THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Volume 62 February 1983 Number 2, Part 1

Digital Communications Over Fading Radio Channels

By G. J. FOSCHINI and J. SALZ

(Manuscript received August 19, 1982)

A major contribution to system outage in a terrestrial digital radio channel is deep fading of the frequency transfer characteristic, which in addition to causing a precipitous drop in received signal-to-noise ratio (s/n) also causes signal dispersion that can result in severe intersymbol interference. Because the temporal variation of the channel is slow compared to the signaling rate, the information theoretic channel capacity and the "Efficiency Index" in bits/cycle—a figure-of-merit we use for the communication techniques considered—can be viewed as random processes. Starting from an established mathematical model characterizing fading channels (derived from extensive measurements), we estimate the probability distribution of channel capacity and the distributions of efficiency indices for different communications techniques. The repertoire of communication methods considered involves quadrature amplitude modulation with adaptive linear and decision feedback equalization, and maximum likelihood sequence estimation. For specific outage objectives the maximum number of bits per cycle achievable by each technique is estimated. The sensitivity of the distributions to bit-error-rate objective and unfaded s/n is assessed. For certain desired operating points the efficacy of adaptive equalization is demonstrated. There are some operating points where adaptive equalization alone is not adequate and therefore space diversity should be considered. An estimate of the effect of frequency diversity is also included.

I. INTRODUCTION AND SUMMARY

Fading of terrestrial digital radio channels owing to multipath reception is a prime cause of system outage. For a specific hop a mathematical model of these fades has been developed by W. D. Rummler[1,2] from extensive measurements of the channel frequency power transfer characteristic over time. The radio channel has a time-varying frequency characteristic, with additive Gaussian noise; however, the temporal variations are sufficiently slow in comparison to the data symbol rate that the characteristics can be represented as a random ensemble of static frequency power transfer functions. The presence of additive noise implies that each member of the ensemble is limited to a maximum rate of transmission of data, depending on the communication method. For each specific communication technique, the stochastic nature of the channel makes it meaningful to consider the probability distribution of data rates that can be supported at a certain bit-error-rate objective.

The purpose of this paper is to explore the relative performance of various communication techniques employing quadrature amplitude modulation (QAM), distinguished by the type of equalization method used. These techniques include variants of linear equalization, decision feedback equalization, and maximum likelihood sequence estimation (MLSE). For these methods a unified set of Chernoff bounds on the probability of error is obtained. Given a communication method, a channel impulse response, an error-rate objective, a received unfaded channel s/n, a channel bandwidth, and a signaling rate we use the Chernoff bounds to estimate the maximum number of bits per cycle of bandwidth (not necessarily integer-valued) for which the constraints are met. By computing the maximum number of bits per cycle supported by each member of a large representative population of channels, we obtain the cumulative probability distribution function. One can use the cumulative distribution curve to determine the probability of outage at a prescribed bit rate.

The information theory bound on the number of bits per cycle attainable is also derived. In a sequence of plots we compare the different schemes with each other and with the information theory limit.

If $F(r)$ is the probability distribution function of data rates associated with a communication method and we set an outage objective, ξ, then the value r_ξ for which $F(r_\xi) = \xi$ represents the maximum data rate at which it is possible to transmit and meet the outage objective. We present and discuss these distributions in the context of desired long-haul and short-haul outage objectives and rates associated with the digital hierarchy constraints. The efficacy of adaptive equalization is established. The advantage of decision feedback and MLSE over

optimum linear equalization is not very substantial. There are some desired operating points for which space diversity should be considered.

For a fair comparison of different communication techniques, the transmitter filter shape must be optimized for a fixed transmitter power. We found the performance to be insensitive to whether or not the transmitter filter is optimized and we provide a theoretical guideline to indicate when this optimization becomes significant.

Our results indicate that optimized equalizer structures yield data rates only a few bits/cycle below channel capacity. It therefore appears that higher dimensional constellations[3] spanning two to four symbol intervals could go a long way toward obtaining that which can be expected practically. Although we did not analyze higher dimensional signal design or optimize the constellation in QAM, it is reasonable to expect that these techniques can offer at most an equivalent few dB increase in s/n. Another method of achieving coding gain "of the order of 3-4 dB" is described in Ref. 4. Moreover, the real limitations on the selection of signal points in a practical system will most likely arise from the nonlinear operation of radio frequency (RF) power amplifiers rather than from s/n limitations.

We argue the merits of adaptive transversal equalization and provide numerical support for our claims. This is not to say that fixed or even adjustable bump and/or slope equalizers in the frequency domain could not provide adequate performance in some cases. However, fluctuating (and sometimes nonminimum) phase distortion associated with fading and other linear filters admits robust and stable compensation via adaptive transversal filters. These structures with adjustable taps can automatically equalize any phase characteristic without noise enhancement and therefore are natural candidates in these applications, especially at a high number of levels where even small amounts of phase distortion can degrade system performance.

Our analysis was carried out with ideal models and an infinite number of taps. The actual number of taps needed in any application would be determined from experiments and/or more detailed analysis.

II. THE EQUALIZED QAM SYSTEM—IDEALIZED MODEL

The use of equalizers to mitigate the effects of intersymbol interference and noise in voiceband data transmission is by now standard practice. We are thus naturally led to consider the application of these techniques in digital data transmission over the radio channel where slowly varying, frequency-selective fading is the predominant impairment. Here we review and derive the applicable mathematical theory that will be used in the sequel to evaluate the system performance indices.

To focus on basics and avoid extensive numerical analysis, we consider idealized equalizer models represented as transversal filters with an infinite number of taps. Tap adjustment algorithms are well established and our formulas are derived under the assumption that the taps have converged to their optimum values.

Our analyses are based on the digital communications model depicted in Fig. 1. To appreciate the applicability and generality of this baseband model to digital radio communications, we observe that, for any bandpass linear channel, the output waveform, when the input is any linearly modulated data wave, can be represented as

$$s(t) = \mathrm{Re}\left\{\sum_n \tilde{a}_n \tilde{h}(t - nT + t_0)\exp[i(2\pi f_0 t + \theta)]\right\},$$

where $\mathrm{Re}\{\cdot\}$ stands for the "real part." The data symbols $\{\tilde{a}_n\}$ transmitted at T-second intervals, are statistically independent and, in two-dimensional modulation systems such as QAM, they assume complex values. The overall equivalent baseband impulse response, $\tilde{h}(t)$, is also complex-valued. The real part represents the in-phase response, while the imaginary part is the quadrature component. The frequency, f_0, is the carrier frequency, θ is the carrier phase, and, t_0 is the timing phase. Ideal demodulation with a known carrier frequency f_0 and carrier phase θ implies a translation of the received bandpass signal to baseband. The real part of the resulting complex signal represents the in-phase modulation, while the imaginary part is the quadrature modulation. This then is the rationale, in addition to economics of notation, for using the complex baseband model depicted in Fig. 1.

We restrict our treatment to ideal Nyquist systems with no excess bandwidth. This permits less cumbersome calculations without loss of physical insights. We also derive our formulas by assuming flat transmitting filters and prove later that in-band optimum shaping yields imperceptible additional benefits. Also neglected is adjacent channel interference, as ideal bandlimiting eliminates this problem.

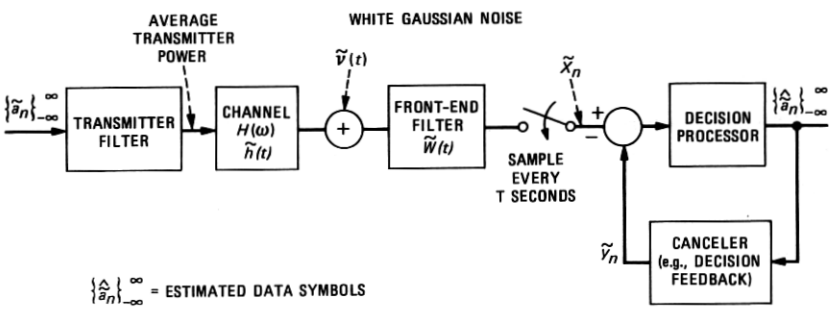

Fig. 1—Complex baseband model for QAM data transmission.

We now return to Fig. 1 and discuss the various functions and notations indicated. Without loss of generality we assume that the complex data symbols, $\{\tilde{a}_n = a_n + ib_n\}_{-\infty}^{\infty}$, take on values on a set of positive and negative odd integers with equal probability. Accordingly,

$$E\tilde{a}_n = 0 \quad \text{and} \quad E|\tilde{a}_n|^2 = 2\frac{L^2 - 1}{3},$$

where $E(\cdot)$ denotes mathematical expectation and L (even) is the maximum number of data levels assumed by a_n and b_n. Thus, in QAM L^2 data points are available for conveying information and the source therefore generates

$$R = \frac{\log_2 L^2}{T} \text{ bits/sec.} \tag{1}$$

For a given channel bandwidth, $\mathscr{W}$, the efficiency index is defined as

$$I = R/\mathscr{W} = \frac{2\log_2 L}{\mathscr{W}T} \text{ bits/cycle.} \tag{2}$$

As we shall see, the relationship among P_e- probability of error, s/n, $\mathscr{W}$, T and $H(\omega)$-channel frequency characteristics is rather complicated. The determination of the relationship for different communication techniques is our chief task in the sequel.

From a mathematical point of view, the fading radio channel is characterized by a slowly varying linear distorting filter whose base-band equivalent complex impulse response is the Fourier transform of the transfer function $H(\omega)$, shifted to zero frequency:

$$\tilde{h}(t) = h_1(t) + ih_2(t) = \int_{-2\pi\mathscr{W}}^{2\pi\mathscr{W}} H(\omega)e^{i\omega t}\frac{d\omega}{2\pi}.$$

At the receiver the added complex noise process, $\tilde{v}(t) = v_1(t) + iv_2(t)$, is assumed to be white Gaussian with $v_1(t)$ independent of $v_2(t)$ and each possessing a double-sided spectral density, $N_0/2$. So,

$$\begin{aligned} E|\tilde{v}(t)|^2 &= Ev_1^2(t) + Ev_2^2(t) \\ &= N_0\delta(0), \end{aligned}$$

where $\delta(\tau)$ is the Dirac delta function. The average transmitted signal power, P_0, for a flat transmitting filter can easily be calculated. However, for our purposes a more relevant quantity is the received, unfaded signal power

$$P = K^2P_0 = 2\frac{L^2 - 1}{3}\frac{K^2}{T^2},$$

where K is a constant that includes the effects of amplifiers, antennas,

and the unfaded channel loss. Also, the added average noise power in the Nyquist band, $\mathscr{W} = 1/2T$, is

$$P_v = \frac{N_0}{T}.$$

Thus the unfaded received s/n, a most important system parameter, is

$$s/n = \rho = 2\frac{L^2 - 1}{3}\frac{K^2}{N_0}\frac{1}{T}. \tag{3}$$

The receiver structures under consideration consist of a perfect demodulator followed by a front-end filter possessing the complex impulse response $\tilde{W}(t)$, a sampler, a decision device, and a canceler. The design of an optimum receiver entails the selection of $\tilde{W}(t)$ and the canceler for a particular channel characteristic. Since the channel characteristics are usually unknown to the receiver, these components must be determined adaptively.

To understand the function of the canceler, consider the signal sample at the output of filter $\tilde{W}(t)$,

$$\tilde{x}_n = \sum_{k=-\infty}^{\infty} \tilde{r}_k \tilde{a}_{n-k} + \tilde{z}_n, \quad -\infty \leq n \leq \infty, \tag{4}$$

where $\tilde{r}_k = \tilde{r}(kT + t_0)$ is the overall complex-sampled system impulse response and $\tilde{z}_n$ is the Gaussian noise output sample. Ideally the canceler strives to synthesize the value

$$\tilde{y}_n = \sum_{k \in S} \tilde{r}_k \tilde{a}_{n-k} \tag{5}$$

and to subtract it from (4) where the set of integers S is defined as $\{k \in S: k = -N_1 \cdots -1, 1 \cdots N_2\}$. The canceler's ability to synthesize these values presumes that some past ($k = 1, \cdots, N_2$) and/or future ($k = -1, \cdots, -N_1$) transmitted data symbols are perfectly detected and, moreover, that the set of complex numbers, $\tilde{r}_k$, are adaptively estimated.

The front-end filter, $\tilde{W}(t)$, is usually determined adaptively by minimizing the mean-squared error (MSE) between the sample, $\tilde{x}_n - \tilde{y}_n$, and the expected data symbol $\tilde{a}_n$:

$$MSE[N_1, N_2, \tilde{W}(t)] = E|\tilde{x}_n - \tilde{y}_n - \tilde{a}_n|^2, \tag{6}$$

and the optimum filter, $\tilde{W}(t)$, is chosen to achieve

$$(MSE)_0 = \min_{\tilde{W}(t)} MSE[N_1, N_2, \tilde{W}(t)] = MSE[N_1, N_2, \tilde{W}_0(t)]. \tag{7}$$

Since (6) is a quadratic functional of $\tilde{W}(t)$, a unique minimum can always be found. It is standard to represent the linear filter $\tilde{W}(t)$ by a

transversal structure and in practice the search for the minimum is accomplished by varying the taps of this filter until a minimum of the time average of the squared error is found. Clearly, to realize such a minimization procedure, estimates of the transmitted data symbols must be used.

III. SYSTEM PERFORMANCE—GENERAL

To get at the efficiency index of a system, the error rate as a function of data rate for any choice of the canceler set, $\{S\}$, and front-end filter, $\tilde{W}(t)$, must be explicitly expressed. Unfortunately, exact relationships are not mathematically tractable for the simplest of systems and so we must employ upperbounds. Fortunately, for the systems under consideration, it is possible to obtain exponentially tight inequalities.

With this approach in mind, note that after perfect cancellation, the decision variable, from (4) and (5) becomes

$$\begin{aligned} s_n &= \tilde{x}_n - \tilde{y}_n \\ &= \tilde{r}_0 \tilde{a}_n + \sum_{k \notin J} \tilde{r}_k \tilde{a}_{n-k} + \tilde{z}_n, \end{aligned} \tag{8}$$

where now the set J is $S \cup 0$, $\{k \in J: k = -N_1 \cdots 0 \cdots N_1\}$. Decisions in QAM are made on the real part of s_n and, separately, on the imaginary part of s_n. Simple calculations give

$$Re(s_n) = \mu_0 a_n - v_0 b_n + \sum_{k \notin J} (\mu_k a_{n-k} - v_k b_{n-k}) + z_{n1},$$

and

$$Im(s_n) = \mu_0 b_n + v_0 a_n + \sum_{k \notin J} (\mu_k b_{n-k} + v_k a_{n-k}) + z_{n2}, \tag{9}$$

where

$$\tilde{r}_k = \mu_k + i v_k,$$

and

$$\tilde{z}_k = z_{k1} + i z_{k2} = \int_{-2\pi W}^{2\pi W} \tilde{v}(t) \tilde{W}(t) dt.$$

For an L-level system, slicing levels are placed at $0 \pm 2\mu_0 \cdots \pm \mu_0(L-2)$ and compared with the received samples $\mathrm{Re}(s_n)$ and $Im(s_n)$. An error occurs whenever the noise plus intersymbol interference (in-phase and quadrature) exceed in magnitude the distance from the transmitted level to the nearest decision threshold, μ_0. However, the outside two levels can be in error in one direction only.

Now denote the event of an error committed in the "real" rail by E_r

and in the "imaginary" rail by E_i. Then the probability of system error, P_e, is the probability of either (or both) E_r or E_i occurring,

$$P_e = P(E_r \cup E_i) \leq P(E_r) + P(E_i), \tag{10}$$

where

$$P(E_r) = \left(1 - \frac{1}{L}\right) \Pr\left[\left| z_{n1} - \sum_{k \notin J} (\mu_k a_{n-k} - v_k b_{n-k}) + v_0 b_n \right| \geq \mu_0\right]$$

and

$$P(E_i) = \left(1 - \frac{1}{L}\right) \Pr\left[\left| z_{n2} - \sum_{k \notin J} (\mu_k G_{n-k} + v_k a_{n-k}) - v_0 a_0 \right| \geq \mu_0\right]. \tag{11}$$

Because of symmetry, $P(E_r) = P(E_i) = P(E)$, and so we only need to upperbound $P(E)$.

We adopt a bounding procedure introduced by B. Saltzberg[5] to analyze the error rate in an unequalized baseband system. We have extended Saltzberg's approach to our systems and it can be shown that

$$P(E; A, B, \delta) \leq 2 \exp\left\{ \frac{\left[\mu_0 - (L-1)\left(\sum_{k \in A} |\mu_k| + |v_k| + \delta v_0\right)\right]^2}{2\left\{\sigma_{z_1}^2 + \frac{L^2 - 1}{3}\left[\sum_{k \in B} \mu_k^2 + v_k^2 + (1-\delta)v_0^2\right]\right\}} \right\}. \tag{12}$$

The set of integers A and B form a partition on the set of integers not included in J. That is,

$$A \cup B = \Omega = \{k: k \notin J\}$$

and

$$A \cap B = \phi.$$

The variable $\delta = 1$ *or* 0, and

$$\sigma_{z_1}^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} |\tilde{W}(t)|^2 dt.$$

The sharpest upperbound is obtained by minimizing (12) with respect to the sets A, B, and δ. Algorithms for carrying out this minimization can be devised readily.

IV. SYSTEM PERFORMANCE

4.1 Discussion

Equation (12) is a rather general upperbound on the error rate for any passband linear data transmission system and it will now be specialized to include the effects of the different choices of equalizers.Before proceeding with the detailed numerical analysis, we need to make a connection between the mean-squared error (MSE), which is minimized by equalizers, and the system probability of error, which, ideally, should be the quantity minimized.

A straightforward but tedious approach for getting at the error rate might be to first determine the filter, $\tilde{W}(t)$, which minimizes the MSE for any particular equalization scheme, calculate the overall resulting impulse response, and then use eq. (12) to upperbound the error rate. This approach can be circumvented by exploiting the explicit relationship between the minimum MSE and the value of the overall impulse response at $t = t_0$ when the optimum filter, $\tilde{W}_0(t)$, is used.

The optimum structure of the minimum mean-squared error receiver can be shown to consist[6] of a matched filter in cascade with a transversal filter combined with a linear intersymbol interference canceler. The implication of this structure is that the resulting overall system transfer function is a real function of frequency. Or, the complex-sampled impulse response, $\{\mu_k + i\upsilon_k\}_{-\infty}^{\infty}$, must be a real number at $k = 0$, which results in $\upsilon_0 = 0$. This follows from the Fourier Transform representation of $\tilde{r}(t)$, from which we see that at $t = 0$ the integrand is real and nonnegative. Indeed the overall phase characteristic has been removed by the matched filter (without enhancing the noise*). Using the fact that $\upsilon_0 = 0$ and careful numerical analysis of the available channel characteristics, our calculations showed that for all practical purposes the bound (12) becomes

$$P(E, S) \leq 2 \exp \left\{ \frac{-\mu_0^2}{2\left[\sigma_{z_1}^2 + \sigma^2(L) \sum_{k \notin S} (\mu_k^2 + \upsilon_k^2)\right]} \right\}, \tag{13}$$

where we set $\sigma^2(L) = \dfrac{L^2 - 1}{3}$.

As will become apparent, the argument of the exponential function in (13) can be directly related to the minimum mean-squared error.

---

* A fractionally spaced ($T/2$) transversal filter can automatically synthesize any matched filter and thus eliminate phase distortion and also compensate for timing phase (see Ref. 7.)

Towards this end we recall a well-known[6] result that states that the best achievable MSE has the simple representation,

$$(MSE)_0 = 2\sigma^2(L)(1 - \mu_0), \tag{14}$$

where μ_0 is the sample at $t = t_0$ at the output of the optimum filter. Also, when the optimum filter, $\tilde{W}_0(t)$, is used, a straightforward calculation of the resulting MSE gives

$$(MSE)_0 = 2\sigma^2(L)(1 - \mu_0)^2 + 2\sigma^2(L) \sum_{k \notin S} (\mu_k^2 + v_k^2) + 2\sigma_{z_1}^2. \tag{15}$$

Relationships (14) and (15) make it possible to write (13) as

$$P(E; S) \leq 2 \exp\left\{ -\frac{1}{(MSE)_0} \left[1 - \frac{(MSE)_0}{2\sigma^2(L)} \right]^2 \right\} \sim 2\exp -\frac{1}{(MSE_0)} \quad \text{for} \quad N_0 \to 0, \tag{16}$$

relating error rate and minimum MSE. This is an extremely useful inequality since (MSE_0) as a function of channel characteristics is often explicitly known for different equalizer structures.

It is also interesting to note (this has been pointed out before[8]) that the filter, $\tilde{W}(t)$, that minimizes MSE also minimizes the upperbound on P_e. This is true because the same quadratic functionals in $\tilde{W}(t)$ are involved in the optimization of both expressions.

We are now in a position to specialize our formulas to the various equalizer structures under investigation.

The six examples that follow do not require the knowledge of channel phase characteristics to compute performance. Implicit in each of these schemes is the complete removal of phase distortion, which can be accomplished without noise enhancement. Only a magnitude characterization of the channel transfer response was available at the time the work reported here was done. While departure from flatness of the magnitude fundamentally affects performance, theoretically, departure of phase from linear has no effect on attainable performance. Therefore, the lack of phase characterization of the channels was not an obstacle to our study. However, a complex characterization of the channel would be useful in determining the minimum number of required taps in the designs of the equalizers.

4.2 Pure phase equalization

In this particular equalizer, $N_1 = N_2 = 0$ (where N_1 and N_2 are the lengths of the precursive and postcursive cancelers, respectively). We choose $\tilde{W}(t)$ so that

$$W(\omega) = e^{-i\phi(\omega)}, \ |\omega| \leq \frac{\pi}{T}$$

$$= 0, \ |\omega| > \frac{\pi}{T},$$

where $W(\omega)$ is the Fourier transform of $\tilde{W}(t)$ and $\phi(\omega)$ is the channel phase characteristic. For this choice of filter, only the magnitude of the channel transfer function enters into the computation of the bound, as shown in eq. (13).

Using the well-known Poisson sum formula along with some algebra, it is possible to write (13) more explicitly, i.e.,

$$P_e \leq 2 \exp\left\{-\frac{\rho}{2\sigma^2(L)} \frac{\langle H \rangle^2}{1 + \rho \langle (H - \langle H \rangle)^2 \rangle}\right\}, \tag{17}$$

where we used the shorthand notation

$$\langle \cdot \rangle = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} [\cdot] d\omega$$

and H is used in place of $|H(\omega)|$.

4.3 Linear equalization

Here again $N_1 = N_2$ (no canceler) and $\tilde{W}(t)$ is chosen to minimize (6). The expression for the optimum MSE in this case has been shown to be[9]

$$(MSE)_0 = 2\sigma^2(L) < \frac{1}{1 + \rho H^2} >. \tag{18}$$

This formula is directly used in (16) to calculate the upperbound on error rate:

$$P_e < 2 \exp\left[-\frac{1}{2\sigma^2(L)} \left(< \frac{1}{1 + \rho H^2} > \right)^{-1}\right].$$

4.4 Inverse equalization

In this case $N_1 = N_2 = 0$ (no cancellation) we choose $\tilde{W}(t)$ to be the inverse of the channel frequency characteristics,

$$W(\omega) = H^{-1}(\omega), \ |\omega| \leq \frac{\pi}{T}$$

$$= 0, \ |\omega| > \frac{\pi}{T}.$$

Here the channel is clearly perfectly equalized so that intersymbol interference is completely eliminated; the penalty is increased output

noise power. For this simple scheme, it is possible to express the error rate exactly but for reasons of uniformity we use the upperbound

$$P_e \leq 2 \exp - \left\{ \frac{\rho}{2\sigma^2(L)} \frac{1}{\frac{1}{<\overline{H^2}>}} \right\}. \tag{19}$$

4.5 Decision feedback

For this equalization system $N_1 = 0$ and $N_2 = \infty$ and again we choose $\tilde{W}(t)$ to minimize (6). In this type of equalizer, the causal or postcursor intersymbol interference is entirely eliminated and an expression for the optimum MSE is known,[10,11] as shown below.

$$(MSE)_0 = \sigma^2(L)\exp\{-\langle \ln[1 + \rho H^2] \rangle\}. \tag{20}$$

This is used in (16) to express an upperbound on error rate.

4.6 The ideal equalizer

In this utopian scheme the precursor and postcursor cancelers become infinite, $N_1 = N_2 = \infty$, so that all the intersymbol interference is eliminated. In this ideal situation we obtain the very best possible result, namely, the matched filter bound, which is a lower bound on P_e. This scheme assumes that it is possible to detect each pulse $\tilde{a}_k \tilde{r}(t - nT)$ optimally by a matched filter without incurring interference from all other pulses. The filter, $\tilde{W}(t)$, in this case is chosen to be matched to the channel characteristic, i.e.,

$$W(\omega) = H^*(\omega), |\omega| \leq \frac{\pi}{T}$$

$$= 0, |\omega| > \frac{\pi}{T},$$

where * denotes the complex conjugate.

For this idealization the upperbound on error rate is simply,

$$P_e \leq 2 \exp\left\{ -\frac{\rho}{2\sigma^2(L)} \langle H^2 \rangle \right\}. \tag{21}$$

No other detection scheme can do better. In the next section we use these formulas to calculate the various efficiency indices.

Before concluding this section, we remark that there is one more easy case and one extremely difficult case that might be considered as candidates for making comparisons. Suppose that no filtering, other than out-of-band elimination of noise, were performed at the receiver. What performance can one expect? While we cannot answer this question exactly because channel phase characteristics are unavailable

at this writing, we would expect performance to be worse than removing the phase characteristic entirely—a situation we will examine.

The second approach, which is a very difficult one to analyze, involves the use of a finite-state Viterbi decoder. Nevertheless, we will report a bound on the performance of this processor. Specifically, the performance of an infinite canceler (the matched filter) is superior to the maximum likelihood (Viterbi) decoder. As shown later, for the channels considered, the matched filter bound is close in performance to decision feedback. Consequently, we shall see that the performance of maximum likelihood sequence estimation is tightly bracketed because it is superior to decision feedback.

4.7 *Information theory bound on communications efficiency index*

In this section we discuss a formula for the maximum number of bits per cycle that can be attained for a given $H(\omega)$. If $H(\omega)$ were constant in frequency the formula for the efficiency index in bits per cycle would be simply

$$I = \log_2(1 + \rho|H|^2).$$

It is reasonable to expect that if $H(\omega)$ is frequency-dependent, the maximum efficiency index would be

$$I = \frac{1}{\Omega}\int \log_2(1 + \rho|H(\omega)|^2)d\omega, \tag{22}$$

where the integral is over a frequency band of size $\Omega = 2\pi W$. Indeed this is the case. To outline a derivation, we note first that A. Kolmogorov has generalized Shannon's notion of capacity to provide a very fundamental definition that gives a useful starting point for developing capacity formulas in nonstandard situations such as the one at hand.[12] M. S. Pinsker[13] was able to derive from the Kolmogorov approach a formula for the amount of information in a stationary Gaussian process about another stationary Gaussian process related to it. Specifically, if $S_x(\omega)$ and $S_y(\omega)$ are the power spectral densities of the processes and $S_{xy}(\omega)$ is the cross-spectral density, the formula for the amount of information is

$$-\int \ln\left(1 - \frac{|S_{xy}(\omega)|^2}{S_x(\omega)S_y(\omega)}\right)d\omega.$$

If we require the transmitter output to be Gaussian, then since the additive noise is Gaussian, Pinsker's formula can be applied to the case where x is the transmitted process and y is the received process to obtain (22). Requiring the transmitter output to be Gaussian is really not a limitation since, when the additive noise is Gaussian, one can prove that capacity is attainable with a Gaussian transmitter output

using the methods discussed in Refs. 14 through 16. Thus, (22) gives the efficiency index formula.

References 16 and 17 also provide approaches to establishing (22).

4.8 Information theory limit on index when transmitter is optimized

So far we have treated ρ, eq. (3), as a constant. In this section we set the stage for exploring the advisability of optimizing the output power spectral density to maximize the efficiency index. Since we are now allowing in-band shaping of the transmitter filter frequency characteristic, we will consider ρ as a function of ω and write $\bar{\rho}$ to denote the previously considered case where ρ is constant over the band.

Although the analysis in this section is focused on the information theory limit on the efficiency index, the decision feedback index involves the identical functional form and so our analysis will be applicable to decision feedback as well.

We will compare the previously discussed index

$$I(\text{flat}) = \frac{1}{\Omega} \int \log_2(1 + \bar{\rho}\,|H(\omega)|^2) d\omega$$

with

$$I(\text{opt}) = \frac{1}{\Omega} \int \log_2(1 + \rho_0(\omega)\,|H(\omega)|^2) d\omega,$$

where $\rho_0(\omega)$ is the function maximizing I under a constraint on $\int \rho(\omega)\, d\omega$, the received signal-to-noise ratio in the absence of fading. This constraint is equivalent to a constraint on the transmitter output power, since in the absence of fading the channel has a flat loss characteristic.

This optimization problem is known[17] and yields easily to the calculus of variations. The solution is called "water pouring." The name stems from the graphical interpretation that if $\bar{\rho}\Omega$ is the constraint on $\int \rho(\omega)\, d\omega$, the optimum $\rho(\omega)$, which we denote by $\rho_0(\omega)$, is obtained by forming a vessel with base $|H(\omega)|^{-2}$ and vertical sides at the band edges. One pours "water," that is, area, of amount $\bar{\rho}\Omega$ into the vessel and $\rho_0(\omega)$ is given by the depth of the water at ω. It is clear that this construction obeys the constraint. Generally, if $\bar{\rho}$ is sufficiently small, the poured water will not touch both of the vertical sides of the vessel. In such situations, it would be advantageous to limit the transmitted power to a frequency band less than Nyquist. In our case, however, the unfaded signal-to-noise ratio is so great that, for the simulated channels, the water level always meets both vertical sides. In other words, the transmitted power always occupies the full Nyquist band. Thus, $\rho_0(\omega) = A - |H(\omega)|^{-2}$, where A is chosen so that $\int \rho_0(\omega) d\omega = \bar{\rho}\Omega$, that is, $A\Omega - \int |H(\omega)|^{-2} d\omega = \bar{\rho}\Omega$ or

$$\rho_0(\omega) = \bar{\rho} + \frac{1}{\Omega}\int |H(v)|^{-2}dv - |H(\omega)|^{-2}.$$

Therefore,

$$\frac{I(\text{flat})}{I(\text{opt})} = \frac{\int \log_2(1 + \bar{\rho}|H(\omega)|^2)d\omega}{\int \log_2\left[\frac{|H(\omega)|^2}{\Omega}\int |H(v)|^{-2}dv + \bar{\rho}|H(\omega)|^2\right]d\omega}.$$

We expand the logarithm for $\bar{\rho}$ large to find an asymptotic representation. We get, after a cumbersome derivation,

$$\frac{I(\text{flat})}{I(\text{opt})} = 1 - \frac{1}{2\bar{\rho}^2\log_2\bar{\rho}}\left(1 + \frac{\int \log_2|H|^2 d\omega}{\Omega \log_2\bar{\rho}}\right)^{-1} \cdot\left[\frac{\int \frac{1}{|H|^4}d\omega}{\Omega} - \left(\frac{\int \frac{1}{|H|^2}d\omega}{\Omega}\right)^2\right] + O\left(\frac{1}{\bar{\rho}^3\log_2\bar{\rho}}\right). \quad (23)$$

The last multiplier in the perturbation expression represents the variance associated with a random sampling of a specific $|H(\omega)|^{-2}$. This multiplier is zero if $|H(\omega)|^{-2}$ is a constant. We note, for later use, that for $\bar{\rho} = 10^{6.3}$, (i.e., a 63-dB s/n in the absence of fading) we have

$$\left(\frac{1}{2\bar{\rho}^2\log_2\bar{\rho}}\right) < 10^{-14}.$$

4.9 *Communication efficiency index*

The relationships in eqs. (17) through (21) are unifying expressions for the error rate for the five equalization cases. From Section 1.0 we have $I = 2\log_2 L$ and $\sigma^2(L) = [(L^2 - 1)/3]$. These two equations in conjunction with the P_e formulas enable us to determine I as a function of the P_e objective, ρ, the channel, and the equalization scheme. Specifically,

$$I = \log_2\left[\frac{G(H, \rho)}{\ln(P_e/2)} + 1\right],$$

where $G(H, \rho)$ is a function that depends on the communication method. With the channel response considered to be a random function, I is a random variable, and we can determine its probability distribution function for each communication scheme. The quantities ρ and P_e are parameters of the distribution.

V. MODEL FOR THE FADING CHANNEL

Now we describe the mathematical model for frequency-selective fading owing to multipath reception. This mathematical model for the random functions $|H(\omega)|^2$ is due to W. D. Rummler[1,2] and is based on measurements of a 26.4-mile hop between Palmetto and Atlanta, Georgia. The measurements of frequency-selective fades were made on a 25.3-MHz channel situated in the 6-GHz band during the heavy fading month of June (1977).

Rummler's model uses a two-ray representation of the signal, which was quite adequate for fitting the experimental records. (The model is not necessarily intended to depict the underlying physical mechanism for a fade. The true mechanism could involve a much more complex ray combination.) Also, it is not possible to deduce the phase characteristic associated with any particular amplitude characteristic. It has been experimentally determined that this kind of a channel cannot always be viewed as minimum phase.[18]

In the model, the $|H(\omega)|^2$ functions are 68-degree sections of scaled, displaced cosine waves. Specifically, conditional on a fade occurring

$$|H(\omega)|^2 = a^2|1 + b^2 - 2b\cos(\omega\tau + \theta)|,$$

where:

(*i*) $b = 1/10^{B/20} > 0$ with B an exponential random variable with mean 3.8.

(*ii*) The parameter, a, is a log normal random variable with dependence on the parameter, b. Specifically, $a = 1/10^{A/20}$, where A is normal with a mean of $24.6(B^4 + 500)/(B^4 + 800)$dB and a standard deviation of 5 dB.

(*iii*) The phase, θ, is independent of a and b and has a constant density on each section $|\theta| > \pi/2$ and $|\theta| < \pi/2$ with $P\{|\theta| < \pi/2\} = 5 \cdot P\{|\theta| > \pi/2\}$.

(*iv*) The scale factor τ is a constant = 6.31 nanoseconds.

In the model, the channel is in the faded state for 8060 seconds in a normal heavy fading month. Thus the channel can be viewed as being in one of two states where:

P {unfaded state} = 0.99689
P {faded state (Rummler model operative)} = 0.00311.

In what follows we employ this model to estimate the outage distributions for various communication methods. The model should be regarded as a working assumption valuable in gaining initial insight into the potential of the communication techniques we consider. However, we emphasize that more measurements may be required to refine Rummler's model to accommodate different geographical situations and wider bandwidths than 25 MHz, and to sharpen the accu-

racy of the representation of the more severe fades that are of major concern in what follows.

VI. OUTAGE OBJECTIVES AND SOME PROSPECTIVE INDEX VALUES

From the proposed performance objectives for the digital transmission network,[19] we have that the round trip system availability objective is 99.98 percent. So the probability of outage is 0.0002 round trip or 0.0001 one way. The 0.0001 breaks down as 0.00005 for fading, 0.000025 for equipment failure, and 0.000025 for maintenance and plant errors. Thus, for a 4000-mile system composed of 156 hops, each with a nominal length of 25.6 miles, we get a per hop outage probability of 3.2×10^{-7} for fading. This corresponds to about 10 seconds of outage per year. If we assume the year is composed of three heavy fading months and nine months with no fading, we obtain that the probability of outage in a heavy fading month is 1.28×10^{-6}. For a 250-mile short-haul system, the outage objective is 16 times less stringent on a hop, namely, 2.05×10^{-5}.

For the purpose of discussion we shall later consider the possibility of accommodating two DS-3 digital signals in a 20-, 30-, and 40-MHz channel. Each DS-3 signal corresponds to 672 64 kb/s voice circuits, so that two DS-3 signals correspond to about 90 Mb/s. Thus, for 20-, 30-, and 40-MHz channels we require 4.5, 3, and 2.25 bits per cycle, respectively. We will use 10^{-4} as the probability of bit-error threshold for registering outage. The sensitivity to this threshold will also be analyzed.

VII. COMPUTER PROGRAM

A comprehensive FORTRAN program was written to compute and display outage distributions. The program is composed of three main segments.

The first segment simulates the power transfer characteristic for the channel in the faded state. It uses a PORT routine to generate random numbers uniformly distributed on [0, 1] and functions of these are evaluated to produce the random variables with the three densities underlying Rummler's model. The variables A and B are appropriately correlated. The random channel characteristics are then computed and a file containing them is produced. The file contains 25,000 characteristics.

The second segment calculates the efficiency index for each channel and then computes the probability distribution function of the indexes. Various options and parameters can be chosen in exercising this stage. These include:

(i) Method of communication (i.e., type of linear equalization, decision feedback equalization, MLSE, and the information theoretic optimum processing)

(*ii*) Transmitter spectrum, i.e., optimization of the transmitter spectral density for a given average power constraint or flat power spectral density

(*iii*) Probability of bit-error objective

(*iv*) Unfaded signal-to-noise ratio at the receiver input

(*v*) Channel bandwith.

Option (*ii*) is only available for the decision feedback and the information theoretic optimum communication schemes. Extending the option to the other schemes seemed inadvisable because of the closeness of the results, as will be seen later.

The number 25,000 was found through computational experience to stabilize the density tail in the range of interest and yet not be wasteful of computer resources. Since the number 25,000 is very close to the number of experimental records of fade characteristics, we could have worked from original experimental data. We elected to work with the Rummler model since it is weighted to track the worse fades, which are our interest here, and since the model is widely accepted.

The final segment of the computer program provides labeled plots of the outage distribution functions. It uses the graphic package DISSPLA.

VIII. PRINCIPAL RESULTS

8.1 Preparatory remarks

For the purpose of presenting our principal results we will need the following notation for the outage distribution functions:

F_{PH}: phase distortion removed

F_{LIN}: optimum linear equalization

F_{DF}: postcursive intersymbol interference (ISI) removed

F_{MF}: all ISI removed (matched filter bound)

F_{IT}: information theory limit (Shannon).

The efficiency index distributions were computed for 30-MHz channels. Strictly speaking, the notion of an index in bits per cycle is imprecise in that $F(I)$ (the probability distribution of bits per cycle) would change if calculated at 20 MHz or at 40 MHz. However, by calculation we established that, for the purposes of the discussion that follows, treating $F(I)$ as invariant over the 20- to 40-MHz range of bandwidths is an adequate approximation.

In the actual development of systems of the kind we have idealized, much more detailed performance analysis is required than that reported here. One important aspect we have not considered is the effect of excess bandwidth associated with practical filter designs with rolloff factors other than zero. To get a preliminary idea of the effect of rolloff of an amount α on the communication efficiency indexes, we would simply scale the distributions abscissas by an amount $1/(1 + \alpha)$.

Equivalently, we would inflate the desired number of bits per cycle by $1 + \alpha$ before going to the curves. Thus, in considering the accommodation of two DS-3 signals, with $\alpha = 1/3$, we would inflate the 4.5-, 3-, and 2.25-bit per cycle values corresponding to 20-, 30-, and 40-MHz channels by 1.333 to obtain 6, 4, and 3 bits per cycle. We would then consult the derived curves at these values to obtain the outage probabilities. For the purpose of discussion in the section that follows, we use these three inflated bits per cycle values along with the long-haul and short-haul objectives of 1.3×10^{-6} and 2.1×10^{-5}, respectively, given in Section VI. Subsequently, an alternative means of accounting for α will be given.

8.2 The graphs

The most striking features of the outage distribution functions $F(I; P_e, \rho)$ are exhibited in Fig. 2. The beneficial effects of adaptive equalization are apparent. The three equalization schemes yield roughly similar results; however, as one looks more closely at the extreme outage tail, F_{LIN}, F_{DF}, and F_{MF} begin to depart from each other. F_{IT} is displaced over two bits per cycle to the right of F_{MF}, while

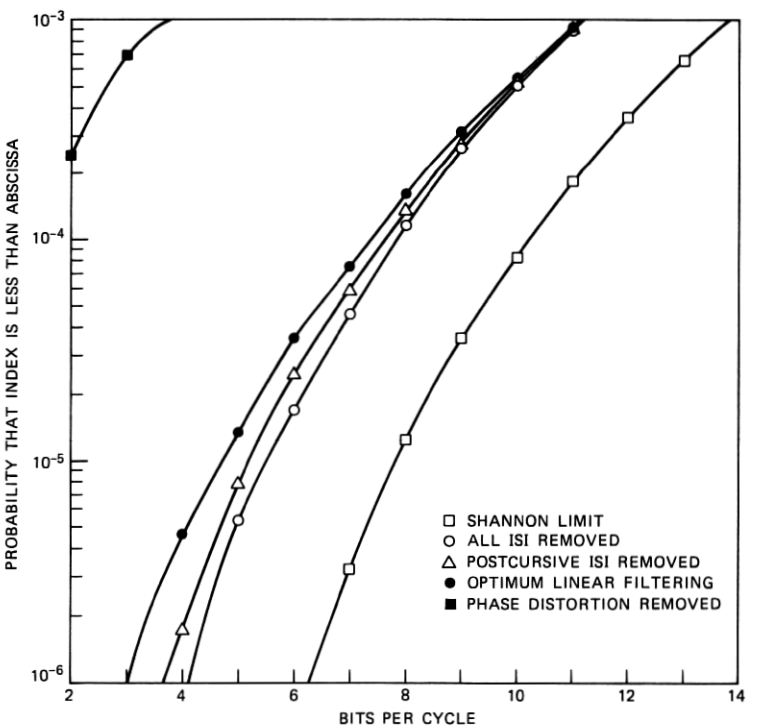

Fig. 2—Comparison of index distribution tails at 63-dB s/n, p.e. $< 10^{-4}$ (p.e. does not apply to the Shannon limit curve).

F_{PH} is very substantially to the left of F_{LIN}. For the 40-MHz band (3 bits/cycle), both outage objectives are met with linear equalization. For the 30-MHz band (4 bits/cycle) and the long-haul objective, linear and decision feedback equalization are not adequate and some coding or use of maximum likelihood sequence estimation are possible solutions. However, it may be practical to overcome the shortfall by some other means, such as improving the amplifier noise figure. For the 20-MHz channel (6 bits/cycle) the long-haul objective is not met. Also, this efficiency is so close to the information theory limit that any attempt to achieve it by coding may be ill-advised because of complexity. On the other hand, with some moderate coding the short-haul objective for 20-MHz channels should be attainable. For the other two bands, short-haul objectives are roundly met.

The plot for the equalizer that inverts the channel is not shown, as it is not perceptibly different from that for the optimum linear equalizer. This is expected since the optimum linear filter is essentially inverting the channel at the high signal-to-noise ratios we are considering.

Figures 3, 4, and 5 show the sensitivity of $F(I, P_e, \rho)$ to P_e. The

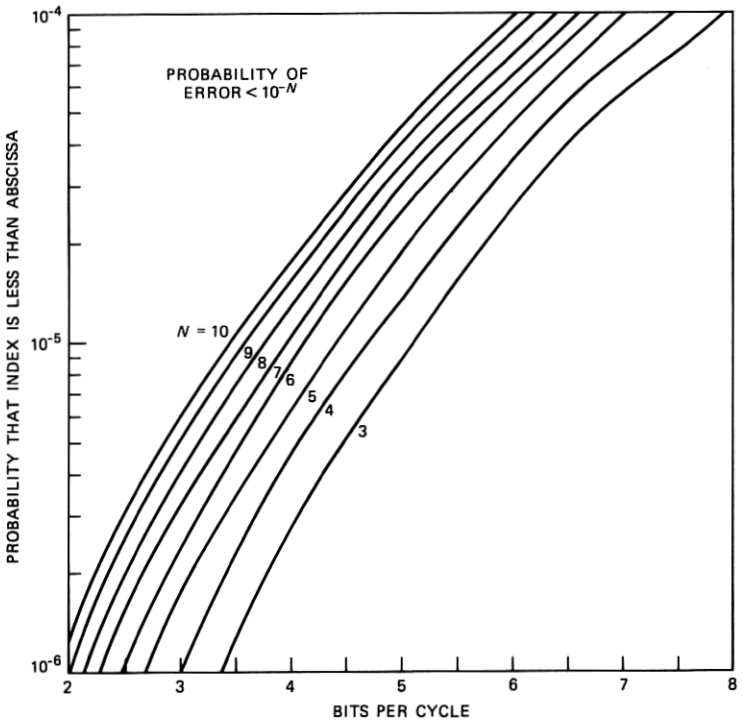

Fig. 3—Index distribution tail sensitivity to probability of error for linear equalization at 63-dB s/n.

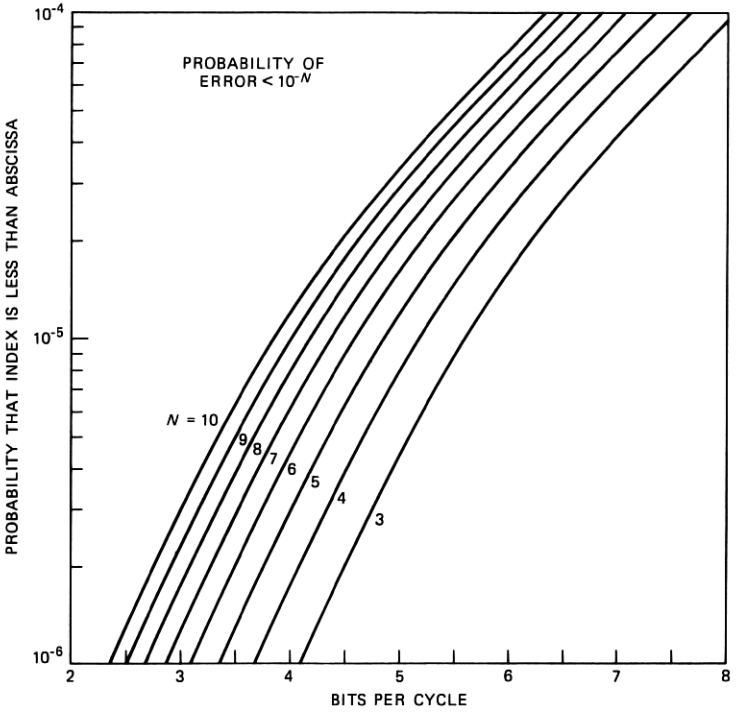

Fig. 4—Index distribution tail sensitivity to probability of error for decision feedback at 63-dB s/n.

sensitivity is especially small at the low error rates needed for data (as opposed to voice) transmission. An asymptotic analysis shows that, for large ρ, the curves translate to the left in accordance with a $\log_2 N$ shift, where 10^{-N} is the P_e objective. This insensitivity is an illustration of the well-known result[20] that once a pulse code modulation (PCM) operating point is achieved it takes a very small improvement to make the error rate an order of magnitude smaller. In fact, if at some operating data rate the probability of error turns out to be $10^{-5} - 10^{-6}$, one should be able to design an error-correcting code of small redundancy and moderate complexity that could improve the error rate by several orders of magnitude.

Figures 6 through 9 illustrate the sensitivity to signal-to-noise ratio. The translation in all cases is roughly 1/3-bit/cycle/dB. Note the curves for the Shannon limit have an ordinate range of 4 to 10 bits per cycle, while the others range from 2 to 8 bits per cycle. No sensitivity for F_{PH} is given since, unlike all the other distributions, there is negligible improvement as ρ increases. This is because, as ρ increases, the effect of intersymbol interference (ISI) remains and nothing is being done to mitigate it. In the other four cases, ISI tends to be

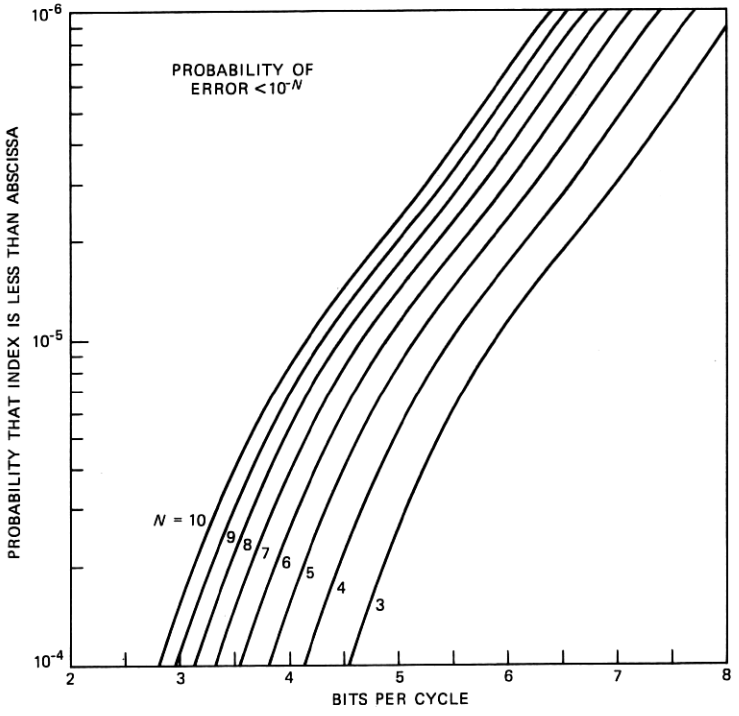

Fig. 5—Index distribution tail sensitivity to probability of error for the case of all ISI removed.

eliminated as ρ increases (the channels can be perfectly equalized without an inordinate amount of noise enhancement).

In Section 8.1 we mentioned the $(1 + \alpha)^{-1}$ scaling as a method of accounting for rolloff. This assumed $\mathscr{W} = 1/T$ so that $(1 + \alpha)\mathscr{W}$ is the actual bandwidth. Suppose instead that the real bandwidth is fixed at $\mathscr{W}$ but the data rate is slowed by an amount $(1 + \alpha)$, leaving the average transmitter power and N_o constant. Then the true s/n is increased by $10 \log_{10}(1 + \alpha)$. From this alternative perspective the suggested $(1 + \alpha)^{-1}$ scaling would be supplemented by a shift to the right of the probability distribution function tail by approximately $(10/3)\log_{10}(1 + \alpha)$ bits per cycle. Whether in estimating the effect of rolloff one takes the perspective that the symbol rate or the bandwidth is fixed is a matter of convenience.

Next, we consider optimization of the transmitter power spectral density. There is a practical question as to whether such an optimization could be achieved since the fade characteristic, which is first determined at the receiver, would need to be relayed back to the transmitter in time to be useful. However, the question is academic since we demonstrated that, even if an optimized transmitter could be

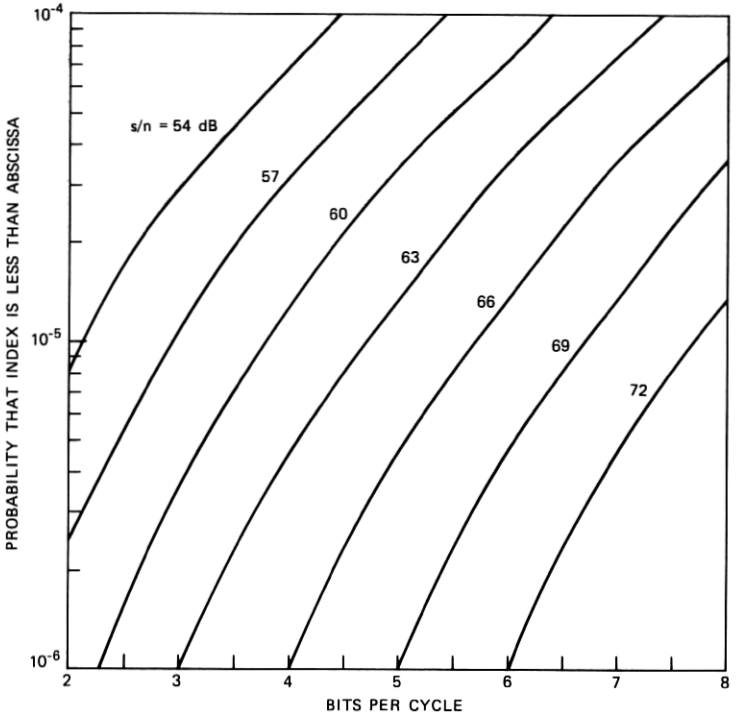

Fig. 6—Index distribution tail sensitivity to s/n for linear equalization, p.e. $< 10^{-4}$.

adapted in real time, the performance benefit would be negligible.

To understand why it is not worthwhile to optimize the transmitted power spectral density, we first consider the information theory limit that was anlayzed in Section 4.7. Our detailed numerical work has shown that a plot of the tail of the index distribution for the information theory limit under the assumption of an optimized transmitter would be imperceptibly different from the F_{DF} tail plotted in Fig. 2. This closeness of the two distributions follows from the fact that, for the severest fades in our data base of 25,000 channels, $|H|^{-2}$ is of the order of 10^{-6} and the terms involving $|H|^{-4}$ ($\sim 10^{-12}$) in the perturbation expression, eq. (23), are not enough to overcome the 10^{-14} multiplier.

Since the decision feedback index has the same form as (22), we can also conclude that the distribution tail corresponding to decision feedback would not be significantly altered if the optimum transmitter were used.

The decision feedback index and information theory limit on the index give imperceptible benefits when the power spectral density is optimized; therefore, it seems extremely unlikely that there is any

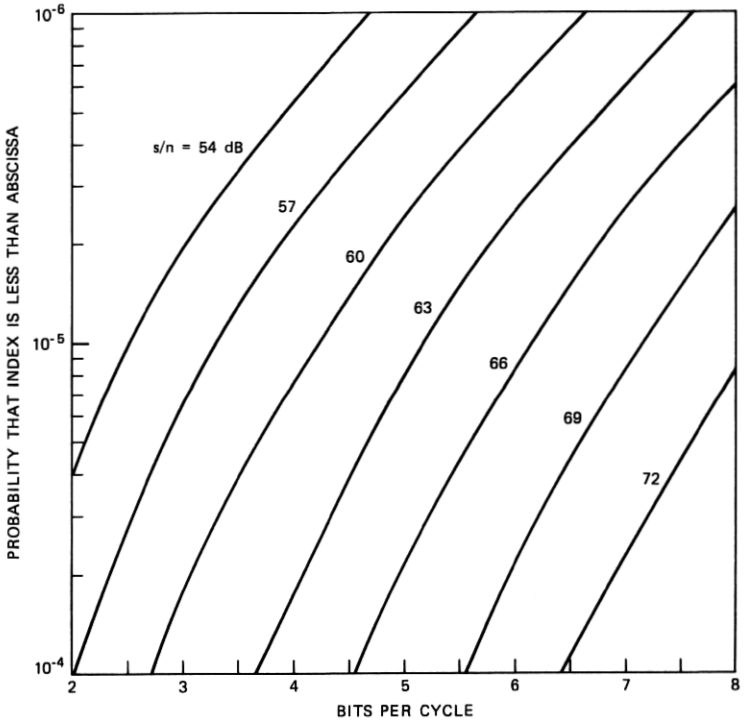

Fig. 7—Index distribution tail sensitivity to s/n for decision feedback, p.e. $< 10^{-4}$.

worthwhile benefit associated with optimizing the transmitter in the other cases.

IX. INITIAL ESTIMATE OF THE EFFECT OF FREQUENCY DIVERSITY

In implementations, digital radio systems are often protected with frequency diversity. In such systems impairments such as fading and equipment outages prompt the switching of communication traffic to a protection channel situated at a different frequency. The notation *mxn* means that *m* protection channels back up *n* working channels. So long as a protection channel is not occupied by an impaired channel, or is not itself impaired, it is available for temporary use in any of the *n* working systems. Some illustrations are 2×10 and 1×11 at 4 GHz, while at 6 GHz 2×6 and 1×7 are examples.

For FM systems the factor expressing the improvement in outage associated with frequency diversity is given by the expression $100/DG \cdot f_0 H^2$ in eq. (24) [corresponding to (34) and (35) in Ref. 21]. The parameter f_0 represents the carrier frequency in gigahertz and D denotes the path distance measured in miles. The parameter G depends

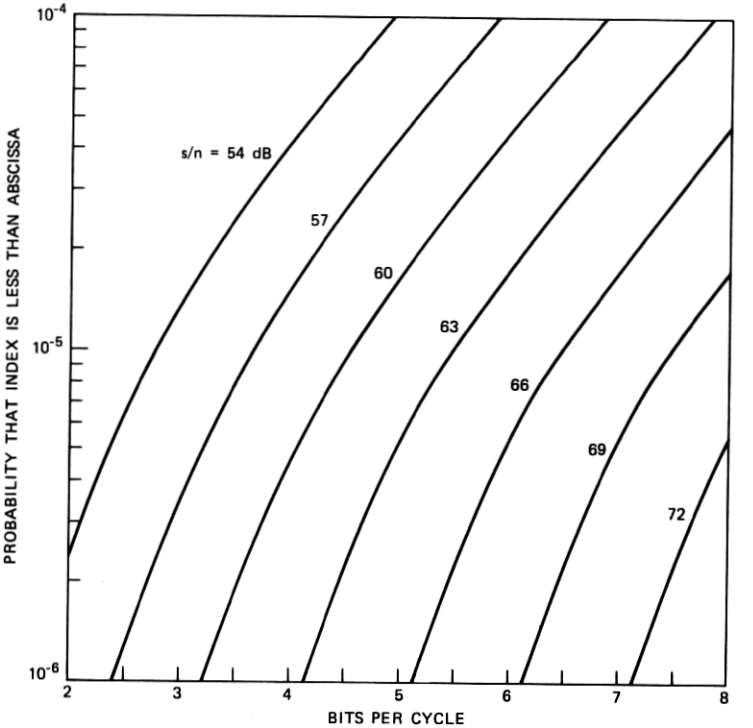

Fig. 8—Index distribution tail sensitivity to s/n for the case of all ISI removed, p.e. $< 10^{-4}$.

on the details of the frequency protection. G incorporates combinatorial effects corresponding to using m channels to back up n as well as empirical expressions involving the individual frequencies of channels involved. The term H is commonly expressed in decibels as $-20 \log H$ and is called fade margin. The fade margin is the smallest loss relative to the unfaded received signal at which the system fails. The notation H for the voltage level agrees with the previous use of H in this paper so long as the channel has a flat characteristic.

As pointed out in Ref. 22 the notation of a flat fade margin is considered meaningless in digital radio systems since the frequency-selective aspect of the fade characteristic appears necessary to describe performance of a channel. As of this writing we are not aware of any method in the extant literature for extending our results to include the effect of frequency diversity. However, for the special case of optimum linear equalization there is a way to introduce an equivalent flat fade margin so as to enable the use of (24) in making a (preliminary) estimate of the diversity effect. The estimation method was discovered in the course of generalizing the computer program to compute index

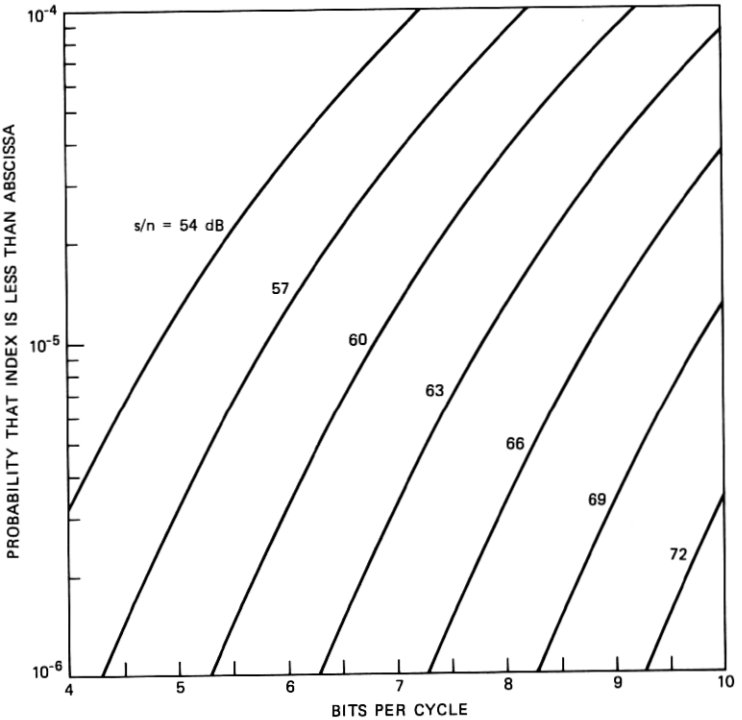

Fig. 9—Distribution tail sensitivity to s/n for the Shannon limit.

distributions for arbitrary bandwidths. The generalization was needed to investigate—and, as it turned out, to substantiate—the bandwidth insensitivity of the distribution in the 20- to 40-MHz range. The generalized program was also exercised for bandwidths an order of magnitude smaller and it was observed that the F_{LIN} distribution changed imperceptibly.* So the F_{LIN} tail in the range of interest here can be correctly obtained by using the univariate samples $|H(0)|^2$. Treating (18) as an equality and solving for $|H(0)|^2$ and then substituting in (24) gives an estimate of the improvement owing to frequency diversity.

For an illustration refer to Fig. 2 for which s/n = 63 dB and $Pe < 10^{-4}$. We see that with linear equalization, at the long-haul outage objective of 1.3×10^{-6}, 3.2 bits per cycle can be supported and the corresponding number for the short-haul objective is 5.3 bits per cycle. Using (24) and the G values from Ref. 21 we show in Table I the

* This is not true of the individual values of I and no claim is made for invariance of F_{DF}, F_{MF}, or F_{IT}.

Table I—Estimates of improved bits/cycle indices using frequency diversity

	Channel at 4 GHz		Channel at 6 GHz	
Short Haul	(1 × 11)	7.1	(1 × 7)	6.5
	(2 × 10)	7.8	(2 × 6)	7.1
Long Haul	(1 × 11)	5.3	(1 × 7)	4.6
	(2 × 10)	5.8	(2 × 6)	5.3

following estimates of improved indices when frequency diversity is employed.

X. ACKNOWLEDGMENT

Discussions with N. Amitay, A. Gieger, L. J. Greenstein, V. K. Prabhu, B. G. King, G. Vannucci, A. Vigants, C. B. Woodworth, W. D. Rummler and Y. S. Yeh were valuable in the course of the work reported here.

XI. POSTSCRIPT

Application of multilevel QAM in the radio channels might be inhibited by the amplitude (AM-AM) and (AM-PM) nonlinearities present in RF power amplifiers. A method for solving this problem without sacrificing amplifier power efficiency will be described in a forthcoming paper.[23]

REFERENCES

1. W. D. Rummler, "A New Selective Fading Model: Application to Propagation Data," B.S.T.J., *58,* No. 5 (May–June 1979), pp. 1032–71.
2. W. D. Rummler, "More on the Multipath Fading Channel Model," IEEE Trans. Commun., *COM-29,* No. 3 (March 1981), pp. 346–52.
3. A. Gersho and V. B. Lawrence, unpublished work.
4. G. Ungerbock, "Channel Coding with Multilevel/Phase Signals," IEEE Trans. Inform. Theory, *IT-28,* No. 1 (January 1982), pp. 55–67.
5. B. R. Saltzberg, "Intersymbol Interference Error Bounds with Application to Ideal Bandlimited Signaling," IEEE Trans. Inform. Theory, *IT-14,* No. 4 (July 1968), pp. 563–8.
6. M. S. Mueller and J. Salz, "A Unified Theory of Data Aided Equalization," B.S.T.J., *60,* No. 9 (November 1981), pp. 2023–38.
7. R. D. Gitlin and S. B. Weinstein, "Fractionally Spaced Equalization: An Improved Digital Transversal Equalizer," B.S.T.J., *60,* No. 2 (February 1981), pp. 275–96.
8. R. D. Gitlin and J. E. Mazo, "Comparison of Some Cost Functions for Automatic Equalization," IEEE Trans. Commun., *COM-21,* No. 3 (March 1973), pp. 233–7.
9. T. Berger and D. W. Tufts, "Optimal Pulse Amplitude Modulation, Part I, Transmitter-Receiver Design and Bounds from Information Theory," IEEE Trans. Inform. Theory, *IT-13,* No. 2 (April 1967), pp. 196–208.
10. J. Salz, "Optimum Mean-Square Decision Feedback Equalization," B.S.T.J., *52,* No. 8 (October 1973), pp. 1341–73.
11. D. D. Falconer and G. J. Foschini, "Theory of Minimum Mean-Square-Error QAM Systems Employing Decision Feedback Equalization," B.S.T.J., *52,* No. 10 (December 1973), pp. 1821–48.
12. S. P. Lloyd, "On a Measure of Stochastic Dependence," Theory of Probability and Its Application, No. 7, 1962, pp. 312–22.

13. M. S. Pinsker, "A Quantity of Information of a Gaussian Random Stationary Process, Contained in a Second Process Connected with it in a Stationary Manner," Doklady Akad. Nank S.S.S.R., New Series, *99*, 1954, pp. 213–16.
14. R. K. Mueller and G. J. Foschini, "The Capacity of Linear Channels with Additive Gaussian Noise," B.S.T.J. *49*, No. 1 (January 1970), pp. 81–94.
15. J. L. Holsinger, "Digital Communication over Fixed Time-Continuous Channels with Memory-with Special Application to Telephone Channels," Lincoln Laboratory, Technical Report 366, Lexington, MA, October 1964.
16. J. B. Thomas, "Statistical Communication Theory," New York: John Wiley and Sons, 1969, Chapter 8.
17. R. M. Fano, "Transmission of Information," New York: M.I.T. Press and John Wiley and Sons, Inc., 1961.
18. B. G. King, unpublished work.
19. M. A. Rezny and J. S. Wu, unpublished work.
20. B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The Philosophy of PCM," Proc. IEEE (November 1948), pp. 1324–31.
21. A. Vigants and M. V. Pursley, "Transmission Unavailability of Frequency—Protected Microwave FM Radio Systems Caused by Multipath Fading," B.S.T.J., *58*, No. 8 (October 1979), pp. 1779–96.
22. A. Gieger and W. T. Barnett, "Effects of Multipath Propagation on Digital Radio," IEEE Trans. Commun., *COM*-29, No. 9 (September 1981), pp. 1345–52.
23. A. A. M. Saleh and J. Salz, "Adaptive Linearization of Power Amplifier Nonlinearity in Digital Radio Systems," B.S.T.J., *62*, No. 4 (April 1983).