In short, to find the capacity of the interleaved channel with perfect side information, average the channel capacity expression conditioned upon a given state against the distribution of states. This seems a little known result and applies for all cases (hard or soft decisions, different channels with memory) provided the compatibility assumption holds. It does for the fading channel under study and for the symmetric p.d.f. situations that arise here, since the equiprobably input distribution optimizes mutual information in the binary and M-ary orthogonal cases studied thus far. No similar interpretation is, however, available for R_0.

Example 4.12 Capacity for Binary Noncoherent FSK with Perfect Side Information

To illustrate this averaging principle, we use the binary FSK example. With hard decisions and side information, the crossover probability is

$$\epsilon(a) = \frac{1}{2 + (a^2 E_s/N_0)} \tag{4.6.16}$$

and the conditional capacity is $C[\epsilon(a)] = 1 - h_2[\epsilon(a)]$. By (4.6.15), we have

$$C_{\text{PSI, hard}} = \int_0^\infty 2ae^{-a^2} C[\epsilon(a)]\, da. \tag{4.6.17}$$

Similarly for the unquantized channel, the conditional capacity is

$$C(s = a) = 1 - \log_2 \int_{y_0} \int_{y_1} \text{Rician } (y_0)\, \text{Rayleigh } (y_1) \log\left[1 + \frac{I_0(\mu y_1/\sigma^2)}{I_0(\mu y_0/\sigma^2)}\right] dy_0\, dy_1. \tag{4.6.18}$$

Numerical calculations presented in Figure 4.6.4 show the hard-decision capacity result for both perfect side information and no side information. The difference is around 3 dB at higher SNR, but at lower rates, where the minimum energy solution exists, the penalty for lack of side information is smaller, 1 to 2 dB. Similar conclusions pertain for the unquantized case.

Unquantized Demodulation, No Side Information (NSI)

If amplitude side information is not available, the decoder must build its metric purely on the basis of $\mathbf{y}$ and uses the channel p.d.f. $f(\mathbf{y}|x)$ without further conditioning on the amplitude a. This p.d.f. can be obtained by averaging:

$$f(\mathbf{y}|x) = \int_0^\infty f(\mathbf{y}|x, a) f_A(a)\, da. \tag{4.6.19}$$

For the case of noncoherent detection on the Rayleigh channel, calculation will show that the optimal decoder is a square-law combiner, that is,

$$\lambda(\mathbf{y}_j, x_j = m) = y_{j_m}^2, \tag{4.6.20}$$

which is outlined in Exercise 4.6.3. The corresponding channel capacity is

$$C_{\text{NSI, UQ}} = \int_{y_0} \int_{y_1} f(y_0, y_1|x_0) \log\left[\frac{2f(y_0, y_1|x_0)}{f(y_0, y_1|x_0) + f(y_0, y_1|x_1)}\right] dy_0\, dy_1 \tag{4.6.21}$$

where the conditional p.d.f.'s are obtained using numerical integration of (4.6.19).

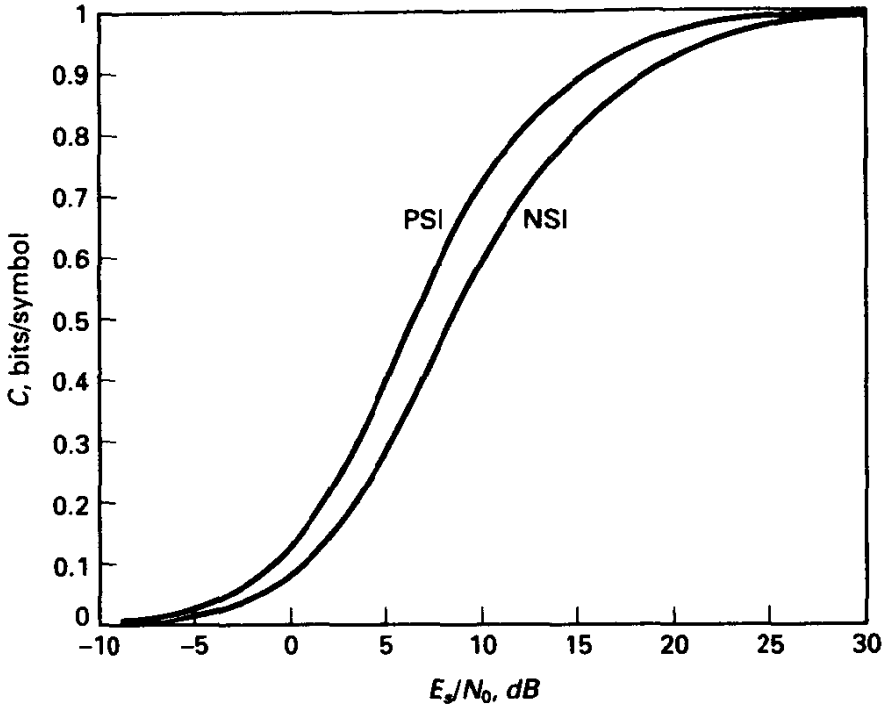

Figure 4.6.4 Effect of side information on Rayleigh channel capacity, binary FSK, noncoherent detection, hard decisions.

4.6.2 *M*-ary Noncoherent Transmission on the Rayleigh Channel

We now consider M-ary extensions of this development, here treating the M-orthogonal noncoherent detection case and the coherently detected QAM case in Section 4.6.3. Noncoherent detection is an important practical case for fading channels, primarily due to the difficulty in maintaining phase coherence with the channel during the deep fading intervals. We will consider M-ary FSK, or more generally M-ary orthogonal signals, with noncoherent detection. Again we invoke a perfect interleaving mechanism and assume the two limiting cases on side information: no side information (NSI) or perfect side information (PSI).

For M-ary FSK, we draw on the analysis of Stark [21], who derives capacity and R_0 for several cases on a general Rician fading channel, which incorporates the Rayleigh channel as a special case.[9] We will not repeat the details, but summarize the results for unquantized demodulation (the hard-decision case is a rather straightforward extension of previous methods and is described in detail in [21]).

With no side information available, the channel capacity is given by, due to symmetry of the p.d.f.'s and the result that the equiprobable input maximizes mutual

[9]Stark works with the squares of the variables we adopt, having chi-squared p.d.f.'s, but the end results are the same in either case.

information,

$$C_{\text{NSI}} = \int_{\mathbf{y}} f(\mathbf{y}|x_0) \log_M \left[\frac{f(\mathbf{y}|x_0)}{f(\mathbf{y})} \right] d\mathbf{y}$$
$$= 1 - \int_{\mathbf{y}} f(\mathbf{y}|x_0) \log_M \left[1 + \sum_{j=1}^{M-1} \frac{f(\mathbf{y}|x_j)}{f(\mathbf{y}|x_0)} \right] d\mathbf{y}, \tag{4.6.22a}$$

where $f(\mathbf{y}|x_0)$ is obtained by averaging (see Exercise 4.6.3):

$$f(\mathbf{y}|x_0) = \int f(\mathbf{y}|x_0, a) f_A(a)\, da$$
$$= \left[\prod_{j=0}^{M-1} \frac{y_j}{\sigma^2} e^{-y_j/2\sigma^2} \right] \frac{\exp\left(\frac{y_0^2}{2\sigma^2} \frac{\mu^2}{\mu^2+2} \right)}{1 + (\mu^2/2)} \tag{4.6.22b}$$

We point out the base M logarithm used in (4.6.22a), following [21]. With this definition all relevant rates, capacities, and the like, will lie between 0 and 1.

We can write (4.6.22a) as

$$C_{\text{NSI}} = 1 - \int \cdots \int \prod_{j=0}^{M-1} \frac{y_j}{\sigma^2} e^{-y_j^2/2\sigma^2} \frac{\exp\left(\frac{\mu^2 y_0^2}{2\sigma^2} \frac{\mu^2}{\mu^2+2} \right)}{1 + (\mu^2/2)}$$
$$\cdot \log_M \left[1 + \sum_{j=1}^{M-1} \exp\left(\frac{\left(y_j^2 - y_0^2 \right)}{2\sigma^2} \left[\frac{\mu^2}{\mu^2 + 2} \right] \right) \right] dy_0\, dy_1 \ldots dy_{M-1}. \tag{4.6.23}$$

This expression is left to numerical integration, but is very time consuming when M increases. Stark develops an equivalent multidimensional integral involving the squares of variables here.

When amplitude side information is available, capacity is given by a similar expression, with the channel amplitude added to the probability statements. Specifically, the conditional p.d.f. required is calculated by

$$f(\mathbf{y}, a|x_i) = f(\mathbf{y}|x_i, a) f(a|x_i) = f(\mathbf{y}|x_i, a) f(a). \tag{4.6.24}$$

Capacity is then found with a modification of (4.6.22):

$$C_{\text{PSI}} = 1 - \int_a \int_{\mathbf{y}} f(\mathbf{y}, a|x_0) \log_M \left[1 + \sum_{j=1}^{M-1} \frac{f(\mathbf{y}, a|x_j)}{f(\mathbf{y}, a|x_0)} \right] d\mathbf{y}. \tag{4.6.25}$$

This is again a task for numerical integration. Analogous to Example 4.12, this capacity measure can be interpreted as the average, over the fading distribution, of the capacity for a fixed-gain channel, although the latter expression is not simple in this situation.

For R_0, we appeal to the generic expression (4.3.32c), merely needing the appropriate p.d.f's. The symmetry of the M-ary orthogonal situation gives that

$$R_{0_{\text{NSI}}} = 1 - \log_M [1 + (M-1) D_{\text{NSI}}] \quad M\text{-ary units/symbol}, \tag{4.6.26}$$

where

$$D_{\text{NSI}} = \left[\int_0^\infty f(\mathbf{y}|x_0)^{1/2} f(\mathbf{y}|x_1)^{1/2} d\mathbf{y}\right]^2. \tag{4.6.27}$$

For the case of Rayleigh fading, the required p.d.f.'s can be found by conditioning on amplitude, giving a product of Rician and Rayleigh p.d.f.'s, and then averaging with respect to the Rayleigh distribution for amplitude. Doing so, with substitution in (4.6.26) gives, after some work,

$$R_{0_{\text{NSI}}} = 1 - \log_M [1 + (M-1)4p(1-p)], \tag{4.6.28a}$$

where

$$p = \frac{1}{2 + (E_s/N_0)}, \tag{4.6.28b}$$

which is just the probability of error for noncoherent binary FSK!

With side information, we can use the expression (4.6.26) with the definition that [21]

$$D_{\text{PSI}} = \int_0^\infty 2ae^{-a^2} D(a)\, da \tag{4.6.29a}$$

and

$$D(a) = R_0 \quad \text{for binary FSK on the AWGN channel}$$

$$= \left[\int_0^\infty \frac{y}{\sigma^2} \exp\left(-\frac{y^2 + a^2\mu^2/2}{2\sigma^2}\right) I_0^{1/2}\left(\frac{a\mu y}{\sigma^2}\right) dy\right]^2 \tag{4.6.29b}$$

This may be recognized from the R_0 calculation for noncoherent binary FSK on a fixed-gain channel in Section 4.5.

As observed by Stark, the expressions for R_0 are easily calculated for different values of M once the appropriate value of D is obtained. This is unfortunately untrue of the capacity calculation and demonstrates one appealing feature of R_0 analysis.

Having obtained the various measures for C or R_0, the minimum energy to noise-density ratio required to keep $R \leq C$ (or R_0) is found by solving, at a given rate,

$$r = C\left(\frac{E_s}{N_0}\right) = C\left(\frac{r \log_2 M E_b}{N_0}\right). \tag{4.6.30}$$

(Recall that our definition of rate here is in M-ary symbols per M-ary channel symbol and is between 0 and 1; the normalized rate r and rate in bits per symbol, R, are related by $R = r \log_2 M$.)

Figure 4.6.5 provides the loci of minimum energy for unquantized noncoherent demodulation of binary FSK on the interleaved Rayleigh channel, with and without side information, for both capacity and R_0. We note that, as with noncoherent demodulation on the AWGN channel, there exist optimum rates in the minimum energy sense, although the optimal rates are smaller than for the AWGN case, and there is somewhat greater dependence on the code rate.

Finally, and perhaps most interesting, is that if the optimal rate is selected then Rayleigh fading costs less than 1.5 dB in the capacity sense, relative to the minimum

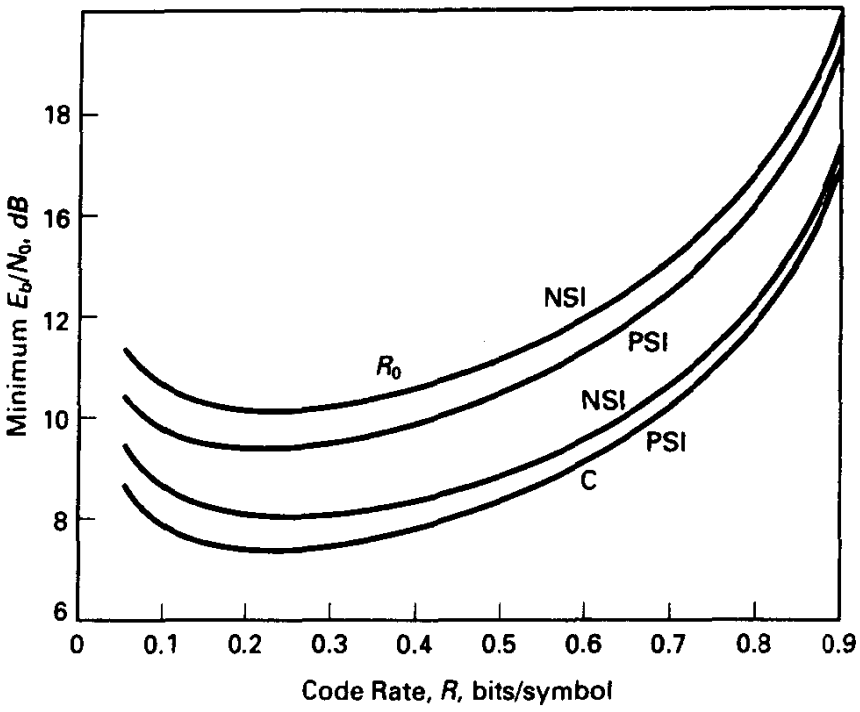

Figure 4.6.5 Minimum energy loci versus rate, Rayleigh channel, binary orthogonal signals, noncoherent detection.

E_b/N_0 for the AWGN channel, as can be seen by comparing Figure 4.5.14. If side information is available, the loss is even less, on the order of 0.6 dB.

R_0 results at least can be easily converted to other values of M as follows [21]: if $e_2(r)$ represents the minimum E_b/N_0 consistent with $R_0 > r$ for *binary* FSK, then $e_M(r)$, the minimum E_b/N_0 allowed by R_0 for M-ary transmission at normalized rate r, is

$$e_M(r) = \frac{1}{r \log_2 M} e_2\left[1 - \log_2\left(\frac{M^{1-r} - 1}{M - 1} + 1\right)\right]\left[1 - \log_2\left(\frac{M^{1-r} - 1}{M - 1} + 1\right)\right]. \tag{4.6.31}$$

This derives from proper normalization of rate and energy when alphabet size changes. Ryan and Wilson [22] illustrate this dependence on M, redrawn in Figure 4.6.6, showing that as M increases beyond 2 the optimal rate drops still further, and the corresponding allowed bit energy-to-noise density ratio diminishes as well. It can be shown that the energy efficiency is approaching, for large M, that of coherent M-orthogonal transmission, paralleling a result of Chapter 3 that noncoherent detection is nearly as efficient as coherent for large alphabet sizes.

The principal messages of this section are as follows:

1. Low-rate codes, in M-ary information symbols per M-ary channel symbol, are warranted on the Rayleigh channel provided significant bandwidth expansion is acceptable.
2. The penalty relative to the AWGN channel is minimal for intelligently designed codes.

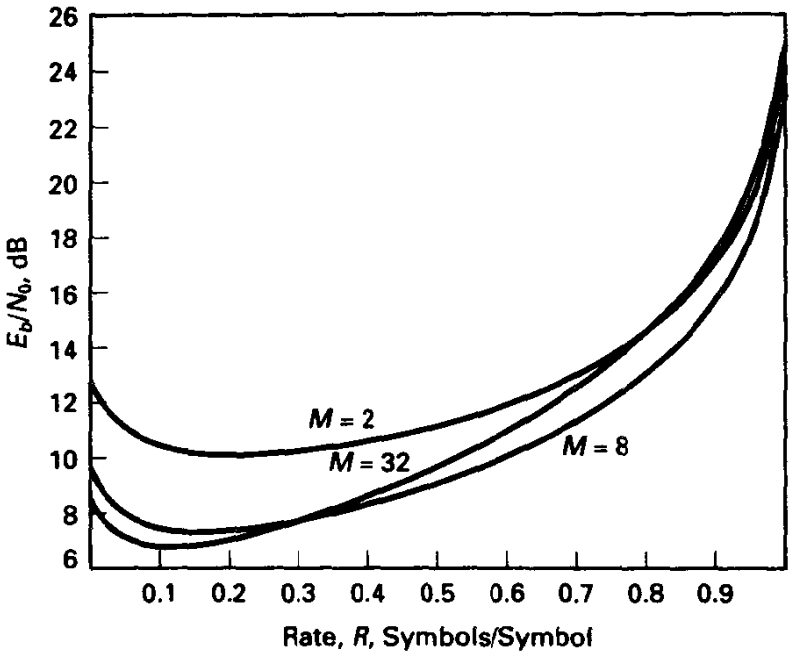

Figure 4.6.6 Minimum energy loci versus rate, Rayleigh channel, M-ary orthogonal signals, noncoherent detection.

4.6.3 Channel Capacity for Bandwidth-efficient Modulation on the Rayleigh Channel

To conclude our study of coding potential for the Rayleigh fading channel, we return to bandwidth-efficient signal sets, such as QAM, and assume coherent detection. The transmitted signal is represented as an N-dimensional signal with signal-space coordinates $\mathbf{s}_i = (s_0, s_1, \ldots, s_{N-1})$. We assume fading is *flat* across the signal bandwidth; that is, all signal-space coordinates of the transmitted signal are identically affected by fading. Thus, the received signal vector is

$$r_{mj} = a_j s_{mj} + n_{mj}, \qquad m = 0, 1, \ldots, N-1, \tag{4.6.32}$$

where j represents the time index, and $\{n_{mj}\}$ is a set of independent, zero-mean Gaussian random variables with variance $N_0/2$. Given a signal vector and the fading amplitude, the receiver outputs are independent and spherically symmetric again; the nuance here is the additional random variable due to fading; this fading expands or shrinks all coordinates of signal space equally.

Provided that the channel is fully interleaved and that the receiver possesses side information regarding the amplitude of the fading process during any symbol interval[10] (and this must be deinterleaved as well), we can employ the theory of the previous subsection to write the channel capacity as

$$C^*_{\text{Rayleigh}} = E_A[C^*(A)], \tag{4.6.33}$$

[10]It is necessary anyway for the demodulator to determine signal strength in order to perform demodulation of QAM sets.

where $C^*(A)$ is the mutual information under an equiprobable input selection, conditioned on a fade amplitude A. Therefore,

$$C^*_{\text{Rayleigh}} = \sum_{i=0}^{M-1} \frac{1}{M} \int_0^\infty 2ae^{-a^2} \int_{\mathbf{y}} f(\mathbf{y}|\mathbf{s}_i, a) \log \left[\frac{f(\mathbf{y}|\mathbf{s}_i, a)}{f(\mathbf{y}|a)} \right] d\mathbf{y}\, da. \tag{4.6.34}$$

In this expression the conditional p.d.f.'s are two dimensional, circularly symmetric p.d.f.'s centered at $a\mathbf{s}_i$, with variance $N_0/2$ in each coordinate.

Figure 4.6.7 illustrates the capacity for the Rayleigh channel and for the AWGN channel for 16-QAM transmission, which is typical of other cases. Again, the fading penalty is modest, vanishing as the code rate in bits per symbol decreases. On the AWGN channel, such analyses suggest that coding to send one bit per symbol less than $\log_2 M$ provides virtually minimum energy design. Here, the suggestion is that smaller rates are desired if the minimum energy solution is sought for the Rayleigh channel. Said another way, for communicating, say, 2 bits/symbol on this channel, it may be appropriate to use a 16-point constellation in two dimensions, having the same bandwidth properties. R_0 analysis would lead to similar conclusions—principally that Rayleigh fading need not imply a huge energy penalty.

In a less constraining manner, Ericsson [23] analyzed the capacity of the interleaved Rayleigh channel with side information subject only to an energy constraint on N-dimensional signals, and in particular not assuming a particular modulator set such as QAM signals selected with equal probability. The finding is that

$$C_{\text{Rayleigh}} = E_A \left[\frac{1}{2} \log \left(1 + A^2 \frac{2E_s}{NN_0} \right) \right] \quad \text{bits/dimension,} \tag{4.6.35}$$

where the expectation is with respect to the fading amplitude A. (This is consistent with our recent discussion for capacity with side-information present.) Equation (4.6.35) is actually very general, holding for N-dimensional sets with arbitrary fading statistics, provided that coherent detection is performed and that perfect interleaving is achieved. One implication of (4.6.35) is that if energy is spread thinly over many signal-space dimensions (although not the case of interest here), then capacity approaches that of the nonfading channel.

In the case of Rayleigh fading and two-dimensional sets, (4.6.35) becomes

$$C_{\text{Rayleigh}} = \int_0^\infty 2ae^{-a^2} \left[\log \left(1 + A^2 \frac{E_s}{N_0} \right) \right] da, \quad \text{bits/dimension} \tag{4.6.36}$$

which can be evaluated numerically. The result is shown in Figure 4.6.8 versus E_s/N_0. We observe that the unconstrained capacity is superior to that of a discrete set as SNR increases, and the departure occurs at lower SNR on the Rayleigh channel. With only an energy constraint on the input, the asymptotic results for low and high SNR show that for low SNR the capacity approaches that of the AWGN channel, $\log(1 + E_s/N_0)$, and at high SNR the capacity penalty is 2.5 dB. These results are developed in Exercise 4.6.6.

Ericsson also determined the R_0 parameter for this same channel model with very similar findings. When E_N/N_0 is small, the R_0 value per signal dimension is essentially that obtained for the nonfading channel. In particular, this R_0 value is asymptotically half the channel capacity in this regime, as we saw earlier for the nonfading channel. However, as the signal-to-noise ratio per dimension increases, R_0 for the Rayleigh channel

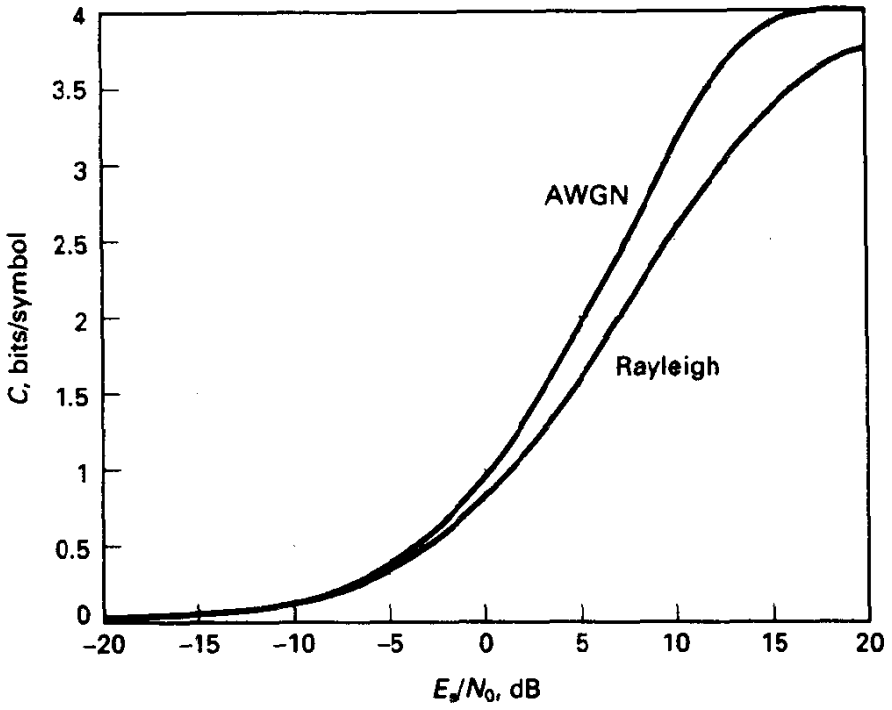

Figure 4.6.7 Capacity for 16-QAM, AWGN and interleaved Rayleigh channel.

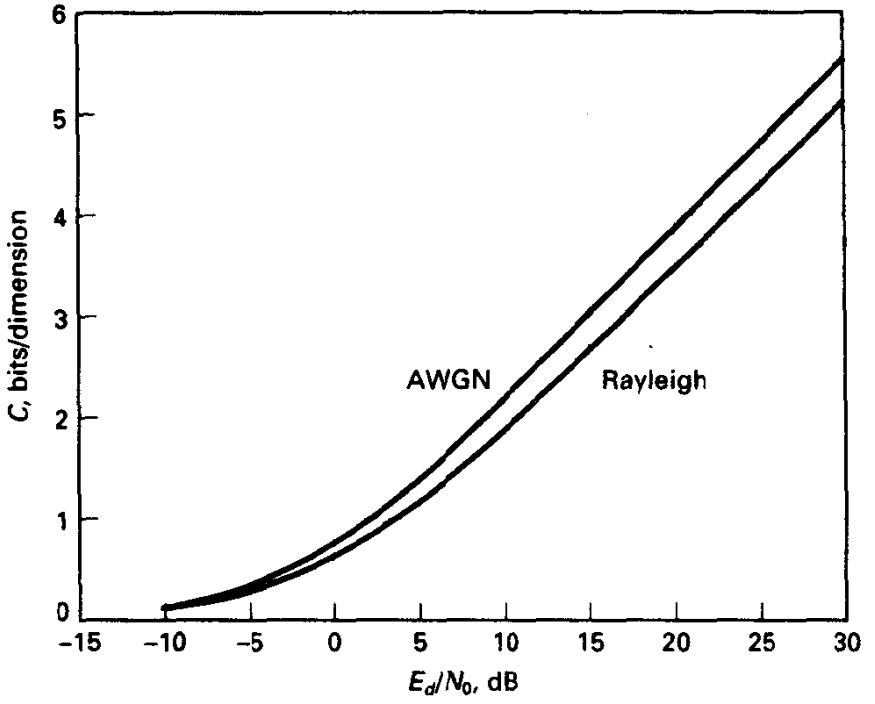

Figure 4.6.8 Capacity per dimension for AWGN and interleaved Rayleigh channels, energy constraint only on input.

experiences a similar degradation, slightly larger in fact than capacity does. Specifically, at $R_0 = 1$ bits/dimension (two bits/two dimensions), the increase in average energy necessary over the nonfading channel is about 2 dB.

We may argue, correctly, that without side information the capacity in bits per symbol increases for the same fading description when we do not interleave, although not by large amounts. We will not pursue this point further, however, since taking full advantage of this requires coding techniques with extremely long memory length. Interleaving remains a viable practical approach on the fading channel.

4.7 FURTHER STUDIES ON CODING POTENTIAL[11]

To conclude the chapter on the potential of channel coding, we study three case studies that involve new channel models, relevant to certain communication engineering situations, and ones that produce some curious results for C and R_0. The channels are (1) the ideal photon counting channel and (2) block interference channels.

4.7.1 Photon Counting (or Direct Detection) Optical Communication

We suppose transmission of information is accomplished by amplitude-modulating an optical source of energy, perhaps a semiconductor laser, having a frequency f typically around $4 \cdot 10^{14}$ Hz, or a wavelength around 0.75 micrometer (μm). Digital information is communicated by breaking a basic signaling interval of duration T_s seconds into slots of width Δt, where ordinarily $\Delta t \ll T_s$, and impressing a set of M modulation patterns (laser intensity patterns) defined on the interval T_s seconds. This forms the modulation scenario, and the inputs to this modulator may be coded to enhance the communication efficiency.

The optical signal travels through free space or a fiber medium and arrives at the detector with *average* power P_r watts. The optical signal will in reality be processed by conversion to an electrical signal by the photoelectric effect; that is, optical photons create hole–electron pairs in a semiconductor, which in turn produce current in the presence of a voltage bias. Such processes, whether occurring in a photomultiplier tube or an avalanche photodiode, are intrinsically noisy. The process of one pair creating many so that currents become suitably large is a random phenomenon. Dark currents exist as well, when no incident radiation is present.

Let's imagine nonetheless that we have available an ideal detector that is able to count the number of photons incident on our optical receiver in every time slot and report this integer-valued count, called $N_i, i = 0, 1, \ldots$. This can be shown to be a sufficient statistic for the problem assuming that the incident wave induces a Poisson stream of photons, with rate dependent on the laser's modulation state. We call such a detector an ideal photon counting detector, also known as direct detection because the signal

[11] No loss of continuity occurs in omitting this section.

processing is on the optical signal directly, rather than employing heterodyne techniques to convert the signal to lower frequencies.

The underlying source of interesting effects on this channel is the quantum nature of the problem; for example, even though the laser intensity in a given slot is a positive real number, the number of observations is a Poisson r.v. with parameter determined by the laser intensity, and it is possible that no photon counts will be registered despite the laser intensity being positive. In essence, the variability of the signal, when observed, is the source of message uncertainty. We will further assume that there is no background radiation to contend with, which can be a good approximation when aperture stops and optical bandpass filters are utilized. The modulation and demodulation scenario is illustrated in Figure 4.7.1.

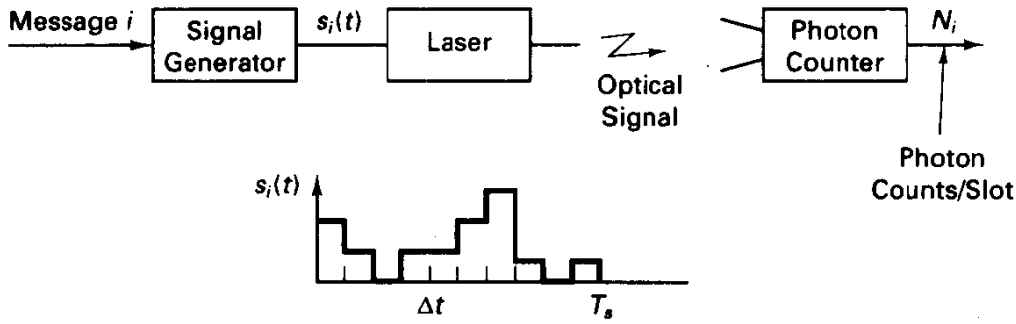

Figure 4.7.1 Optical signaling model with photon counting.

Since each photon carries hf joules, the average number of photons available per second is

$$\lambda = \frac{P_r}{hf} \quad \text{photons/second.} \tag{4.7.1}$$

If the available energy to a single interval is completely allocated to one slot, for example, the intensity during this slot is

$$\lambda_s = \frac{P_r}{hf}\frac{T_s}{\Delta t} \quad \text{photons/second,} \tag{4.7.2}$$

also implying that the peak laser power is much greater than the average power, a possible technological difficulty.

In studying this scenario, Pierce [24] first considered ideal amplification of a signal of frequency f observed in a black-body radiation field with temperature T and, citing Gordon [25], observed that the equivalent noise power spectral density is

$$G_n(f) = \frac{hf}{2(e^{hf/kT} - 1)} + \frac{(G-1)hf}{2} \quad \text{W/Hz (two-sided),} \tag{4.7.3}$$

where G is the ideal amplifier power gain. (An ideal amplifier in this context is one that increases the intensity of the radiation by G units.)

For microwave and lower operating frequencies, or more precisely when the energy per photon is relatively small and $hf \ll kT$, the noise model appears as an additive white Gaussian noise contamination with noise level $N_0/2 = kT/2$. The channel capacity for

the unlimited bandwidth AWGN channel is related to received power and noise level by (Section 2.9)

$$C_{\text{AWGN}} = \frac{P_r}{N_0} = \frac{P_r}{kT} \quad \text{nats/second} = \frac{1}{kT} \quad \text{nats/joule} = \frac{hf}{kT} \text{ nats/photon.} \tag{4.7.4}$$

[Note that we will frequently employ units of nats for information measures here; a nat is equivalent to $1/\log_e 2 = 1.44$ bits.]

On the other hand, if $hf \gg kT$, as a quick calculation will show pertains for optical frequencies, the second noise term, the quantum noise or partition noise, will dominate in (4.7.3), and if the amplifier gain G is large, the noise source will be Gaussian with spectral density $G_n(f) \approx Ghf/2$, assuming that $G \gg 1$. Assuming that the noise spectrum is roughly constant over a narrow range occupied by the optical signaling process[12] and noting that the average signal power is also increased by the factor G, the corresponding AWGN channel capacity expression applied here yields

$$C = \frac{P_r}{hf} \quad \text{nats/second} = \frac{1}{hf} \quad \text{nats/joule} = 1 \quad \text{nat/photon} \qquad \text{(ideal amplification).} \tag{4.7.5}$$

It is concluded that the capacity for ideally amplified signals is 1 nat per photon, or about 1.44 bits per photon, in the high-frequency regime.

Pierce went on to argue that if the optical detector is able to count photons directly in intervals of length Δt, rather than use an ideal optical amplifier, subsequent heterodyning, and typical radio-frequency signal processing, the channel capacity, in bits per photon, is

$$C_{ph} = \frac{hf}{kT} \quad \text{bits/photon} \qquad \text{(photon counting),} \tag{4.7.6}$$

and in principle if the detector is cooled to sufficiently low temperature and the optical background contributes no thermal photons, the capacity per photon is indefinitely large! [The same could be claimed in the low-frequency regime as $T \to 0$ from (4.7.4), although the actual value of hf/kT will be much smaller for a given T when operation is in the radio-frequency range, say 10^9 Hz.)

One way to achieve this efficiency is to use a laser source that is pulsed exactly once per T_s seconds for a duration of $\Delta t = T_s/M$; this is referred to as M-ary optical PPM. We let the intensity of the optical source be characterized by a mean rate of photons at the detector of λ_s, specified in (4.7.2). Thus, the mean number of photons per slot is $\lambda_s \Delta t = P_r T_s/hf$.

For a detection model, we assume the detector counts the number of arriving photons, N_i, in each slot. If one or more counts register in a given slot, the detector outputs the index of the time slot, $i = 0, 1, \ldots, M-1$. Notice that it is not possible for the wrong slot to register (because background radiation sources are neglected), but it is possible for no counts to register in the proper slot, due to the quantum nature of the problem. In this case the detector outputs an erasure symbol, say E. The probability of an erasure, is just the probability that a Poisson r.v. is zero when the Poisson parameter is $\lambda_s \Delta t$. This is

$$P(E) = 1 - e^{-\lambda_s \Delta t}. \tag{4.7.7}$$

[12]Note that even a 10 GHz modulation rate produces a narrowband signal centered at $4 \cdot 10^{14}$ Hz.

We thus have an M-input symmetric, discrete memoryless channel model for each basic signaling interval, corresponding to the M-ary erasure channel shown in Figure 4.7.2. The capacity in bits per channel use is found by computing the difference between $H(X)$ and $H(X|Y)$ under an equiprobable input model. The result is

$$C_{\text{PPM}} = (1 - e^{-\lambda_s \Delta t}) \log_e M \quad \text{nats/channel use} \tag{4.7.8a}$$

or

$$C_T = \frac{(1 - e^{-\lambda_s \Delta t}) \log_e M}{M \Delta t} \quad \text{nats/second.} \tag{4.7.8b}$$

Because the average number of photons per channel use, or per symbol, is $\lambda_s \Delta t$, we argue that the channel capacity, per photon, is

$$C_{ph} = \frac{(1 - e^{-\lambda_s \Delta t}) \log_e M}{\lambda_s \Delta t} \quad \text{nats/photon.} \tag{4.7.9}$$

For any fixed laser power constraint, which fixes $\lambda_s \Delta t$ through $\lambda_s \Delta t = P_r T_s$, we can make the capacity *per photon* arbitrarily large by letting M become large. In fact, the most efficient operation is obtained when $\lambda_s \Delta t = P_r T_s$ is small, in which case the capacity per photon approaches $\log_e M$ nats/photon (not nats/symbol!).

Technological limits eventually set in—we cannot transmit with arbitrarily small slot width, detectors do not have infinite bandwidth, and the channel medium may be time dispersive. Furthermore, holding the average power fixed while Δt shrinks means the peak laser power is becoming unbounded. Nonetheless, the fact remains that ideal photon detection on the optical channel, or any quantum-limited channel, has potentially very large capacity per unit of energy (photon). The most obvious penalty is the exponential increase in bandwidth attached to the increase in information efficiency.

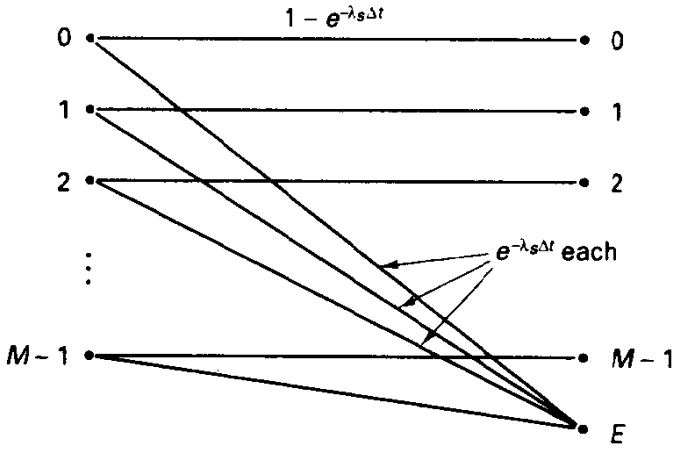

Figure 4.7.2 M-ary erasure channel model for optical PPM with photon counting.

McEliece [26] further studied the general intensity-modulated model, allowing the intensity in each time interval to be a real number λ_s, while the detector still reports the registered number of photon counts, possibly zero, in each time interval. Each interval is a new use of the channel. The channel capacity is developed by maximizing mutual information between input and output, under the constraint that we expend μ photons

per interval, on average. Under an average power constraint, μ becomes

$$\mu = \frac{P_r \Delta t}{hf}. \tag{4.7.10}$$

The mutual information between the real-valued input intensity μ and the output count N is upperbounded by the unconditional entropy of the channel output count. This entropy is in turn maximized over all nonnegative integer-valued random variable distributions having constrained mean m when [25]

$$P(N_i = n) = p^n(1-p), \qquad n = 0, 1, \ldots, \tag{4.7.11}$$

with $p = \mu/(1+\mu)$. (This says that the output counts should have a geometric distribution.) Substitution of this result into the expression for entropy of the output shows that channel capacity is upper-bounded by

$$C \leq \log(1+\mu) + \mu \log\left(1 + \frac{1}{\mu}\right) \quad \text{nats/channel use.} \tag{4.7.12}$$

Normalizing by μ still allows C_{ph} to be arbitrarily large (when μ shrinks toward zero). However, since the actual rate of transmission is bounded by C and $\log(1+\mu) \leq \mu$, we have that $R/\mu < 1 + \log(1 + 1/\mu)$. McEliece used this result to show that for *any* intensity modulation scheme, including of course PPM, the bandwidth expansion factor grows exponentially at the same time as C_{ph} shrinks toward zero.

Returning to M-ary PPM, the R_0 parameter can be easily established from the erasure channel model of Figure 4.7.2. Recalling that the definition of R_0 is, for the symmetric channel,

$$R_0 = -\log_e \sum_y \left[\sum_x \frac{1}{M} P(y|x)^{1/2} \right]^2, \tag{4.7.13}$$

we can determine that

$$\begin{aligned} R_0 &= -\log_e \left[\frac{M(1 - e^{-\lambda_s \Delta t})}{M^2} + e^{-\lambda_s \Delta t} \right] \\ &= -\log_e \left[e^{-\lambda_s \Delta t} + \frac{1 - e^{-\lambda_s \Delta t}}{M} \right] \quad \text{nats/symbol.} \end{aligned} \tag{4.7.14}$$

Taking the limit as $\lambda_s \Delta t \to 0$ shows that, in units of nats per photon, R_0 can be made arbitrarily close to $(M-1)/M \approx 1$ by lowering the laser intensity and utilizing arbitrarily short time slots! McEliece also shows that more general signaling, that is, allowing signaling in each slot, but with an average power constraint, produces at most $R_0 = 1$ nat/photon.

In contrast to results obtained on the AWGN channel, where C and R_0 did not differ by more than a factor of 2, we have here a puzzling contrast between the assessments offered by channel capacity and R_0 analysis: the former is unbounded, while the latter saturates at 1 nat/photon. The belief that R_0 represents a limit on reliable communication rate with feasible complexity would suggest that reliable throughput much in excess of a few bits per photon will be difficult.

Butman et al. [27] have studied the optical PPM channel from a different perspective, observing first that the usual communication concern is achieving a given throughput

R_T in *bits per second*, and thus the optimization should require that a given C_T (or R_0 in nats per second) be obtained. [Notice that if our only concern is maximizing information capacity per photon the maximum throughput C_T vanishes to zero by (4.7.8b)!] Furthermore, it is realistic to establish a bandwidth constraint or, equivalently, a minimum pulse width Δt. So the slot width Δt is specified, and the capacity in bits per second, given by (4.7.8b), is set equal to some target value, and we ask for the best value of M to maximize C_{ph}. In keeping the value of C_T fixed, while varying Δt, it is obvious that λ_s must vary, and this allows the best utilization of energy. Maximizing C_{ph} numerically leads to a result that depends on the specified Δt and the target throughput C_T. If, for example, $\Delta t = 1$ ns, and we seek a modest throughput of $C_T = 10^6$ nats/s, then the optimum M is about 2000, resulting in an energy efficiency of 6.5 nats/photon. For the same desired throughput and a pulse width $\Delta t = 10^{-12}$, which is present state of the art for laser pulse width, the optimal M is about $2 \cdot 10^6$, allowing communication with about 13.5 nats/photon. The plot of C_{ph} versus M for these conditions is shown in Figure 4.7.3.

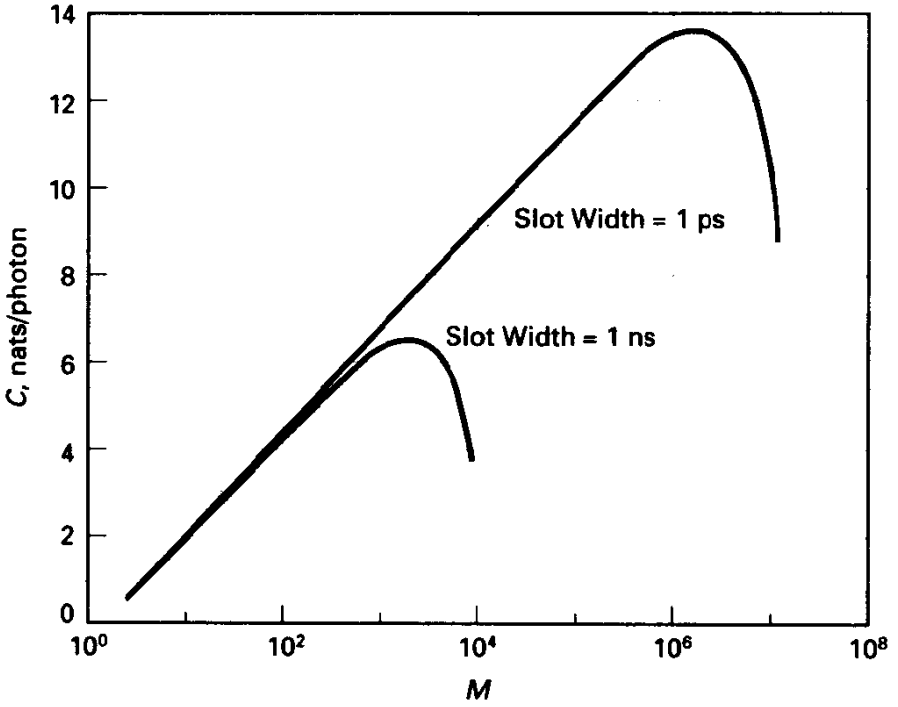

Figure 4.7.3 Capacity per photon, optical PPM, throughput = 10^4 nats/s.

In the latter case, if, through channel coding, we push the transmission rate up close to capacity, then the signaling interval is $10^{-12}(2 \cdot 10^6) = 2 \cdot 10^{-6}$ s, corresponding to an information rate of 2 nats/interval, despite the large alphabet size. (Lesh [28] later showed that if we adhere to this optimizing policy the capacity in nats per channel use is upper bounded by 2, approaching 2 as the bandwidth expansion factor $M/\log_2 M$ increases.) Recalling that the $M \approx 2 \cdot 10^6$, we find that the suggested code rate is about $2/\log_e(2 \cdot 10^6) \approx 0.14$. The bandwidth expansion factor would be $M/(R \log_2 M) \approx 7 \cdot 10^5$.

4.7.2 Block Interference Channels

McEliece and Stark [29] analyze a class of ***block interference channels*** wherein blocks of m channel symbols are affected in statistically identical, but memoryless, manner, constituting a special kind of channel with memory. On this particular channel, it is found that capacity increases with memory or the length of the interference interval, while R_0 actually drops!

An example, which will be studied shortly, is the situation where we can communicate by one of two BSCs: a good channel or a bad (high interference) channel with larger crossover probability, for a duration of m bits. Then, a new channel presents itself. In each condition, the channel behaves as a memoryless channel with fixed statistical description. The assumed mechanism on channel selection is Bernoulli trials; that is, that choice of a successive block's model is independent of other blocks. The probability that the channel is the good channel will be designated P_0.

Such channels are *phased* in the sense that the behavior is always in blocks of length m, but in [29] it is observed that this may be a good model for jamming or multiple-access channels that suffer from bursts of interference of fixed duration m or as a model for concatenated coding scheme (see Chapter 5), wherein an inner decoder failure presents the outer decoder with a poor *superchannel* for a fixed interval.

We are interested both in the channel capacity and the R_0 parameter that are measures forecasting the success of channel coding on this kind of channel. Two cases are analyzed: first, the decoder may possess side information on which channel state is in effect in a given block, or, conversely, the decoder may not have channel state information. In either case, the *encoder* does not know the channel state.

The communication process may be viewed as having superchannel inputs given by m-tuples over the channel input alphabet and with outputs given by m-tuples over the channel output alphabet. If side information is provided, the decoder is also supplied the index of the channel state, s, for the given block. In [29] it is assumed that the input probability distribution achieving capacity is the same for all channel states, referred to as a compatibility requirement, which holds in the examples considered.

By treating the communication process as a DMC at the m-tuple level, the capacity with side information, $\overline{C}(m)$, is readily found. Let $C(s)$ denote the per-symbol channel capacity for a channel in state $S = s$. Then, because the channel is memoryless within a block,

$$\overline{C}(m) = \frac{1}{m} \max_{P(\mathbf{X})} I(\mathbf{X}; \mathbf{Y}, S) \tag{4.7.15}$$

where $I(\mathbf{X}; \mathbf{Y}, S)$ is the mutual information between input vectors and the output collection $(\mathbf{Y}, S)$. This mutual information may be decomposed as [9, 29]

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}, S) &= I(\mathbf{X}; S) + I(\mathbf{X}; \mathbf{Y}|S) \\ &= I(\mathbf{X}; \mathbf{Y}|S) = E_S[I(\mathbf{X}; \mathbf{Y}^{(s)})], \end{aligned} \tag{4.7.16}$$

where $\mathbf{Y}^{(s)}$ is the m-tuple of channel outputs when state s is in effect and when the common (compatible) input distribution is used. [We used the assumption that $\mathbf{X}$ and S

are independent random variables and hence have zero mutual information in (4.7.16).] Now, we know that $I(\mathbf{X}; \mathbf{Y}^{(s)}) \leq mC(s)$, with equality if the input distribution achieves capacity in a given state and inputs are selected equiprobably. Thus, from (4.7.15) the channel capacity, with side information, is

$$\overline{C}(m) = \overline{C} = E[C(S)], \tag{4.7.17}$$

independent of block memory length! The expectation is over the set of channel states, and in simple terms, the channel capacity is just the average of the capacities of the component channels, each of which is readily found for each DMC.

Now, regarding the channel capacity in the absence of side information, $C(m)$, we first note that

$$C(m) \leq \overline{C}(m) = \overline{C}, \tag{4.7.18}$$

since

$$I(\mathbf{X}; \mathbf{Y}) \leq I(\mathbf{X}; \mathbf{Y}, S).$$

Furthermore,

$$I(\mathbf{X}; \mathbf{Y}, S) \leq I(\mathbf{X}; \mathbf{Y}) + I(\mathbf{X}, \mathbf{Y}; S) \tag{4.7.19}$$

with equality holding if Y, S are independent. Dividing by m and taking the limit as m grows gives

$$\lim_{m \to \infty} C(m) = \lim_{m \to \infty} \frac{1}{m} I(\mathbf{X}; \mathbf{Y}) = \overline{C} \tag{4.7.20}$$

since $\lim_{m \to \infty} I(\mathbf{X}, \mathbf{Y}; S)/m = 0$, provided S has finite entropy itself, which we assume. This reveals that the penalty for lacking channel state is small if the block length is sufficiently long. Intuitively, we may probe the channel every block with test symbols for each block to learn the channel state at the receiver, and this overhead can be made negligibly small as the block length increases.

If we again define R_0 in terms of two-codeword error probability averaged over an ensemble of codes, it is shown in [28] that, with side information supplied to the decoder,

$$\overline{R_0}(m) = -\frac{1}{m} \log_2 E_S \left[2^{-mR_{0,S}} \right], \tag{4.7.21}$$

where $R_{0,s}$ is the R_0 parameter for the channel in state s and where expectation is again over channel states. Thus, we do not average the individual channel R_0 parameters, as for capacity, but instead $\overline{R_0}(m)$ satisfies, from (4.7.21),

$$2^{-m\overline{R_0}(m)} = \sum_i P(S_i) 2^{-mR_{0,S_i}}. \tag{4.7.22}$$

Experience with various examples shows this *decreases* with m. Analysis of the no-side-information case shows that $R_0(m) \leq \overline{R_0}(m)$, as we would expect, and it is conjectured that the former is also strictly decreasing in m.

Let's interpret all this through a simple example of a block interference channel.

Example 4.13 Block Interference Channel with Two BSCs

Consider the example introduced at the outset, where the channel states have either zero error probability or error probability $\epsilon = \frac{1}{2}$. (This model is generalized in the exercises.) The probability that the good channel is in effect is $P_0 = 0.9$, and $P_1 = 1 - P_0$. The memory length is again m channel symbols.

The two channels have capacities of 1 and 0 bit/channel use, respectively, so the overall capacity with side information is $\overline{C} = P_0 = 0.9$ bits per symbol, independent of memory length. This result would certainly be anticipated if both encoder and decoder knew the channel state, for we can communicate with rate 1 during the good channel states and rate 0 elsewhere, attaining a throughput of 0.9, but it is perhaps surprising that this throughput can be achieved reliably without encoder clairvoyance.

The two channels also have respective $R_{0,i}$ values 1 and 0 bits/channel use, and so

$$\begin{aligned}\overline{R_0}(m) &= \frac{1}{m}\left[-\log_2[(1-P_0)2^0 + P_0 2^{-m(1)}]\right] \quad \text{bits/channel symbol} \\ &= -\frac{1}{m}\log_2[0.1 + (0.9)2^{-m} + 0.1].\end{aligned} \tag{4.7.23}$$

This *decreases* monotonically to zero as memory increases.

The corresponding expressions for the no-side-information case are calculated by taking an m-tuple view of the communication problem and are extracted from [29]:

$$\begin{aligned}C(m) = &[P_0 + (1-P_0)2^{-m}] \\ &- \frac{1}{m}[h_2(P_1 - P_1 2^{-m}) + (P_1 - P_1 2^{-m})\log(1 - 2^{-m})]\end{aligned} \tag{4.7.24}$$

and

$$R_0(m) = -\frac{1}{m}\log J_0, \tag{4.7.25}$$

where

$$J_0 = \frac{2^m - 1}{2^m} J + \frac{1}{2^m} \tag{4.7.26a}$$

and

$$J = \frac{2^m - 2}{2^m} P_1 + 2^{1-m}\left[P_1(P_1 + 2^m(1 - P_1))^{1/2}\right]. \tag{4.7.26b}$$

Table 4.2 provides the results of the calculation of these four quantities for selected block lengths of a power of 2. Notice that the opposing trends of capacity and R_0 result.

TABLE 4.2 CODING PARAMETERS VERSUS BLOCK LENGTH FOR BLOCK INTERFERENCE CHANNEL

m	$R_0(m)$	$\overline{R_0}(m)$	$C(m)$	$\overline{C}(m)$
1	0.48	0.86	0.71	0.90
4	0.45	0.67	0.80	0.90
16	0.21	0.21	0.87	0.90
64	0.052	0.052	0.89	0.90
256	0.013	0.013	0.90	0.90

These channel descriptors forecast the coding potential on this channel if we treat the problem as one of communicating m-tuples, with or without side information. On the other hand, the code alphabet for such a system would be large, $|A|^m$, where $|A|$ is the input alphabet size for the basic channel, and practical realization of the coding potential is unlikely. Interleaving is a common practice of scrambling or permuting the order of symbols output by an encoder, prior to transmission, when the channel has memory. The effect we hope to achieve is to produce a memoryless channel, as seen by the encoder/decoder, with a first-order probability structure given by the expected channel behavior averaged over its modes. (Permutation of channel symbols, then reassembling them in proper order, does not change the true capacity at all; it is just that the encoder/decoder proceed on the supposition that the channel is memoryless.) In other words we form a channel with memory $m = 1$, having the same marginal behavior as the real channel. The traditional justification is that codes designed for memoryless channels will have better performance and not be overwhelmed by bursts of more errors than the code can accommodate.

If side information is present (and interleaved/deinterleaved along with channel symbols), the fact that capacity is independent of memory length suggests that interleaving is not helpful, but, more surprisingly perhaps, not harmful either. On the other hand, when side information is lacking, the improvement of $C(m)$ with m suggests that interleaving may be harmful; this is, in fact, the usual case. For the two-state in Example 4.13, the corresponding memoryless channel would be a BSC with parameter $P_1/2$, having capacity

$$C_{\text{interleaved}} = 1 - h_2\left(\frac{P_1}{2}\right) \quad \text{bits/channel use} \tag{4.7.27}$$

which is less than the capacity of the channel with memory. R_0 analysis may suggest that interleaving is useful, with or without side information, since both R_0 parameters decrease with memory length m. The example of the two BSCs shows dramatic difference between values for $m = 1$ and $m = 256$. A clever "smart interleaving" scheme in [29] involving test patterns attached to each block and erasure of the block if the correct test pattern is not received leads to further improvement in the channel parameters, rendering the channel nearly as good as having full side information. The gain in R_0 is especially large. We should caution that the block memory assumed may not be correct or even that this particular model for channel memory is accurate.

McEliece and Stark offer a reconciliation of the paradoxical findings of C and R_0: rather than a measure of complexity (code block length) needed to achieve a given level of performance on a certain channel, R_0 should be regarded as a reciprocal measure of required *delay*, this delay being incurred through interleaving of codewords.

APPENDIX 4A1: DECODING ON CHANNELS WITH MEMORY

We imagine the scenario depicted in Figure 4A.1, wherein input vectors $\mathbf{x}$ are injected into a channel, and the aggregate channel output is a vector (of vectors perhaps) $\bar{\mathbf{y}} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$. We suppose that the channel has an internal mechanism called a state, and the sequence of states will be denoted $\mathbf{s} = (s_1, s_2, \ldots, s_n)$. We make the further

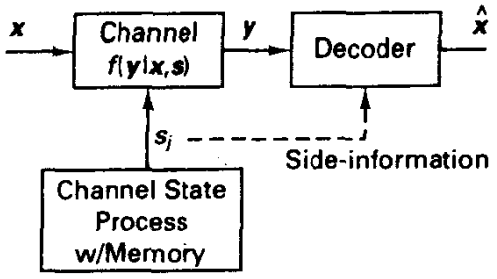

Figure 4A.1 Channel model for channel with memory.

assumption that, conditioned on knowledge of the state and the input sequence, the output sequence is memoryless; that is, we assume that

$$f(\bar{\mathbf{y}}|\mathbf{x}, \mathbf{s}) = \prod_{j=1}^{n} f(\mathbf{y}_j|x_j, s_j). \tag{4A1.1}$$

We are interested in optimal decoding on such a channel under one of two cases. First, we decode only observing $\bar{\mathbf{y}}$ (the no-side-information case). Second, we may have access to $\mathbf{s}$ as well (side information).

No Side Information

Given only $\bar{\mathbf{y}}$, the optimal decoder will maximize the *a posteriori probability* $P(\mathbf{x}_i|\bar{\mathbf{y}})$, which can be written as

$$P(\mathbf{x}_i|\bar{\mathbf{y}}) = \frac{f(\bar{\mathbf{y}}|\mathbf{x}_i)P(\mathbf{x}_i)}{f(\bar{\mathbf{y}})}. \tag{4A1.2}$$

If we assume that messages are equiprobable and note that the denominator in (4A1.2) is a scale factor that is constant for all hypothesized $\mathbf{x}_i$, we may use the maximum-likelihood rule:

$$\max_{\mathbf{x}_i} f(\bar{\mathbf{y}}|\mathbf{x}_i). \tag{4A1.3}$$

This looks familiar, but we should realize that this conditional p.d.f. must incorporate a channel state effect with memory. This can be expressed by

$$f(\bar{\mathbf{y}}|\mathbf{x}_i) = \int f(\bar{\mathbf{y}}|\mathbf{x}_i, \mathbf{s}) f_S(\mathbf{s}) d\mathbf{s}. \tag{4A1.4}$$

In general, the corresponding metric is difficult to formulate for arbitrary channels with memory, although Markov models for the state sequence simplify the results. More important, however, is the fact that unless the decoder operates over many modes of the channel, rather than just basing message decisions on a short span of activity, the *error performance* will be very poor. This is basically the rationale for interleaving, discussed in Sections 4.6 and 4.7; the channel is rendered memoryless by a scrambling operation. In that case at least, assuming that the state process is stationary, the per-symbol metric becomes

$$\lambda(\mathbf{y}_j x_j) = \log f(\mathbf{y}_j|x_j) = \log\left[\int_s f(\mathbf{y}_j|x_j, s) f_S(s)\, ds\right]; \tag{4A1.5}$$

that is, a state-averaged p.d.f. is employed.

Perfect Side Information

Suppose now that s is observed as well. Beginning with the MAP optimality rule, we have that

$$P(\mathbf{x}_i|\bar{\mathbf{y}}, \mathbf{s}) = \frac{f(\bar{\mathbf{y}}|\mathbf{x}_i, \mathbf{s})\, P(\mathbf{x}_i|\mathbf{s})}{f(\bar{\mathbf{y}}|\mathbf{s})}. \tag{4A1.6}$$

We assume that the message input x is independent of the channel state and, moreover, equiprobable. Thus, $P(\mathbf{x}_i|\mathbf{s})$ can be neglected in the metric. Also, the denominator does not depend on the index i and can be removed as a simple scale factor. Thus, because we assumed that $f(\bar{\mathbf{y}}|\mathbf{x}_i, \mathbf{s})$ was a product form, in the perfect-side-information case we find that the per-symbol metric should be

$$\lambda(\mathbf{y}_j, x_j; s_j) = \log f(\mathbf{y}_j|x_j, s_j). \tag{4A1.7}$$

In other words, we use a sequence of metrics appropriate for the various (known) channel states.

Despite the fact that the metric looks very familiar in this case, we should not infer that we have taken care of all the memory difficulties of the channel. In particular, unless we use codewords whose lengths span many channel decorrelation times, performance is dominated by the very poor channel modes, for example, deep fades on a Rayleigh fading channel. Thus, there may be good reason to employ interleaving on the channel *even* when side information *is* available. This interleaving will not change the metric when side information is used (after all, we are just rearranging the order of time), but performance can be drastically different.

BIBLIOGRAPHY

1. Wozencraft, J. M., and Kennedy, R. S., "Modulation and Demodulation for Probabilistic Coding," *IEEE Transactions on Information Theory*, July 1966.
2. Massey, J. L., "Coding and Modulation in Digital Communications," Proceedings of Zurich Seminar on Digital Communication, March 12–15, 1974.
3. Forney, G. D., Jr., "The Viterbi Algorithm," IEEE Proceedings, 1972.
4. Kailath, T., "Bhattacharyya Distance and Signal Selection," *IEEE Transactions on Information Theory*, 1967.
5. Omura, J. K., and Levitt, B. K., "Coded Error Probability Evaluation for Antijam Communication Systems," *IEEE Transactions on Communications*, vol. 30, pp. 896–903, May 1982.
6. Wozencraft, J. M., and Reiffen, B., *Sequential Decoding*. Cambridge, MA: MIT Press, 1961.
7. Jacobs, I. M., "Sequential Decoding for Efficient Communication from Deep Space," *IEEE Transactions on Communication Technology*, vol. COM-15, pp. 492–501, August 1967.
8. Gallager, R., "A Simple Proof of the Coding Theorem," *IEEE Transactions on Information Theory*, 1966.
9. Gallager, R., *Information Theory and Reliable Communication*. New York: Wiley, 1968.
10. Blahut, R., *Principles of Information Theory*. Reading, MA: Addison-Wesley.

11. Reiffen, B., "A Note on Very Noisy Channels," *Information and Control*, vol. 6, pp. 126–130, June 1963.

12. Viterbi, A. J., and Omura, J. K., *Principles of Digital Communications and Coding*, New York: McGraw-Hill, 1979.

13. Viterbi, A. J., "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 260–269, April 1967.

14. Jordan, K. L., "The Performance of Sequential Decoding in Conjunction with Efficient Modulation," *IEEE Transactions on Communications*, vol. 14, pp. 283–291, June 1966.

15. Miller, L. E., and Lee, J. S., "The Probability Density Function for the Output of an Analog Cross-Correlator with Correlated Bandpass Inputs," *IEEE Transactions on Information Theory*, vol. IT-20, pp. 433–440, July 1974. (See also Pawula?)

16. Lee, L. N., "On Optimal Soft-Decision Demodulation," *IEEE Transactions on Information Theory*, vol. IT-22, pp. 437–444, July 1976.

17. Ungerboeck, G., "Channel Coding with Amplitude/Phase Signals," *IEEE Transactions on Information Theory*, pp. 55–65, January 1982.

18. Forney, G. D., Jr., "Trellis Shaping," *IEEE Transactions on Information Theory*, vol. 38, pp. 281–300, March 1992.

19. Bello, P. A., "Channel Capacity for DPSK," unpublished memorandum, 1987.

20. Hagenauer, J., "Viterbi Decoding of Convolutional Codes for Fading and Burst Channels," Proceedings of International Zurich Seminar on Digital Communications, 1980.

21. Stark, W. E., "Capacity and Cutoff Rate of Noncoherent FSK with Nonselective Rician Fading," *IEEE Transactions on Information Theory*, vol. 33, pp. 1153–1159, November 1985.

22. Ryan, W. E., and Wilson, S. G., "Two Classes of Convolutional Codes over GF(q)," *IEEE Transactions on Communications*, January 1991.

23. Ericsson, T., "A Gaussian Channel with Slow Fading," *IEEE Transactions on Information Theory*, pp. 353–356, May 1970.

24. Pierce, J. N., "Optical Channels: Practical Limits with Photon Counting," *IEEE Transactions on Communications*, vol. 26, pp. 1819–1820, December 1978.

25. Gordon, J. P., "Quantum Effects in Communication," Proceedings of IRE, September 1962.

26. McEliece, R. J., "Practical Codes for Photon Communication," *IEEE Transactions on Information Theory*, vol. 27, pp. 393–397, July 1981.

27. Butman, S. A., Katz, J., and Lesh, J. R., "Bandwidth Limitations on Noiseless Optical Channel Capacity," *IEEE Transactions on Communications*, vol. 30, pp. 1262–1264, May 1982.

28. Lesh, J. R., "Capacity Limit of the Noiseless, Energy-Efficient Optical PPM Channel," *IEEE Transactions on Communications*, vol. 31, pp. 546–548, April 1982.

29. McEliece, R. J., and Stark, W. E., "Channels with Block Interference," *IEEE Transactions on Information Theory*, vol. 30, pp. 44–53, January 1984.

30. Viterbi, A. J., "Spread Spectrum Communications—Myths and Realities," *IEEE Communications Magazine*, vol. 4, pp. 11–18, May 1979.

31. Wilson, S. G,. and Parsons, R. D., "Polar Quantization for Coded M-PSK," *IEEE Transactions on Communications*. 1990.

EXERCISES

4.2.1. A famous block code, which will be studied in Chapter 5, is the (23, 12) code due to Golay. The encoder maps 12-bit binary information sequences to 23-bit binary code sequences.
(a) What is the size and rate of this code?
(b) Repeat for a code that maps 9-tuples from an alphabet of size 16 to length 15 sequences from the same alphabet.

4.2.2. Discuss the usage of the term *rate* of a code and when it is actually the information, or entropy, rate of the code in bits per symbol.

4.2.3. M-ary orthogonal signaling is to be analyzed for use with coded transmission with noncoherent detection of code symbols. The demodulator produces for each symbol the vector of M noncoherent matched filter outputs, designated $\mathbf{y}_j$, as described in Section 3.4.
(a) Show that the ML metric for scoring symbol $x_j = m$ at time j, given that the vector observation $\mathbf{y}_j$ is

$$\lambda(\mathbf{y}_j, x_j = m) = \log_e I_0\left(\frac{\mu y_{j_m}}{\sigma^2}\right),$$

where $\mu = E_s^{1/2}$ is the (known) signal strength parameter involved in the Rician p.d.f. at the receiver. In other words, the ML decision maker, in scoring a codeword, uses only the demodulator output corresponding to the hypothesized code symbols in the codeword and ignores the remaining data. This follows from writing the complete likelihood expression $f(\mathbf{y}|x_i)$ for a given transmission and then simplifying by eliminating terms common to all such expressions.
(b) By finding a series expansion for $I_0(x)$ for small argument x and then using an expansion of the logarithm for arguments near 1, show that for small SNR cases the square-law metric

$$\lambda(\mathbf{y}_j, x_j = m) = y_{j_m}{}^2$$

is near optimal and avoids use of special functions.

4.2.4. Let a set of binary codewords be given by (000), (110), (101), and (011). Map these onto vertices of a cube through the usual antipodal relation, '0' $\rightarrow -1$,'1' $\rightarrow 1$. Sketch or otherwise describe the decision regions in the manner of Figure 4.2.3 for ML and binary hard-decision decoding on the AWGN channel. Shade the regions as black, white, or gray (flip a coin) if the likelihoods are equal.

4.2.5. Show that for the memoryless binary erasure channel, regardless of erasure probability, the ML decision rule is "find the codeword that has zero Hamming distance to the observed vector in nonerased positions; if ties exist, make an equiprobable choice."

4.2.6. It is appropriate to remember that decision procedures that result from addition of AWGN to signals can be inappropriate if the noise mechanism is non-Gaussian. Suppose that the noise p.d.f. appears as in Figure P4.2.6, which might correspond to the mixture of two possible p.d.f.'s, a low-noise condition and a high-noise condition. In a two-codeword situation with codewords $(-1, -1, -1)$ and $(1, 1, 1)$, suppose that the received vector is $(-0.9, -1.1, 2.1)$. Based on independent Gaussian noise assumptions, the decision would be in favor of $(1, 1, 1)$, since the sum of observations exceeds 0. Which codeword has greater likelihood under the adopted noise model?

4.2.7. Suppose 16-QAM is to be employed for coding on an AWGN channel. One option would be to perform hard symbol decisions, with P_s given by the theory of Chapter 3. (Note that symbol error types are not equal here).

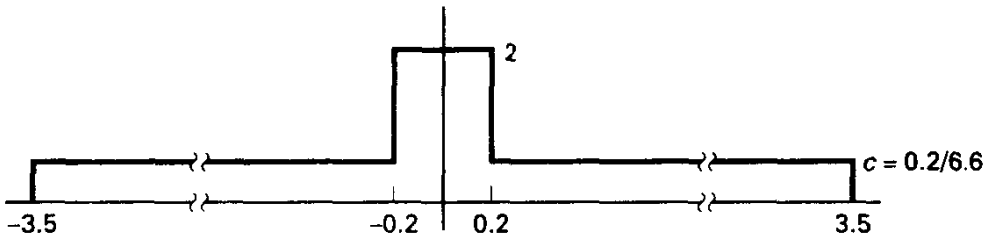

Figure P4.2.6 Non-Gaussian noise p.d.f.

(a) What per-symbol metric would you suggest for measuring demodulator output y against code symbol hypothesis x?

(b) If unquantized two-dimensional observations were produced by the demodulator, derive the optimal per-symbol metric for decoding. (Ans.: It is negative Euclidean distance, which can be simplified to a correlation plus an energy-related correction.)

4.3.1. Show the Bhattacharyya distance $d_B(\mathbf{x}, \mathbf{y})$ as formulated in Section 4.2 has two of the properties of a metric, that $d_B(\mathbf{x}, \mathbf{y}) \geq 0$ and $d_B(\mathbf{x}, \mathbf{y}) = d_B(\mathbf{y}, \mathbf{x})$, but does not satisfy the triangle inequality, $d_B(\mathbf{x}, \mathbf{y}) \leq d_B(\mathbf{x}, \mathbf{z}) + d_B(\mathbf{z}, \mathbf{y})$.

4.3.2. Demonstrate that Hamming distance between two n-tuples has the necessary properties of a metric as listed in Exercise 4.3.1.

4.3.3. Consider two codewords of length 10, $\mathbf{x}_1 = (0000000000)$ and $\mathbf{x}_2 = (1111111111)$. One codeword is transmitted by a memoryless binary erasure channel with erasure probability 0.1.

(a) Formulate the ML decision rule, invoking a tie-breaking rule if necessary. Show that the decoding error probability is $P_E = 5 \times 10^{-11}$. Calculate the Bhattacharyya upper bound on codeword error probability given by (4.3.4).

(b) Repeat for the Z-channel, with crossover parameter 0.1. Here the actual conditional error probabilities for the two codewords are different. Formulate the ML rule, and note that an error may occur only if the all-ones codeword is transmitted and 10 crossovers occur. Calculate the two-codeword bound by (4.3.4). Notice that the bound does not reveal the asymmetry of the problem.

(c) Change the codewords to (0000011111) and (1111100000). Notice that the Hamming distance is still 10. Performance will remain unchanged on the BEC, but what about the Z-channel?

4.3.4. Randomly generate a code of size $T = 4$ codewords of length 8, and thus rate $R = \frac{1}{4}$, to be used on a BSC with $\epsilon = 0.1$. To avoid natural biases toward good codes, use a coin or random number generator to produce the codewords. List the four words and find the Hamming distance between all pairs.

(a) Evaluate the probability of error for your code using a union bound. Note that some codewords may have poorer reliability than others.

(b) Formulate the random coding viewpoint for this set of parameters by using R_0 analysis to upper bound the error probability for the ensemble of such codes. Evaluate at $R = \frac{1}{4}$, and determine how your code fares.

(c) Slepian (1956) produced a code with $T = 4$ and $n = 8$ that is equivalent to $\mathbf{x}_1 = (00000000)$, $\mathbf{x}_2 = (01001111)$, $\mathbf{x}_3 = (10111100)$, and $\mathbf{x}_4 = (11110011)$. Note this code has at least five disagreements between every pair of vectors and is said to be double-error correcting. Using the approximation that any three or more error patterns in 8 bits causes a decoding error, find the error probability for this code.

4.3.5. Consider a pulse position modulation strategy for a laser communication system. In each symbol interval, we use a pulse of optical energy in one of M slots, each of length T_s/M seconds. At the receiver, we assume that background radiation and dark-current noise are

negligible, so we have a perfect photon-counting channel. Let λ be the mean number of photons in a slot. Then the probability of receiving zero photons when the pulse is present is $e^{-\lambda}$. The resulting channel model is the M-ary erasure channel shown in Figure P4.3.5.

(a) Calculate the channel capacity and R_0 in bits per interval.

(b) Investigate what happens as M becomes arbitrarily large.

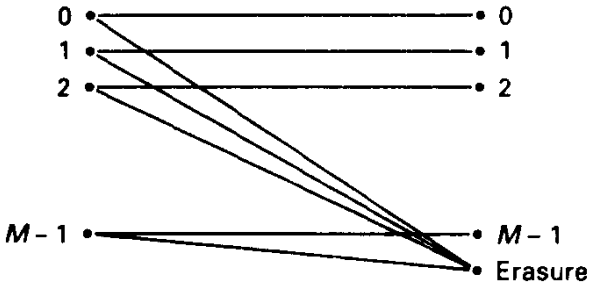

Figure P4.3.5 Channel model for optical PPM.

4.3.6. For the Z-channel of Exercise 4.3.3, calculate C and R_0. Note that an equiprobable input distribution does not maximize the related functions, so attack this numerically by evaluating the expressions as a function of a single parameter describing the binary input distribution.

4.3.7. Write a program to evaluate R_0 for a constellation built from the $\mathbf{D}_4$ lattice. Take 24 points of the form $(\pm 1, \pm 1, 0, 0)$ together with 8 points of the form $(\pm 2, 0, 0, 0)$ to construct a 32-ary four-dimensional signal set. Let E_s be the average energy per two-dimensional coordinate, and graph R_0 versus E_s/N_0. Determine the E_s/N_0 that is required to attain $R_0 = 4$ bits/four-dimensional symbol (or 2 bits/two-dimensional symbol), and compare with the corresponding number for the 8-QAM (box) two-dimensional constellation. This may demonstrate a modest coding advantage when using higher-dimensional signal sets. (Note that this corresponds to "constellation doubling," since we only require 16 four-dimensional signals to send 4 bits/four dimensions. The peak energy of the four-dimensional design is slightly less as well, for a given average energy.)

4.4.1. For the binary symmetric channel with crossover probability 0.1, assume an input symbol distribution that is equiprobable, and compute $E_0(\rho, P)$ as given by (4.4.18). Plot as a function of ρ between 0 and 1. By graphing tangent lines and by visual inspection of your sketch, verify the properties of (4.4.21). Sketch the line ρR for $R = \frac{1}{3}$, and graphically determine the value of ρ that maximizes the Gallager exponent at this rate. What is the value of $E(R)$ at this rate?

4.4.2. Consider the BSEC shown in Figure P4.4.2. Calculate the channel capacity C and the channel parameter R_0. Make a plot of $E_0(\rho, P)$ versus ρ for the equiprobable input condition, and label this plot with the quantities C, R_0, and the critical rate R_{cr}.

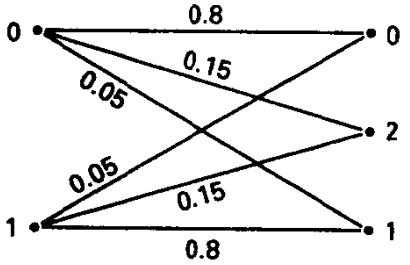

Figure P4.4.2 BSEC channel.

4.4.3. (*Very Noisy Channels*) Consider the case of antipodal signaling with $E_s/N_0 = -10$ dB, and suppose the demodulator output is subjected to four-level quantization with thresholds set at 0 and $0.8E_s^{1/2}$. You may choose $N_0/2 = 1$ for computation purposes.

(a) Using a table of the Q-function, determine the channel transition probabilities for the 2-input, 4-output channel.

(b) Calculate C in bits per channel symbol.

(c) Calculate R_0 in bits per channel symbol. You should find that $R_0 \approx C/2$, which is a general result for very noisy channels, those for which the conditional probability of a given symbol differs by only a small amount as the input symbol label is varied. An equivalent interpretation of the factor of 2 difference between C and R_0 is that in the low SNR regime 3 dB additional SNR is required to achieve a given value of R_0 than for C.

4.4.4. Repeat the calculations of Exercise 4.4.3 if $E_s/N_0 = 0$ dB.

4.4.5. **(a)** Construct the trellis coding exponent $E_t(R)$ using (4.4.38) for a BSC with $\epsilon = 0.1$.

(b) For rate $\frac{1}{4}$ trellis codes with channel constraint length of $n_E = 20$ bits (memory order $m = 4$), determine the form of a bound on the ensemble average bit error probability at the decoder output.

(c) Repeat part (b) when the encoder constraint length is $n_E = 40$.

4.5.1. Consider antipodal signaling on the AWGN channel. The R_0 expressions for ML decoding and for the binary-quantized decoding are, respectively,

$$R_0 = 1 - \log_2(1 + e^{-E_s/N_0})$$

and

$$R_0 = 1 - \log_2\left[1 + (4\epsilon(1-\epsilon))^{1/2}\right].$$

Calculate R_0 for both cases, assuming that $E_b/N_0 = 4$ dB = 2.5 and the code rate is $R = \frac{1}{2}$. Apply this in the random-coding bound (union-Bhattacharyya bound) to determine the block length of $R = \frac{1}{2}$ codes needed to achieve ensemble average probability of error of 10^{-5}. (This exercise should convince you of the benefits of avoiding binary quantizing and of the often disappointing absolute numbers of random coding bounds. Fortunately, the best codes perform quite a bit better than the ensemble average.)

4.5.2. We discussed in Chapter 2 the simple (3, 1) code using antipodal signaling on the additive white Gaussian noise channel and examined the effect of receiver quantization. This problem is intended to demonstrate the effect experimentally. Let $E_b/N_0 = 4.3$ dB = 2.7, so $E_s/N_0 = 0.9$. Assume the encoder selects codeword (000); then the matched filter receiver will produce at its output

$$r_i = -\left(\frac{2E_s}{N_0}\right)^{1/2} + n_i, \qquad i = 1, 2, 3,$$

where n_i are normal, zero-mean, independent, Gaussian r.v.'s with unity variance.

(a) For ML decoding, we form $T = r_1 + r_2 + r_3$ and compare with a zero threshold. If $T \geq 0$, decide (incorrectly) that (111) was sent. Perform 5000 trials of this process on a computer, calculating the empirical error probability. Note that the theoretical error probability for this antipodal arrangement is $Q[(2E_b/N_0)^{1/2}] = 0.01$, so we should experience a number of errors on the order of 50.

(b) For binary decoding, first make a sign decision on each of the three symbols in a codeword; then do majority voting (which minimizes Hamming distance). Repeat the 5000 trials with the same random numbers, and recalculate the error probability.

You might also try increasing E_b/N_0 by 3/2, or 1.8 dB, to test the calculated difference in energy efficiencies between these two approaches.

(c) Finally, do three-level decoding by declaring an erasure if a symbol has a magnitude that falls in the range of $(-0.5\sigma, 0.5\sigma)$, or one-half the noise standard deviation on either side of zero. To decode, if you observe types (1, 1, 1), (1, 1, E), or (1, E, E), decide in favor of (111). If types (1, 0, E) or (E, E, E) are observed, flip a fair coin. Otherwise, decode to (000). Repeat the experiment and calculate the error probability. It should, of course, be intermediate to those obtained previously.

4.5.3. Repeat the calculation of Example 4.7 pertaining to binary DPSK with hard decisions, using instead channel capacity as the ultimate limit on rate R. You should find a lower threshold on E_b/N_0 at any designated rate. What is the implication for $R = \frac{1}{2}$ codes?

4.5.4. (a) Plot the R_0 expression versus E_s/N_0 for noncoherent FSK on a linear scale, and observe that the function is not convex, in contrast to, say, the expression for PSK with coherent detection. Experiment with the graphical solution of Figure 4.5.3 to show that an optimum rate exists for coded noncoherent FSK.

(b) Define $E_s/N_0 = g(R_0)$ to be the inverse relation of $R_0 = f(E_s/N_0)$ and argue that if we can find a rate R for which

$$\frac{dg(R)}{dR} = \frac{g(R)}{R}$$

this rate is optimal in minimizing energy. Interpret this graphically.

4.5.5. Suppose that your company is charged with developing a coding system to be coupled with 8-PSK modems. As delivered, the demodulator outputs are hard (8-level) decisions. At reasonable SNR, you can assume that the channel transitions only to neighboring decision zones, so the channel model of Figure P4.5.5 results, with $2\delta = P_s$.

(a) Is the channel symmetric?

(b) Show that $\delta = Q[(0.29E_s/N_0)^{1/2}]$.

(c) Calculate and plot R_0.

(d) If a code rate of $R = 2$ bits per symbol is adopted, what is the minimum E_b/N_0 that maintains $R_0 > R$?

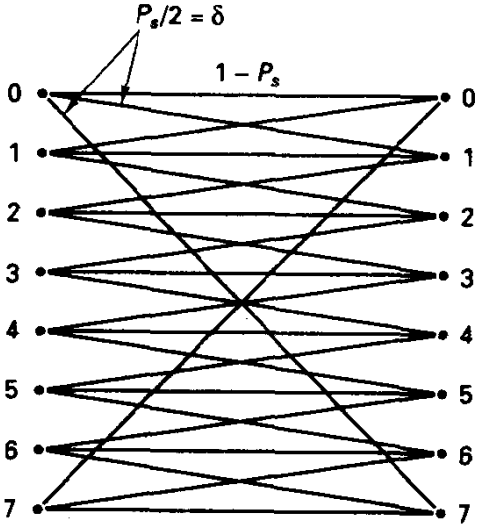

Figure P4.5.5 8-PSK hard decision channel model; adjacent symbol errors allowed.

(e) Make a case to your management for retrofitting the demodulator to provide finely quantized in-phase and quadrature channel outputs to the decoder. How much savings in E_b/N_0 do you expect based on unquantized R_0?

4.5.6. (a) Compute R_0 for the 32-ary QAM constellation of Section 3.3. Use the expression of (4.3.38) together with an equiprobable distribution on inputs, and exploit constellation symmetries to write R_0 of the form

$$R_0 = \log_2 32 - \log\left[1 + N_1 e^{-d_1^2/4N_0} + N_2 e^{-d_2^2/4N_0} + \ldots\right],$$

where N_i are multipliers dependent on numbers of signals at various distances and d_i are intrasignal distances.

(b) Suppose that we wish to send 4 bits per modulator symbol using coding. What is the smallest E_b/N_0 consistent with $R_0 > R$?

(c) Uncoded 16-QAM would have the same dimensionality as the coded technique in part (b). What E_b/N_0 is necessary to maintain $P_s = 10^{-5}$ for this uncoded signaling? You should conclude that coding can gain about 8 dB in energy efficiency with the *same* dimensionality.

4.5.7. Consider 8-PSK modulation on the AWGN channel. The unquantized R_0 behavior is shown in Figure 4.5.11 versus E_s/N_0. If we believe that $R = R_0$ is a practical limit on information rate, what is the minimum E_b/N_0 allowed by $R = 2$ bit/symbol coding? What is the minimum specified by the channel capacity limit, $C = \log_2(1 + E_s/N_0)$, for two-dimensional signaling?

4.5.8. (a) We could use two consecutive symbols of 8-PSK in coded fashion to send $R = 5$ (not 6) bits of information per four-dimensional signal pair (see Wilson, *IEEE Transactions on Communications*, October 1986). Formulate R_0 for the 64-point constellation corresponding to pairs of 8-PSK symbols, and determine the minimum E_b/N_0 limited by $R_0 = 5$ bits/four-dimensional symbol.

(b) Compare these findings with the building of codes directly onto 8-PSK symbols, at $R = 2.5$ bits/symbol.

4.5.9. (a) Ternary ($M = 3$) PSK forms a simplex signal set. Write an expression for R_0 as a function of E_s/N_0. Determine the minimum E_b/N_0 as specified by R_0 capable of providing $R_0 > 1$ bit per symbol. Compare this result with that of 3-level AM or 3-orthogonal transmission. (You should find that the ternary design is slightly superior to both.)

(b) If $R = 1$ bit/two-dimensional symbol is desired as a code rate, how would you rate 3-PSK relative to 4-PSK?

4.5.10. Lee [16] established necessary conditions for a Q-zone quantizer to be optimal: for every point y on the boundary between decision zones D_1 and D_2

$$\sum_{m=0}^{M-1}\left[\sum_{i=0}^{M-1}\left(\left[\frac{P(D_2|i)}{P(D_2|m)}\right]^{1/2} - \left[\frac{P(D_1|i)}{P(D_1|m)}\right]^{1/2}\right)\right] f(\mathbf{y}|m) = 0$$

(a) Use this to find the best three-level quantizer for antipodal signals in AWGN when $E_s/N_0 = -3$ dB.

(b) Show that for M-PSK signaling a $2M$-ary quantizer with $2M$ pie-shaped regions having boundaries through signal points meets the necessary condition (Parsons and Wilson [31]).

4.5.11. With 4-PSK modulation, find R_0 for the following quantization strategies as a function of E_s/N_0:

(a) $Q = 4$, decision regions being the four quadrants. This yields a 4-ary symmetric channel.

(b) $Q = 5$, with a square region at the origin providing the fifth decision region, called an erasure. The sides of the square are $0.5(N_0/2)^{1/2}$ (not necessarily optimal). Probabilities of various regions can be found using products of Q-functions.

(c) $Q = 8$, with eight pie-shaped sectors, with sector boundaries lying on signal points. Assume that noise may only move the received vector into the two adjacent zones to a signal.

4.5.12. Prove that an M-ary simplex is the constellation with M points maximizing R_0 for a given SNR on the additive white Gaussian noise channel, assuming that dimensionality is unconstrained (see Massey [2]).

4.6.1. From the definition of mutual information and use of rules for conditional p.d.f.'s such as $f(y, a|x) = f(y|x, a)f(x|a)$, verify (4.6.15) pertaining to the channel capacity for an interleaved channel with perfect side information.

4.6.2. Given channel observations $\mathbf{y}$ and the corresponding fading amplitudes $\mathbf{a}$, show that maximizing the *a posteriori probability* $P(\mathbf{x}|\mathbf{y}, \mathbf{a})$ in order to minimize probability of message error corresponds to maximizing the likelihood

$$\max_{\mathbf{x}} f(\mathbf{y}, \mathbf{a}|\mathbf{x})$$

If the amplitude is slowly varying, the p.d.f cannot be factored into product form. Show, however, that use of $f(\mathbf{y}, \mathbf{a}|\mathbf{x}) = f(\mathbf{y}|\mathbf{x}, \mathbf{a})f(\mathbf{a}|\mathbf{x})$ allows us to write the likelihood in the form

$$\lambda(\mathbf{y}, \mathbf{x}; \mathbf{a}) = \sum_{j=0}^{n-1} \log f(y_j|x_j, a_j) + \log f(\mathbf{a})$$

since the fading is independent of the input $\mathbf{x}$. Then, because the last term does not involve $\mathbf{x}$, it is sufficient to use the symbol metric $\log f(y_j|x_j, a_j)$.

4.6.3. This exercise concerns coding/decoding with noncoherent detection of M-ary orthogonal signals on a Rayleigh fading channel when side information is lacking. To obtain the metric for decoding and to calculate channel parameters such as C, we need $f(\mathbf{y}|x = m)$. This can be found by

$$f(\mathbf{y}|x = m) = \int f(\mathbf{y}|x = m, a) f_A(a)\, da.$$

(a) By substitution of the relevant p.d.f's, and integration over the variable a, show that

$$f(\mathbf{y}|x = m) = \left[\prod_{j=m} \frac{y_j}{\sigma^2} \exp\left[-\frac{y_j^2}{2\sigma^2}\right]\right] \frac{\exp\left[\frac{y_m^2}{2\sigma^2}\left(\frac{\mu^2}{\mu^2+2}\right)\right]}{1 + \mu^2/2}.$$

(A key step is manipulating the integrand into a Rician p.d.f. form whose integral is 1.)

(b) Show then that the optimal metric to be applied at time j is the square-law metric

$$\lambda(\mathbf{y}_j, x_j = m) = y_{j_m}{}^2.$$

(c) What is the optimal metric if side information is available?

4.6.4. In Section 4.6, the R_0 analysis was applied to M-ary orthogonal signaling on the Rayleigh fading channel. Figure 4.6.6 presents the minimum E_b/N_0 allowed as a function of coding rate R, without regard to bandwidth or dimensionality constraints. On this set of curves, plot contours of equal dimensionality per information bit by using the fact that the dimensionality

per information bit

$$\frac{D}{R} = \frac{M}{\log_2 M}.$$

Determine the best signal set size and best code rate if $D/R = 16$ is tolerable.

4.6.5. Your telecommunications firm has the task of utilizing some existing 4-ary orthogonal FSK modems, which use noncoherent detection and supply hard decisions and no other side information. The expected channel environment is slow Rayleigh flat fading with additive white Gaussian noise. You suggest that coding would be a useful add-on and that interleaving would be useful.

(a) What rate of coding would you adopt, as suggested by R_0 considerations? What is the minimum E_b/N_0 allowed?

(b) Suppose the allowable bandwidth expansion constrains the ratio of signal-space dimensions per symbol, D, to rate R bits/symbol, to be no larger than 25. What code rate should be adopted now?

4.6.6. **(a)** Beginning with the expression for capacity in (4.6.34) for the energy-constrained two-dimensional Rayleigh fading channel with side information, show that as E_s/N_0 becomes small the capacity approaches $E_s/N_0/(\log_e 2)$ bits/symbol, the same as that for the nonfading channel.

(b) Show that at *high* E_s/N_0 the energy penalty implied by fading amounts to 0.577 nats/symbol, or $10\log_{10}\exp(0.577) = 2.5$ dB in E_s/N_0.

4.6.7. Suppose a fading channel exhibits slow, flat fading but instead of Rayleigh amplitude variations, the channel gain is Rician:

$$f_A(a) = \frac{a}{\beta}\exp(-a^2 + \alpha^2)I_0(\alpha a/\beta^2), \qquad a \geq 0,$$

where α and β are fading distribution parameters. (Note that the case with $\alpha = 0$ is the Rayleigh situation.) Formulate expressions for channel capacity with binary noncoherent FSK when side information on amplitude is present for both hard decisions and unquantized demodulation.

4.7.1. Consider optical PPM, with ideal photon counting and no background radiation. Let the detector observe photon counts n_i in each slot. In uncoded transmission, the demodulator would output the index of that slot registering *any* counts, assuming one exists. If all counts are zero, a random decision could be made. Show that in coded PPM transmission, for which the channel model is the M-ary erasure channel, that it is sufficient to report only the index of the slot registering counts, or an erasure if none does; that is, it is no loss of information to discard the actual value of the integer count in the slot producing a nonzero count.

4.7.2. Design an optical PPM system for linking Earth and Mars that has a throughput of $C_T = 10 \cdot 10^6$ bits/s. We are allowed the use of an argon laser (wavelength $= 0.6\,\mu$m), which can be pulsed with a width of 10 ns.

(a) Use the procedure outlined in the text to optimize the capacity of the PPM channel, in bits per photon, under the given constraints. What is the optimal value of M and the resulting capacity in bits per photon? Be mindful of the logarithm base.

(b) What is the resulting R_0 in bits per photon?

(c) What channel coding rate is implied to achieve the given throughput, together with the modulation parameters you have found?

(d) What is the average number of photons received per slot at the detector, and what would be the corresponding *average* power in watts?

4.7.3. Analyze the following binary block interference channel. Blocks of three binary channel bits either are transmitted correctly or are each acted on by a purely random memoryless channel with error probability $\frac{1}{2}$ due to interference (which could be to block signal dropouts or block noise bursts). The probability of either condition is $\frac{1}{2}$, and the condition is chosen independently for each block. We can view the channel as memoryless at the level of 3-tuples, and the channel diagram is either of the two channels depicted in Figure P4.7.3.

(a) First, assume that the decoder knows the channel state for a given 3-tuple. Find the capacity $\overline{C}(3)$ and $\overline{R_0}(3)$, and express in units of bits per binary channel symbol.

(b) Now suppose that the decoder does not know the channel state. Find the average channel transition probabilities $P(\mathbf{y}|\mathbf{x})$ and draw the corresponding channel diagram. Recalculate $C(3)$ and $R_0(3)$.

One approach to communicating over such a channel is precisely that suggested: view 3-tuples as symbols in the code alphabet. Another is to design a binary code capable of correcting 3-bit bursts. Still another is to scramble (or interleave) the binary channel transmissions so that the channel appears like a DMC with effective error probability of 1/4. Suppose that side information is also available for each symbol. Does interleaving increase or decrease channel capacity here? How about R_0?

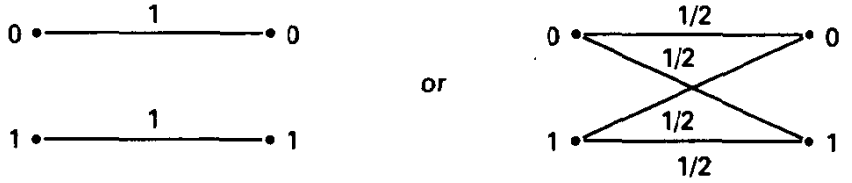

Figure P4.7.3 Two BSCs.

4.7.4. Consider a simple two-mode channel that has a good state and a bad state, each having probability 1/2. In the good state the channel is a memoryless BSC with zero error rate. In the bad state the BSC error probability is 0.1.

(a) Assume that the channel is interleaved, making the channel states independent in time. In the first case, assume that the receiver has access only to the binary channel output y_j. Show that the channel capacity for this case is $C = 1 - h_2\ (0.05)$.

(b) Now suppose that in addition to y_j you can have access to the channel state at a given time. Show that the capacity now is $\frac{1}{2}C(\epsilon = 0) + \frac{1}{2}C(\epsilon = 0.1)$.

(c) Conclude that knowing the channel state does indeed improve the capacity.

4.7.5. Suppose that the channel of Exercise 4.7.4 is interleaved two symbols at a time, making a memoryless channel over successive pairs, but within a pair the channel resides in a given state. Recalculate channel capacity here with and without side information. Does capacity increase or decrease?

4.7.6. Consider the interference or jammed channel shown in Figure P4.7.6. Antipodal signals are inputs to the channel, and with probability ρ, an independent Gaussian noise variate is added to the input. When noise is present, the variance is σ^2/ρ so σ^2 is the average noise energy per channel use. The jammer is free to select ρ and knows the resources E_s and σ^2 available to both parties. Four cases for reception are to be analyzed, corresponding to switch positions in the diagram. In case I, the receiver has hard decisions available, but no channel state information. In case II, hard decisions and channel state are available. For case III, the analog channel output is provided with no channel state, while for case IV, the analog channel output and channel state are provided.

(a) For each case, formulate the ML decoding metric to be applied to each symbol received.

(b) Evaluate R_0 as a function of ρ. For each case determine the jammer's best strategy.

(c) Find the union-Chernoff bound on performance for coded systems of rate R when the jammer applies its best strategy for each situation. You should find that the use of unquantized reception without channel state information is a poor performer!

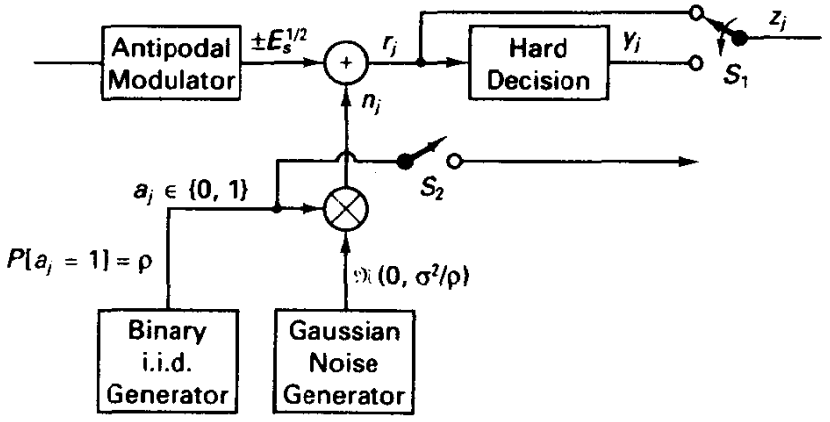

Figure P4.7.6 Pulsed interferer model.