<table>
<thead>
<tr>
<th>Title</th>
<th>Channel capacity and error exponents of variable rate adaptive channel coding for Rayleigh fading channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Lau, KN</td>
</tr>
<tr>
<td>Citation</td>
<td>IEEE Transactions on Communications, 1999, v. 47 n. 9, p. 1345-1356</td>
</tr>
<tr>
<td>Issued Date</td>
<td>1999</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/42849">http://hdl.handle.net/10722/42849</a></td>
</tr>
<tr>
<td>Rights</td>
<td>©1999 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Abstract—We have evaluated the information theoretical performance of variable rate adaptive channel coding for Rayleigh fading channels. The channel states are detected at the receiver and fed back to the transmitter by means of a noiseless feedback link. Based on the channel state informations, the transmitter can adjust the channel coding scheme accordingly. Coherent channel and arbitrary channel symbols with a fixed average transmitted power constraint are assumed. The channel capacity and the error exponent are evaluated and the optimal rate control rules are found for Rayleigh fading channels with feedback of channel states. It is shown that the variable rate scheme can only increase the channel error exponent. The effects of additional practical constraints and finite feedback delays are also considered. Finally, we compare the performance of the variable rate adaptive channel coding in high bandwidth-expansion systems (CDMA) and high bandwidth-efficiency systems (TDMA).

I. INTRODUCTION

ERROR correction codes have been widely used to combat the effect of Rayleigh fading in mobile radio channels. In traditional FEC schemes [1], [2], fixed rate codes were used which failed to explore the time varying nature of the channel. To keep the performance at a desirable level, they were designed for the average or worst case situation. To better exploit the time varying nature of the channel, adaptive channel coding based on feedback channel state has been proposed. The performance of uncoded variable rate and power transmission schemes for Rayleigh fading channel based on the feedback of channel state information has been considered in [3]–[7]. Many practical adaptive error correction codes have been proposed in recent years to reduce the bit error rate and to increase throughput of the mobile radio channels [8]–[14], [15]. In this paper, we model a general scheme of variable rate adaptive channel coding which varies the code rate according to the channel condition and explore the fundamental reasons why there is a performance improvement over fixed-rate coding.

We investigate the information theoretical performance, namely the channel capacity and the error exponent, of Rayleigh fading channels using variable rate adaptive channel coding (VRAECC) with constant transmitted power. Channel capacity describes the maximum allowable bit rate for reliable transmission across a channel. Error exponent describes how fast error probability drops w.r.t. block length. Based on the channel state informations, the transmitter can adjust the rate of the channel coding scheme accordingly. We try to answer the following questions in this paper.

• Is channel capacity or error exponent increased by using VRAECC?
• What are the optimal rate control functions that maximize the error exponent?
• Does VRAECC perform better in high or low bandwidth expansion?

An equivalent discrete time channel model is developed in Section II. For simplicity, coherent detection and ideal interleaving are assumed. The error exponent and channel capacity of a Rayleigh fading channel with feedback of channel state using constant input VRACE are evaluated in Sections III and IV, respectively. Numerical results are presented and discussed in Section V. Finally, we conclude with a brief summary in Section VI.

II. CHANNEL MODELING AND INDUCED STATE DISTRIBUTION

A. Physical Channel Model

The physical Rayleigh fading channel is a bandlimited continuous-time channel in which the channel input can be modeled by a bandlimited complex random process. The random process is segmented into a number of channel symbols with the $i$th channel symbol $X_i(t)$ having a variable duration $T_i$, as shown in Fig. 1. To maintain generality, no modulation format is specified. Variable rate channel encoder is integrated

1In particular, channel capacity of downlink fading channels with variable power schemes has been investigated in [16]. It is shown that the optimal power distribution that maximize the channel capacity is achieved by water-filing in time domain. However, due to the rapid power control required to compensate for the channel fading in the scheme, it is not feasible with nowadays power-amplifier technology. Variable rate transmission, on the other hand, is more feasible and hence, we focus on the optimal variable rate schemes in this paper. As shown in Section IV, variable rate schemes cannot increase the channel capacity. Instead, we aim at increasing the error exponent with variable rate schemes. Optimal rate control rule is in the sense of maximizing the error exponent.
Fig. 1. Segmentation of a bandlimited random process $\tilde{X}(t)$ into a channel symbol $\tilde{X}$.

$$\tilde{X}(t)$$

Fig. 2. Equivalent discrete-time channel model for variable duration scheme.

into variable throughput modulator with an average power constraint $P_0$ only. The $i$th channel output (in complex low-pass equivalent) $\tilde{Y}_i(t)$ is given by

$$\tilde{Y}_i(t) = C_i(t)\tilde{X}_i(t) + \tilde{N}_i(t)$$

where $C_i(t)$ and $\tilde{N}_i(t)$ are channel fading attenuation and complex white Gaussian noise for the $i$th channel symbol, respectively.

It has been shown that a continuous-time complex signal which is approximately time-limited to $T_i$ and bandlimited to $W$ can be represented by a $2WT_i$-dimensional vector in the signal space spanned by the Prolate spheroidal wave functions [17]. Hence, the $i$th channel symbol is represented by a $2WT_i$-dimensional vector $\tilde{X}_i$. Assume that $T_i$ and $W$ are both much smaller than the coherence time, and the coherence bandwidth $\gamma_i$ can be considered as a constant in every dimension of the signal space. Hence, the continuous-time model is reduced to a discrete-time model.

2 Since coherent detection is assumed, channel phase variation is corrected by the receiver and hence, WLOG, the channel phase reference is set to zero.

3 We have a flat fading channel.

B. Equivalent Discrete-Time Channel Model

The equivalent channel is a discrete-time, continuous-input and continuous-output channel with feedback. There is a channel state $C_i$ associated with the $i$th channel symbol. The channel state is available to the receiver and known to the transmitter via a feedback channel with a certain unavoidable delay, $\Delta$ seconds. For each $C_i$, there is a corresponding prediction, denoted by $\hat{Z}_i$, at the transmitter. The channel states $C_i$ and $\hat{Z}_i$ (and hence, $\tilde{Z}_i$ and $\tilde{Z}_j$) are correlated but through ideal interleaving, they become i.i.d. and the channel becomes a memoryless channel.

The $i$th channel output is given by

$$\tilde{Y}_i = C_i\tilde{X}_i + \tilde{N}_i$$

where $\tilde{N}_i$ is an uncorrelated Gaussian noise variable with variance $\mathcal{E}[\tilde{N}_i^2] = \gamma_0WT_i$ and $\gamma_0$ is the white noise spectral density. We assume $\mathcal{E}[\tilde{N}_i^2] = 1$.

For the variable duration scheme, symbol duration of the $i$th channel symbol $T_i$ is varying according to the predicted channel state $\hat{Z}_i$. The channel model is illustrated in Fig. 2. Each channel symbol carries a constant number of information bits $R$ with a varying dimension $2WT_i$, which is a function
Fig. 3. Equivalent discrete-time channel model for variable input scheme.

of predicted channel state $\hat{Z}_i$. Random block coding [18] with block length $N$ is used. An index $\Omega \in [1,2^N-1]$ is fed into the variable dimension channel encoder, giving out a codeword of $N$ channel symbols. Each channel symbol in a codeword $X_i$ is generated randomly according to a continuous distribution $Q(\tilde{X}_i | \hat{Z}_i)$. This forms a random codebook of size $2^N$ which is known both to the transmitter and the receiver. The channel can be described by a channel transition density $p_0(\tilde{X}_i | \hat{Z}_i)$. The dependence of $Q$ and $p_0$ on $\hat{Z}_i$ comes from the dependence of the symbol dimension $2WT_0$ on $\hat{Z}_i$.

For the variable input scheme, the $i$th symbol duration $T_0$ is a constant given by $T_0$. Fig. 3 illustrates the equivalent discrete-time channel model. The $i$th channel symbol carries $R_0$ information bits and variable throughput is achieved by varying $R_0$ w.r.t. the predicted state $\hat{Z}_i$. Random block coding is used and the $i$th channel symbol is generated randomly according to a continuous time distribution $Q(\tilde{X}_i | \hat{Z}_i)$. The overall size of the random codebook is $2^{WRT_0}$ which is known both to the transmitter and the receiver. The channel can be described by a transition probability $p_0(\tilde{Y}_i | \hat{Z}_i, C_i)$. Note that due to the constant symbol duration, the channel transition probability is independent of the predicted state $\hat{Z}_i$.

For both schemes, channel outputs, together with channel states and predicted states, are fed into a deinterleaver and a maximum-likelihood decoder at the receiver. 4 The decoder produces an estimate of the transmitted index $\hat{\Omega}$ and an error occurs when $\Omega \neq \hat{\Omega}$.

C. Induced State Distribution

For the variable duration scheme, the sequence of symbol duration $\{T_1(\hat{Z}_1), \ldots, T_N(\hat{Z}_N)\}$ is varying according to the sequence of predicted channel states $\hat{Z}_N = \{\hat{Z}_1, \ldots, \hat{Z}_N\}$, and hence, it induces a probability density on $\hat{Z}_N$ which is different from the original fading density in general. For simplicity, we assume a simple prediction rule $\hat{Z}(t) = c(t-D)$. Hence, $\hat{Z}(t)$ is an ergodic random process and it is shown in Appendix A that the induced probability density on $\hat{Z}_N$, denoted by $P_N(\hat{Z}_N)$, is given by

$$P_N(\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_N) = P(\hat{Z}_1)P(\hat{Z}_2) \cdots P(\hat{Z}_N)$$  (3)

where

$$P(\hat{Z}) \triangleq \lim_{N \to \infty} \frac{N_0}{N} = \frac{f(\hat{Z})}{E_fT(\hat{Z})}.$$  (4)

$N_0$ is the number of symbols with $\hat{Z} \in [\hat{Z}, \hat{Z} + \Delta]$ in a sequence of $N$ symbols, $f(\hat{Z})$ is the fading density and $E_f = \int_{\hat{Z}} (1/T(\hat{Z})) f(\hat{Z}) d\hat{Z}$. Furthermore, $E_f$ is shown to be the average symbol rate (number of channel symbols per second). Since given $\hat{Z}$, the symbol duration $T_i$ is constant, the conditional density $P(c | \hat{Z})$ is not affected by the varying symbol duration and is given by [21]

$$P(c | \hat{Z}) = f(c | \hat{Z})$$

$$= \frac{2\sigma}{(1-\lambda^2)} \exp \left[ -\frac{\sigma^2 + \lambda^2}{(1-\lambda^2)} \right] I_0 \left( 2\sigma^2 \frac{\lambda}{1-\lambda^2} \right)$$  (5)

where $\lambda^2 = \int_0 f_d(\Delta) d\Delta$, $f_d$ is the Doppler spread, $I_0$ is the zeroth-order Bessel function, and $I_0$ is the zeroth-order modified Bessel function.

For the variable input scheme, the symbol duration is constant and hence, the induced density reduces to original standing fading density.

III. ERROR EXPONENT FOR VARIABLE RATE SCHEMES

We shall bound the average codeword error probability by an average error exponent using Gallager’s approach [22]. Given a sequence of channel states $\hat{X}_N$ and a sequence of
predicted channel states \( \tilde{z}_N \), the channel is equivalent to an additive white Gaussian noise (AWGN) channel. Hence, the conditional codeword error probability \( P_e(\tilde{c}_N, \tilde{z}_N) \) is bounded by [22]

\[
P_e(\tilde{c}_N, \tilde{z}_N) \leq 2^{-N[E(\rho, Q) - \rho R_b]} \tag{6}
\]

for all \( N \geq 1 \) and \( \rho \in [0, 1] \). \( E(\rho, Q, \tilde{c}_N, \tilde{z}_N) \) is called the Gallager’s error exponent and is given by

\[
E(\rho, Q, \tilde{c}_N, \tilde{z}_N) = \frac{1}{N} \sum_{j=1}^{N} \log_2 \left\{ \int_{x_j} Q(x \mid \tilde{z}_j) \right\} \tag{7}
\]

To determine the average error probability, we have to uncondition (6) w.r.t. \( \tilde{c}_N \) and \( \tilde{z}_N \). The average codeword error probability is given by

\[
P_e = \int_{\tilde{c}_N} \int_{\tilde{z}_N} P_e(\tilde{c}_N, \tilde{z}_N) \, d\tilde{c}_N \, d\tilde{z}_N \tag{8}
\]

where \( P(\tilde{c}_N, \tilde{z}_N) \) is the joint density of \( c_i \) and \( z_i \).

By symmetry, the 4th channel symbol \( X_i \) has 2WTl i.i.d. components \( [X_{i,1}, \ldots, X_{i,2WTl}] \). Hence, \( Q(x_i \mid \tilde{z}_i) \) and \( p_0(\tilde{y}_j \mid x_{i,2i}\tilde{z}_i) \) can be expressed using product forms as

\[
Q(x_i \mid \tilde{z}_i) = \prod_{j=1}^{2WTl} Q(x_{i,j} \mid \tilde{z}_i) \tag{9}
\]

\[
p_0(\tilde{y}_j \mid x_{i,2i}\tilde{z}_i) = \prod_{j=1}^{2WTl} p_0(y_{i,j} \mid x_{i,j}\tilde{z}_i). \tag{10}
\]

Separating (8) into product of \( N \) integrals and using (6), (9), and (10), the average error probability \( P_e \) is given by

\[
P_e \leq 2^{-N[E(\rho, Q) - \rho R_b]} \tag{11}
\]

for all \( \rho \in [0, 1] \) and \( Q(x \mid \tilde{z}) \) where \( E(\rho, Q) \) is given by

\[
E(\rho, Q) = -\log_2 \left\{ \int_{x} P(\tilde{z}) \right\} \left\{ \int_{x} Q(x \mid \tilde{z}) \right\} \tag{12}
\]

\[
\times p_0(y \mid x)^{1/\rho} \, dx \right\} \tag{12}
\]

where \( R_b \) is the joint density of \( c_i \) and \( z_i \).

A. Variable Duration Scheme

Since \( \{T_1(\tilde{z}_1), \ldots, T_N(\tilde{z}_N)\} \) is an i.i.d. sequence and by the weak law of large numbers [23]

\[
\frac{1}{N}(T_1(\tilde{z}_1) + \cdots + T_N(\tilde{z}_N)) \rightarrow \mathbb{E}[T] = 1/\tilde{E}_f \text{ w.h.p.}
\]

will converge in probability to \( \mathbb{E}[T] = 1/\tilde{E}_f \), where the expectation is taken w.r.t. the induced density \( P(\tilde{z}) \). The average information bit rate (bits per second) \( \tilde{R}_b \) is given by

\[
\tilde{R}_b = \frac{NR}{T_1 + T_2 + \cdots + T_N} \rightarrow \infty \text{ RE}_f. \tag{13}
\]

Hence, we have

\[
\tilde{R} = \frac{\tilde{R}_b}{\tilde{E}_f} \tag{14}
\]

Expressing (11) and (12) in terms of \( \tilde{R}_b \), the average error probability \( P_e \) is given by

\[
P_e \leq 2^{-\tilde{R}_b [E(\rho, Q) - \rho \tilde{R}_b]} \tag{15}
\]

where \( E(\rho, Q) \) is given by

\[
E(\rho, Q) = E_f \tilde{E}(\rho, Q) \tag{15}
\]

\[
= -E_f \log_2 \left\{ \frac{1}{E_f} \int_{x} \frac{f(x \mid \tilde{z}) f(\tilde{z})}{T(\tilde{z})} \right\} \tag{15}
\]

\[
\times \left[ \int_{x} Q(x \mid \tilde{z}) p_0(y \mid x)^{1/\rho} \, dx \right] \frac{1}{2WTl} \tag{15}
\]

To obtain a tight error bound, we have to minimize \( P_e \) w.r.t. \( \rho \) and \( T(\tilde{z}) \). Since \( \rho \) is a parameter which is not measurable in practice, the functions \( Q(x \mid \tilde{z}) \) and \( T(\tilde{z}) \) are independent of the parameter \( \rho \) and hence, the optimization is decoupled.

1) Optimization w.r.t. \( \rho \): Define the average error exponent \( \mathbb{E}_e(\rho, Q, \tilde{R}_b) \) as

\[
\mathbb{E}_e(\rho, Q, \tilde{R}_b) = [E(\rho, Q) - \rho \tilde{R}_b]. \tag{17}
\]

We first prove the following lemmas.

Lemma 1: For any \( T(\tilde{z}) \) and \( Q(X \mid \tilde{Z}) \), \( \tilde{E}(\rho, Q) \) in (16) satisfies the following properties:

(i) \( 0 < \frac{\partial E}{\partial \rho} \leq E_f I(X; \tilde{z}^T \mid \tilde{z}) \). Equality holds iff \( \rho = 0 \).

(ii) \( \frac{\partial^2 E}{\partial \rho^2} < 0 \).

Proof: Refer to Appendix C-1.

Let

\[
R_0 = \frac{\partial E}{\partial \rho} \bigg|_{\rho = 0} \tag{18}
\]

\[
R_1 = \frac{\partial E}{\partial \rho} \bigg|_{\rho = 1}. \tag{19}
\]

The following summarizes the general result of the optimization w.r.t. \( \rho \).

Lemma 2: For any \( T(\tilde{z}) \) and \( Q(X \mid \tilde{Z}) \), the optimal error exponent is given by

\[
\tilde{E}_e(\rho, \tilde{R}_b) \triangleq \min_{\rho} \mathbb{E}_e(\rho, Q, \tilde{R}_b) = E(\rho^*, Q) - \rho^* \tilde{R}_b \tag{20}
\]

where \( \rho^* \in [0, 1] \), a function of \( \tilde{R}_b \), is given by

\[
\left\{ \begin{array}{ll}
\frac{\partial \tilde{E}_e}{\partial \rho} \bigg|_{\rho^*} = \tilde{R}_b & \text{if } R_1 \leq \tilde{R}_b \leq R_0, \\
\rho^* = 1 & \text{if } R_1 \leq \tilde{R}_b \leq 0.
\end{array} \right. \tag{21}
\]

Proof: Refer to Appendix C-2.

Collecting the above results, we have the following theorem.

Theorem 1: For any \( Q(X \mid \tilde{Z}) \) and \( T(\tilde{z}) \), if \( \tilde{R}_b < R_0 \), then \( \lim_{N \to \infty} P_e = 0 \).

Proof: Consider

\[
\frac{\partial \tilde{E}_e(\rho, \tilde{R}_b)}{\partial \tilde{R}_b} = \frac{\partial \tilde{R}_b}{\partial \rho} \frac{\partial \tilde{E}_e(\rho, \tilde{R}_b)}{\partial \rho} = \frac{\partial R_1}{\partial \rho} \frac{\partial \tilde{E}_e(\rho, \tilde{R}_b)}{\partial \rho} = -\rho < 0 \tag{22}
\]

for \( \tilde{R}_b \in [R_1, R_0] \), \( \frac{\partial \tilde{E}_e(\rho, \tilde{R}_b)}{\partial \tilde{R}_b} = -1 \) for \( \tilde{R}_b \in [0, R_1] \).
Hence, $E_p^*$ is strictly decreasing w.r.t. $R_b$. At $R_b = R_0$, $ho^* = 0$ and $E_p^*(Q^*, R_b) = 0$. If $R_b < R_0$, we have $E_p^*(R_b) > E_p^*(R_0) = 0$ and hence from (15), $P_e \to 0$ as $N \to \infty$. \hfill \square

2) Optimization w.r.t. $Q(x \mid \hat{z})$: For simplicity, take $Q$ to be a capacity achieving distribution, $Q^*$, which maximizes the mutual information. Given a channel state $C_t$ and a predicted state $\hat{z}_t$, the channel is memoryless and is equivalent to an AWGN channel. By symmetry, the capacity achieving distribution [22], [11] $Q^*(x_{i,j} \mid \hat{z}_i)$ would be a zero-mean Gaussian density with variance $\sigma_X^2 = R_0/(2W)$. Since the variance $\sigma_X^2$ is independent of $\hat{z}_t$, $Q^*(x_{i,j} \mid \hat{z}_i)$ is independent of $\hat{z}_t$; we shall drop the conditional notation of $\hat{z}$ in $Q^*$ hereafter.

The remaining problem is to minimize $P_e$ w.r.t. $T(\hat{z})$.

3) Optimization w.r.t. $T(\hat{z})$: In this section, we minimize $P_e$ w.r.t. $T(\hat{z})$. For any given $R_b$, take $\rho = \rho^*$ as in (21) and $Q(x) = Q^*(x)$. Two situations, namely negligible feedback delay and significant feedback delay, are considered as follows.

a) Small feedback delay: We assume feedback delay is small relative to the channel coherence time. Therefore, $\hat{z}(t) = c(t - \Delta) \approx c(t)$ and

$$f(c \mid \hat{z}) \to \delta(c - \hat{z}).$$

By (22), rewrite $P_e$ in (15) as

$$P_e \leq 2^{-N \rho^*} \frac{1}{E_f} \prod_{i=1}^{N} \left\{ \int_{\hat{z}_i} f(\hat{z}_i) \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \right. \right. \right.$$

$$\times \left. p_c(y \mid x, \hat{z}_i)^{1/\rho^*} \left[ \right. \int_{x} Q^*(x) \right]^{\rho^*+1} \right] dx \right. \left. \int_{x} Q^*(x) \right\}^{2W T(t)} \left. d\hat{z}_j \right\}. \tag{23}$$

The optimization problem is equivalent to choosing $\{T_1, T_2, \ldots, T_N\}$ that minimize $P_e$ in (23) under the constraints

$$\frac{1}{E_f} \int_{\hat{z}_i} f(\hat{z}_i) \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \right. \right. \right.$$

$$\times \left. p_c(y \mid x, \hat{z}_i)^{1/\rho^*} \left[ \right. \int_{x} Q^*(x) \right]^{\rho^*+1} \right] dx \right. \left. \int_{x} Q^*(x) \right\}^{2W T(t)} \left. d\hat{z}_j \right\} = 1 \quad \forall i \in [1, N] \tag{24}$$

$$T(\hat{z}_i) \geq T_1 \quad \forall i \in [1, N] \tag{25}$$

$$T(\hat{z}_i) \leq T_p \quad \forall i \in [1, N]. \tag{26}$$

Constraint (24) is due to the fact that total area under the induced density $P(\hat{z}_i)$ should be equal to one. Constraint (25) is to set a lower limit on $T(\hat{z}_i)$ so that $2WT_i \geq 2$. Constraint (26) is to set a peak limit on $T(\hat{z}_i)$ so that symbol duration is smaller than the channel coherence time.

By the Calculus of Variations, it is shown in Appendix B that the optimal $T(\hat{z})$ is given by

$$T(\hat{z}) = \frac{A_0(\rho^*)}{\log_2 \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \left( x, \hat{z}_i \right)^{1/\rho^*+1} \right] \right]} \tag{27}$$

where $\hat{z}_b$ is given by the solution of

$$T_p = \frac{A_0(\rho^*)}{\log_2 \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \left( x, \hat{z}_i \right)^{1/\rho^*+1} \right] \right]} \tag{28}$$

$$\hat{z}_b \text{ is given by the solution of } \tag{29}$$

$$T_i = \frac{A_0(\rho^*)}{\log_2 \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \left( x, \hat{z}_i \right)^{1/\rho^*+1} \right] \right]} \tag{29}$$

Intuitively, a longer symbol duration should be used to encode $R$ information bits when the predicted state $\hat{z}$ is small. Substituting (27) into the constraints (24), (25), and (26), the constant $A_0(\rho^*)$ is given by

$$A_0(\rho^*) = \frac{1}{E_f} \left\{ \int_{\hat{z}_p} f(\hat{z}_i) \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \right. \right. \right.$$

$$\times \left. p_c(y \mid x, \hat{z}_i)^{1/\rho^*} \left[ \right. \int_{x} Q^*(x) \right]^{\rho^*+1} \right] dx \right. \left. \int_{x} Q^*(x) \right\}^{2W T(t)} \left. d\hat{z}_j \right\} \cdot f(\hat{z}) d\hat{z} + \log_2 \int_{x_j} \left[ \int_{x} Q^*(x) \right. \right.$$

$$\times \left. p_c(y \mid x, \hat{z}_i)^{1/\rho^*} \left[ \right. \int_{x} Q^*(x) \right]^{\rho^*+1} \right] dx \right. \left. \int_{x} Q^*(x) \right\} \tag{30}$$

The optimal error exponent $E_p^*(Q^*, R_b)$ is found by solving the simultaneous equations of (20), (21) in Lemma 2, as well as (27) numerically for any given $R_b \in [0, R_0]$. 

b) Large feedback delay: When feedback delay is large, (22) no longer holds. Since the integrand of $P_e$ is not separable w.r.t. $c$ and $\hat{z}$, it is not possible to obtain a closed-form expression for the optimal $T(\hat{z})$. We shall investigate the effect of feedback delay on the performance using the control rule in (27) instead.

4) Overall Result of Optimizations for Variable Duration Scheme: Given $R_b$, take $\rho^*$, $Q^*(x \mid \hat{z})$ and $T(\hat{z})$, $R_b$ to be the optimal parameter, the capacity achieving distribution and the optimal symbol duration control rule, respectively. The solution of the average error exponent $E_p^*(Q^*, R_b, \Delta)$ with delay $\Delta$ is given by the following.

i) $0 \leq R_b \leq R_1$: Using $T(\hat{z}, R_b)$ in (27) with $\rho^* = 1$, the solution is given by

$$E_p^*(Q^*, R_b, \Delta) = \frac{1}{2} \log_2 \left[ \frac{1}{E_f} \left\{ \int_{\hat{z}_p} f(\hat{z}_i) \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \right. \right. \right.$$

$$\times \left. p_c(y \mid x, \hat{z}_i)^{1/\rho^*+1} \left[ \right. \int_{x} Q^*(x) \right]^{\rho^*+1} \right] dx \right. \left. \int_{x} Q^*(x) \right\}^{2W T(t)} \left. d\hat{z}_j \right\} \tag{31}$$

ii) $R_1 \leq R_b \leq R_0$: Using $T(\hat{z}, R_b)$ in (27), the solution is given by the following nonlinear parametric equations:

$$E_p^*(Q^*, R_b, \Delta) = \frac{1}{2} \log_2 \left[ \frac{1}{E_f} \left\{ \int_{\hat{z}_p} f(\hat{z}_i) \left[ \int_{x_j} \left[ \int_{x} Q^*(x) \right. \right. \right.$$

$$\times \left. p_c(y \mid x, \hat{z}_i)^{1/\rho^*+1} \left[ \right. \int_{x} Q^*(x) \right]^{\rho^*+1} \right] dx \right. \left. \int_{x} Q^*(x) \right\} \tag{32}$$

B. Variable Input Schemes

Because of the constant symbol duration, the induced density in Section II-C is reduced to the standard fading density $f(c, \hat{z})$. Using similar techniques to Section III-A, the average codeword error probability $P_e$ is bounded by

$$P_e \leq 2^{-N E_p(\rho^* Q_0 R)} \tag{33}$$
where
\[
\begin{align*}
\bar{E}_r(\rho, Q, \mathcal{R}) & = -\log_2 \left\{ \int_{c} \int_{z} 2^{-[E(\rho Q(c, z)) - \rho \mathcal{R}(z)]} f(c, z) \, dc \, dz \right\} \\
E(\rho, Q, c, z) & = 2WT_2 \log_2 \left\{ \int_{y} \int_{c} Q(x | z) p_0(y | xc)^{\frac{1}{\rho+1}} \, dx \, dy \right\},
\end{align*}
\]
(35)
To obtain a tight error bound, we have to minimize \( \bar{P}_e \) w.r.t. \( \rho, Q(x | z) \), and \( \mathcal{R}(z) \). Since \( Q(x | z) \), \( \mathcal{R}(z) \) and \( \rho \) are independent of each other, the optimization problem is uncoupled.

1) Optimization w.r.t. \( \rho \): Minimization of \( \bar{P}_e \) is equivalent to maximization of \( \bar{E}_r(\rho, Q^*, \mathcal{R}) \). Define the optimal error exponent \( E^*_r(Q^*, \mathcal{R}) \) as
\[
E^*_r(Q^*, \mathcal{R}) = \max_{\rho} E_r(\rho, Q^*, \mathcal{R}) = E_r(\rho^*, Q^*, \mathcal{R}).
\]
(36)
We first introduce the following lemma.

Lemma 3: For any \( \mathcal{R}(z) \), \( \partial^2 E_r(\rho, Q^*, \mathcal{R}) \)/\( \partial \rho^2 < 0 \) for all \( \rho \).

Proof: Refer to Appendix D-1.

The result of the optimization is summarized in Lemma 4.

Lemma 4: For any \( \mathcal{R}(z) \), let \( \rho_1 \) be the roots of the equation
\[
\int_{0}^{\infty} \int_{0}^{\infty} f(c, z) 2^{-[E(\rho Q(c, z)) - \rho \mathcal{R}(z)]} \left( \frac{\partial E}{\partial \rho} - \mathcal{R}(z) \right) \, dc \, dz = 0.
\]
(37)
The optimal parameter \( \rho^* \) is given by
\[
\rho^* = \begin{cases} 
\rho_1, & \rho_1 \in [0, 1] \\
1, & \rho_1 > 1.
\end{cases}
\]
(38)
Proof: The necessary condition for \( \rho_1 \) to be the optimal parameter is \( (\partial E_r(\rho_1, Q^*, \mathcal{R}))/\partial \rho \big|_{\rho=\rho_1} = 0 \). By Lemma 3, \( \partial^2 E_r(\rho, Q^*, \mathcal{R}) \)/\( \partial \rho^2 < 0 \), and hence, the stationary point obtained, \( \rho_1 \), corresponds to the absolute maximum point. Furthermore, \( \partial E_r(\rho, Q^*, \mathcal{R}) \)/\( \partial \rho \) is a decreasing function of \( \rho \). Suppose \( \rho_1 > \rho_2 \), we have \( \partial E_r(\rho_1, Q^*, \mathcal{R}) > \partial E_r(\rho_2, Q^*, \mathcal{R}) = 0 \). Therefore, \( \bar{E}_r \) is an increasing function w.r.t. \( \rho \) for \( \rho < \rho_1 \). Since the error bound in (33) is valid only when \( \rho \in [0, 1] \), the optimal parameter \( \rho^* \) is given by \( \rho^* = 1 \) if \( \rho_1 > 1 \).

Collecting the results from the two lemmas, we have the following theorem.

Theorem 2: For any \( \mathcal{R}(z) \), if the average code rate \( \mathcal{R} \) is equal to \( \int_{0}^{\infty} \mathcal{R}(z) f(z) \, dz \), then \( \lim_{N \to \infty} \bar{P}_e = 0 \).

Proof: Refer to Appendix D-2.

2) Optimization w.r.t. \( Q(x | z) \): Similar to Section III-A-1, take \( Q(x | z) \) as the capacity achieving distribution \( Q^*(x) \), which is a Gaussian distribution with variance \( \sigma^2_z = P_0/2W \). The remaining problem is to minimize \( \bar{P}_e \) w.r.t. \( \mathcal{R}(z) \).

3) Optimization w.r.t. \( \mathcal{R}(z) \): In this section, we minimize \( \bar{P}_e \) w.r.t. \( \mathcal{R}(z) \). Using a normalized rate control, \( \mathcal{R}(z) = \bar{R}(\mathcal{R}) \), we find the optimal normalized rate control function, \( \bar{r}(\mathcal{R}) \). For any given \( \bar{R} \), take \( \rho = \rho^* \) as in (37) and \( Q(x | z) = Q^*(x) \). Unlike Section III-A-3, a general expression for the optimal rate control function for both negligible and nonnegligible feedback delays is derived.

Using the normalized rate control, \( \bar{R}(\mathcal{R}) = \bar{R}(\mathcal{R}) \), \( \bar{P}_e \) in (33) is given by
\[
\bar{P}_e \leq \left\{ \int_{z_1}^{z_2} f(c, z) 2^{-[E(\rho^* Q^*(z)) - \rho^* \mathcal{R}(z)]} \, dc \, dz \right\} \cdots \left\{ \int_{z_N}^{\infty} f(c, z) 2^{-[E(\rho^* Q^*(z)) - \rho^* \mathcal{R}(z)]} \, dc \, dz \right\}.
\]
(39)
The optimization problem is equivalent to choosing \( \{r_1, r_2, \ldots, r_N\} \) that minimizes \( \bar{P}_e \) in (39) under the constraints
\[
\int_{z_i}^{z_i} f(\hat{z}_i) r_i(\hat{z}_i) d\hat{z}_i = 1 \quad \forall i \in [1, N]
\]
\[
r_i(\hat{z}_i) \geq 0 \quad \forall i \in [1, N].
\]
(40)
(41)
By the Calculus of Variations, it is shown in Appendix E that the optimal control rule is given by
\[
r(\hat{z}) = \begin{cases} 
\frac{1}{\rho^*} \mathcal{R}(G(\rho^*, Q^*, \hat{z})) - G(\rho^*, Q^*, \mathcal{R}(\hat{z})), & \mathcal{R}(\hat{z}) \geq \hat{z}_i \\
0, & 0 < \hat{z}_i < \hat{z}_i
\end{cases}
\]
(42)
where \( G(\rho^*, Q^*, \mathcal{R}(\hat{z})) \) is given by
\[
G(\rho^*, Q^*, \mathcal{R}(\hat{z})) = \log_2 \left\{ \int_{0}^{\infty} f(c | \mathcal{R}(\hat{z})) 2^{-E(\rho^* Q^*)} \, dc \right\}
\]
(43)
and \( \hat{z}_i \) is given by the solution of
\[
\frac{1}{\rho^*} \mathcal{R} \left\{ G(\rho^*, Q^*, \hat{z}_i) \right\} \int_{\hat{z}_i}^{\infty} f(\hat{z}) d\hat{z} = \frac{1}{\rho^*} \mathcal{R} \left\{ G(\rho^*, Q^*, \hat{z}_i) \right\} \int_{\hat{z}_i}^{\infty} G(\rho^*, Q^*, \hat{z}) f(\hat{z}) d\hat{z} = 1.
\]
(44)
Intuitively, at small feedback delay, \( r(\hat{z}) \) is an increasing function of \( \hat{z} \) and the control law implies that more information bits per symbol should be carried if the predicted state \( \hat{z} \) is good. On the other hand, at large feedback delay, \( \hat{z} \) and \( c \) are independent and \( r(\hat{z}) \) tends to be independent of \( \hat{z} \), suggesting that fixed-rate control will be optimal if the predicted state is not accurate.

4) Overall Result of Optimizations for Variable Input Scheme: Given an average code rate \( \mathcal{R} \), take \( \rho^*, Q^*, \) and \( \mathcal{R}(\hat{z}) \) to be the optimal parameter, the capacity achieving distribution, and the optimal rate control law, respectively. Let \( R_0 = I(X, \hat{Y} \mid C, \hat{Z}) \) and \( R_1 > 0 \) be the solution of the equation
\[
\int_{0}^{\infty} \int_{0}^{\infty} f(c, z) 2^{-[E(\rho^* Q^*(z)) - \rho^* \mathcal{R}(\hat{z})]} \left( \frac{\partial E}{\partial \rho} - \mathcal{R}(\hat{z}) \right) \, dc \, dz = 0.
\]
(45)
The optimal error exponent \( E^*_r(Q^*, \mathcal{R}, \Delta) \) at a feedback delay \( \Delta \) is given by the following.

i) \( R_1 \geq \bar{R} \geq 0 \): Using the rate control rule in (42) with \( \rho^* = 1 \), the solution is given by
\[
E^*_r(Q^*, \bar{R}, \Delta) = \bar{E}_r(1, Q^*, \bar{R}, \Delta).
\]
(46)
Using the rate control rule in (42), the solution is obtained by solving the following system of three nonlinear simultaneous equations:

\[
\frac{1}{\rho^* R} \left\{ G(\rho^*, Q^*, \bar{z}_i) \int_{\bar{z}_i}^{\infty} f(\bar{z}) d\bar{z} - \int_{0}^{\bar{z}_i} G(\rho^*, Q^*, \bar{z}) f(\bar{z}) d\bar{z} \right\} = 0
\]

\[
\int_{0}^{\infty} \int_{0}^{\infty} f(c, \bar{z}) g(\rho, Q, z; \bar{z}) - \rho^{R(\bar{z})} \left( \frac{\partial E^*}{\partial \rho} - \bar{R}_{\varphi}(\bar{z}) \right) dc d\bar{z} = 0
\]

\[
E^*_q(Q^*, \bar{R}, \Delta) = E_{\varphi}(\rho^*, Q^*, \bar{R}, \Delta).
\]

IV. CHANNEL CAPACITY FOR VARIABLE RATE SCHEMES

In this section, we derive a general expression for channel capacity of Rayleigh fading channel with variable rate transmission. Channel capacity is defined as follows.

**Definition 1:** A channel is said to have a channel capacity \( C \) if

(i) for every \( \epsilon > 0 \) and \( \delta > 0 \), \( \exists \) a channel code of rate \( \bar{R} = C - \epsilon \) with block length \( N \) such that the error probability is bounded above by \( \delta \) \( \forall N > n_0 \) for some \( n_0 > 0 \).

(ii) for every \( \epsilon > 0 \), all codes with rate \( \bar{R} = C + \epsilon \) cannot have asymptotically zero error probability as \( N \rightarrow \infty \).

**Lemma 5 (Converse):** The channel capacity \( C \) (in bits/symbol) of a Rayleigh fading channel, with feedback of channel states to the transmitter using variable rate transmission, is upper bounded by \( C_0 \)

\[
C \leq C_0 \overset{\text{def}}{=} \max_{Q \in \mathcal{X}} I(X; Y | C \bar{Z}).
\]

**Proof:** Refer to Appendix C-3.

**Lemma 6 (Achievability):** The channel capacity \( C \) (in bits/symbol) of a Rayleigh fading channel, with feedback of channel states to the transmitter using variable rate transmission, is lower bounded by \( C_0 \).

**Proof:** Assume that \( C < C_0 \) and let \( C < \bar{R} < C_0 \). For the variable duration scheme, by Theorem 1, there is at least a code of rate \( \bar{R} \) that has asymptotically zero error probability as \( N \rightarrow \infty \). However, this contradicts the definition of channel capacity which states that no such code exists with \( \bar{R} > C \). For the variable input scheme, by Theorem 2, there is at least a code of average rate \( \bar{R} \) that has asymptotically zero error probability as \( N \rightarrow \infty \). However, this contradicts the definition of channel capacity which states that no such code exists with \( \bar{R} > C \). Hence, the result follows.

Combining the above two lemmas, we have the following theorem.

**Theorem 3:** For any symbol duration control law \( T(\bar{z}) \) or rate control law \( \bar{R}(\bar{z}) \), the channel capacity of a Rayleigh fading channel with feedback of channel states to the transmitter using variable rate transmission is equal to \( C_0 \) in (48).

Using the capacity achieving distribution \( Q^* \), the mutual information is given by

\[
I(X_{a,i}; Y_{a,i} | C_i \bar{Z}_i) = \frac{1}{2} \int_{c \in \mathcal{C}} \int_{\bar{z} \in \mathcal{Z}} \log_2 \left( 1 + \frac{c^2\bar{P}_0}{W\bar{R}_0} \right) f(c | \bar{z}) \mathcal{D}(\bar{z}) d\bar{z} dc.
\]

For the variable duration scheme, the feedback channel capacity \( C \) (in bits/sec) becomes

\[
C = W E_f \int_{c \in \mathcal{C}} \int_{\bar{z} \in \mathcal{Z}} T(\bar{z}) \log_2 \left( 1 + \frac{c^2\bar{P}_0}{W\bar{R}_0} \right) f(c | \bar{z}) \mathcal{D}(\bar{z}) d\bar{z} dc
\]

\[
= W \int_{c \in \mathcal{C}} f(c) \log_2 \left( 1 + \frac{c^2\bar{P}_0}{W\bar{R}_0} \right) \int_{\bar{z} \in \mathcal{Z}} f(\bar{z} | c) \mathcal{D}(\bar{z}) d\bar{z} dc
\]

\[
= C(\bar{R}, P_0)
\]

where \( C(\bar{R}, P_0) \) is the fixed-duration channel capacity without feedback. For the variable input scheme, \( C \) (in bits/sec) becomes \( (I(X; Y | C \bar{Z}))/T_s \) which is equal to the fixed-rate channel capacity, \( C(\bar{R}, P_0) \) as well. Hence, the variable rate schemes cannot increase the channel capacity of Rayleigh fading channels.

V. RESULTS AND DISCUSSIONS

In a microcellular environment at 2 GHz with mobiles moving at a maximum speed of 75 km/h, the coherence time is around 1 ms and the coherence bandwidth is around 2 MHz. We choose the symbol rate \( E_f \) to be 40 ksym/s and the system bandwidth to be 800 kHz. Hence, the system bandwidth is smaller than the coherence bandwidth and the average symbol duration (1/E\(_f\) = 0.025 ms) is much smaller than the coherence time. These justify the flat-fading assumption made in the channel model. Since \( W = 800 \) kHz, \( T_s = 5 \mu s \) is sufficient to ensure \( 2W T_s > 2 \). For a fixed-duration system, the symbol duration \( T(\bar{z}) \) is constant and is equal to \( 1/E_f \). Hence, \( T_s = 1/E_f \) is taken to be the reference symbol duration. The channel normalized fading rate is \( f(E_f T_s) = 2.5 \times 10^{-3} \).

A. Variable Rate Channel Capacity

It is shown in Section IV that channel capacity is not increased by variable rate control. This is intuitively correct since in a large block, we have either the total block duration approaches a constant value of \( N/E_f \) [refer to (13)] for the variable duration scheme or the total number of information bits transmitted in a large block approaches the average code rate, \( \bar{R} \) for the variable input scheme. Hence, there is no difference with fixed-rate coding schemes asymptotically. The channel capacity in the example is equal to 616 kb/s at reference SNR, \( E_b/\bar{N}_0 = 3 \) dB.
B. Variable Rate Error Exponent

Although channel capacity is not increased by variable rate channel coding, the error exponent is increased significantly compared with the fixed-duration error exponent. For the variable duration scheme from Fig. 4, the improvement in error exponent is three times the fixed-duration case at $\mathcal{R}_b = 0.5\mathcal{C} = 308$ kbit/s under ideal situations ($\Delta = 0$, $T_p = \infty$, $T_i = 5 \mu s$). The performance improvement is degraded to 2.1 times and 1.13 times if the feedback delay is 18 symbols and 25 symbols, respectively. The effect of peak time constraint $T_p$ is shown in Fig. 5. Define the ratio of peak to average symbol duration $\xi$ as $\xi = T_p/\mathbb{E}[T] = E_fT_p$. With $\xi = 2.98$ and $\xi = 1.21$, the improvement in the error exponent is 2.46 and 1.40 times, respectively.

For the variable input scheme from Fig. 6, the improvement in error exponent is 2.67 times at $\mathcal{R}_b = 0.5\mathcal{C} = 308$ kbit/s with negligible feedback delay compared with fixed-rate code. The performance is degraded to 2.27 times and 1.73 times if the feedback delay is 18 symbols and 23 symbols respectively. At a feedback delay of 100 symbols, error exponent of variable input scheme approaches the fixed-rate error exponent because the optimal rate control rule would be a fixed rate control at such large feedback delay. The bandwidth expansion used in the above calculation is $W/\mathcal{R}_b = 1.3$. Note that a 2-time increase in error exponent means a 2-time reduction in coding complexity (e.g. block length $N$) to achieve the same error probability.

C. Bandwidth Expansion Consideration

We consider two extreme cases, a bandwidth expansion of 0.25 which models TDMA systems and a bandwidth expansion of 20 which models CDMA systems. Error exponents of the variable duration scheme for small and large bandwidth expansion systems are shown in Fig. 7(a) and (b). For the system with small bandwidth expansion (TDMA), there is a significant 5.62-time increase in error exponent at $\mathcal{R}_b = 0.8\mathcal{C}$. For the system with large bandwidth expansion (CDMA), there is only a 1.5-time improvement in error exponent relative to fixed-duration scheme at the same $\mathcal{R}_b$.

The error exponents for the variable input scheme with small (0.25) and large bandwidth expansion (20) systems are shown in Fig. 8(a) and (b). For systems with small bandwidth expansion (TDMA), there is a significant 5-time increase in error exponent at $\mathcal{R}_b = 0.8\mathcal{C}$. For systems with large bandwidth expansion (CDMA), there is only a 1.1-time improvement in error exponent relative to fixed-rate schemes at the same $\mathcal{R}_b$.

Therefore, variable rate channel coding is more effective in high bandwidth-efficiency systems. This means that only a limited gain can be achieved in high bandwidth-expansion systems when very powerful capacity achieving codes are used as the component codes in the construction of variable rate adaptive codes. However, a significant gain should be expected in high bandwidth-efficiency systems even when very powerful component codes are used.
VI. SUMMARY

In this paper, we have evaluated the channel capacity and the error exponent of variable rate Rayleigh fading channel using variable duration and variable input schemes. Optimal symbol duration control law and optimal input rate control law are derived taking into account of feedback delay. Performance degradation w.r.t. feedback delay is also investigated. We found that channel capacity was not increased by variable rate coding schemes for any control law. On the other hand, there was a significant increase in error exponent for both schemes. This means that less complex codes can be found to achieve the same using variable rate adaptive coding.

Hence, instead of aiming at maximizing the channel capacity by previous approaches, we should aim at maximizing the error exponent with variable rate adaptive channel coding.

For the dependence of the improvement on bandwidth expansion, we found that improvement was limited at large bandwidth expansion. On the contrary, significant gain resulted when bandwidth expansion was small. This suggests that variable rate channel coding schemes have limited intrinsic gains in CDMA systems compared with TDMA systems.

APPENDIX A

INDUCED PROBABILITY DENSITY ON $\hat{Z}_N$

A. One Dimension

We first prove (4) in Section II.

Proof: Define the induced probability density $P(\hat{z})$ as

$$P(\hat{z}) \delta \equiv \lim_{N \to \infty} \frac{N_{\hat{z}}}{N}$$

where $N_{\hat{z}}$ is the number of symbols with $\hat{Z} \in [\hat{z}, \hat{z} + \delta]$ in a sequence of $N$ symbols, and $\delta$ is a small increment in $\hat{z}$.

Observe $\hat{Z}(t)$ over a long time duration $T_0$. Suppose $\hat{Z}(t)$ is an ergodic random process, we have

$$\lim_{T_0 \to \infty} \frac{T_0}{T_0} = f(\hat{z})\delta$$

where $T_0$ is the total time that $\hat{Z}(t) \in [\hat{z}, \hat{z} + \delta]$ and $f(\hat{z})$ is the fading density. Hence, $N_{\hat{z}} = T_0/T(\hat{z}) = T_0 f(\hat{z})\delta/T(\hat{z})$
where $T(\xi)$ is the symbol duration (a function of $\xi$) and $\mathcal{N} = \int_{\mathcal{Z}} \mathcal{N}(\xi) d\xi = T_0 \int_{\mathcal{Z}} f(\xi)/T(\xi) d\xi$. Therefore,
\[
P(\xi) = \frac{f(\xi)}{E_f T(\xi)},
\]
where
\[
E_f = \int_{\mathcal{Z}} \frac{1}{T(\xi)} f(\xi) d\xi.
\]

We shall show that $E_f$ is the average symbol rate. Suppose we observe $\hat{Z}(t)$ over a very long time interval $T_0$, then $\mathcal{Z} = f(\xi)T_0$ and there are $\mathcal{Z}/T(\xi)$ symbols during this time duration. By definition, the average symbol rate is the number of symbols transmitted per unit time and is given by $\int_{\mathcal{Z}} \mathcal{Z}/T(\xi) d\xi$ which is equal to $E_f$.

We shall illustrate the use of this induced probability density in the following example.

Example 1: Suppose the symbol error probability $P_e(\xi)$ is conditioned on $\hat{\xi}$. To obtain the unconditional error probability, we have to use the induced density in the integration given by
\[
P_e = \int_{\mathcal{Z}} P_e(\xi) P(\xi) d\xi.
\]

Proof: The average number of symbol errors given that $\hat{Z} \in [\xi, \xi + \delta]$ is given by $\mathcal{N}_e P_e(\xi)$. By definition, the average error probability is the total number of symbol errors divided by the total number of symbols transmitted and is given by
\[
P_e = \lim_{N \to \infty} \frac{\int_{\mathcal{Z}} \mathcal{N}_e P_e(\xi) d\xi}{\mathcal{N}} = \int_{\mathcal{Z}} P(\xi) P_e(\xi) d\xi.
\]

B. N Dimensions

For the ideal interleaved channel, $\hat{Z}$ and $\hat{\xi}$ are independent. Hence, the $N$-dimensional induced probability density on $\mathcal{Z}_N$ is the product of $N$ one-dimensional induced densities
\[
P_N(\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_N) = \prod_{i=1}^{N} \frac{1}{E_f^i} \frac{f(\xi_i) f(\xi_i) \cdots f(\xi_i)}{T(\xi_i) T(\xi_i) \cdots T(\xi_i)},
\]
Extending Example 1 to the $N$-dimensional and ideal interleaving case, the unconditional error probability is given by
\[
P_e = \lim_{N \to \infty} \frac{\int_{\mathcal{Z}_N} \mathcal{N}_e P_e(\hat{\xi}_1, \ldots, \hat{\xi}_N) d\mathcal{Z}_N}{\mathcal{N}} = \int_{\mathcal{Z}_N} P_N(\hat{\xi}_1, \ldots, \hat{\xi}_N) P_e(\hat{\xi}_1, \ldots, \hat{\xi}_N) d\mathcal{Z}_N.
\]

APPENDIX B

OPTIMAL CONTROL RULE FOR VARIABLE DURATION SCHEME

The problem is to choose $\{T_1(\hat{\xi}_1), \ldots, T_N(\hat{\xi}_N)\}$ that minimize $P_e$ under the constraints of (24), (25), and (26). We form the $\beta$th Lagrange multiplier as
\[
L_\beta(T_i) = \frac{2 N \mathcal{N}_e \mathcal{Z}_i}{E_f} \sum_{j=1}^{N} \left\{ \int_{0}^{\infty} \frac{f(\xi_j)}{T_j} G(\rho^s, \hat{Z}_j) T_j d\xi_j \right\} \times \frac{f(\xi_j)}{T_j} G(\rho^s, \hat{Z}_j) T_j
\[
- \beta \frac{1}{E_f} \frac{f(\xi_j)}{T_j}
\]
where
\[
G(\rho^s, \hat{Z}_j) = \int_{y} \left( \int_{x} Q^s(x) \mathcal{P}(y | x \hat{Z}_j) \sigma^{-\frac{1}{2}} dx \right)^{i+1} dy.
\]

By the Calculus of Variations [24], the necessary condition for $T_\beta(\hat{Z}_j)$ to be the optimal control is $\frac{\partial L_\beta}{\partial T_j} = 0$ for all $i \in [1, N]$. This implies
\[
G(\rho^s, \hat{Z}_j) = K_0
\]
where $K_0$ is a constant independent of $\hat{Z}_j$. This means that $G(\rho^s, \hat{Z}_j) T_j$ must be a constant for every $i$. Hence, the optimal $T(\xi)$ is given by
\[
T(\xi) = \frac{K}{E_f K_0 \int_{y} \left( \int_{x} Q^s(x) \mathcal{P}(y | x \hat{Z}_j) \sigma^{-\frac{1}{2}} dx \right)^{i+1} dy}
\]
where $K$ is determined by substituting back into the constraints (24)–(26).

APPENDIX C

PROOF OF LEMMASES FOR VARIABLE DURATION SCHEME

A. Proof of Lemma 1

Proof: From (16) and (7), $\mathcal{E}(\rho, Q)$ can be expressed as
\[
\mathcal{E}(\rho, Q) = \int_{c}^{d} \mathcal{E}(c, \hat{\xi}) | P(\hat{\xi}) | d\hat{\xi}.
\]

(i) Since $\partial E_N/\partial \rho > 0$ [22], [1], we have $\partial \mathcal{E}/\partial \rho > 0$. Since
\[
\frac{\partial E_N}{\partial \rho} \bigg|_{\rho=0} = I(\hat{X}^c; Y^c | C = c, \hat{Z} = \hat{\xi})
\]
and $E_N(0, Q, C, \hat{Z}) = 0$, we have
\[
\frac{\partial \mathcal{E}}{\partial \rho} \bigg|_{\rho=0} = E_f \int_{c}^{d} \mathcal{E}(c, \hat{\xi}) | P(\hat{\xi}) | d\hat{\xi}.
\]

(ii) Since [22], [1]
\[
E_N(\alpha_1 \rho_1 + \alpha_2 \rho_2, Q, C, \hat{Z}) > \alpha_1 E_N(\rho_1, Q, C, \hat{Z}) + \alpha_2 E_N(\rho_2, Q, C, \hat{Z})
\]
for any $\alpha_1 + \alpha_2 = 1$, we have
\[
E\left(\alpha_1 p_1 + \alpha_2 p_2, Q\right) > -E_f \log_2 \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(z)f(c \mid z) \right.
\times 2^{-\left[ E_f (p_1) + \alpha_2 E_f (p_2) \right] dc dz} \right] = -E_f \log_2 \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(z)f(c \mid z) 2^{-\left[ E_f (p_1) \right] \alpha_1} \times \left[ P(z)f(c \mid z) 2^{-\left[ E_f (p_2) \right] \alpha_2} \right] dc dz \right\} \]
\geq -E_f \log_2 \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(z)f(c \mid z) 2^{-\left[ E_f (p_1) \right] \alpha_1} \times \left[ P(z)f(c \mid z) 2^{-\left[ E_f (p_2) \right] \alpha_2} \right] dc dz \right\} \]
\geq \alpha_1 E_f (p_1, Q) + \alpha_2 E_f (p_2, Q)
\]
where the last inequality follows from the Holder's inequality. Hence, $E\left(\rho, Q\right)$ is a convex function in $\rho$ and the result follows.

B. Proof of Lemma 2

Proof: The necessary condition for $\rho = \rho^*$ to maximize $E\left(\rho, Q\right)$ is $\frac{dE}{d\rho} |_{\rho = \rho^*} = 0$. This gives
\[
\frac{dE}{d\rho} |_{\rho = \rho^*} = R_b.
\]

By Lemma 1(ii), we have $\frac{d^2 E}{d\rho^2} < 0$. This verifies that the stationary point obtained, $\rho^*$, corresponds to the absolute maximum and $R_b$ in (56) is a strictly decreasing function of $\rho^*$. Hence, as $\rho^*$ increases from 0 to 1, $R_b$ decreases from $R_0$ to $R_1$ and $E_r$ maximizes at $\rho^*$. For $R_b < R_0$, $\rho^* > 1$. Since $\frac{d^2 E}{d\rho^2}$ is a decreasing function of $\rho$, we have
\[
\frac{dE}{d\rho} |_{\rho = \rho^*} > \frac{dE}{d\rho} |_{\rho = \rho^*} = R_b > 0
\]
for $0 < \rho < \rho^*$. Since the error bound in (6) is valid only when $\rho \in [0, 1]$, $E_r$ maximizes at $\rho_1 = 1$.

C. Proof of Lemma 5

Proof: The estimated index $\hat{\Omega}$ is given by
\[
\hat{\Omega} = g(\hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N)
\]
where $g(\hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N)$ is a general decoding function. The average codeword error probability $P_e$ is given by
\[
P_e = P(\hat{\Omega} \neq \Omega).
\]

By Fano's inequality
\[
H(\hat{\Omega} \mid \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N) \leq 1 + P_e NR.
\]

Since $I(\Omega; \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N) = H(\Omega) - H(\hat{\Omega} \mid \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N), H(\Omega) = NR$ because of equiprobable input and
\[
I(\Omega; \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N) = I(\Omega; \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N) + I(\Omega; \hat{C}_N, \hat{Z}_N)
\]
we have
\[
R \leq \frac{1}{N} I(\Omega; Y_1, \ldots, Y_N; \hat{C}_N, \hat{Z}_N) + \epsilon
\]
where $\epsilon = \frac{1}{N}[1 + P_e NR]$. Given any particular $\hat{C}_N = \hat{C}_N$ and $\hat{Z}_N = \hat{Z}_N$, we have $\Omega = \{Y_1, \ldots, Y_N\} \rightarrow \{\hat{Y}_1, \ldots, \hat{Y}_N\}$ and by the data processing inequality [18], we have
\[
I(\Omega; \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N) \leq I(\hat{X}_1, \ldots, \hat{X}_N; \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N)
\]
Hence
\[
I(\Omega; \hat{Y}_1, \ldots, \hat{Y}_N; \hat{C}_N, \hat{Z}_N) = \int_{\hat{C}_N, \hat{Z}_N} P(\hat{C}_N, \hat{Z}_N) I(\Omega; \hat{Y}_1, \ldots, \hat{Y}_N)
\]
which follows from the Holder's inequality. Hence, $E(\rho, Q)$ is convex in $\rho$ and the result follows.

Appendix D

Proof of Lemmas and Theorem for Variable Input Scheme

A. Proof of Lemma 3

Proof: Since [22], [11]
\[
E(\alpha_1 p_1 + \alpha_2 p_2, Q, c, \hat{z}) > \alpha_1 E(p_1, Q, c, \hat{z}) + \alpha_2 E(p_2, Q, c, \hat{z})
\]
for any $\alpha_1 + \alpha_2 = 1$, we have
\[
P_e(\alpha_1 p_1 + \alpha_2 p_2, Q)
\]
\[
> -\log_2 \left[ \int_{c_1}^{c_2} \int_{z_1}^{z_2} f(c, \hat{z}) \right.
\times 2^{-\left[ E_f (p_1) + \alpha_2 E_f (p_2) \right] dc dz} \right] = -\log_2 \left\{ \int_{c_1}^{c_2} \int_{z_1}^{z_2} f(c, \hat{z}) 2^{-\left[ E_f (p_1) + \alpha_2 E_f (p_2) \right] dc dz} \right\}
\]
\[
\geq -\log_2 \left\{ \left[ \int_{c_1}^{c_2} f(c, \hat{z}) 2^{-\left[ E_f (p_1) + \alpha_2 E_f (p_2) \right] dc} \right] \left[ \int_{z_1}^{z_2} 2^{-\left[ E_f (p_1) + \alpha_2 E_f (p_2) \right] dz} \right] \right\} \]
\[
= \alpha_1 E_f (p_1, Q) + \alpha_2 E_f (p_2, Q)
\]
However, this contradicts (61) which states that $\lim_{N \to \infty} P_e = 0$.

Hence, the lemma follows.
where the last inequality follows from the Holder’s inequality. Hence, $E_r(\rho, Q)$ is a convex function in $\rho$ and the result follows.

\[ (62) \]

**B. Proof of Theorem 2**

**Proof:** Express the code rate $R(\xi)$ as

\[ R(\xi) = \tilde{R}(\xi) \]

(39) under the constraints of (40) and (41). The optimization problem is to choose $\{R_1, \ldots, R_N\}$ that minimizes $\tilde{P}_e$ in (39) under the constraints of (40) and (41). The $i$th Lagrange multiplier is given by

\[ \frac{\partial \tilde{P}_e}{\partial \tilde{R}_i} = -\frac{1}{E^*_{\xi}} \int \xi f(\xi) \tilde{R}(\xi) d\xi \]

where $\tilde{P}_e$ is the root of the equation

\[ \frac{1}{\rho^* \tilde{R}} \left\{ G(\rho^*, Q^*, \xi) \int_{\xi}^{\infty} f(\xi) d\xi \right\} = 1. \]

**ACKNOWLEDGMENT**

The author would like to thank Dr. S. V. Maric and Dr. M. D. Macleod for their advice in this work.

**REFERENCES**


Vincent K. N. Lau (M’92), for photograph and biography, see p. 589 of the April 1999 issue of this TRANSACTIONS.