

Optimal Allocation of Bandwidth for Source Coding, Channel Coding, and Spreading in CDMA Systems

Qinghua Zhao, Pamela Cosman, *Senior Member, IEEE*, and Laurence B. Milstein, *Fellow, IEEE*

Abstract—This paper investigates the tradeoffs between source coding, channel coding, and spreading in code-division multiple-access systems, operating under a fixed total bandwidth constraint. We consider two systems, each consisting of a uniform source with a uniform quantizer, a channel coder, an interleaver, and a direct-sequence spreading module. System A is quadrature phase-shift keyed modulated and has a linear block channel coder. A minimum mean-squared error receiver is also employed in this system. System B is binary phase-shift keyed modulated. Rate-compatible punctured convolutional codes and soft-decision Viterbi decoding are used for channel coding in system B. The two systems are analyzed for both an additive white Gaussian noise channel and a flat Rayleigh fading channel. The performances of the systems are evaluated using the end-to-end mean squared error. A tight upper bound for frame-error rate is derived for nonterminated convolutional codes for ease of analysis of system B. We show that, for a given bandwidth, an optimal allocation of that bandwidth can be found using the proposed method.

Index Terms—Bandwidth allocation, convolutional codes, direct-sequence code-division multiple access (DS-SS), frame-error rate (FER), multiuser system, Rayleigh fading, transmission over wireless channels.

I. INTRODUCTION

SOURCE coding, channel coding, and spread spectrum are three of the main components in a code-division multiple-access (CDMA) communication system. They compete for the major shared-resource, bandwidth. Source coding frees up bandwidth for both forward-error correction (FEC) and spreading. Allocating more bandwidth to source coding allows more information from the source to be transmitted, but reduces the bandwidth available for both FEC and spreading. For different compression methods and rates, the bit stream coming out of the source encoder will be more or less sensitive to different types of error patterns. FEC and spreading protect the transmitted bits from noise and interference. Depending on the channel conditions and the characteristics of the source-coded bit stream, the system will perform better with either more FEC or more spreading.

Related studies in the literature are limited to the tradeoff between either source coding and channel coding [1], [2], or

channel coding and spreading [3], [4]. In each case, research topics included analyzing a given system to find the optimal bandwidth allocation to each module as in [1] and [3], and joint design of coding algorithms or transmitter/receiver schemes for each category [2], [4]. In [5], we studied the bandwidth-allocation tradeoff for a direct-sequence (DS)-SS system that incorporated an image coder, a rate-compatible punctured convolutional (RCPC) [6] channel coder, and a RAKE receiver. Due to the complexity of the system, we obtained most of the results through simulations. In this paper, we investigate the tradeoffs using a combination of analytical and numerical techniques.

Let r_s , r_c , and N_s denote the source-code rate (in bits per source symbol), the channel-code rate, and the spreading factor, respectively. If the source produces U symbols per second, for a given chip rate of C chips per second, we have the following constraint:

$$U \cdot r_s \cdot \frac{1}{r_c} \cdot N_s \leq C \implies r_s \cdot \frac{1}{r_c} \cdot N_s \leq C_0 \quad (1)$$

where $C_0 = C/U$ is a constant that constrains the number of chips available for each source symbol. We will find the bandwidth allocation $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$ that optimizes system performance. We will also address the question of how sensitive the optimal allocation is to changes in the channel conditions, transmission rate, or bandwidth constraint.

Note that, in reality, C_0 is determined by the spread bandwidth, the pulse shaping, and the source-symbol rate. For example, consider a wideband CDMA system with a chip rate 2×10^7 chips/s, operating in conjunction with a video conference application of 10 frames per second (fps), with 176-by-224-pixel frames. This requires a source-symbol rate of $10 \times 176 \times 224 = 3.9424 \times 10^5$ symbols/s. Therefore, $C_0 \approx 51$. In our paper, the values of C_0 are generally larger than those of a practical system. This is because our system uses a simple high-rate uniform quantizer, whereas a practical system would use a more complicated source code that allows the source to be transmitted at much lower rates.

In multiuser CDMA systems, each user has its own performance requirements and bandwidth constraints, depending on the applications (e.g., video, voice, image). Without changing the transmission power, our optimization allows individual users to tune their own parameters (or a base station to tune the parameters for its users), independent of the performance of all other users. Our optimization setup can be restated as: Given a constraint on the chip rate available per source symbol, find the optimal $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$ that minimizes the end-to-end distortion. This optimization setup can be easily converted to the following two alternative optimization setups.

Paper approved by G. E. Corazza, the Editor for Spread Spectrum of the IEEE Communications Society. Manuscript received February 3, 2003; revised October 27, 2003; December 30, 2003; March 3, 2004; and March 31, 2004. This work was supported in part by the Center for Wireless Communications of UCSD, in part by the CoRe program of the State of California, and in part by Ericsson Wireless Communications Incorporated.

The authors are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: qizhao@ucsd.edu; pcosman@ucsd.edu; lmlstein@ucsd.edu).

Digital Object Identifier 10.1109/TCOMM.2004.836432

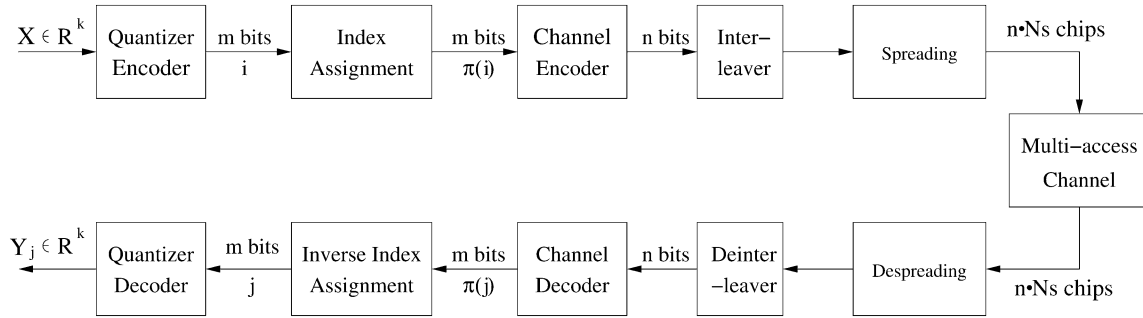


Fig. 1. System overview.

- Assuming all users are identical, for a given end-to-end distortion requirement, find the optimal $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$ that allows the largest number of users in the system.
- Assuming the transmitted signal energy is constant, given an end-to-end distortion requirement, what is the optimal $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$ that enables the mobile to have the largest coverage radius?

Which of these three optimization criteria is the most useful depends upon the specific scenario being considered. In a system that is transmitting scientific or medical images, the goal would often be to minimize distortion subject to the bandwidth constraint. This goal might also apply to any undersubscribed system, for which maximizing capacity is not currently an issue, and minimizing distortion is. On the other hand, if one is concerned with communicating in a dense urban environment, maximizing the capacity is probably the most useful criterion. Alternatively, if the environment is rural, maximizing coverage is often the most meaningful criterion. Some examples of these two alternative optimizations will be given in the results of Section III.

The paper is organized as follows. Section II gives an overview of the systems and our approach to the bandwidth-allocation problem. Sections III and IV present the analysis for two different systems. Representative results of tradeoffs among the three components are also given at the end of these two sections. The conclusions are given in Section V.

II. SYSTEM OVERVIEW

We consider a multiuser scenario. The system for each user is similar and is shown in Fig. 1. The source input vector $X \in \mathbf{R}^k$ has closed bounded support. Each component of the vector is considered to be one source symbol. The output of the source encoder is an m -bit binary index, and so the source-code rate $r_s = m/k$. The source encoder is a quantizer with distortion

$$D_m = \sum_{i=0}^{2^m-1} \int_{S_i} \|x - y_i\|^p f(x) dx \quad (2)$$

where $\{S_i\}_{i=0}^{2^m-1}$ is a partition of \mathbf{R}^k into disjoint regions, each of which is represented by code vector $y_i \in \mathbf{R}^k$, $\|\cdot\|^p$ represents the p th power of the usual Euclidean l_2 norm, and $f(x)$ is the probability density function (pdf) of X .

In our system, we take X to be a one-dimensional (1-D) ($k = 1$) uniform source over $[0, 1]$. We also take $p = 2$, so that the end-to-end distortion is the mean squared error (MSE).

The quantizer we use here is designed to be optimal for a noiseless channel, and it has partitions and code points

$$S_i = [i \cdot 2^{-m}, (i+1) \cdot 2^{-m}), \quad y_i = \frac{2^{-m}}{2}(2i+1) \quad (3)$$

respectively, where $i = 0, \dots, 2^m - 1$, and y_i is the centroid of S_i . In Appendix A, we prove that even though the analysis is done for a uniform source, the results can be applied to a wide variety of source distributions. Since $r_s = m/k = m$, we will use m and r_s interchangeably.

The m -bit binary representation $i \in \{0, 1, \dots, 2^m - 1\}$ of a source symbol is mapped to an m -bit index $\pi(i) \in \{0, 1, \dots, 2^m - 1\}$ by the index-assignment block.¹ Its purpose is to rearrange the indexes so that those with small Hamming distances between them represent quantization levels which are close. This way, the distortion caused by the most probable errors is small, and thus, the total distortion is small. There are many different index-assignment schemes possible for a scalar quantizer, such as the natural binary code (NBC), the Gray code (GC), and the two's complement code (TCC) [7]. We pick a random index assignment [1], where the mapping $\pi(\cdot)$ is a one-to-one mapping from indexes ranging from 0 to $2^m - 1$ to a random permutation of the same indexes. Since the permutation is random, the index assignment can be good or bad. To measure its distortion, we must average over all possible permutations, i.e., we use the expectation of the distortion to evaluate its performance. The use of random indexing simplifies the analysis, although the method proposed in this paper will work for any specific index assignment. In practice, a good index assignment (e.g., a NBC or a GC) should give a better performance than that of a random index assignment.

A channel encoder with rate $r_c = m/n$ codes the indexes and passes them to the interleaver. The interleaver output is multiplied by the spreading sequence assigned to the given user, with spreading factor N_s . The output of the spreading is modulated and passed to the channel. Here we consider DS-SS systems, with channel-symbol rate $1/T_s$, chip rate $1/T_c$, and spreading factor $N_s = T_s/T_c$. The system has K active users, with the zeroth user taken as the reference user.

At the receiver, the received signal is demodulated, despread, and decoded by the channel decoder to m -bit indexes. These indexes are mapped by the inverse index-assignment block and decoded by the quantizer decoder. By comparing the reconstructed source with the original source symbol, we can

¹The index-assignment block is actually a part of the source coding. We separated it out for ease of analysis.

calculate the end-to-end distortion. In actual applications, such as image compression and video compression, the end-to-end distortion is generally measured by the MSE, or equivalently, peak signal-to-noise ratio (PSNR), where PSNR is defined as $10 \log_{10}(\text{Peak}^2/\text{MSE})$, and Peak is the peak value of the source. In this paper, we will use the MSE criteria.

From [8], the expected MSE of a system with a uniform source, a uniform scalar quantizer, and a random index assignment, is

$$D(m, P_e) = \frac{2^{-2m}}{12} + \frac{P_e}{6}(1 + 2^{-m}) \approx \frac{2^{-2m}}{12} + \frac{P_e}{6} \quad (4)$$

where P_e is the probability of index error, i.e., at least one bit of the m -bit index is in error, so the index is incorrect. In earlier work, without proof, [9] gives a similar result for an uncoded memoryless binary symmetric channel. Equation (4) works for all channel codes and channels.

The value of P_e depends on the channel code, modulation scheme, channel conditions, and receiver structure. Generally, finding the expression for P_e is no trivial task. In this paper, we will find a close upper bound for P_e , and thus an upper bound for the distortion $D(m, P_e)$, as a function of all three parameters r_s, r_c , and N_s . We denote this upper bound by $D_U(r_s, r_c, N_s)$, and find the optimal bandwidth allocation triplet $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$ for D_U . The true optimal bandwidth allocation for the system could be different from $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$, but by operating the system based upon the optimal allocation of the upper bound, we can guarantee that the system performs no worse than the best performance of the upper bound. In the rest of the paper, we use the term ‘‘optimal allocation’’ to refer to the optimal allocation based upon the upper bound.

Since we use nontrivial channel codes, P_e , and thus $D(m, P_e)$, are decreasing functions of $1/r_c$, i.e., if both r_s and N_s are given, the larger $1/r_c$ (for a given level of complexity) is, the better the performance of the system. Therefore, we can replace the inequality constraint (1) by an equality constraint

$$r_s \cdot \frac{1}{r_c} \cdot N_s = C_0. \quad (5)$$

Hence, the problem we need to solve is to minimize $D_U(r_s, r_c, N_s)$ under constraint (5).

In the next two sections, we introduce two different systems. For each of these two systems, we first find an upper bound for the end-to-end distortion $D_U(r_s, r_c, N_s)$, and then determine the optimal bandwidth allocation for this upper bound.

III. SYSTEM A

In system A, the bit stream out of the index-assignment block is encoded by a linear block (m, n) code with code rate $r_c = m/n$. At the receiver, the n -bit codeword is decoded with a hard-decision decoder. We consider an asynchronous CDMA system employing quaternary phase-shift keying (QPSK) modulation, and a minimum MSE (MMSE) receiver is implemented to suppress the multiple-access interference (MAI) [10]. Since we are using an MMSE receiver, the spreading sequence for each user is periodic, with a period equal to the channel symbol duration T_s . We study the performance for both the additive

white Gaussian noise (AWGN) channel and the flat Rayleigh fading channel.

A. Upper Bound on the End-to-End Distortion

For ease of analysis, we assume infinite interleaving, i.e., each bit into the decoder experiences an independent fade. For hard-decision decoding, we have

$$P_e \leq \sum_{i=t+1}^n \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \quad (6)$$

where ϵ is the raw bit-error rate (BER), $t = \lfloor (d_{\min})/(2) \rfloor$ is the number of correctable errors, and d_{\min} is the minimum distance of the block code. From the Gilbert–Varshamov bound, [11, p. 463], binary block codes exist for $d_{\min} \geq n\alpha$, where α is related to the code rate r_c through the equation

$$r_c = 1 - H(\alpha) = 1 + \alpha \log_2 \alpha + (1 - \alpha) \log_2 (1 - \alpha) \quad (7)$$

$\alpha \leq (1/2)$, and $H(\cdot)$ is the binary entropy function. In this paper, we assume $d_{\min} = \lfloor n \cdot \alpha \rfloor$.

For small ϵ , we can use the following approximation:

$$\begin{aligned} P_e &\leq \sum_{i=t+1}^n \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \approx \binom{n}{t+1} \epsilon^{t+1} \\ &= \frac{n! \cdot \epsilon^{t+1}}{(n-t-1)!(t+1)!}. \end{aligned} \quad (8)$$

From the Stirling approximation for $n!$, we have

$$P_e \leq \frac{n^{n+\frac{1}{2}}}{\sqrt{2\pi}} \cdot (n-t-1)^{-(n-t-\frac{1}{2})} \cdot (t+1)^{-(t+\frac{3}{2})} \cdot \epsilon^{t+1}. \quad (9)$$

Substituting the above into (4), and replacing m by nr_c , we have the end-to-end distortion

$$\begin{aligned} D(n, r_c, \epsilon) &\leq \frac{1}{12} \left(2^{-2nr_c} + \underbrace{\sqrt{\frac{2}{\pi}} \cdot n^{n+\frac{1}{2}}}_{K} \cdot \underbrace{(n-t-1)^{-(n-t-\frac{1}{2})}}_{K_1} \right. \\ &\quad \left. \cdot \underbrace{(t+1)^{-(t+\frac{3}{2})}}_{K_2} \cdot \epsilon^{t+1} \right) \\ &\triangleq D_U(n, r_c, \epsilon) \triangleq \frac{1}{12} D_1(n, r_c, \epsilon) \end{aligned} \quad (10)$$

where $t = \lfloor (n\alpha)/(2) \rfloor$, and α is related to r_c through (7). Note that (9) and (10) are actually approximate upper bounds, due to the fact that the Stirling approximation was used. In the rest of this section, the term ‘‘upper bound’’ is used to refer to the approximation of the upper bound.

1) *BER ϵ for MMSE Receiver:* After down conversion, the received signal is despread by being passed through a chip matched filter (MF), and then through a linear, adaptive tapped-delay line with N_s taps, as shown in [10]. Hard decisions are made in both the in-phase (I) and quadrature (Q) channels of the output of the adaptive receiver. The multiplexed binary stream is then fed into the block deinterleaver and a hard-decision channel decoder.

Assuming equal probability for the i th symbol of the reference user, (i.e., with probability $1/2$, d_0^i takes the value of $1/\sqrt{2}$ or $-1/\sqrt{2}$), and using $Q(\cdot)$ to denote the standard Q -function, it

is shown in [12] that the conditional bit-error probability (BEP) is given by

$$P_1(\text{bit error}|\{\mathbf{c}_{\text{opt}}^i\}, \{\alpha_0^i\}) = Q(\sqrt{\gamma_d/2})$$

$$\gamma_d = (\alpha_0^i)^2 \frac{\mathbf{a}_0^T \tilde{\mathbf{R}}_i^{-1} \mathbf{a}_0}{2}. \quad (11)$$

In [12], \mathbf{a}_0 is a vector containing the components of the reference user's spreading sequence, α_0^i is the magnitude of the Rayleigh fading parameter for the i th symbol of the reference user, with $E[(\alpha_0^i)^2] = 1$, and $\tilde{\mathbf{R}}_i$ is defined by (9) in [12], calculated by using the channel state information (CSI) of the reference user. We assume the receiver has perfect CSI for the reference user.

In deriving (11), we assume, as in [12], a rapidly varying channel in which the adaptive algorithm is not able to track the fading on any of the interfering users in the system. Also, we assume that the delays experienced by each user remain constant throughout a decoding interval. These assumptions result in $\tilde{\mathbf{R}}_i$ being independent of i . Thus, γ_d reduces to the following equation:

$$\gamma_d = \frac{\mathbf{a}_0^T \tilde{\mathbf{R}}_i^{-1} \mathbf{a}_0}{2} \cdot (\alpha_0^i)^2 \triangleq \gamma_c \cdot (\alpha_0^i)^2. \quad (12)$$

For the AWGN channel, we have $\alpha = 1$ (i.e., a constant); thus, the conditional BEP is given by

$$P_2(\text{bit error}|\{\tau_k\}) = Q(\sqrt{\gamma_c/2}) \quad (13)$$

where τ_k is the delay of the k th user relative to the reference signal, assumed to be independent and identically distributed (i.i.d.) for each k with a uniform distribution in the interval $[0, T_s)$, and $\tau_0 = 0$.

For a flat Rayleigh fading channel, averaging over the fading on the desired user, we can reduce the conditional BEP to a simple closed form given by [11, (14-3-7)]

$$P_2(\text{bit error}|\{\tau_k\}) = \frac{1}{2} \left[1 - \sqrt{\frac{\gamma_c}{1 + \gamma_c}} \right]. \quad (14)$$

Finally, the BER ϵ can be obtained by sample averaging P_2 over the $\{\tau_k\}$.

B. Optimization

From the previous subsection, calculation of P_2 involves evaluating an N_s -by- N_s matrix $\tilde{\mathbf{R}}$ which is determined by the spreading sequences of all the users. To obtain the coded BER, we have to average the conditional BEP P_2 in (13) or (14) over many realizations of the delays $\{\tau_k\}$. Since the spreading factor, N_s , over which we intend to optimize, does not appear in a simple manner in $D_1(n, P_e, \epsilon)$ [see (10)], we use a procedure which combines numerical evaluation and analysis to determine the optimal bandwidth allocation under a given bandwidth constraint.

Given a bandwidth constraint $r_s/r_c \cdot N_s = n \cdot N_s \leq C_0$, we calculate the ϵ 's for a given set of possible N_s 's by numerically averaging P_2 over a large number of sets of randomly generated delays $\{\tau_k\}$ for each N_s . Then we substitute the numerical value of $\epsilon(N_s)$ into D_1 , and find the optimal $(n^*(N_s), r_c^*(N_s))$ and corresponding optimal distortion $D_1(n^*(N_s), r_c^*(N_s), \epsilon(N_s))$

under the bandwidth constraint $n \leq C_0/N_s$. Here we use $(\cdot)^*$ to represent the optimal parameters for a given N_s , and use $(\hat{\cdot})$ to represent the overall optimal parameters for a given bandwidth constraint. By comparing the optimal D_1 's, we can find the optimal 3-tuple $(\hat{n}, \hat{r}_c, \hat{N}_s)$, and thus $(\hat{r}_s, \hat{r}_c, \hat{N}_s)$ among all possible bandwidth allocations of interest.

To find the optimal pair $(n^*(N_s), r_c^*(N_s))$ for a specific value of $\epsilon(N_s)$, and the constraint $n \leq C_0/N_s$, note that we always want to use all the bandwidth available. Thus, $n^*(N_s) = \lfloor C_0/N_s \rfloor$. Substituting $n^*(N_s)$ and $\epsilon(N_s)$ into D_1 , the expression can be simplified to a single variable function $D_{N_s}(r_c) \triangleq D_1(n^*(N_s), r_c, \epsilon(N_s))$, for which we want to find the optimal r_c^* .

Note that $D_{N_s}(r_c)$ is not a continuous function, since it is only meaningful when $t = \lfloor (n \cdot \alpha) / (2) \rfloor$ is an integer. However, for the sake of obtaining an analytically tractable solution, we treat this function as being continuous, with t taking any value between 0 and $\lfloor (n/4) \rfloor$. Note also that if \tilde{r} is a local optimum, and its corresponding \tilde{t} is not an integer, then we need to consider both $\lfloor \tilde{t} \rfloor$ and $\lfloor \tilde{t} \rfloor + 1$. For a given t , there are multiple values of r_c corresponding to it; among all these possible r_c 's, the largest r_c will give the smallest D_{N_s} . This is because for a constant t , the second term of (10) is constant, and the first term is a decreasing function of r_c . We denote the largest r_c which has error-correction capability t by $r_c = \chi(t)$. Therefore, the optimal \hat{r}_c is either $\chi(\lfloor \tilde{t} \rfloor)$ or $\chi(\lfloor \tilde{t} \rfloor + 1)$ for each locally optimum \tilde{t} .

For small ϵ , which holds for most channels of interest, the continuous version of $D_{N_s}(r_c)$ can be shown to be a convex function [13], which indicates there is only one global optimum \tilde{r}_c . To find \tilde{r}_c , we differentiate D_{N_s} as follows:

$$\frac{\partial D_{N_s}(r_c)}{\partial r_c} = 2^{-2nr} \cdot (-2n \ln 2) + K K_1 K_2 \cdot \epsilon^{t+1}$$

$$\cdot \left[\frac{1}{K_1} \frac{\partial K_1}{\partial r_c} + \frac{1}{K_2} \frac{\partial K_2}{\partial r_c} + \ln \epsilon \cdot \frac{\partial t}{\partial r_c} \right] = 0 \quad (15)$$

where

$$\frac{1}{K_1} \frac{\partial K_1}{\partial r_c} = \left(\ln(n-t-1) - \frac{-n+t+1/2}{n-t-1} \right) \frac{\partial t}{\partial r_c}$$

$$\frac{1}{K_2} \frac{\partial K_2}{\partial r_c} = \left(-\ln(t+1) + \frac{-t-3/2}{t+1} \right) \frac{\partial t}{\partial r_c}$$

$$\frac{\partial t}{\partial r_c} = \frac{\partial \frac{n\alpha}{2}}{\partial r_c} = \frac{n}{2 \log_2 \frac{\alpha}{1-\alpha}}.$$

Equation (15) can be solved by any good root-finding algorithm.

C. Results

The possible spreading factors we considered are 15, 31, 63, and 127. We used Kasami sequences for $N_s = 15$ and Gold sequences for $N_s = 31, 63, 127$. For each case, the ϵ were obtained by averaging the conditional error probability P_2 in (13) or (14) over 3000 realizations of sets of delays. We define E_s as the energy per coded bit, and E_c as the energy per chip.

We studied many different cases for both AWGN and flat Rayleigh fading channels. In this section, we give two sets of representative results, both for AWGN channels. For results for flat Rayleigh channels, please refer to [8].

TABLE I
OPTIMAL BANDWIDTH ALLOCATIONS FOR 3 CASES.
AWGN CHANNEL, $E_c/N_o = -7$ dB, $C_0 = 1270$

K	P_k/P_0	N_s	n	ϵ	r_c^*	$D(n, r_c^*, \epsilon)$	r_s^*
9	0dB	15	84	4.39e-2	0.0833	2.41e-5	7
		31	40	2.34e-3	0.3	<u>2.45e-8</u>	12
		63	20	5.60e-6	0.55	3.47e-8	11
		127	10	1.53e-11	1	7.95e-8	10
15	0dB	15	84	8.13e-2	0.0357	9.27e-3	3
		31	40	8.58e-3	0.2	2.64e-6	8
		63	20	3.75e-5	0.55	8.01e-8	11
		127	10	1.45e-10	1	<u>7.97e-8</u>	10
9	6dB	15	84	9.00e-2	0.0238	2.33e-3	1
		31	40	4.69e-3	0.3	2.70e-7	12
		63	20	9.67e-6	0.55	<u>3.67e-8</u>	11
		127	10	2.05e-11	1	7.95e-8	10

Table I gives the results for an AWGN channel with the chip energy-to-noise ratio $E_c/N_o = -7$ dB, and the bandwidth constraint $r_s \cdot (1)/(r_c) \cdot N_s \leq C_0 = 1270$ (e.g., for a block length $n = 127$, the largest spreading factor possible is 10). Results are shown for three different cases: nine active users with no near-far problem, 15 users with no near-far problem, and nine users with $(P_k)/(P_0) = 6$ dB. For each case, the lowest distortion is underlined. For the first case, we can see that a channel code rate of 0.3 and a spreading factor of 31 give the smallest distortion. The corresponding source-code rate $r_s = n \cdot r_c = 40 \times 0.3 = 12$, i.e., the optimal 3-tuple is $(\hat{r}_s, \hat{r}_c, \hat{N}_s) = (12, 0.3, 31)$.

For the second case, the optimal 3-tuple is (10, 1, 127). As the number of users increases, the interference from other users to the desired user increases, so we need a larger spreading factor for interference suppression. At the same time, the bandwidth allocated to both source coding and channel coding is decreased. The last case incorporates a near-far situation where all the other interfering users have a stronger power. Here the optimal 3-tuple is (11, 0.55, 63). Like the second case, as the interference from other users increases, we need to allocate more bandwidth to spreading to have better interference suppression.

When the spreading factor increases, not only can the MMSE receiver more effectively eliminate the interference from other users, but also the coded bit energy-to-noise ratio, $(E_s/N_o) = -7 + 10 \log_{10}(N_s)$ dB, increases. Note also that for all three channel conditions above, for $N_s = 127$, the error rates are different, but both the optimal r_c^* 's and the total distortions are almost the same. When ϵ is extremely small, almost all the indexes are received reliably without FEC, so $r_c^* = 1$. Thus, all the bandwidth available for both source coding and channel coding is allocated to source coding.

Table II shows the results for an AWGN channel with $E_c/N_o = -5$ dB, and $C_0 = 1905$. The three cases shown are the same as those in Table I. The optimal allocations are (22, 0.3607, 31), (21, 0.7, 63), and (20, 0.6667, 63), respectively. Note that when $N_s = 127$, all three cases have very low channel-error rates, so the total distortion is solely due to

TABLE II
OPTIMAL BANDWIDTH ALLOCATIONS FOR 3 CASES.
AWGN CHANNEL, $E_c/N_o = -5$ dB, $C_0 = 1905$

K	P_k/P_0	N_s	n	ϵ	r_c^*	$D(n, r_c^*, \epsilon)$	r_s^*
9	0dB	15	127	0.024	0.1260	1.4e-10	16
		31	61	3.2e-4	0.3607	<u>3.3e-14</u>	22
		63	30	4.0e-8	0.6667	4.2e-13	12
		127	15	5.5e-17	1	7.8e-11	15
15	0dB	15	127	0.06	0.0630	5.4e-5	8
		31	61	2.8e-3	0.2951	1.1e-10	18
		63	30	8.7e-7	0.7	<u>5.7e-11</u>	21
		127	15	2.8e-15	1	7.8e-11	15
9	6dB	15	127	0.065	0.0472	1.3e-4	6
		31	61	8.5e-4	0.3607	3.6e-12	22
		63	30	6.6e-8	0.6667	<u>6.3e-13</u>	20
		127	15	1.0e-16	1	7.8e-11	15

quantization error. Therefore, since the number of source bits per source symbol is the same for each case, the final distortions are the same.

In practice, the source-code rate r_s varies, depending on the application. For example, speech-coding systems can achieve 0.3–4 bits per sample, i.e., compressed rates of 2.4–32 kb/s for 8-kHz samples. A high-quality image coder compresses by a factor of at least 10, and sometimes up to 30, which translates to 0.8–2.4 bits per pixel for a color image. In medical imaging, where sometimes no compression is allowable, a grayscale image could have a source-code rate of 12 or more bits per source symbol.

In general image-compression applications, a PSNR of around 32 dB results in a good compressed image (this corresponds to an MSE of 6.31×10^{-4} when normalized to a uniform quantizer with peak value 1, as in this paper). In medical image compression, a PSNR of 40–55 dB or higher is desirable [14] (corresponding to a normalized MSE of 1.0×10^{-4} to 3.16×10^{-6}), depending on the application. In some cases, when a medical image starts out as an analog image, it is initially compressed by digitizing it, in which case, one might use 12 or more bits per pixel. Often there is no further compression beyond this, and the PSNR is ≥ 83 dB (a normalized MSE of 4.97×10^{-9}). These values serve to motivate our selection of the fixed distortion requirement in what follows.

As mentioned in the Introduction, our system optimization can be converted to two alternative optimization problems. Table III shows the maximum number of users with different bandwidth allocations for a given system end-to-end distortion requirement of $D \leq 3.00 \times 10^{-6}$. We assume that the transmission power for each user is the same. From the table, we can see that for $N_s = 15$, the system capacity is $K = 7$ users, with the optimal (m^*, r_c^*) for this N_s . Similarly, the system can achieve a capacity of 15, 23, and 45 users, for $N_s = 31, 64,$ and 127, respectively. Note the largest system capacity overall is 45 users with $(\hat{r}_s, \hat{r}_c, \hat{N}_s) = (10, 1, 127)$. Thus, in this case, where the chip energy-to-noise ratio is low, the best system performance is achieved by using the largest spreading factor, which leads to

TABLE III
OPTIMAL SYSTEM CAPACITY, AWGN CHANNEL, $E_c/N_o = -7$ dB,
 $P_k/P_0 = 0$ dB, $C_0 = 1270$, AND $D \leq 3.00 \times 10^{-6}$

N_s	n	ϵ	r_c^*	$D(n, r_c^*, \epsilon)$	r_s^*	K^*
15	84	3.31e-2	0.11905	1.36e-6	10	7
31	40	8.58e-3	0.2	2.64e-6	8	15
63	20	2.79e-4	0.55	2.61e-6	11	23
127	10	1.68e-6	1	2.88e-6	10	45

TABLE IV
OPTIMAL COVERAGE RADIUS, AWGN CHANNEL, $K = 9$,
 $P_k/P_0 = 0$ dB, $C_0 = 1270$, AND $D \leq 8.00 \times 10^{-8}$

N_s	n	ϵ	r_c^*	$D(n, r_c^*, \epsilon)$	r_s^*	$E_c/N_o(\text{dB})$	Radius
15	84	2.29e-2	0.14286	6.91e-8	12	-4.8	d
31	40	3.48e-3	0.3	7.37e-8	12	-7.5	1.17d
63	20	3.44e-5	0.55	7.27e-8	11	-8.0	1.20d
127	10	2.54e-10	1	7.99e-8	10	-7.6	1.17d

a higher SNR. Table IV shows the optimal coverage radius for the reference user for a given end-to-end distortion requirement of $D \leq 8.00 \times 10^{-8}$. In the table, we assume $P_k/P_0 = 0$ dB. Coverage radii are normalized to the radius of the system with the smallest spreading factor (i.e., $N_s = 15$).

Assuming the transmission power is constant, the distances in the table are calculated by assuming that the received signal power is inversely proportional to the fourth order of the distance, i.e., a 12-dB drop in E_c/N_o indicates a doubling of the distance. We can see that the bandwidth allocation $(\hat{r}_s, \hat{r}_c, \hat{N}_s) = (11, 0.55, 63)$ results in the largest coverage area. The required MSE values in Tables III and IV are arbitrarily picked, but nevertheless are realistic.

IV. SYSTEM B

In this system, the bit stream out of the index-assignment block is encoded by a nonterminated convolutional encoder with code rate r_c . At the receiver, a soft-decision Viterbi decoder decodes the noise-contaminated bit stream to indexes. The output of the interleaver is multiplied by a long pseudorandom sequence assigned to the given user and transmitted using binary phase-shift keying (BPSK) modulation.

A. Upper Bound on the End-to-End Distortion

Since we transmit the indexes by sequentially passing them through a nonterminated convolutional code, the m -bit index-error rate is also the frame-error rate (FER) of this convolutional encoder. A frame of size m consists of m consecutive information bits. The error rate of an information frame of size m is the probability that at least one of the m bits in the frame is decoded incorrectly. In [15], an upper bound for the frame-error probability (FEP) was given heuristically, but a requirement of very large m/r_c was posed. In Appendix B, we derive a tight upper

bound for FER for any coded frame lengths which are larger than the constraint length

$$P_e = P_f(m) \leq \sum_d ((l_m - 1)a_d + b_d)P_1(d) \quad (16)$$

where m is the information frame size, l_m is the number of branches of the trellis that are in a frame, $P_1(d)$ is the pairwise probability of two sequences that have Hamming distance d , and a_d and b_d are defined in (44). Values of b_d for memory $M = 4$ RCPC codes in [6] are listed in Table V. For both AWGN and Rayleigh fading channels, we calculate $P_1(d)$ and then optimize the end-to-end distortion of the system. For all the RCPC codes in [6], the number of information bits I_b on each trellis branch is one. In these cases, $l_m = m/I_b = m$. This property will be used in our optimization.

1) *AWGN Channel*: For a DS-CDMA system with a large number of users, the pairwise error probability (PEP) for the AWGN channel is approximately given by

$$\begin{aligned} P_1(d) &= Q\left(\sqrt{\frac{E_s \cdot d}{g(K-1)E_s/N_s + N_o/2}}\right) \\ &= Q\left(\sqrt{\frac{E_c \cdot dN_s}{g(K-1)E_c + N_o/2}}\right) \triangleq Q(\sqrt{HdN_s}) \end{aligned} \quad (17)$$

where

$$H \triangleq \frac{E_c}{g(K-1)E_c + N_o/2} = \frac{1}{g(K-1) + N_o/2E_c}. \quad (18)$$

Also E_s is the energy-per-channel bit, $(N_o/2)$ is the power spectral density of the white Gaussian noise, g is a constant which depends on the pulse shape and equals $2/3$ when we use rectangular-shaped chips, K is the total number of users, N_s is the spreading factor, and $E_c = (E_s)/(N_s)$ is the energy per chip (which is kept constant). Substituting (17) into (16) and then into (4), we have

$$\begin{aligned} D(m, r_c, N_s) &\leq \frac{1}{12} \cdot 2^{-2m} + \frac{1}{6} \sum_d ((l_m - 1)a_d + b_d) \\ &\quad \cdot Q(\sqrt{HdN_s}) \\ &\triangleq D_U(m, r_c, N_s). \end{aligned} \quad (19)$$

2) *Flat Rayleigh Fading Channel*: Assume $E[\alpha^2] = 1$, where α is the fade amplitude and has a Rayleigh density, and assume the fading seen by the channel decoder is uncorrelated from bit to bit. For a DS-CDMA system, the conditional PEP, conditioned on the fading parameters, is given by [11]

$$\begin{aligned} P_2(d|\{\alpha_i\}) &= Q\left(\sqrt{\frac{E_s \sum_0^{d-1} \alpha_i^2}{g(K-1)E_s/N_s + N_o/2}}\right) \\ &= Q\left(\sqrt{\frac{E_c \cdot N_s \sum_0^{d-1} \alpha_i^2}{g(K-1)E_c + N_o/2}}\right) \\ &\triangleq Q\left(\sqrt{HN_s \sum_0^{d-1} \alpha_i^2}\right). \end{aligned}$$

TABLE V
 b_d FOR RCPC CODES WITH MEMORY 4, PERIOD 8

	8/9 8/9	8/10 4/5	8/12 2/3	8/14 4/7	8/16 1/2	8/18 4/9	8/20 2/5	8/22 4/11	8/24 1/3	8/26 4/13	8/28 2/7	8/30 4/15	8/32 1/4
	1111 0111 1000 1000 0000 0000	1111 1111 1000 1000 0000 0000	1111 1111 1010 1010 0000 0000	1111 1111 1110 1110 0000 0000	1111 1111 1111 1111 0000 0000	1111 1111 1111 1111 1000 1000	1111 1111 1111 1111 1100 1100	1111 1111 1111 1111 1110 1110	1111 1111 1111 1111 1111 1111	1111 1111 1111 1111 1111 1111	1111 1111 1111 1111 1111 1111	1111 1111 1111 1111 1111 1111	1111 1111 1111 1111 1111 1111
3	494	88								1000 1000	1010 1010	1110 1110	1111 1111
4	8170	568	20										
5	120693	5276	0	10									
6	1661611	41600	1092	148									
7	21741344	293712	0	308	104	10							CODE
8		2078020	22260	810	216	100	10						10011
9			0	4258	320	146	96	28					11101
10				11024	1408	196	76	32					10111
11					3960	820	162	98	40	10			11011
12					8168	1964	526	180	120	20			
13							852	274	184	140	68	10	
14								544	136	106	84	48	
15									240	100	92	104	96
16									624	274	64	0	48
17											144	118	0
18											472	196	72
19													128
20													264

Averaging $P_2(d|\{\alpha_i\})$ over the distribution of all $\alpha_i, i = 0, \dots, d-1$ yields the PEP [11, (14.4.15)]

$$P_1(d) = \left(\frac{1-\mu}{2}\right)^d \sum_{i=0}^{d-1} \binom{d-1+i}{i} \left(\frac{1+\mu}{2}\right)^i$$

where

$$\mu = \sqrt{\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}} \quad (20)$$

and $\bar{\gamma}_c = (1/2)HN_sE[\alpha_i^2]$. When $\bar{\gamma}_c \gg 1$, (i.e., > 10 dB), $P_1(d)$ can be further simplified to

$$\begin{aligned} P_1(d) &\approx \left(\frac{1}{4\bar{\gamma}_c}\right)^d \cdot \binom{2d-1}{d} = \left(\frac{1}{2HN_s}\right)^d \binom{2d-1}{d} \\ &= \left(\frac{1}{2H}\right)^d \binom{2d-1}{d} \cdot N_s^{-d} \triangleq h(d) \cdot N_s^{-d} \end{aligned} \quad (21)$$

where

$$h(d) = \left(\frac{g(K-1) + N_0/2E_c}{2}\right)^d \binom{2d-1}{d}. \quad (22)$$

Substituting (21) into (16) and then into (4), we have

$$\begin{aligned} D(m, r_c, N_s) &\leq \frac{1}{12} \cdot 2^{-2m} + \frac{1}{6} \sum_d ((l_m - 1)a_d + b_d) \\ &\quad \cdot h(d) \cdot N_s^{-d} \\ &\triangleq D_U(m, r_c, N_s). \end{aligned} \quad (23)$$

B. Optimization

In the equations for the upper bound of the distortion, (19) and (23), $D_U(m, r_c, N_s)$ is not a simple function of r_c , i.e., for a given set of RCPC codes, the spectrum a_d and b_d cannot be

written as a function of r_c . Thus, we cannot find the optimal bandwidth allocation by taking derivatives of D_U with respect to r_c . We will use a similar approach to that for system A to find the optimal bandwidth-allocation triplet $(\hat{m}, \hat{r}_c, \hat{N}_s)$. We first fix r_c , and find the optimal allocation (m^*, N_s^*) and the minimum distortion $D_U^*(m)$ for this r_c . Then by comparing the minimum $D_U^*(m)$'s for different r_c , we find the best triplet. For a given channel-code rate r_c , we can use the bandwidth constraint and substitute $N_s = C_0 \cdot r_c / m$ into $D_U(m, r_c, N_s)$, so that the upper bound D_U becomes a function of a single variable m . We denote this new function $D_U(m)$.

1) *AWGN Channel*: Substituting $N_s = C_0 \cdot r_c / m$ into (19), we have

$$\begin{aligned} D_U(m) &= \frac{1}{12} \cdot 2^{-2m} + \frac{1}{6} \sum_d ((l_m - 1)a_d + b_d) \\ &\quad \times Q(\sqrt{HdC_0r_c/m}). \end{aligned} \quad (24)$$

Differentiating $D_U(m)$ with respect to m , and setting it equal to zero, results in

$$\begin{aligned} 0 &= 12 \frac{\partial D_U(m)}{\partial m} = 2^{-2m} \cdot (-2 \ln 2) \\ &\quad + 2 \sum_d a_d \cdot \frac{1}{I_b} Q(\sqrt{HdC_0r_c/m}) \\ &\quad + 2 \sum_d ((l_m - 1)a_d + b_d) \cdot \left(-\frac{e^{-HdC_0r_c/2m}}{\sqrt{2\pi}}\right) \cdot \left(-\frac{\sqrt{HdC_0r_c}}{2\sqrt{m^3}}\right) \\ &\approx 2^{-2m} \cdot (-2 \ln 2) + 2 \sum_d \frac{a_d}{I_b} \frac{e^{-HdC_0r_c/2m}}{\sqrt{2\pi} \sqrt{HdC_0r_c/m}} \\ &\quad + 2 \sum_d ((l_m - 1)a_d + b_d) \cdot \frac{e^{-HdC_0r_c/2m}}{\sqrt{2\pi}} \cdot \frac{\sqrt{HdC_0r_c}}{2\sqrt{m^3}}. \end{aligned} \quad (25)$$

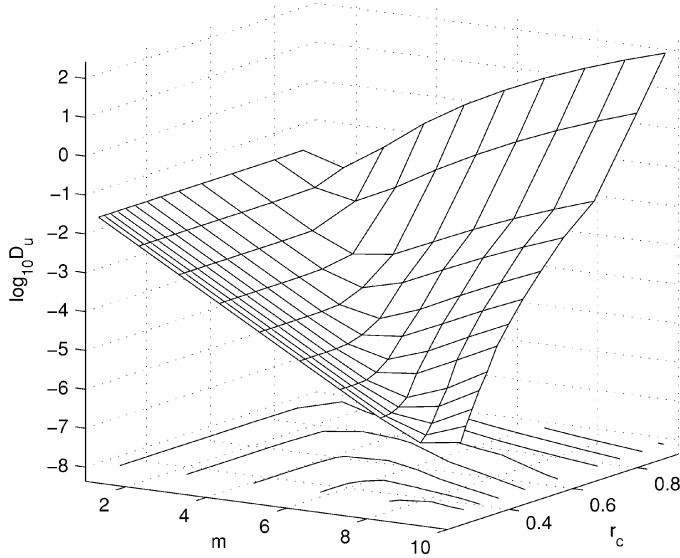


Fig. 2. Distortion D_U versus source-code rate m and channel-code rate r_c . $C_0 = 800$, $E_c/N_o = -6$ dB, $K = 10$.

The approximation in the last step of (25) is valid when $HdC_0r_c/m = HdN_s$ is large. It is easy to show that $D_U(m)$ is convex cup, so solving (25) numerically with any good root-finding algorithm gives the optimal m^* for an AWGN channel.

2) *Flat Rayleigh Channel*: Substituting $N_s = C_0 \cdot r_c/m$ into (23) results in

$$D_U(m) = \frac{1}{12} \cdot 2^{-2m} + \frac{1}{6} \sum_d h(d)(C_0 r_c)^{-d} \cdot ((l_m - 1)a_d + b_d)m^d. \quad (26)$$

Upon setting the derivative of $D_U(m)$ equal to zero, we obtain

$$\begin{aligned} 0 &= 12 \frac{\partial D_U(m)}{\partial m} \\ &= 2^{-2m} \cdot (-2 \ln 2) + 2 \sum_d a_d h(d)(C_0 r_c)^{-d} \\ &\quad \cdot \left(\frac{m^d}{l_b} + ((l_m - 1)a_d + b_d)d \cdot m^{d-1} \right). \end{aligned} \quad (27)$$

As was the case with (25), (27) needs to be solved numerically. Note that for large SNRs $(E_s)/(N_o)$, we can ignore non-minimum-distance error events, and thus use simpler forms of $(\partial D_U)/(\partial m)$ for both cases above.

C. Results

Fig. 2 shows the upper bound for the end-to-end distortion D_U versus the source-code rate and channel-code rate for an uncorrelated Rayleigh fading channel, under the bandwidth constraint $(r_s)/(r_c) \cdot N_s \leq C_0 = 800$. Here $(E_c)/(N_o)$ is -6 dB, and the number of active users in the system is $K = 10$. The RCPC codes used are from [6, Table I]; their spectra are listed in Table II of this paper. From Fig. 2, we can see that, for each given channel-code rate r_c , there is an optimal source-code rate m that achieves the smallest D_U for this r_c . The global optimum falls at the smallest r_c , i.e., the strongest channel coding. This was true for both AWGN and flat Rayleigh fading channels when no interference suppression was implemented.

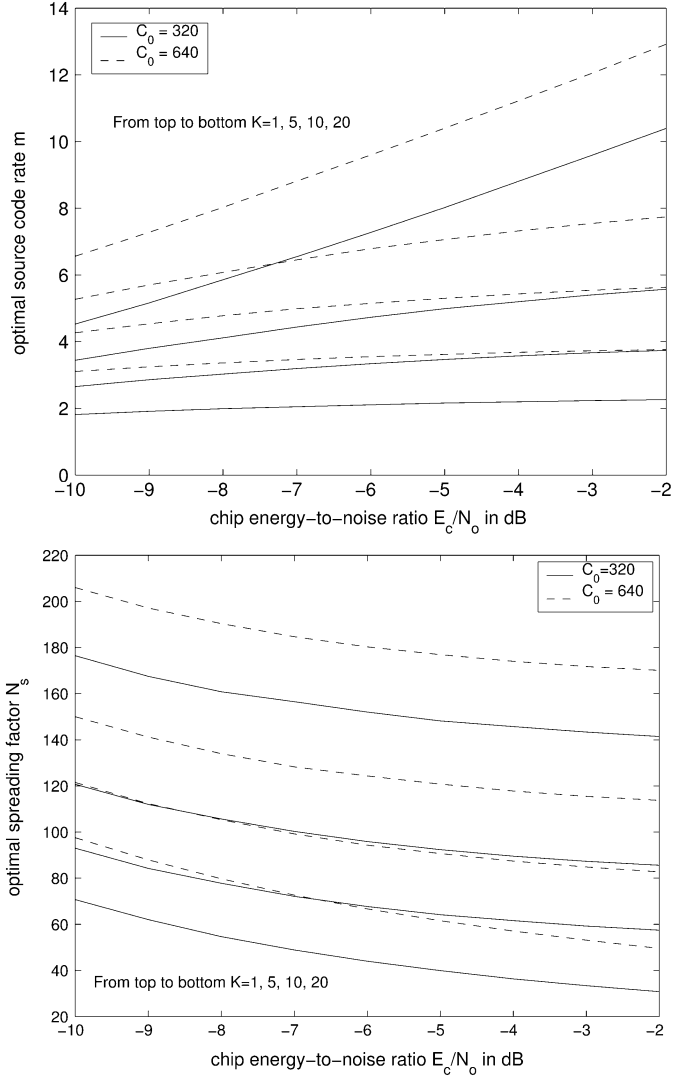


Fig. 3. m^* and N_s^* versus chip energy-to-noise ratio E_c/N_o . Flat Rayleigh fading channel.

For any fixed r_c , by solving (25) and (27), we also show in the following figures how m^* and N_s^* vary when the channel conditions change.

Fig. 3(a) shows the variation of the optimal m with the chip energy-to-noise ratio $(E_c)/(N_o)$, and Fig. 3(b) shows analogous results for the optimal value of N_s . There are two sets of curves on each figure, one for bandwidth constraint $C_0 = 320$, and the other for $C_0 = 640$. The curves are parameterized by the number of users $K = 1, 5, 10, 20$. Also, all curves correspond to the use of the memory 4, rate-1/2 code in [6], and an uncorrelated flat Rayleigh fading channel.

For each curve in Fig. 3, where the number of users K is fixed, we see that as $(E_c)/(N_o)$ increases, m^* increases, and N_s^* decreases. This is because the spreading factor N_s has two effects on the performance of the system: 1) A larger N_s suppresses more interference from other users; 2) a larger N_s leads to a larger $E_b = (N_s \cdot E_c)/(r_c)$, and thus reduces the raw error rate into the channel decoder. As $(E_c)/(N_o)$ increases, we do not need as large an N_s for a given E_b , and thus, we can allocate more of the available bandwidth to source coding, i.e., increase m .

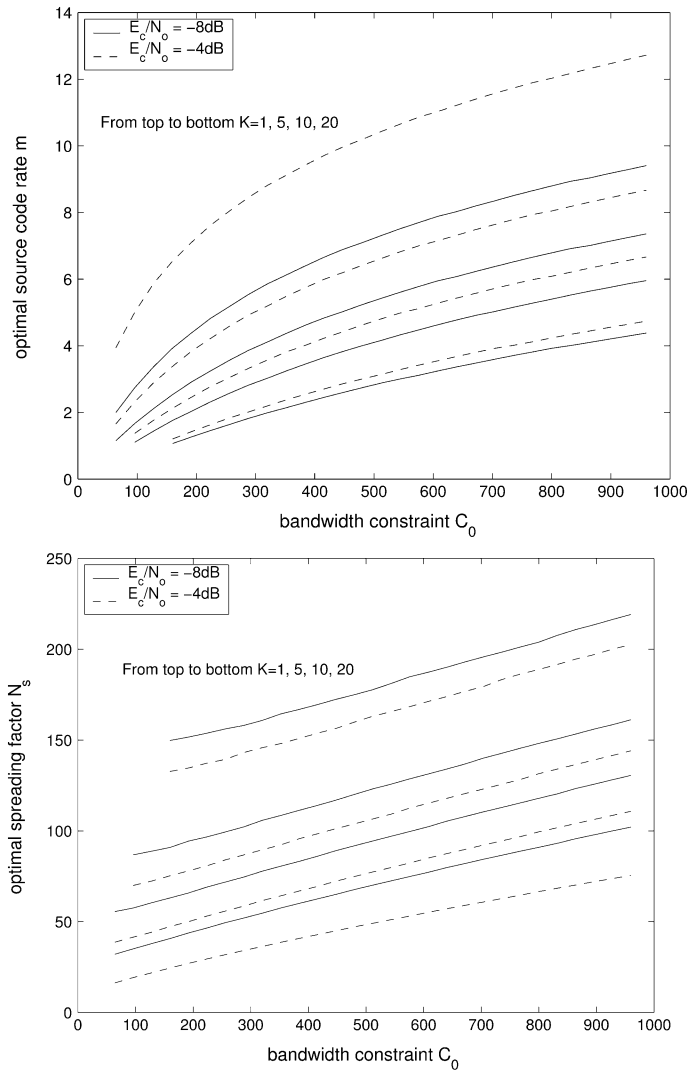


Fig. 4. m^* and N_s^* versus bandwidth constraint C_0 . Flat Rayleigh fading channel.

Alternatively, for each set of curves which have the same bandwidth constraint, we see that as the number of users increases, m^* decreases. This is because we need to allocate more bandwidth to spreading to suppress the multiuser interference (MUI). We can also see that the decrease of $m^*(N_s^*)$ is slower for a larger number of users. This is because with more users, the MUI dominates the thermal noise, while the effects of the change of the $(E_c)/(N_o)$ are comparably less significant.

Fig. 4 illustrates how m^* and N_s^* change as the bandwidth constraint C_0 changes. The system used in this figure is the same as in Fig. 3. From this figure, we see that as C_0 increases, m^* increases, and m^* increases faster when there are fewer users in the system. This is because when there is less interference, as $(E_c)/(N_o)$ increases, the channel condition improves faster than when there is more interference. Thus we do not need as large a spreading factor N_s , and we can afford to allocate more bandwidth to the source coding. Similar results for the AWGN channel are presented in [13].

V. CONCLUSION

In this paper, we studied the bandwidth-allocation problem for two CDMA systems. System A employed hard-decision block channel coding, and an MMSE receiver for interference suppression. System B employed RCPC channel coding and soft-decision Viterbi decoding. Under a bandwidth constraint, we optimized the system performance by combining analytical and numerical techniques. In addition, we derived a tight upper bound for the FEP of convolutional codes. We also proved that our optimal allocation results apply to a large class of source distributions in Appendix A.

For the former system, when there is a large number of active users in the system, or when the interference from other users is high, we have an interference-limited system. In this case, it is desirable to allocate less bandwidth to source coding and channel coding, and more bandwidth to spreading, so that the MMSE receiver can effectively suppress the interference. On the other hand, when the MAI is small, it is always beneficial to allocate less bandwidth to spreading, so that there is more bandwidth available to the rest of the system. Also, for a given spreading factor, when the raw BER is small, we should allocate more bandwidth to the source coding, so that we have a good representation of the original source; when the channel BER is large, we should allocate more bandwidth to the channel coding, so that we can get more source symbols correctly through the channel.

For the latter system, where no interference-suppression scheme is employed, our results show that for both AWGN and flat Rayleigh fading channels, it is beneficial to use the strongest channel code possible when the complexity of the system is not a concern. We also showed, for a given r_c , how the optimal allocation between $(r_s^*$ and $N_s^*)$ changes when the channel conditions (number of interfering users, channel noise, or bandwidth constraints) change.

APPENDIX A

In this appendix, we show the main steps of a proof showing that with a uniform quantizer, the optimal allocation \hat{m} , for a nonuniform source, with a uniform source encoder, is within ± 0.5 bits away from the optimal \hat{m} of the uniform source, if the following conditions hold:

- the \hat{m} -bit quantizer can be regarded as a high-density quantizer for this nonuniform source;
- the derivative of the index-error probability, $(\partial P_e(m))/(\partial m)$, is monotonically increasing in m .

A detailed proof can be found in [13].

Note that if \hat{m} is the optimal source-code rate for the upper bound for a system with a uniform source, then \hat{m} refers to the optimal source code rate for the upper bound for a system with a nonuniform source.

Proof: Assume a continuous source X has a density function $f(x)$ with support on $[0, 1)$.

For an m -bit uniform quantizer on $[0, 1)$, the code points and partitions are given in (3). Let the probability of y_i, p_i be defined as $p_i \triangleq \int_{x \in S_i} f(x) dx$. Also, let $q(j|i)$ be the probability that the source decoder input is the j th index when the source

encoder outputs the i th index. From the definition of $P_e(m)$, we know $q(i|i) = 1 - P_e(m)$. Further, since we are using a random index assignment, $q(j|i) = q(k|i) \forall j \neq i, k \neq i$, and so we have $q(j|i) = (P_e(m))/(2^m - 1) \forall i, j, i \neq j$. Thus, the end-to-end distortion is given by

$$\begin{aligned}
\tilde{D}(m) &= \sum_{i=0}^{2^m-1} \sum_{j=0}^{2^m-1} q(j|i) \int_{S_i} (x - y_j)^2 f(x) dx \\
&= \sum_{i=0}^{2^m-1} \sum_{j=0}^{2^m-1} q(j|i) \int_{S_i} (x - y_i)^2 f(x) dx \\
&\quad + 2 \sum_{i=0}^{2^m-1} \sum_{j=0}^{2^m-1} (y_i - y_j) q(j|i) \int_{S_i} (x - y_i) f(x) dx \\
&\quad + \sum_{i=0}^{2^m-1} \sum_{j=0}^{2^m-1} q(j|i) (y_i - y_j)^2 \cdot \int_{S_i} f(x) dx \\
&= \sum_{i=0}^{2^m-1} \int_{S_i} (x - y_i)^2 f(x) dx + \frac{2P_e(m)}{2^m - 1} \\
&\quad \times \underbrace{\sum_{i=0}^{2^m-1} \int_{S_i} (x - y_i) \sum_{j=0}^{2^m-1} (y_i - y_j) f(x) dx}_{T_2} \\
&\quad + \underbrace{\frac{P_e(m)}{2^m - 1} \sum_{i=0}^{2^m-1} \sum_{j=0}^{2^m-1} (y_i - y_j)^2 \cdot p_i}_{T_3}. \tag{28}
\end{aligned}$$

The first term of (28) is the quantization distortion of a uniform quantizer for a nonuniform source. When the first condition is satisfied, i.e., the m -bit quantizer is a high-resolution quantizer for the given source, each region can be regarded as a uniform region, and the overall quantization error is approximately $2^{-2m}/12$.

The second term of (28) can be bounded as follows [13]:

$$\left| \frac{2P_e(m)}{2^m - 1} \cdot T_2 \right| \leq P_e(m) 2^{-m-1} (1 + o(1)) \tag{29}$$

where $o(\cdot)$ is the *little-o* notation, i.e., if $a(m) = o(b(m))$, then $\lim_{m \rightarrow \infty} a(m)/b(m) = 0$. The third term of (28) is given by [13]

$$\frac{P_e(m)}{2^m - 1} T_3 = P_e(m) \left(E[x^2 - x] + \frac{1}{3} + o(1) \right) \tag{30}$$

where $E[\cdot]$ indicates the expected value.

Substituting (29) and (30) into (28), we have

$$\tilde{D}(m) \leq \frac{1}{12} 2^{-2m} + P_e(m) \left(E[x^2 - x] + \frac{1}{3} \right). \tag{31}$$

Since $-(1/4) \leq x(x-1) \leq 0$, we have $(1/12) \leq E[x^2 - x] + (1/3) \leq (1/3)$, and we can write (31) in the following form:

$$\tilde{D}(m) = \frac{1}{12} 2^{-2m} + \frac{k}{6} P_e(m) \tag{32}$$

where $k = 6(E[x^2 - x] + (1/3))$ is a constant, and $k \in [(1/2), 2]$. Let $\hat{D}(m)$ denote the end-to-end distortion for the

uniform source. Note that $\hat{D}(m)$ is a special case of $\tilde{D}(m)$, where $k = 1$. We have

$$\begin{aligned}
&\tilde{m} \text{ optimizes } \tilde{D}(m) \\
&\iff \left. \frac{\partial \tilde{D}(m)}{\partial m} \right|_{\tilde{m}} = 0 \\
&\iff 2^{-2\tilde{m}} \cdot (-2 \ln 2) + 2k \left. \frac{\partial P_e(m)}{\partial m} \right|_{\tilde{m}} = 0 \\
&\iff 2^{-2\tilde{m}} \cdot \ln 2 = k \cdot \left. \frac{\partial P_e(m)}{\partial m} \right|_{\tilde{m}}. \tag{33}
\end{aligned}$$

For a uniform source, we have

$$2^{-2\hat{m}} \cdot \ln 2 = 1 \cdot \left. \frac{\partial P_e(m)}{\partial m} \right|_{\hat{m}}. \tag{34}$$

Dividing both sides of (33) by (34), we obtain

$$2^{-2(\tilde{m}-\hat{m})} = k \cdot \frac{\left. \frac{\partial P_e(m)}{\partial m} \right|_{\tilde{m}}}{\left. \frac{\partial P_e(m)}{\partial m} \right|_{\hat{m}}} \triangleq k \cdot C. \tag{35}$$

When the second condition is satisfied, i.e., $(\partial P_e(m))/(\partial m)$ is a monotonic increasing function of m , we can show that \tilde{m} is within 0.5 b away from \hat{m} .

• If $\tilde{m} > \hat{m}$, the second condition $\Rightarrow C > 1$. From (35)

$$2^{-2(\tilde{m}-\hat{m})} = k \cdot C > k > \frac{1}{2} \Rightarrow \tilde{m} - \hat{m} < 0.5 \Rightarrow \tilde{m} < \hat{m} + 0.5.$$

• If $\tilde{m} < \hat{m}$, the second condition $\Rightarrow C < 1$. From (35)

$$2^{-2(\tilde{m}-\hat{m})} = k \cdot C < k < 2 \Rightarrow \tilde{m} - \hat{m} > -0.5 \Rightarrow \tilde{m} > \hat{m} - 0.5.$$

Therefore, it follows that $\hat{m} - 0.5 < \tilde{m} < \hat{m} + 0.5$. ■

APPENDIX B FER FOR CONVOLUTIONAL CODES

We present here the key steps of the derivation (details are in [13]). If we let I_b be the number of information bits on one trellis branch, and let l_m be the number of trellis branches in a frame, then $l_m = \lceil (m/I_b) \rceil$, or $l_m = \lceil (m/I_b) \rceil + 1$, depending on whether or not the frame starts at the beginning of a branch. For convolutional codes, we can assume the all-zero information sequence is transmitted. We denote the path taken by the decoder for a frame as $B_i^{i+l_m-1} = \{b_i, b_{i+1}, \dots, b_{i+l_m-1}\}$, where each b_i is a branch with starting node i and ending node $i+1$.

Define event $\alpha_i \triangleq$ node i is in the all-zero state, event $\beta_i \triangleq$ an error event starts at node i , and event $\gamma_i \triangleq$ an error event ends at node i . Then $P(\beta_i)$ and $P(\alpha_i)$ are given by

$$\begin{aligned}
P(\beta_i) &= P(\beta_i|\alpha_i)P(\alpha_i) + P(\beta_i|\bar{\alpha}_i)P(\bar{\alpha}_i) \\
&= P(\beta_i|\alpha_i)P(\alpha_i) + 0 \cdot P(\bar{\alpha}_i) = P(\beta_i|\alpha_i)P(\alpha_i) \tag{36}
\end{aligned}$$

$$\begin{aligned}
P(\alpha_i) &= P(b_{i-1} \text{ not on an error path}) + P(\gamma_i) \\
&= 1 - P(b_{i-1} \text{ on an error path}) + P(\gamma_i). \tag{37}
\end{aligned}$$

In [16], the *error-event probability* P_E is defined as the probability that the decoder is off the correct path at a given branch b_i , which is equal to the second term of (37). Reference [16] also defines the *first error probability* $P_{E,1}$ as the probability that an error event begins at a node i , given that node i is in all-zero state, i.e., $P_{E,1} \triangleq P(\beta_i|\alpha_i)$.

Note that all error probabilities that we defined so far are independent of the node index i . Also, since for every error event,

there is one starting node and one ending node, $P(\beta_i) = P(\gamma_i)$. Substituting all definitions and (37) into (36), we have

$$P(\beta) = P_{E,1} \cdot (1 - P_E + P(\gamma)) = P_{E,1} \cdot (1 - P_E + P(\beta)).$$

Solving the above equation, we obtain

$$P(\beta) = \frac{1 - P_E}{1 - P_{E,1}} \cdot P_{E,1} \quad (38)$$

and comparing (38) with (36), we see that

$$P(\alpha_i) \triangleq P(\text{node } i \text{ in all-zero state}) = \frac{1 - P_E}{1 - P_{E,1}}. \quad (39)$$

$P_{E,1}$ can be bounded using the *complete path enumerator* [16] $T(x, y) = \sum_{d,s} T_{d,s} x^d y^s$, where d is the Hamming weight of the encoder output of a path, s is the length of a path, and d, s both go from 1 to $+\infty$. Then $P_{E,1}$ is upper bounded by $P_{E,1} \leq \sum_{d,s} T_{d,s} P_1(d)$, where $P_1(d)$ is the pairwise probability of two sequences that have Hamming distance d . Similarly, we can bound P_E as $P_E \leq \sum_{d,s} s T_{d,s} P_1(d)$.

Now for the frame $B_i^{i+l_m-1}$, the FEP is

$$\begin{aligned} P_f(l_m) &\leq \sum_{j=i}^{i+l_m-1} P(\text{an error event starts at node } j) \\ &\quad + \sum_{j=1}^{\infty} P(\text{an error event with more than } j \\ &\quad \text{branches starts at node } i - j) \\ &= l_m \cdot P(\beta) + \sum_{j=1}^{\infty} P(\text{an error event with more than } \\ &\quad j \text{ branches starts at node } i - j) \\ &\triangleq l_m \cdot P(\beta) + P_2. \end{aligned} \quad (40)$$

Using the same techniques, we can upper bound the second term of (40) by

$$\begin{aligned} P_2 &\leq \sum_{j=1}^{\infty} \sum_{d,s=j+1,\dots,\infty} T_{d,s} P_1(d) \cdot P(\alpha_{i-j}) \\ &= P(\alpha) \cdot \sum_{d,s} \sum_{j=1}^{s-1} T_{d,s} P_1(d) \\ &= P(\alpha) \cdot \sum_{d,s} (s-1) T_{d,s} P_1(d). \end{aligned} \quad (41)$$

Substituting (41) into (40), we have the union bound of the FER

$$P_f(l_m) \leq l_m \cdot P(\beta) + P(\alpha) \cdot \sum_{d,s} (s-1) T_{d,s} P_1(d). \quad (42)$$

After some simplifications, we have

$$\begin{aligned} P_f(l_m) &\leq \sum_d (l_m - 1) P_1(d) a_d + \sum_d P_1(d) b_d \\ &= \sum_d ((l_m - 1) a_d + b_d) P_1(d) \end{aligned} \quad (43)$$

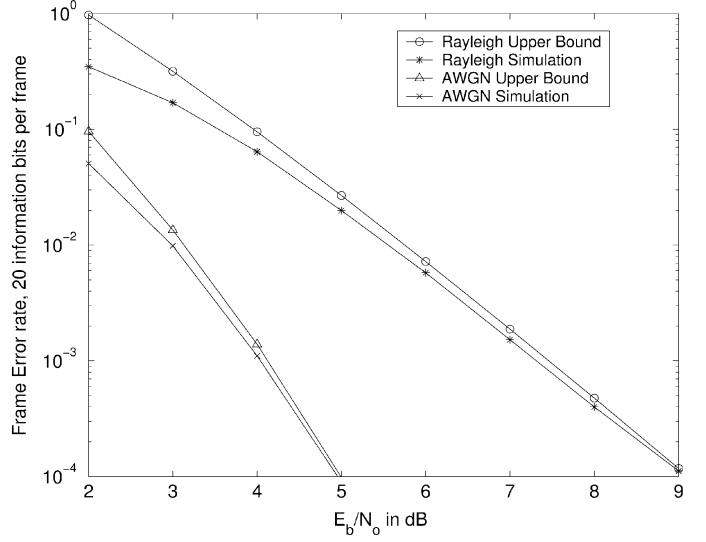


Fig. 5. Upper bound on FER for rate-1/2 convolutional code in Table V.

where a_d is defined, as in [6], by

$$\begin{aligned} a_d &\triangleq \sum_s T_{d,s}, \quad \text{and} \\ b_d &\triangleq \sum_s s T_{d,s} = \left. \frac{\partial \sum_s T_{d,s} x^d y^s}{\partial y} \right|_{x=y=1}. \end{aligned} \quad (44)$$

The values of b_d for the memory length 4 RCPC code of [6, Table I] are listed in Table V of this paper. More listings of b_d for other RCPC codes in [6] can be found in [13]. In Fig. 5, we compare the bound of (43) with simulation results for the rate-1/2 code in Table V. From the plot, we can see that the theoretical upper bound is quite tight.

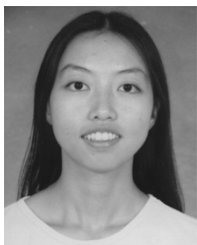
ACKNOWLEDGMENT

The authors wish to thank Professor K. Zeger for valuable discussions and suggestions. We also wish to thank the anonymous reviewers for constructive suggestions.

REFERENCES

- [1] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sept. 1997.
- [2] M. Zhao, A. A. Alatan, and A. N. Akansu, "A new method for optimal rate allocation for progressive image transmission over noisy channels," in *Proc. IEEE DCC*, Mar. 2000, pp. 213–222.
- [3] K. H. Li and L. B. Milstein, "On the optimum processing gain of a block-coded direct-sequence spread-spectrum system," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 618–626, May 1989.
- [4] D. J. V. Wyk, I. J. Oppermann, and L. P. Linde, "Performance tradeoff among spreading, coding, and multiple-antenna transmit diversity for high capacity space-time coded DS-CDMA," in *Proc. MILCOM*, vol. 1, Sept. 1999, pp. 393–397.
- [5] Q. Zhao, P. C. Cosman, and L. B. Milstein, "Tradeoffs of source coding, channel coding, and spreading in frequency selective Rayleigh fading channels," *J. VLSI Signal Processing—Syst. Signal, Image, Video Technol.*, vol. 30, pp. 7–20, Feb. 2002.

- [6] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCP codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.
- [7] A. Mehes and K. Zeger, "Binary lattice vector quantization with linear block codes and affine index assignments," *IEEE Trans. Inform. Theory*, vol. 44, pp. 79–94, Jan. 1998.
- [8] Q. Zhao, P. C. Cosman, and L. B. Milstein, "On the optimal allocation of bandwidth for source coding, channel coding and spreading in a coherent DS-CDMA system employing an MMSE receiver," in *Proc. Eur. Wireless Conf.*, vol. 2, Feb. 2002, pp. 663–668.
- [9] Y. Yamaguchi and T. S. Huang, "Optimum binary fixed-length block codes," MIT Res. Lab. Electron., Cambridge, MA, Quarterly Progress Rep. 78, July 1965.
- [10] S. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, pp. 1746–1775, Oct. 1995.
- [11] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [12] J. R. Foerster and L. B. Milstein, "Coding for a coherent DS-CDMA system employing an MMSE receiver in a Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, pp. 1012–1021, June 2000.
- [13] Q. Zhao, "Bandwidth allocation and tradeoffs of source coding, channel coding, and spreading for CDMA systems," Ph.D. dissertation, Univ. California San Diego, La Jolla, CA, 2004.
- [14] P. Cosman, C. Tseng, R. Gray, R. Olshen, L. Moses, H. Davidson, C. Bergin, and E. Riskin, "Tree-structured vector quantization of CT chest scans: Image quality and diagnostic accuracy," *IEEE Trans. Medical Imaging*, vol. 12, pp. 727–739, Dec. 1993.
- [15] G. Caire and E. Viterbo, "Upper bound on the frame error probability of terminated trellis codes," *IEEE Commun. Lett.*, vol. 2, pp. 2–4, Jan. 1998.
- [16] R. J. McEliece, *The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977.



Qinghua Zhao was born in Shannxi, China, in 1976. She received the B.S. degree in electrical engineering from Xi'an JiaoTong University, Xi'an, China, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, San Diego (UCSD), in 2000 and 2004, respectively.

She is currently with Vativ Technologies Inc., San Diego, CA. Her research interests are in the area of communication theory, information theory, channel coding, and spread spectrum.



Pamela Cosman (S'88–M'93–SM'00) received the B.S. degree with Honor in electrical engineering from the California Institute of Technology, Pasadena, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1989 and 1993, respectively.

She was a National Science Foundation Postdoctoral Fellow at Stanford University and a Visiting Professor at the University of Minnesota, Minneapolis, during 1993–1995. In 1995, she joined the faculty of the Department of Electrical and Computer Engineering, University of California, San Diego, where she is currently a Professor. Her research interests are in the areas of image and video compression and processing.

Dr. Cosman is the recipient of the ECE Departmental Graduate Teaching Award (1996), a Career Award from the National Science Foundation (1996–1999), and a Powell Faculty Fellowship (1997–1998). She was an Associate Editor of the *IEEE COMMUNICATIONS LETTERS* (1998–2001), a Guest Editor of the June 2000 special issue of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, and was the Technical Program Chair of the 1998 Information Theory Workshop in San Diego. She is currently an Associate Editor of the *IEEE SIGNAL PROCESSING LETTERS*, and a Senior Editor of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*. She is a member of Tau Beta Pi and Sigma Xi.



Laurence B. Milstein (S'66–M'68–SM'77–F'85) received the B.E.E. degree from the City College of New York, New York, NY, in 1964, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1966 and 1968, respectively.

From 1968 to 1974, he was with the Space and Communications Group of Hughes Aircraft Company, and from 1974 to 1976, he was a member of the Department of Electrical and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Since

1976, he has been with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, where he is a Professor and former Department Chairman, working in the area of digital communication theory with special emphasis on spread-spectrum communication systems. He has also been a consultant to both government and industry in the areas of radar and communications.

Dr. Milstein was an Associate Editor for Communication Theory for the *IEEE TRANSACTIONS ON COMMUNICATIONS*, an Associate Editor for Book Reviews for the *IEEE TRANSACTIONS ON INFORMATION THEORY*, an Associate Technical Editor for the *IEEE Communications Magazine*, and the Editor-in-Chief of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*. He was the Vice President for Technical Affairs in 1990 and 1991 of the IEEE Communications Society, and has been a member of the Board of Governors of both the IEEE Communications Society and the IEEE Information Theory Society. He is a former Chair of the IEEE Fellows Selection Committee, and a former Chair of ComSoc's Strategic Planning Committee. He is a recipient of the 1998 Military Communications Conference Long Term Technical Achievement Award, an Academic Senate 1999 UCSD Distinguished Teaching Award, an IEEE Third Millennium Medal in 2000, the 2000 IEEE Communication Society Armstrong Technical Achievement Award, and the 2002 MILCOM Fred Ellersick Award.