code; and figures similar to Fig. 6 that summarize effective coding gains versus decoding complexity are shown.

### G. Sequential Decoding in the High-SNR Regime

In the high-SNR regime, the cutoff rate is a factor of $4/e$ (1.68 dB) away from capacity [95]. Therefore, sequential decoders should be able to achieve an effective coding gain of about 5.8 dB at $P(\mathcal{E}) \approx 10^{-6}$. Experiments have confirmed that sequential decoders can indeed achieve such performance [111].

### H. Multilevel Codes and Multistage Decoding

To approach the Shannon limit even more closely, it is clear that much more powerful codes must be used, together with decoding methods that are simpler than optimal maximum-likelihood (ML) decoding, but with near-ML performance. Multilevel codes and multistage decoding may be used for this purpose [60], [75]. Multilevel coding may be based on a chain of sublattices of $\mathbb{Z}^N$

$$\Lambda_0 \supset \Lambda_{n-1} \supset \cdots \supset \Lambda_1 \supset \Lambda_0 = \mathbb{Z}^N$$  \hspace{1cm} (4.18)

which lead to a chain of lattice partitions $\Lambda_{j-1}/\Lambda_j$, $1 \leq j \leq n$. A different trellis encoder as in Fig. 5 may be used independently on each such lattice partition.

Remarkably, it follows from the chain rule of mutual information that the capacity $C(\Lambda_0/\Lambda_n)$ that can be obtained by coding over the lattice partition $\Lambda_0/\Lambda_n$ is equal to the sum of the capacities $C(\Lambda_{j-1}/\Lambda_j)$ that can be obtained by independent coding and decoding at each level [46], [58], [67], [110]. With multistage decoding, decoding is performed separately at each level; at the $j$th level the decisions at lower levels $(0, 1, \cdots, j-1)$ are taken into account, whereas no coding is assumed for the higher levels $(j+1, \cdots, n)$. If the partition $\Lambda_0/\Lambda_n$ is “large enough” and appropriately scaled, then $C(\Lambda_0/\Lambda_n)$ approaches the capacity of the ideal AWGN channel.

The lattices may be the one- or two-dimensional integer lattices $\mathbb{Z}$ or $\mathbb{Z}^2$, and the standard binary partition chains

$$\cdots \subset 8\mathbb{Z} \subset 4\mathbb{Z} \subset 2\mathbb{Z} \subset \mathbb{Z};$$

$$\cdots \subset 4\mathbb{Z}^2 \subset 2R\mathbb{Z}^2 \subset 2\mathbb{Z}^2 \subset R\mathbb{Z}^2 \subset \mathbb{Z}^2$$  \hspace{1cm} (4.19)

may be used. Then a powerful binary code with rate close to $C(\Lambda_j/\Lambda_{j+1})$ can be used at each level to approach the Shannon limit. In particular, by using turbo codes of appropriate rate at each level, it has been shown that reliable transmission can be achieved within 1 dB of the Shannon limit [109].

Powerful probabilistic coding methods such as turbo codes are really needed only at the lower levels. At the higher levels, the channels become quite clean and the capacity $C(\Lambda_{j-1}/\Lambda_j)$ approaches $\log_2 |\Lambda_{j-1}/\Lambda_j|$, so that the desired redundancy approaches zero. For these levels, algebraic codes and decoding methods may be more appropriate [46], [110].

In summary, multilevel codes and multistage decoding allow the Shannon limit to be approached as closely in the high-SNR regime as it can be approached in the low-SNR regime with binary codes. The state of the art in multilevel coding is reviewed in [110].

### I. Other Capacity-Approaching Nonbinary Coded-Modulation Schemes

Various capacity-approaching alternative schemes to multilevel coding have been proposed.

One scheme, called bit-interleaved coded modulation (BICM) [15], has been developed primarily for fading channels, but has turned out to be capable of approaching the capacity of high-SNR AWGN channels as well. A BICM transmitter comprises an encoder for a binary code $C$, a “bit interleaver,” and a signal mapper. The output sequence of the bit interleaver is segmented into $m$-bit blocks, which are mapped into a $2^m$-point signal constellation using Gray coding. Although BICM cannot operate arbitrarily closely to capacity, good performance close to capacity can be obtained if $C$ is a long parallel or serially concatenated turbo code, and the decoder is an iterative turbo decoder.

Other schemes that are more straightforward extensions of standard binary turbo codes are called turbo trellis-coded modulation (TTCM) [89] and parallel concatenated trellis-coded modulation (PCTCM) [8]. Instead of two binary systematic convolutional encoders, two rate-$k/(k+1)$ trellis encoders are used. The interleaver is arranged so that the two encoded output bits of the two encoders are aligned with the same $k$ information bits. To avoid excessive rate loss, a special puncturing technique is used in TTCM: only one of the two coded bits is transmitted with each symbol, and the two encoder outputs are chosen alternately.

### V. CODING AND EQUALIZATION FOR LINEAR GAUSSIAN CHANNELS

In this section we discuss coding and equalization methods for a general linear Gaussian channel. We first present the “water-pouring” solution for the capacity-achieving transmit spectrum. Then we describe the multicarrier-achieving transmit spectrum. The remaining parts of the section are devoted to serial transmission. We show how the linear Gaussian channel may be converted, without loss of optimality, to a discrete-time AWGN channel with intersymbol interference. We then give a pictorial illustration of the coding principles required to approach capacity. Finally, we review practical
precoding methods that have been developed to implement these principles.

A. Water Pouring

The general linear Gaussian channel was characterized in Section II as a real channel with input signal $s(t)$, impulse response $g(t)$, and additive Gaussian noise $n(t)$ with one-sided p.s.d. $N(f)$. The channel output signal is $r(t) = s(t)g(t) + n(t)$ (see Fig. 8 below). In addition, the transmit signal $s(t)$ has to satisfy a power constraint of the form

$$\int_{f \geq 0} P_s(f) df \leq P \tag{5.1}$$

where $P_s(f)$ is the one-sided p.s.d. of $s(t)$. The channel SNR function for this channel is defined as

$$\text{SNR}_c(f) = |G(f)|^2/N(f). \tag{5.2}$$

Intuitively, the preferred transmission band is where SNR is largest.

The optimum $P_s(f)$ maximizes the mutual information between channel input and output subject to the power constraint (5.1) and $P_s(f) \geq 0$. A Lagrange multiplier argument shows that the largest mutual information is achieved by the “water-pouring” solution [94], illustrated in Fig. 7 for a typical $\text{SNR}_c(f)$ is largest.

The optimum $P_s(f)$ maximizes the mutual information between channel input and output subject to the power constraint (5.1) and $P_s(f) \geq 0$. A Lagrange multiplier argument shows that the largest mutual information is achieved by the “water-pouring” solution [94], illustrated in Fig. 7 for a typical $\text{SNR}_c(f)$ is largest.

The optimum $P_s(f)$ maximizes the mutual information between channel input and output subject to the power constraint (5.1) and $P_s(f) \geq 0$. A Lagrange multiplier argument shows that the largest mutual information is achieved by the “water-pouring” solution [94], illustrated in Fig. 7 for a typical $\text{SNR}_c(f)$ is largest.

B. Approaching Capacity with Multicarrier Transmission

One way of approaching capacity is suggested directly by the water-pouring argument. The capacity-achieving band $B$ may be divided into disjoint subbands of small enough width $\Delta f$ that $\text{SNR}_c(f)$ and thus $P_s(f)$ are nearly constant over each subband. Each subband can then be treated as an ideal AWGN channel with bandwidth $\Delta f$. Multicarrier transmission methods that can be used for this purpose are discussed in Appendix I.

The transmit power allocated to a subband centered at $f$ should be approximately $P_s(f) \Delta f$. Then the signal-to-noise ratio in that subband will be approximately $P_s(f) \text{SNR}_c(f)$, and the subband capacity will be approximately

$$C_{B/f}[\log_2(1 + P_s(f) \text{SNR}_c(f))]. \tag{5.8}$$

As $\Delta f$ is made sufficiently small, the aggregate power in all subbands approaches $P$ and the aggregate capacity approaches $C_B$ as given in (5.4). To approach this capacity, powerful coding in each subband at a rate $R_{B/f}[\log_2(1 + P_s(f) \text{SNR}_c(f))].$ near the subband capacity $C_B$ is needed. To minimize delay, coding should be applied “across subbands” with variable numbers of bits per symbol in each subband [14], [90].

Multicarrier transmission is inherently well suited for channels with a highly frequency-dependent channel SNR function, particularly for channels whose capacity-achieving band $B$ consists of multiple intervals. The symbol length in each subchannel is of the order of $1/\Delta f$, so that symbols become long, generally much longer than the response $g(t)$ of the channel. This makes multicarrier transmission less sensitive to moderate impulsive noise than serial transmission, and simplifies equalization. However, the resulting delay is undesirable for some applications.

As the transmitted signal is the sum of many independent components, its distribution tends to be Gaussian. The resulting large peak-to-average ratio (PAR) is a potential disadvantage of multicarrier transmission. However, recently several methods of controlling the PAR of multicarrier signals have been developed [81], [82].

C. Approaching Capacity with Serial Transmission

Alternatively, the capacity of a linear Gaussian channel may be approached by serial transmission. In this subsection, we...
first examine the conversion of the waveform channel to an equivalent discrete-time channel with intersymbol interference. This is achieved by using a transmit filter that shapes the transmit spectrum to the optimal water-pouring form, and employing a sampled matched filter (MF) or whitened matched filter (WMF) in the receiver. With the WMF, a canonical discrete-time channel with trailing intersymbol interference (ISI) and i.i.d. Gaussian noise is obtained, as depicted in Fig. 8. We show that the contribution of ISI to channel capacity vanishes at high SNR. This suggests that combining an ISI-canceling equalization technique with ideal-channel coding and shaping will suffice to approach capacity. In the following two subsections, we will then discuss the implementation of this approach.

We assume that the capacity-achieving band $B$ is a single positive-frequency interval $[f_1, f_2]$ of width $W = f_2 - f_1$. If $B$ consists of several frequency intervals, then the same approach may be used over each interval separately, although the practical attractiveness of this approach diminishes with multiple intervals. We concentrate on minimum-bandwidth passband transmission of complex symbols at modulation rate $1/T = W$. The transmission of complex signals over a real channel has been discussed in Section II.

In Fig. 8, the transmitted real signal is

$$s(t) = \text{Re} \left( \sum_i a_i g_T(t - iT) \right)$$

(5.9)

where ideally $\{a_i\}$ should be a sequence of i.i.d. complex Gaussian symbols with variance (average energy) $\sigma_a^2$ per symbol, and $g_T(t)$ is a positive-frequency transmit symbol response with Fourier transform $G_T(f)$. The transmit filter is chosen so that the p.s.d. of $s(t)$ is the water-pouring p.s.d. $P_s(f)$; i.e.,

$$\frac{\sigma_a^2}{T} |G_T(f)|^2 = P_s(f).$$

(5.10)

In the receiver, the received real signal $r(t)$ is first passed through a filter which suppresses negative-frequency components and whitens the noise in the positive-frequency band. This filter may be called a noise-whitening Hilbert splitter, because in terms of one-dimensional signals it “splits” the positive-frequency component of a real signal into its real and imaginary parts. The transfer function $G_w(f)$ of this filter is chosen such that $|G_w(f)|^2 = 1/N(f)$ at least over the band $B$, and $G_w(f) = 0$ for $f < 0$. The resulting complex signal is then

$$r(t) = \sum_i a_i v(t - iT) + w(t)$$

(5.11)

where $v(t)$ is a symbol response with Fourier transform $V(f) = G_T(f)G(f)G_w(f)$, and $w(t)$ is normalized AWGN with p.s.d. 1 over the signal band $B$.

Because the set of responses $\{v(t - iT)\}$ represents a basis for the signal space, by the principles of optimum detection theory [118] the set $\{\gamma_i\}$ of $T$-sampled matched-filter outputs

$$\gamma_i = \int r(t)v(t - iT)dt$$

(5.12)

of a matched filter (MF) with response $r_T(-t)$ is a set of sufficient statistics for the detection of the symbol sequence $\{a_i\}$. Thus no loss of mutual information or optimality occurs in the course of reducing $r(t)$ to the sequence $\{\gamma_i\}$. The composite receive filter consisting of the noise-whitening filter and the MF has the transfer function

$$G_{MF}(f) = G_w(f)V^*(f) = G_T^*(f)G_w(f)/N(f).$$

(5.13)

The Fourier transform of the end-to-end symbol response $Q(f) = G_T(f)G(f)G_{MF}^*(f) = |G_T(f)G(f)|^2/N(f)$ has the properties of a real nonnegative power spectrum that is band-limited to $B$. Moreover, $Q(f)$ is equal to the noise p.s.d. $P_n(f)$ at the MF output, because

$$P_n(f) = N(f)|G_{MF}^2 = |G_T(f)G(f)|^2/N(f) = Q(f).$$

(5.15)

The sampled output sequence of the MF is given by

$$\gamma_i = \sum_{\ell} q_{\ell} a_{i-\ell} + n_i$$

(5.16)

where the coefficients $q_{\ell} = q_{\ell}(T)$ are the sample values of the end-to-end symbol response $Q(f)$. In general, the discrete-time Fourier transform of the sequence $\{q_{\ell}\}$ is the $1/T$-aliased spectrum

$$\mathcal{Q}(f) = \frac{1}{T} \sum_{m \in \mathbb{Z}} Q(f + m/T).$$

(5.17)
In our case, because $Q(f)$ is limited to a positive-frequency band $B$ of width $W = 1/T$, there is no aliasing; i.e., the $1/T$-periodic function $\tilde{Q}(f)$ is equal to $(1/T) Q(f)$ within $B$. Because $P_h(f) = Q(f)$, $\{\eta_t\}$ is a Gaussian noise sequence with autocorrelation sequence $\{R_{\eta,m}\} = \{\eta_t\}$. Note that $\{\eta_t\}$ is Hermitian-symmetric, because $Q(f)$ is real.

So far, we have obtained without loss of optimality an equivalent discrete-time channel model which can be written in D-transform notation as

$$\hat{z}(D) = a(D)q(D) + n(D)$$  \hspace{1cm} (5.18)

where $q(D)$ is Hermitian-symmetric. We now proceed to develop an alternative, equivalent discrete-time channel model with a causal, monic, minimum-phase (“canonical”) response $h(D)$ and white Gaussian noise $\nu(D)$. For this purpose, we appeal to the discrete-time spectral factorization theorem:

**Spectral Factorization Theorem (Discrete-Time) [83]:** Let $\{\eta_t\}$ be an autocorrelation sequence with $D$-transform $Q(D) = \sum q_k D^k$, and assume its discrete-time Fourier transform $\hat{Q}(f) = q(e^{-2\pi i T})$ satisfies the discrete-time Paley–Wiener condition

$$\int_{1/T} |\log(\hat{Q}(f))| df < \infty$$

where $\int_{1/T}$ denotes integration over any interval of width $1/T$. Then $Q(D)$ can be factored as follows:

$$Q(D) = h(D^{-1})A^2 h(D)$$

$$\hat{Q}(f) = \hat{H}(f) A^2 \hat{H}(f)$$  \hspace{1cm} (5.19)

where $h_l = 1 + h_l D + \cdots$ is causal ($h_l = 0$ for $\ell < 0$, monic ($h_0 = 1$), and minimum-phase, and $\hat{H}(f) = h(e^{-2\pi i T})$ is the discrete-time Fourier transform of $h(D)$. The factor $A^2$ is the geometric mean of $\hat{Q}(f)$ over a band of width $1/T$; i.e.,

$$\log A^2 = T \int_{1/T} \log \hat{Q}(f) df$$  \hspace{1cm} (5.20)

where the logarithms may have any common base.

The discrete-time Paley–Wiener criterion implies that $\hat{Q}(f)$ can have only a discrete set of algebraic zeroes. Spectral factorization is further explained in Appendix III.

A causal, monic, minimum-phase response $h(D)$ is called canonical. The spectral factorization above is unique under the constraint that $h(D)$ be canonical. If $\hat{Q}(f)$ satisfies the Paley–Wiener condition, then (5.18) can be written in the form

$$\hat{z}(D) = a(D) A^2 h(D) h^*(D^{-1}) + \nu(D) A h^*(D^{-1})$$  \hspace{1cm} (5.21)

in which $\nu(D)$ is an i.i.d. Gaussian noise sequence with symbol variance 1. Filtering $\hat{z}(D)$ by $1/A^2 h^*(D^{-1})$ yields the channel model of the equivalent canonical discrete-time Gaussian channel

$$\hat{z}(D) = a(D) h(D) + \nu(D)$$  \hspace{1cm} (5.22)

where $\nu(D)$ is an i.i.d. Gaussian noise sequence with variance $1/A^2$. This requires that $h(D)$ has a stable inverse, and hence $h^*(D)$ has a stable anti-causal inverse, which is true if $q(D)$ has no spectral zeroes.

However, the invertibility of $h(D)$ is not a serious issue because $\hat{z}(D)$ can be obtained directly as the sequence of sampled outputs of a whitened matched filter (WMF) [41]. The transfer function of the composite receive filter consisting of the noise-whitening filter and the WMF is given by

$$G_R(WMF) = \frac{G_R^{WMF}(f)}{A^2 H^*(f)} = \frac{G_R(f)G^*(f) \hat{H}(f)}{N(f)} \hat{Q}(f).$$  \hspace{1cm} (5.23)

The only condition for the stability of this filter is the Paley–Wiener criterion. It is shown in [41] that the time response of this filter is always well defined.

The discrete-time Paley–Wiener criterion implies that $\hat{Q}(f)$ can have only a discrete set of algebraic zeroes. Spectral factorization is further explained in Appendix III.

A causal, monic, minimum-phase response $h(D)$ is called canonical. The spectral factorization above is unique under the constraint that $h(D)$ be canonical. If $\hat{Q}(f)$ satisfies the Paley–Wiener condition, then (5.18) can be written as

$$Q(f) = \frac{|G_{R}(f)|^2}{N(f)} \frac{T}{\sigma^2_{\nu}} P_\nu(f)$$  \hspace{1cm} (5.24)

and $\hat{Q}(f) = (1/T)Q(f)$, the capacity $C_{[\nu]}$ and its high-SNR approximation can be written as

$$C_{[\nu]} = \int_{B} \log_2 (1 + \sigma^2_{\nu} \hat{Q}(f)) df \approx \int_{B} \log_2 (\sigma^2_{\nu} \hat{Q}(f)) df.$$  \hspace{1cm} (5.25)

We now show that at high SNR the capacity $C_{[\nu]}$ of the linear Gaussian channel is approximately equal to the capacity of the ideal ISI-free channel that would be obtained if somehow ISI could be eliminated from the sequence $\hat{z}(D)$. In other words, ISI does not contribute significantly to capacity. This observation was first made by Price [85].

Suppose that somehow the ISI-causing “tail” of the response $h(D)$ could be eliminated, so that in the receiver $\hat{z}(D) = a(D) + \nu(D)$ rather than $\hat{z}(D) = a(D) h(D) + \nu(D)$ could be observed. The signal-to-noise ratio of the resulting ideal AWGN channel would be $SNR_{ISI-free} = \sigma^2_{\nu} A^2$. Thus the capacity of the ISI-free channel, and its high-SNR approximation, would be

$$C_{ISI-free[\nu]} = W \log_2 (1 + \sigma^2_{\nu} A^2) \approx W \log_2 (\sigma^2_{\nu} A^2).$$  \hspace{1cm} (5.26)
Price observed that at high SNR \( C_{[y/s]} \approx C_{\text{ESF-free}}[y/s] \), i.e.,
\[
C_{[y/s]} \approx \int_B \log_2 \sigma_\alpha^2 Q箕(j)\,dj = W \log_2 \sigma_\alpha^2 A^2 \approx C_{\text{ESF-free}}[y/s]
\]
where the equality in (5.27) follows from (5.20).

An alternative “minimum-mean-squared error” (MMSE) form of this result has been obtained in [22] and [23]. This result shows that if both ISI and bias are eliminated in an MMSE-optimized receiver, then the resulting signal-to-noise ratio \( \text{SNR}_{\text{ESF-free,MMSE}} \) is equal to \( \text{SNR}_{\text{eff}} \) for any linear Gaussian channel at any signal-to-noise ratio. If the interference in the MMSE-optimized unbiased receiver were Gaussian and uncorrelated with the signal, then this would imply \( C_{[y/s]} = C_{\text{ESF-free}}[y/s] \) with equality at all SNR’s; however, in general, this relation is again only approximate.

Price’s result and the alternative MMSE result suggest strongly that the combination of ISI-canceling equalization method and powerful ideal-channel coding and shaping will suffice to approach channel capacity on any linear Gaussian channel, particularly in the high-SNR regime.

D. Coding Objectives for Linear ISI-AWGN Channels

An intuitive explanation of the coding objectives for linear Gaussian channels is presented in Fig. 9.

Signal sequences are shown here as finite-dimensional vectors \( \mathbf{a}, \mathbf{y}, \mathbf{w}, \) and \( \mathbf{z} \). The convolution \( y(D) = a(D)h(D) \) that determines the noise-free output sequence becomes a vector transformation \( \mathbf{y} = \mathbf{aH} \) by a channel matrix \( \mathbf{H} \). Because the canonical response \( h(D) \) is causal and monic (\( h_0 = 1 \)), the matrix \( \mathbf{H} \) is triangular with an all-ones diagonal, so it has unit determinant: \( \det \mathbf{H} = 1 \). It follows that the linear transformation from \( \mathbf{a} \) to \( \mathbf{y} \) is volume-preserving [23].

To achieve shaping gain (minimize transmit power for a given volume), at the channel input the constellation points \( \mathbf{a} \) should be uniformly distributed within a hypersphere. The channel matrix \( \mathbf{H} \) transforms the hypersphere into a hyperellipsoid of equal volume containing an equal number of constellation points \( \mathbf{y} \).

To achieve coding gain, the noise-free channel output vectors \( \mathbf{y} \) should be points in an infinite signal set \( \Lambda \) with good distance properties (figuratively shown as an integer lattice). If the transmitter knows the channel matrix \( \mathbf{H} \), it can predistort the input vectors \( \mathbf{a} \) to be points in an appropriately predistorted signal set \( \Delta \mathbf{H}^{-1} \) (figuratively shown as a skewed integer lattice). The volumes \( V(\Lambda) \) and \( V(\Delta \mathbf{H}^{-1}) \) per point in the two signal sets are equal.

The noise-free output vectors are observed in the presence of i.i.d. Gaussian noise \( \mathbf{w} \). A minimum-distance detector for the infinite signal set \( \Lambda \) can therefore obtain the coding gain of \( \Lambda \) on an ideal AWGN channel.

In summary, coding and shaping may take place in two different Euclidean spaces, which are connected by a known volume-preserving linear transformation. Coding and shaping may be separately optimized, by choosing an infinite signal set with a large coding gain (density) for an ideal AWGN channel in the coding space, and by choosing a shaping scheme with a large shaping gain (small normalized second moment) in the shaping space. The channel intersymbol interference may be eliminated at the channel output by predistortion of the infinite signal set at the input, based on the known channel response.

E. Precoding and Trellis Coding for Linear Gaussian Channels

Precoding is a pre-equalization method that achieves the coding objectives set out in Section V-D, assuming that the canonical channel response \( h(D) \) is known at the transmitter. A pre-equalized sequence \( a(D) = y(D)/h(D) \) is transmitted such that

a) the output sequence \( z(D) = y(D) + w(D) \) is the output of an apparently ISI-free ideal channel with input sequence \( y(D) \);

b) the noise-free output sequence \( y(D) \) may therefore be a coded sequence from a code designed for the ideal AWGN channel, and may be decoded by an ideal-channel decoder;

c) redundancy in \( y(D) \) may be used to minimize the average power of the input sequence \( a(D) = y(D)/h(D) \), or to achieve other desirable characteristics for \( a(D) \).

Precoding was first developed for uncoded \( M \)-PAM transmission in two independent masters’ theses by Tomlinson [97] and Harashima [54]. Its application was not pursued at that time because decision-feedback equalization was, and still is, the preferred ISI-canceling method for uncoded transmission.
With trellis coding, however, decision-feedback equalization is no longer attractive, because reliable decisions are not available from a Viterbi decoder without significant delay. As we shall see, Tomlinson–Harashima precoding can be combined with trellis coding, but not with shaping.

Trellis precoding [35] was the first technique that allowed combining trellis coding, precoding, and shaping. However, in trellis precoding, coding and shaping are coupled, which to some degree inhibits control of constellation characteristics such as peak-to-average ratio (PAR). Therefore, during the development of the V.34 modem standard techniques called “flexible precoding” were proposed independently by Eyuboglu [80], Cole and Goldstein [52], and Laroia et al. [73]. In flexible precoding, coding and shaping are decoupled. Later, Laroia [71] proposed an “ISI precoding” technique that used feedback trellis coding (see Section IV-D) to reduce “dither power.” Further improvements were made by Cole and Eyuboglu [53] and Betts [4], and the [53] scheme was ultimately adopted for the V.34 standard [62]. Recently, feedback trellis coding has been combined with Tomlinson–Harashima precoding [20].

We will describe Tomlinson–Harashima precoding and flexible precoding, with and without trellis coding. The general principle applied in these schemes is as follows. Let the canonical channel response $h(D)$ be written as $h(D) = 1 + h_1 D + h_2 D^2 \ldots$ be written as $h(D) = 1 + Dh_i(D)$. Then the noise-free output sequence of the channel is given by

$$y(D) = a(D) + a(D)Dh_i(D) = a(D) + p(D)$$  \hspace{1cm} (5.28)

where $p(D)$ is the sequence of traling intersymbol interference. In all types of precoding, the precoder determines $y_i = \sum_{j \in A} h_{ij} a_{i-n}$ from past signals and sends $a_i = y_i - p_i$ so that the channel output will be a desired signal $y_i$.

Tomlinson–Harashima (TH) precoding is illustrated in Fig. 10 for one-dimensional transmission with $M$-PAM signal constellations as defined in Section II-C. In the language of lattice constellations introduced in Section IV-A, an $M$-PAM signal constellation with a signal spacing of $\sqrt{M}$ is defined as

$$A = C(\mathbb{Z}, \mathbb{R}(M\mathbb{Z})) = (\mathbb{Z} + t) \cap \mathbb{R}(M\mathbb{Z})$$

where $t = 1/2$ for $M$ even and $t = 0$ for $M$ odd, and $\mathbb{R}(M\mathbb{Z}) = (-M/2, M/2]$ is the fundamental Voronoi region of the sublattice $M\mathbb{Z}$. The translates of $\mathbb{R}(M\mathbb{Z})$ by integer multiples of $M$ then cover (“tile”) all of real signal space, so every real signal may be expressed uniquely as $r = a + Mn$ for $a \in \mathbb{R}(M\mathbb{Z})$ and some integer $n \in \mathbb{Z}$. The unique $a \in \mathbb{R}(M\mathbb{Z})$ so determined is called “$r$ mod $M\mathbb{Z}$.”

Fig. 10 is drawn such that one can easily see that the noiseless channel-output sequence is $y(D) = x(D) - Mn(D)$, where $x(D) \in A$ is the sequence of PAM signals to be transmitted, and $n(D)$ is an integer sequence generated by the “$r\mod M\mathbb{Z}$” unit in the precoder so that the elements of the transmitted sequence $a(D)$ are contained in $(-M/2, M/2]$, and the elements of $y(D)$ are in $\mathbb{Z} + t$. The decoder operates on the noisy sequence $z(D) = y(D) + w(D)$ and outputs the sequence $\hat{y}(D)$, which is then reduced by the “$r\mod M\mathbb{Z}$” unit to the sequence $\hat{r}(D) = \hat{y}(D) + Mn(D) \in A$. In the absence of decoding errors, $\hat{y}(D) = y(D)$ and therefore $\hat{r}(D) = x(D)$. Note that “inverse precoding” is memoryless, so no error propagation can occur. Because the channel response $h(D)$ does not need to be inverted in the receiver, $h(D)$ may exhibit spectral nulls, e.g., contain factors of the form $(1 \pm D)$.

The combination of trellis coding with TH precoding requires that $y(D) = x(D) - Mn(D)$ is a valid code sequence. For practically all trellis codes, when $M$ is a multiple of 4 and $x(D)$ is a code sequence, then $y(D)$ is also a code sequence (in particular, this holds for “$\mod 4$” or “$\mod 2$” coset codes [42], [43]). Thus trellis coding and Tomlinson–Harashima precoding are easily combined.

With a more complicated constellation $A$, trellis coding can be combined with TH precoding by using the idea of feedback trellis coding [20], provided that a) the first-level subsets $B(0)$ and $B(1)$ of $A$ are congruent; b) the constellation shape region $\mathbb{R}(A)$, namely, the union of the Voronoi regions of the signals in $A$, is a space-filling region (has the “tiling” property). Condition b) obviously holds for $M$-PAM and square $M \times M$-QAM constellations. It also holds for a 12-QAM “cross” constellation, for example, but not for a 32-QAM cross constellation. We note that signal constellations whose sizes are not powers of two have become practically important in connection with mapping techniques such as shell mapping and modulus conversion (see Section IV-B).

In summary, Tomlinson–Harashima precoding permits the combination of trellis coding with ISI-canceling (DFE-equivalent) equalization. Its main problem is the requirement that the signal space can be “tiled” with translates of the
constellation shape region. Tomlinson–Harashima precoding results in a small increase in transmit signal power, equal to the average energy of a random “dither variable” uniformly distributed over the Voronoi region around the constellation signals. If the Voronoi region is \( \mathbb{R}(\mathbb{Z}) \), then the increase in signal power is \( 1/12 \), which is negligible for large constellations.

Flexible (FL) precoding is illustrated in Fig. 11, again for one-dimensional \( M \)-PAM transmission. The general concept of FL precoding is to subtract from a sequence of \( M \)-PAM signals a sequence of smallest “dither” variables \( q(D) \) such that the channel-output sequence \( y(D) \) is a valid uncoded or coded sequence with elements in \( \mathbb{Z} \). The constellation shape region of the signal constellation \( A \) can be arbitrary, so any shaping scheme can be used. At time \( t \), given the ISI term \( p_t \), the “\( \text{mod} \) \( \mathbb{Z} \)” unit in the precoder determines the integer \( n_t \in \mathbb{Z} \) that is closest to \( p_t \). The dither variable is the difference \( q_t = p_t - n_t \). Clearly, \( q_t \) lies in the Voronoi interval \( \mathbb{R}(\mathbb{Z}) = (-1/2, 1/2] \). It will typically be uniformly distributed over \( \mathbb{R}(\mathbb{Z}) \), and thus increases the average energy of the transmit signal \( a_t = x_t + q_t \) by \( 1/12 \).

As can be seen from Fig. 11, the sequence of noiseless channel-output signals becomes

\[
y(D) = x(D) - q(D) + n(D) = x(D) + n(D),
\]

The decoder operates on the noisy sequence \( z(D) = y(D) + w(D) \) and produces the sequence \( \hat{y}(D) \). To obtain the sequence \( \hat{x}(D) = \hat{y}(D) - \hat{n}(D) \), the sequence \( \hat{n}(D) \) is recovered by channel inversion as shown in Fig. 11. A “bit-identical” realization of the two functional blocks shown in dashed blocks in Fig. 11 is absolutely essential.

The combination of flexible precoding with feedback trellis coding works as follows. At time \( t \), the transmitter knows the integer \( n_t \) and the current trellis-code state \( s_t \). To continue a valid code sequence \( y(D) \) at the noiseless channel output, the transmitter selects a data symbol \( x_t \) such that \( y_t = x_t + n_t \) is in the first-level infinite subset \( B(0) \) or \( B(1) \) determined by \( s_t \). The symbol \( x_t \) is chosen to lie in a desired constellation shape region \( \mathbb{R}(A) \), or according to some other shaping scheme. The symbol \( y_t \) then determines the next state \( s_{t+1} \), as explained in Section IV-D.

The main advantage of flexible precoding is that any constellation-shaping method can be used, whereas with Tomlinson–Harashima precoding or trellis precoding, shaping is connected with coding. The main disadvantage is that decoding errors tend to propagate due to the channel inversion in the receiver, which is particularly troublesome on channels with spectral nulls. A technique to mitigate error propagation in the inverse precoder is described in [37]. The main application of flexible precoding so far has been in V.34 modems. The application of precoding in digital subscriber lines has been studied in [38].

In conclusion, at high SNR’s, precoding in combination with powerful trellis codes and shaping schemes allows the capacity of an arbitrary linear Gaussian channel to be approached as closely as capacity can be approached on an ideal ISI-free AWGN channel, with about the same coding and shaping complexity.

VI. FINAL REMARKS

Shannon’s papers established ultimate limits, but gave no constructive methods to achieve them. They thereby threw down the gauntlet for the new field of coding. In particular, they established fundamental benchmarks for linear Gaussian channels that have posed tough challenges for subsequent code inventors.

The discouraging initial progress in developing good constructive codes was captured by the folk theorem [119]: “All codes are good, except those that we know of.” However, by the early 1970’s, good practical codes had been developed for the low-SNR regime:

- moderate-constraint-length binary convolutional codes with Viterbi decoding for 3–6-dB coding gain with moderate complexity;
- long-constraint length convolutional codes with sequential decoding to reach \( R_0 \);
- oncatenation of RS outer codes with algebraic decoding and moderate-complexity inner codes to achieve very low error rates near \( R_0 \).

By the 1980’s, the invention of trellis codes enabled similar progress to be made in the high-SNR regime, at least for the ideal band-limited AWGN channel. In the 1990’s, these
gains were extended to general linear Gaussian channels by use of multicarrier modulation, or alternatively by use of single-carrier modulation with transmitter precoding.

Notably, many of these techniques were developed by practically oriented engineers in standards-setting groups and in industrial R&D environments. Moreover, as a consequence of advances in VLSI design and technology, the lag between invention and appearance in commercial products has often been very short.

By now, it certainly appears that the field of modulation and coding for Gaussian channels has reached a mature state, at least below the cutoff rate $R_0$, even though very few systems reaching $R_0$ have ever actually been implemented. In the beyond-$R_0$ regime, however, the invention of turbo codes has touched off intense activity that seems likely to continue well into the next half-century.

APPENDIX I
MULTICARRIER MODULATION

The history of multicarrier modulation began more than 40 years ago with an early system called Kineplex [30] designed for digital transmission in the HF band. Work by Holsinger [57] and others followed.

The use of the discrete Fourier transform (DFT) for modulation and demodulation was proposed in [115]. DFT-based multicarrier systems are now referred to as orthogonal frequency-division multiplexing (OFDM) or discrete multitone (DMT) systems [2], [14], [56], [90]. Basically, OFDM/DMT is a form of frequency-division multiplexing (FDM) in which modulation symbols are transmitted in individual subchannels using QAM or CAP modulation (see Section II-B). Current applications of OFDM/DMT include digital audio broadcasting (DAB) [33] and asynchronous digital subscriber lines (ADSL) [1].

Another multicarrier transmission technique of the FDM type became more recently known as discrete wavelet multitone (DWMT) modulation [91], [102]. DWMT modulation has its origin in filter banks for subband source coding. Efficient implementations of DWMT modulation and demodulation employ the discrete cosine transform (DCT).

A comprehensive treatment of multirate systems and filter banks is given in [106]. Recent reviews of theory and applications of filter banks and wavelet transforms are presented in [99].

The coding aspects of multicarrier systems have been discussed in Section IV-B. In this appendix, we give brief descriptions of DMT and DWMT modulation (together with some background information). For the purposes of this appendix, we assume transmission over a noise-free discrete-time real linear channel which is modeled by $y(D) = x(D)h(D)$, where $h(D)$ is the channel response.

A. DFT-Based Filter Banks and Discrete Multitone (DMT) Modulation

Before specifically addressing DMT modulation, we examine the general concept of DFT-based multicarrier modulation. These systems subdivide the channel into $N$ narrowband subchannels whose center frequencies are spaced by integer multiples of the subchannel modulation rate $1/T$. Let $\{A_k(\gamma)\}$ be the sequence of generally complex data symbols transmitted in the $k$th subchannel, and let $A_k = [A_k(\gamma), 0 \leq k \leq N-1]$ denote the $N$-vector transmitted over these channels at time $t'$. A DFT-based “synthesis” filter bank generates at rate $N/T$ the transmit signals

$$x_i = \sum_{k=0}^{N-1} \sum_{\ell} A_k(\gamma) v_{i-\ell N} \exp(j2\pi ki/N)$$  \hspace{1cm} (AI.1)

where $\{v_i, i \geq 0\}$ is a real causal symbol sequence with a baseband spectrum $V(f)$. This $N/T$-sampled response is sometimes called the “prototype response.” The least integer $\gamma \geq 0$ such that $v_i = 0$ for all $i \geq (\gamma + 1)N$ is called the “overlap factor.” Note that the sequence $\{x_i\}$ has an $N/T$-periodic spectrum and is composed of subchannels whose center frequencies are located at integer multiples of $1/T$, as illustrated in Fig. 12.

If we express the time index as $i = mN+n$, with $m \in \mathbb{Z}$ and $0 \leq n \leq N-1$, then (AI.1) becomes

$$x_mN+n = \sum_{\ell} \left( \sum_{k=0}^{N-1} A_k(\gamma) \exp(j2\pi kn/N) \right)v_{m-\ell N+n}$$

$$= \sum_{\ell} a_{e}(n) x_{m-\ell N+n}$$

$$= \sum_{\ell=0}^{\gamma} p_{e}(n) x_{m-\ell N+n}, \hspace{1cm} 0 \leq n \leq N-1$$  \hspace{1cm} (AI.2)

where $a_{e}(n), 0 \leq n \leq N-1$ is the inverse discrete Fourier transform (IDFT) of $A_k = [A_k(\gamma), 0 \leq k \leq N-1]$, and $p_{e}(n) = v_{nN+n}, 0 \leq \ell \leq \gamma$, for $0 \leq n \leq N-1$, is the $1/T$-sampled $\gamma$-phase component of the prototype response. Thus (AI.2) shows that the signals $\{x_i\}$ may be generated by $N$ independent convolutions of sequences of the time-domain symbols $\{a_{e}(n)\}$ with the corresponding phase components $\{p_{e}(n)\}$ of the prototype response. Such a system is called a DFT-based polyphase filter bank [7], [106].

To obtain real transmit signals, the frequency-domain modulation symbols must exhibit Hermitian symmetry; i.e., $A_k(\gamma) = A_k^{*}(N-k)$. Then the IDFT coefficients $a_{e}(n)$ are real and hence the signals $x_i$ are real. The restriction to Hermitian symmetry allows for valid transformations of sequences of the form $\{a_{e}(n), 0 \leq k \leq N-1\}$ into $N$ complex symbols $\{A_k(\gamma), 0 \leq k \leq N-1\}$ for $N$ even, one possible mapping is

$$A_k(0) = a_{e}(0)$$

$$A_k(N/2) = a_{e}(N-1)$$

$$A_k(\gamma) = a_{e}(2k-\gamma) + j a_{e}(2k) = A_k^{*}(N-k),$$

$$1 \leq k \leq N/2-1.$$  \hspace{1cm} (AI.3)

In the absence of distortion, the IDFT coefficients can be recovered from $\{y_i\} = \{x_i\}$ by a polyphase “analysis” filter bank, which performs the matched filter operations

$$\sum_{m} x_{mN+n} p_{m-\ell N+n} = a_{e}(n).$$  \hspace{1cm} (AI.4)
If the orthogonality conditions
\[
\sum_m p_m(n)p_{m-\ell}(n) = \delta_\ell
\] (AI.5)
are satisfied for all \(\ell \in \mathbb{Z}\) and \(0 \leq n \leq N - 1\), then \(\hat{a}_\ell(n) = a_\ell(n)\). Finally, the \(N\)-vector \(\mathbf{A}_\ell = [a_k(\ell), 0 \leq k \leq N - 1]\) is obtained as the DFT of \(\mathbf{a}_\ell = [a_k(n), 0 \leq n \leq N - 1]\).

We note from (AI.5) that in an ideal system each of the \(N\) phase components of the prototype response must individually satisfy the Nyquist condition for zero-ISI transmission in that subchannel. Then, summing (AI.4) over \(n = 0, 1, \cdots, N - 1\) shows that also the prototype response satisfies the orthogonality condition for transmission at rate \(T/1\).

The orthogonality conditions are trivially satisfied if the prototype response is a sampled rectangular pulse of length \(T\), i.e., \(v_i = \text{const.}, 0 \leq i \leq N - 1\) \((\gamma = 0)\). In this case, no polyphase filtering is required. However, because the spectra of the subchannels will then overlap according to the \(\sin(x)/x\) shape of the prototype response spectrum, small amounts of channel distortion can be sufficient to destroy orthogonality and cause severe interchannel interference (ICI). For some applications, a higher spectral concentration of the subchannel spectra will therefore be desirable. This leads to a filter design problem, which consists in minimizing for given filter length, i.e., for given overlap factor \(\gamma > 0\), the spectral energy of the prototype response outside of the band \(|f| < 1/2T\) while approximately maintaining orthogonality.

For DMT modulation, \(v_i = 1\) for \(0 \leq i \leq N - 1\) is used, and the transmitted signals could simply be the unfiltered IDFT coefficients:
\[
x_i = x_{IN+n} = a_m(n).
\] (AI.6)
In the absence of channel distortion, this simple scheme would be sufficient. However, if the channel response \(h_k\) has length \(L'\) (i.e., \(h_0 \neq 0, \cdots, h_{L'} \neq 0\)), then the \(L'\) last IDFT coefficients of every transmitted block of \(N\) IDFT coefficients would interfere with the first \(L'\) coefficients of the next block (assuming \(L' \leq N\)). In principle, combinations of pre-equalization and post-equalization (before and after the DFT operation in the receiver) could be used to mitigate the effects of this interference.

Practical DMT systems employ the simpler method of “cyclic extension,” originally suggested in [86], to cope with distortion. If the channel response has at most length \(L\), then every block of IDFT coefficients is cyclically extended by \(L\) coefficients. For example, with “cyclic prefix extension” the \(m\)th block is extended from length \(N\) to length \(L + N\) as follows:
\[
\{a_m(N - L), \cdots, a_m(N - 1); a_m(0), a_m(1), \cdots, a_m(N - 1), a_m(N - 1)\}.
\] (AI.7)
Assume the actual length of the channel response is \(L' < L\). Then the received sequence \(y_k\) contains cyclic signal repetitions within windows of length \(L' \leq L\) for every received block of extended length \(L + N\). The receiver selects blocks of signals of length \(N\) such that the received signals at the beginning and end of these blocks are cyclically repeating. The DFT of these blocks is then
\[
\{A_m(n)H(n/T), 0 \leq n \leq N - 1\}
\] (AI.8)
where \(H(n/T)\) is the Fourier transform of the channel response at \(f = n/T\). Multiplication with the estimated inverse spectral channel response yields the desired symbols \(A_m(n)\).

With this “one tap per symbol” equalization method, distortion is dealt with in a simple manner at the expense of a rate loss by the factor of \(L'/L\). For example, the ADSL standard [1] specifies \(N = 512\) and \(L = 32\), which results in a loss of 5.8%.

In practical DMT systems, one may not rely entirely on the cyclic-extension method to cope with distortion. An adaptive pre-equalizer may be employed to minimize the channel-response length prior to the DFT-based receiver operations described above. The problem is essentially similar to adjusting the forward filter in a decision-feedback equalizer or an MLSD receiver such that the channel memory is truncated to a given length [21], [36].

B. Cosine-Modulated Filter Banks and Discrete Wavelet Multitone (DWMT) Modulation

Cosine-modulated filter banks were first introduced for subband speech coding. The related concept of quadrature mirror filters (QMF) was probably first mentioned in [32]. In the field of digital communications, multicarrier modulation by means of cosine-modulated filter banks is referred to as DWMT modulation [91], [102]. The main feature of DWMT is that all signal processing is performed on real signals.
DWMT can be understood as a form of multiple carrierless single-sideband modulation (CSSB, see Section II-B).

The sequences \( \{A_k(k)\}, 0 \leq k \leq N-1 \), are now sequences of real modulation symbols. DWMT employs a real prototype response \( \{v_k\} \) with a baseband spectrum \( V(f) \) that satisfies the Nyquist criterion for ISI-free transmission at symbol rate \( 1/2T \). Also, \( V(f) \) has no more than 100\% spectral roll-off (i.e., \( V(f) = 0 \) for \( |f| > 1/2T \)). Furthermore, it is convenient to assume that \( V(f) \) has zero phase.

A cosine-modulated “synthesis” filter bank generates at rate \( N/T \) the transmit signals

\[
  x_i = \sum_{k=0}^{N-1} \sum_{\ell} A_k(\ell) p_{i-\ell N}(k)
\]

where the samples of the symbol response for the \( k \)th sub-channel are given by

\[
  p_i(k) = 2v_i \cos(\pi(k+0.5)/N + \varphi_k)
\]

\[
  \varphi_k = (-1)^{k+1} \pi/4.
\]

The corresponding spectral symbol response is then

\[
  P^k(f) = e^{-j\varphi_k} V\left(f + \frac{k+0.5}{2T}\right) + e^{j\varphi_k} V\left(f - \frac{k+0.5}{2T}\right).
\]

The spectrum of the transmitted signal is shown in Fig. 13. One can verify that any two overlapping frequency-shifted prototype spectra differ in phase by \( \pm \pi/2 \). This is the so-called QMF condition, which ensures orthogonality [106].

A cosine-modulated “synthesis” filter bank generates at rate \( N/T \) the transmit signals

\[
  x_i = \sum_{k=0}^{N-1} \sum_{\ell} A_k(\ell) p_{i-\ell N}(k)
\]

where the samples of the symbol response for the \( k \)th sub-channel are given by

\[
  p_i(k) = 2v_i \cos(\pi(k+0.5)/N + \varphi_k)
\]

\[
  \varphi_k = (-1)^{k+1} \pi/4.
\]

The corresponding spectral symbol response is then

\[
  P^k(f) = e^{-j\varphi_k} V\left(f + \frac{k+0.5}{2T}\right) + e^{j\varphi_k} V\left(f - \frac{k+0.5}{2T}\right).
\]

The spectrum of the transmitted signal is shown in Fig. 13. One can verify that any two overlapping frequency-shifted prototype spectra differ in phase by \( \pm \pi/2 \). This is the so-called QMF condition, which ensures orthogonality [106].

For a DCT-based polyphase realization of cosine-modulated filter banks, see [106].

**APPENDIX II**

**UNIFORMITY PROPERTIES OF EUCLIDEAN-SPACE CODES**

A binary linear code has a useful and fundamental uniformity property: the code “looks the same” from the viewpoint of any of its sequences. The number of neighbors at each distance is the same for all code sequences, and, provided the channel has appropriate symmetry, the error probability is independent of which code sequence is transmitted. If this uniformity property holds, then one has to consider only the set of code distances from the all-zero sequence (i.e., the set of code weights), and in performance analysis one may assume that the all-zero sequence was transmitted.

Similar uniformity properties are helpful in the design and analysis of Euclidean-space codes. A variety of such properties have been introduced, from the original “quasilinearity” property of Ungerboeck [104] to geometric uniformity [44] (a property that generalizes the group property of binary linear codes), and finally rotational invariance. In this section we briefly review these properties.

**A. Geometrical Uniformity**

As we have seen in Section IV, for coding purposes it is useful to regard a large constellation based on a dense packing such as a lattice translate \( \Lambda + t \) as an effectively infinite constellation consisting of the entire packing. This approximation eliminates boundary effects, and the resulting infinite constellation usually has greater symmetry. As we have seen, shaping methods (selection of a finite subconstellation) may be designed and analyzed separately in the high-SNR regime.

The symmetry group of a constellation is the set of isometries of Euclidean \( n \)-space \( \mathbb{R}^n \) (i.e., rotations, reflections, translations) that map the constellation to itself. For a lattice translate \( \Lambda + t \), the symmetry group includes the infinite set \( \{t_\lambda, \lambda \in \Lambda\} \) of all translations \( t_\lambda \) by lattice elements \( \lambda \in \Lambda \). It will also generally include a finite set of orthogonal transformations; e.g., the symmetry group of the QAM lattice translate \( \mathbb{Z}^2 + (1/2, 1/2) \) includes the eight symmetries of the square (the dihedral group \( D_8 \)). A constellation is **geometrically uniform** if it is the orbit of any of its points under its symmetry group. A lattice translate \( \Lambda + t \) is necessarily geometrically uniform, because it can be generated by the translation group \( \{t_\lambda, \lambda \in \Lambda\} \). An \( M \)-PSK constellation is geometrically uniform, because it can be generated by rotations of multiples of \( 2\pi/M \).

A fundamental region \( R(\Lambda) \) of a lattice \( \Lambda \) is a compact region that contains precisely one point from each translate (coset) \( \Lambda + t \) of \( \Lambda \). For example, any Voronoi region of \( \Lambda \) is a fundamental region if boundary points are appropriately divided between neighboring regions. The translates \( \{R(\Lambda) + \lambda, \lambda \in \Lambda\} \) of any fundamental region \( R(\Lambda) \) tile \( n \)-space.
be uniquely expressed as $x = r + \lambda$ for some $r \in R(\Lambda)$ and $\lambda \in \Lambda$, where $r$ is the unique element of $R(\Lambda)$ in the coset $\Lambda + x$. The volume of any fundamental region is $V(\Lambda)$.

As discussed in Section IV, trellis codes are often based on lattice partitions $\Lambda/\Lambda'$. A useful viewpoint that has emerged from multilevel coding is to regard such codes as being defined on a fundamental region $R(\Lambda')$ of the sublattice $\Lambda'$. The $[\Lambda/\Lambda']$ cosets of $\Lambda$ in $\Lambda + t$ are represented by their unique representatives in $R(\Lambda')$. At the front end of the receiver, each received point $x \in \mathbb{R}^n$ is mapped to the unique point of $\Lambda' + x$ in $R(\Lambda')$. Although such a “map” is not information-lossless, it has been shown that it does not reduce capacity [46]. Moreover, it decouples coding and decoding implementation and performance from any coding that may occur at other levels. Conceptually, this viewpoint reduces infinite- constellation coding and decoding to coding and decoding using a finite constellation of $[\Lambda/\Lambda']$ points on the compact region $R(\Lambda')$.

A geometrically uniform code based on a lattice partition $\Lambda/\Lambda'$ may be constructed as follows. Let $G$ be a label group that is isomorphic to the quotient group $\Lambda/\Lambda'$; or, more generally, let $G$ be a group of $[\Lambda/\Lambda']$ symmetries that can generate a set of $[\Lambda/\Lambda']$ subsets congruent to $\Lambda'$ whose union is $\Lambda + t$. Let $T$ be a time axis (e.g., $T = \{1, \ldots, n\}$ for block codes, or $T = \mathbb{Z}$ for convolutional codes), let $GT$ be the set of all sequences $(g_t, t \in T)$ of elements of $G$ defined on $T$, which form a group under the componentwise group operation of $G$, and let $C \subset GT$ be any subgroup of $GT$. Then the orbit of any sequence of subsets of $\Lambda + t$ under $C$ is a geometrically uniform Euclidean-space code. In particular, for any geometrically uniform code, the set of distances from any code sequence to all other code sequences is the same for all code sequences, and the probability of error on an ideal AWGN channel does not depend on which sequence was sent. This observation has stimulated research into group codes, namely codes $C$ that are subgroups of sequence groups $GT$, and their geometrically uniform Euclidean-space images. Although the ultimate goal of this research has been to construct new classes of good codes, the main results so far have been to show that most known good codes constructed by other methods are geometrically uniform. Ungerboeck’s original 1D and 2D codes are easily shown to be geometrically uniform. Trotter showed that the nonlinear Wei 4D 8-state trellis code used in the V.32 modem standard may be represented as a group code whose state space is the non-Abelian dihedral group $D_8$ [100]. The 4D 16-state Wei code and the 32-state Williams code used in the V.34 modem standard are both geometrically uniform; however, Sarvis and Trotter have shown that there is no geometrically uniform code with the same parameters as the 4D 64-state Wei V.34 code [92].

**B. Quasilinearity**

A weaker uniformity property introduced in [104] is quasilinearity. This property permits the minimum squared distance $d_{\text{min}}^2(C)$ of a trellis code $C$ to be found by simply computing the weights of a set of error sequences, rather than by pairwise comparison of all pairs of possible code sequences.

Quasilinearity depends only on the use of a rate-$(n-1)/n$ binary linear convolutional code in an encoder like that in Fig. 5, and the fact that the subsets $B(0)$ and $B(1)$ associated with the first-level partition are geometrically congruent to each other.

Let $\{c_i\}$ and $\{c'_i\}$ be any two binary convolutional code sequences, and let their difference be the error sequence $\{e_i\} = \{c_i + c'_i\}$, which by linearity is also a code sequence. For each error term $e_i$, define the squared Euclidean weight

$$w^2(e_i | c'_i) = \min \{w(e_i) - a(c_i + e_i), a(c_i) - a(e_i)\}$$

where $e_i$ ranges over all binary $n$-tuples $e_i = [c_i^0 \cdots c_i^n]$ with the given bit $c_i^j$, and $a(c_i)$ and $a(c_i + e_i)$ range over all signals in the subsets labeled by $c_i$ and $c_i + e_i$, respectively.

If $B(0)$ and $B(1)$ are congruent, then it is easy to see that $w^2(e_i | 1) = w^2(e_i | 0)$. We may then simply write $w^2(e_i)$ for this common minimum.

It then follows that the minimum squared distance $d_{\text{min}}^2(C)$ between any two trellis-code sequences that correspond to different convolutional-code sequences $\{c_i\}$ and $\{c'_i\}$ is

$$d_{\text{min}}^2(C) = \min_{\{c_i\} \neq \{0\}} \sum_i w^2(e_i).$$

The proof is that $\{c_i\}$ and $\{c'_i\}$ may differ arbitrarily in the $n-1$ label bits $[c_i^1 \cdots c_i^n]$ and $[c'_i^1 \cdots c'_i^n]$, to which the encoder adds appropriate parity-check bits $c_i^0$ and $c'_i^0$, and for any $e_i$ the minimum $[a(c_i) - a(c_i + e_i)]^2$ is achievable with either choice of $c_i^0$. Thus quasilinearity permits $d_{\text{min}}^2(C)$ to be computed just as the minimum Hamming distance of binary linear convolutional codes is computed; one simply replaces Hamming weights $w_F(e_i)$ by the squared Euclidean weights $w^2(e_i)$ [104], [120] in trellis searches. This property has been used extensively in searches for the trellis codes reported in [104] and by subsequent authors.

The number of codes to be searched for a code with the highest value of $d_{\text{min}}^2(C)$ can be further reduced significantly if $w^2(e_i)$ attains for all $e_i$ the set-partitioning lower bound $\Delta_{q(e_i)}^2$, where $\Delta_{q(e_i)}^2$ is the minimum intrasubset squared distance at the first partitioning level $q$ for which $c_i^0 \neq 0$ [104]. This property usually holds for large geometrically uniform constellations.

**C. Rotational Invariance**

In carrier-modulated complex-signal transmission systems, the absolute carrier phase of the received waveform signal is generally not known. The receiver may demodulate the received signal with a constant carrier-phase offset $\Delta \varphi \in \Phi$, where $\Phi$ is the group of phase rotations that leave the two-dimensional signal constellation invariant. Then, if a signal sequence $\{a_i\}$ was transmitted, the decoder operates on the sequence of rotated signals $\{a_i \exp(i \Delta \varphi)\}$. For example, for $M$-QAM we have $\Phi = \{0, 90^\circ, 180^\circ, 270^\circ\}$. Rotational invariance is the property that rotated code sequences are also valid code sequences. This is obviously the case for uncoded modulation with signal constellations that exhibit rotational
symmetries. To achieve transparency of the transmitted information under phase offsets, phase-differential encoding and decoding may be employed.

A trellis code $C$ is fully rotationally invariant if for all coded signal sequences $\{a_i\} \in C$ and all phase offsets $\Delta \varphi \in \Phi$ the rotated sequences are code sequences, i.e., $\{a_i \exp(\Delta \varphi)\} \in C$. Note that signals $a_i$ here always denote two-dimensional signals, which may be component signals of a higher dimensional signal constellation.

Trellis codes are not necessarily rotationally invariant under the same set of symmetries $\Phi$ as their two-dimensional signal constellations. It has been shown that trellis codes based on linear binary convolutional codes with mapping into two-dimensional $M$-PSK or $M$-QAM ($M > 2$) signals can at most be invariant under $180^\circ$ rotation. Fully rotationally invariant linear trellis codes exist only for higher dimensional constellations. Such codes are given for $K \times M$-PSK in [105], [113], $(K = 2, 4)$, and for $K \times M$-QAM in [87], [114], $(M = 8, 16; K = 2, 3, 4)$.

Fully rotationally invariant trellis codes over two-dimensional PSK or QAM constellations must be nonlinear [88], [112]. We briefly describe the analytic approach presented in [88] for the design of nonlinear two-dimensional PSK and QAM trellis codes by an example. Let a QAM signal constellation be partitioned into eight subsets with subset labels $c = [c^1, c^2, c^3]$, where $c^2 = 2c^2 + c^4 \in \mathbb{Z}_4$ (ring of integers modulo 4). The labeling is chosen such that under $90^\circ$ rotation the labels are transformed into $c' = [c^1, (c^2 + 1) \mod 4]$. A fully rotationally invariant trellis code can then be found by requiring that for all label sequences $\{c'_i\} = \{c^1_i, c^2_i\}$ that satisfy a given parity-check equation, the sequences $\{c''_i\} = \{c^1_i, (c^2_i + 1) \mod 4\}$ also satisfy this equation. This is achieved by a binary nonlinear parity-check equation of the form

$$h^2(D)e^{\beta(D)} \equiv [h^r(D)e^{\beta'(D)} \mod 4] = 0(D)$$

where the coefficients of $h^2(D)$ are binary, and the coefficients of $h^r(D)$ and the elements of $c^r(D)$ are elements of $\mathbb{Z}_4$. The notation $\{\psi(D)\}^T$ means that from the binary representation of every element $c_{i\ell} = 2c^1_i + c^2_i \in \mathbb{Z}_4$ in $a(D)$ the most significant bit $c^2_i \in \{0, 1\}$ is chosen. Let $1(D)$ denote the all-ones sequence. The codes defined by the parity-check equation are rotationally invariant if $h^r(D)[c^r(D) + 1(D)] \equiv h^r(D)e^{\beta'(D)} \mod 4$ for all $c^r(D)$, which requires that the coefficient sum of $h^r(D)$ must be $0 \mod 4$.

Code tables for fully rotationally invariant nonlinear trellis codes over two-dimensional QAM and $M$-PSK ($M = 4, 8, 16$) signal constellations are given in [88].

Alternative methods for the construction of rotationally invariant trellis codes are discussed in [10] and [101].

**APPENDIX III**

**DISCRETE-TIME SPECTRAL FACTORIZATION**

The spectral factorization $\tilde{Q}(f) = \tilde{H}^*(f)A^2 \tilde{H}(f)$ given by (5.19) may be obtained by expressing $\log \tilde{Q}(f)$ as a Fourier series

$$\log \tilde{Q}(f) = \sum_{\ell} \alpha_\ell e^{-j2\pi f/T}$$

whose coefficients are given by

$$\alpha_\ell = T \int_{1/T} \log \tilde{Q}(f)e^{j2\pi f/T} df.$$  (AIII.1)

The Paley–Wiener condition ensures that this Fourier series exists. Grouping the terms with negative, zero, and positive indices of this Fourier series yields the desired factorization

$$\log \tilde{H}^*(f) = \sum_{\ell>0} \alpha_\ell e^{-j2\pi f/T}$$

$$\log A^2 = \alpha_0 = T \int_{1/T} \log \tilde{Q}(f) df$$

$$\log \tilde{H}(f) = \sum_{\ell<0} \alpha_\ell e^{-j2\pi f/T}.$$  (AIII.3)

In particular, this yields (5.20).

To obtain an explicit expression for the coefficients of $h(D)$, define the formal power series

$$\psi(D) = \sum_{\ell \geq 0} \alpha_\ell D^\ell$$

where $D$ is an indeterminate. Because $\tilde{H}(f) = h(e^{-j2\pi f/T})$, we have

$$\psi(D) = \log h(D),$$

Taking formal derivatives yields

$$\psi^{(i)}(D) = h^{(i)}(D)/h(D)$$

or, equivalently,

$$h^{(i)}(D) = \psi^{(i)}(D)h(D).$$  (AIII.7)

Repeated formal differentiation of (AIII.7) yields the recursive relation for $k \geq 1$

$$h^{(k)}(D) = \sum_{i=0}^{k-1} \binom{k-1}{i} \psi^{(k-i)}(D)h^{(i)}(D),$$

Finally, noting that

$$h^{(k)}(0) = (k!)h_k$$

we obtain the explicit expressions

$$h_0 = 1$$

$$h_k = \sum_{i=0}^{k-1} \frac{k-i}{k} h_{i+k}, \quad k \geq 1.$$  (AIII.10)

This shows that $h(D)$ is causal and monic, and that $h(D)$ is uniquely determined by $q(D)$. For a proof that $h(D)$ is minimum-phase, see [83].
ACKNOWLEDGMENT

G. D. Forney wishes to acknowledge the education, stimulation, and support offered by many mentors and colleagues over the past 35 years, particularly R. G. Gallager, J. L. Massey, J. M. Wozencraft, A. Kohlenberg, S. U. Qureshi, G. R. Lang, F. M. Longstaff, L.-F. Wei, and M. V. Eyuboglu.

G. Ungerboeck wishes to thank K. E. Drangeid, Director of the IBM Zurich Research Laboratory, for his essential support of this author’s early work on new modem concepts and their implementation. He also gratefully acknowledges the stimulation and encouragement that he has received throughout his career from numerous colleagues, in particular R. E. Blahut, G. D. Forney, J. L. Massey, and A. J. Viterbi.

REFERENCES


