Constellation Mappings for Two-Dimensional Signaling of Nonuniform Sources

Glen Takahara, *Member, IEEE*, Fady Alajaji, *Senior Member, IEEE*, Norman C. Beaulieu, *Fellow, IEEE*, and Hongyan Kuai

Abstract—The design of two-dimensional constellation mappings for the transmission of binary nonuniform memoryless sources over additive white Gaussian noise channels using standard M-ary PSK and QAM modulation schemes is investigated. The main application of this problem is the incorporation of an adaptive mapping assignment in modem devices that employ fixed PSK/QAM modulation schemes for the transmission of heterogenous data (such as multimedia information) containing various levels of nonuniformity. In general, the optimal mapping depends on both the probability distribution of the input signals and the signal-to-noise ratio (SNR) in the channel, in addition to the geometry of the signal constellation. We show that constellation mappings which follow the objective of minimizing the average symbol energy and, given this, maximizing the decoding probability of the most likely signals, can yield symbol-error-rate and bit-error-rate performance that is substantially better than Gray encoding maps. Gains as high as 3.5 dB in SNR E_b/N_0 are obtained for highly nonuniform sources. Finally, we note that the mappings techniques result in nonzero mean constellations and briefly consider their performances when they are converted to zero mean constellations by shifting. In this case, we observe that the shifted zero-mean Gray map outperforms our shifted maps for small- to medium-sized constellations ($M \leq 32$), but not for larger sizes.

Index Terms—Error rate, Gray map, joint source-channel coding, maximum *a posteriori* (MAP) and maximum-likelihood (ML) decoding, nonuniform sources, phase-shift keying (PSK) and quadrature amplitude modulation (QAM), signal mapping.

I. INTRODUCTION

F OR EQUALLY likely signals, Gray mapping in two-dimensional (2-D) signaling is widely known as optimal for minimizing bit-error rate (BER). However, many data sources (e.g., image or speech signals) are nonuniformly distributed; thus, they contain substantial amounts of natural redundancy. Furthermore, even when they are compressed, they still exhibit

a residual redundancy due to the suboptimality of the compression scheme. This redundancy, inherent or residual (after source coding), has two forms: redundancy due to the nonuniform distribution of the source symbols and redundancy due to source memory [1]. The redundancy due to the source nonuniformity can be statistically characterized by modeling the source (before or after compression) as an independent identically distributed (i.i.d.) nonuniform process. The total redundancy (due to both nonuniformity and memory) can be described using a Markov process model.

The redundancy of speech and image signals has been investigated in several recent works (e.g., [2]–[8]). In [2], the residual redundancy present in the line spectral parameters (LSPs) at the output of the U.S. Federal Standard 4.8 kb/s CELP vocoder was quantified using Markov processes. It was shown that 17.5% of the LSP bits are redundant due to their nonuniformity and that 24% of them are redundant due to their statistical dependency, resulting in a total redundancy of 41.5%. In [8], the residual redundancy reflected in discrete cosine transform (DCT) coded images was studied. In [3], the natural redundancy embedded in (uncompressed) bitonal images was examined; it was observed that many binary images (e.g., text and medical images) were up to 80% redundant due to nonuniformity. Similar results were obtained in [7].

In cases where the information is transmitted over a noisy channel using a standard signaling constellation, its redundancy can be appropriately exploited by using a maximum *a posteriori* (MAP) detector instead of a maximum-likelihood (ML) detector (e.g., [1], [9]–[13]), hence, reducing the error rate of the communication system. A simple example is shown in Fig. 1, in which the use of MAP decoding improves system performance over ML decoding for a highly nonuniform source and in which the gain obtained from nonuniformity in the source and the use of MAP decoding is illustrated. In this example, the BER is examined for a 16-quadrature amplitude modulation (QAM) modulated (with Gray mapping) nonuniform binary i.i.d. source $\{X_n\}$ (taking values in the set $\{0,1\}$) with $p = 0.5$ and $p = 0.9$ transmitted over an additive white Gaussian noise (AWGN) channel, where $p = \Pr\{X = 0\}$. The BER, which is obtained by simulating four million source bits, is plotted for both MAP decoding and ML decoding.

In addition to the gain from MAP decoding for nonuniform signals transmitted using a given 2-D signal constellation [e.g., QAM or phase-shift keying (PSK)], it is no longer clear that Gray mapping will provide the optimal BER performance. Indeed, we show that for nonuniform sources Gray mapping is not
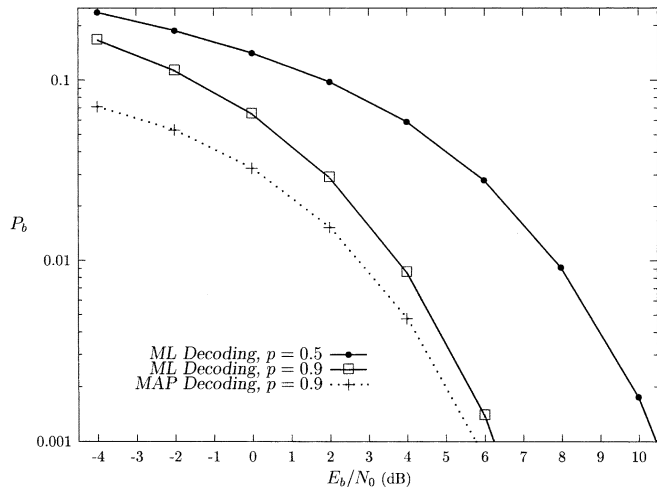
Fig. 1.   BER of a 16-QAM modulated nonuniform source with $p = 0.5, 0.9$.

necessarily optimal for minimizing BER or symbol error rate (SER).

In this paper, we propose criteria for the construction of good mappings from a set of signals derived from a nonuniform binary source $\{X_n\}$ to points of a given 2-D M-ary constellation. We illustrate this in the context of an uncoded communication system with QAM or PSK modulated nonuniform signals sent over an AWGN channel and decoded using MAP decoding. Motivated by the existing modem infrastructure where PSK or QAM modulation constellations are usually preset or fixed (for example, hardwired into widely used chipsets), we focus mainly on the mapping design without allowing the constellation to be altered. This scenario is particularly suitable for the communication of multimedia data whose information content is heterogenous (e.g., speech, images, and text) and exhibits various levels of nonuniformity. Hence, as different types of data arrive for transmission at a modem with a preset modulation scheme, the modem adapts to the new data by simply selecting (via a software operation) the appropriate mapping. Performance comparisons with Gray mapping schemes for different values of the constellation size M and signal-to-noise ratio (SNR) E_b/N_0 are provided. We also observe that, when using MAP decoding and fixed signal constellations for highly nonuniform signals, the BER performance of both 16-QAM and 64-QAM can be better than that of conventional 2-QAM, even though 16-QAM has four times higher spectral efficiency than that of 2-QAM, and 64-QAM has six times higher spectral efficiency than that of 2-QAM.

Finally, we remark that, by not allowing the modulation constellations to be altered (as they are assumed to be preset in the modem), our above mapping schemes result in nonzero mean constellations since the modulated signals are nonuniformly distributed. Therefore, the standard signal constellations are not minimum energy when used with nonuniform data. Minimum-energy constellations can be constructed from the given constellations and the symbol probabilities by shifting the centroid of the constellation to the origin of the signal space [14, pp. 247–248]. This may be undesirable in some systems where peak-to-average power ratios are of concern. Furthermore, redesigning such nonzero-mean PSK and QAM constellations

into zero-mean constellations is impractical for some systems, especially as a mechanism for adapting to varying levels of data redundancy. We study this issue in the sequel and provide some quantitative examples of the energy penalties incurred in exchange for the benefits of retaining a fixed standard 2-D constellation. We also show that, in some cases of higher order modulation formats, our mappings provide gains relative to the minimum-energy (zero-mean) Gray map constellation.

The remainder of this paper is organized as follows. The general problem formulation is introduced in Section II. In Section III, design criteria for 2-D constellation mappings for the transmission and MAP decoding of binary nonuniform memoryless sources over M-ary PSK and QAM modulated AWGN channels are presented. Performance results and discussions are given in Section IV. In Section V, zero-mean shifted constellations are considered. Finally, conclusions are stated in Section VI.

II. PROBLEM STATEMENT

We consider a source $\{X_n\}$ that generates independent binary symbols $\{0, 1\}$ according to a nonuniform probability $p \triangleq \Pr\{X_n = 0\} > 1/2$. It is desired to transmit the information to a user over an AWGN channel. We assume that an M-ary 2-D modulation scheme is to be used without the introduction of additional delay from source or channel coding and that we want to maximize the efficiency of the transmission in terms of achieving the maximum average data throughput per transmission at the smallest possible SER and/or BER. It is assumed that M is a power of two for the convenience of implementation and analysis. The binary source symbols are grouped into sequences having $\log_2 M$ bits. Thus, we form a new channel symbol sequence $\{Y_n\}$, where the independent channel symbols assume one of M distinct values $\{s_0, s_1, \cdots, s_{M-1}\}$ with probabilities $\{p_0, p_1, \cdots, p_{M-1}\}$. The channel symbols $\{Y_n\}$ are mapped onto signal points in an M-point signal constellation for transmission over the physical channel. Our objective is then to determine the optimal constellation/mapping pair that will yield the smallest SER and/or BER for a given number of constellation points M and a fixed E_b/N_0.

The special, restricted case of this problem for $M = 2$ [note that for $M = 2$, the problem is one-dimensional (1-D)] has been analytically considered in [15]. In this case, a closed-form expression for the BER is available [14], making the solution tractable. In particular, in the case of antipodal signaling [15] shows that the asymmetric constellation with the most likely signal (say s_0) mapped to the point $(\sqrt{E_s(1 - p)/p}, 0)$ and the least likely signal, s_1, mapped to the point $(-\sqrt{E_s p/(1 - p)}, 0)$, where $p = \Pr\{s_0\} > 1/2$ is optimal. This constellation/mapping pair gives as much as 4.5-dB gain in E_b/N_0 at a BER of 10^{-7} over conventional binary PSK (BPSK) when $p = 0.9$ [15]. Note that the above asymmetric constellation has a zero mean (since it is optimal), while the conventional BPSK constellation has nonzero mean. One can show that, when the conventional BPSK constellation is converted to a zero-mean constellation (by applying an appropriate translation without changing the distance between

the points), the gain of the asymmetric constellation over the shifted BPSK is decreased to around 0.06 dB.

Tractable closed-form expressions for the BER under MAP decoding for other values of $M(M > 2)$ are not available for general 2-D constellations. Intuitively, one expects that a strategy for assigning the least-energy signal point to the most likely channel symbol is desirable, and it is this principle that is followed in [15] for the 1-D case. The problem of finding the optimal constellation/mapping pair for $M > 2$ in two dimensions is likely infeasible (note that, in the special case of equally likely signals, Foscini *et al.* implement in [16] a gradient search method to determine a locally optimal constellation based on an asymptotic criterion function as the SNR approaches infinity). Moreover, constellations with large peak-to-average energy ratio or a high degree of asymmetry (which are likely properties of optimal or near-optimal constellations; see, for example, [16]–[18]) are undesirable in practical implementations. Also, in systems with dynamically changing source statistics (such as with heterogenous data), it is much simpler to switch between mappings of a *fixed* constellation than to switch between different constellations. For these reasons, and also because of their widespread use in practice (as discussed in Section I), our focus here is on considering fixed PSK and QAM constellations; that is, we consider the problem of determining optimal or near-optimal mappings for fixed PSK and QAM constellations.

III. CONSTELLATION MAPPINGS FOR MAP DECODING

In this section, we consider the construction of mappings from a set of M nonuniformly distributed symbols to the points of a fixed 2-D constellation with the objective of minimizing SER/BER. For a given M-ary constellation, there are $M!$ possible mappings from the source symbol set to the constellation points, and thus, despite the fact that in many cases many of the mappings are equivalent to one another in terms of SER/BER performance (in the case of a binary source), this problem has exponential search complexity.

A further complication is the following: tractable closed-form analytical expressions for the SER or BER under MAP decoding are not known for a general given constellation and a given mapping. Moreover, it is not feasible to search through, for example, thousands of mappings via simulation. To partially address this problem, we employ new upper and lower bounds recently derived in [19] to assess the BER/SER performance for any given mapping. The bounds are very accurate and computationally efficient. They are useful for performing exhaustive searches for small values of M (i.e., $M \leq 16$ for SER and $M \leq 8$ for BER) and for conducting limited searches for moderate values of M (i.e., $M = 32$ for SER and $M = 16$ for BER). In addition, we propose criteria for constructing a heuristic map that performs well compared to optimal maps when known (obtained using our bounds) and locally optimal maps obtained via a search algorithm. For large values of M (i.e., $M = 64$), we compare the performance of the heuristic map to that of a Gray map (optimized for a given p).

In the case of equally likely signals, the average energy per symbol is the same for all mappings. However, for signals that

are not equally likely this is not true, and it is intuitively desirable to have mappings which minimize the average energy per symbol. Also, intuition suggests that one should minimize the probability of error when the most likely signals are sent. We propose the following criteria, listed in order of priority, as guidelines for constructing a mapping with the objective of jointly minimizing SER [criteria 1) and 2)] and BER [criterion 3)]:

1) minimize the average energy per symbol for the M given symbol probabilities;
2) successively minimize the conditional symbol decoding error probabilities (conditioned on the transmitted symbol), proceeding from the most likely to the least likely symbol;
3) when several mappings all satisfy criteria 1) and 2) equally, choose from these mappings any mapping which is a Gray map or, if not possible, choose a mapping that is closest to a Gray map in the sense of maximizing the number of adjacent pairs of signals that differ in only one bit.

The following condition determines a mapping which satisfies criterion 1) up to permutations within sets of symbols with the same average energy. Given M signals with energies $E_1 \leq \ldots \leq E_M$ and probabilities $\{p_i\}_{i=1}^{M}$, and given any permutation π of the energy indices $\{1, 2, \ldots, M\}$, the sum $\sum_{i=1}^{M} E_i p_{\pi(i)}$ is minimized if and only if the permutation satisfies $p_{\pi(1)} \geq \ldots \geq p_{\pi(M)}$. This condition is valid because if violated (e.g. $p_{\pi(i)} > p_{\pi(j)}$ with $i < j$) then the average energy can be reduced by interchanging the violating mappings (interchange the mappings of channel signals s_i and s_j).

Subject to criterion 1), we next consider criterion 2). Let $s_1, \ldots, s_M$ denote the signals listed from most likely to least likely. The SER is given by

$$\text{SER} = \sum_{i=1}^{M} \Pr\{\text{Symbol Error} \mid s_i \text{ sent}\}\Pr\{s_i\}.$$

A reasonable approach to minimizing the SER is the greedy one [criterion 2)], which consists of first minimizing $\Pr\{\text{Symbol Error} \mid s_1 \text{ sent}\}$ subject to the average energy per symbol being minimized, then minimizing $\Pr\{\text{Symbol Error} \mid s_2 \text{ sent}\}$ subject to the average energy per symbol and the first conditional error probability being minimized, etc. At any given stage, after selecting a mapping to minimize the average energy per symbol, we are faced with a set of permutations of that mapping which is more restricted than at the previous stage. From this new restricted set of permutations, we choose a mapping to minimize the conditional error probability of the current stage.

However, there is a practical limitation to the above procedure: finding the set of maps which minimizes a given conditional error probability [as suggested in criterion 2)] is only slightly less complex than finding the overall optimal map for any moderately sized M. Indeed, this may require considering mapping choices over very large sets.

Therefore, we propose the following two simple heuristics in accordance with criteria 1)-3) for constellations with several

energy levels (such as QAM with $M > 4$) and constellations with a single energy level (such as PSK), respectively.

1) QAM Heuristic Mapping: For each energy level, starting with the lowest and in increasing order of energy levels, we choose a set of symbols to fill up that energy level from the set of unallocated symbols to satisfy criterion 1') below and then allocate the symbols in this chosen set to the signal points of that energy level to satisfy criterion 2') below.

> 1') The symbols in this set are the most likely symbols among the unallocated symbols (note that this set is not unique in general).
>
> 2') The least likely symbols of the current set are the nearest neighbors of the most likely symbols of the set allocated for the previous energy level.

When there is more than one set/allocation of symbols that satisfies criteria 1') and 2'), use a Gray mapping procedure [as discussed in 3) above] to identify a single set and allocation.

2) PSK Heuristic Mapping: For constellations with a single energy level (such as PSK), we propose the following simpler heuristic.

1) Place the (single) most likely symbol arbitrarily.
2) Place the next $m = \log_2 M$ most likely symbols such that these first $m + 1$ most likely symbols are as uniformly spread as possible, with the most likely symbol being the most isolated (in Euclidean distance sense) when possible.
3) Place the two least likely symbols next to the most likely symbol and the next $2m$ least likely symbols as nearest neighbors to the m symbols placed in step 2), adhering to the Gray mapping principle whenever possible.
4) Repeat steps 2) and 3) until all the remaining symbols are placed.

3) Search Algorithm: Criteria 1')–2') [or 1)–4)] can be implemented in the following algorithm to search for locally optimal mappings that have good SER and BER performance.

- Set $n = 1$, $E_b/N_0 = $ lowest SNR of interest, and $\Delta = $ increment.
- Use criteria 1')–2') [or 1)–4)] to determine a mapping M_1.
- While $E_b/N_0 \leq$ highest SNR of interest, do the following:
— Input mapping M_n as the initial point in a steepest descent search algorithm, which searches locally at each iteration for the mapping which minimizes the objective function (SER or BER) by interchanging the assignment of a pair of symbols. Let M_{n+1} denote the final (locally optimal) mapping obtained by this search.
— Let $n = n + 1$ and $E_b/N_0 = E_b/N_0 + \Delta$.

Note that the mapping provided by the above search algorithm is, in general, a function of E_b/N_0.

IV. NUMERICAL RESULTS AND DISCUSSION

We consider a nonuniform binary source for transmission over an AWGN channel with 4-, 8-, 16-, and 32-PSK modulations as well as 16- and 64-QAM modulations and MAP decoding. Some SER and BER performance results are presented in Figs. 2–7. For $M \leq 16$, the upper and lower bounds of [19]
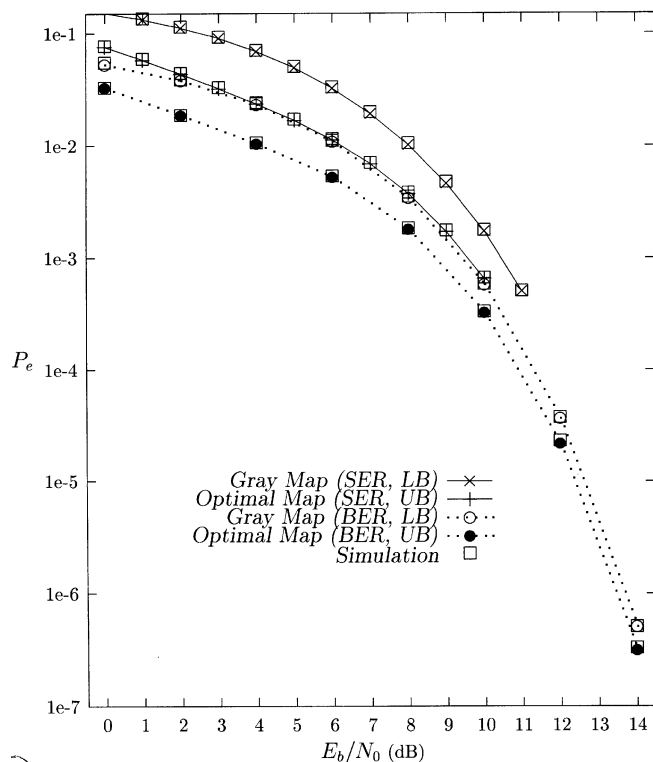

Fig. 2. BER and SER of 8-PSK $p = 0.9$ using the optimal and Gray maps. Gray (optimal) map curves were obtained using the lower (upper) bound in [19]. Simulated values are also shown.
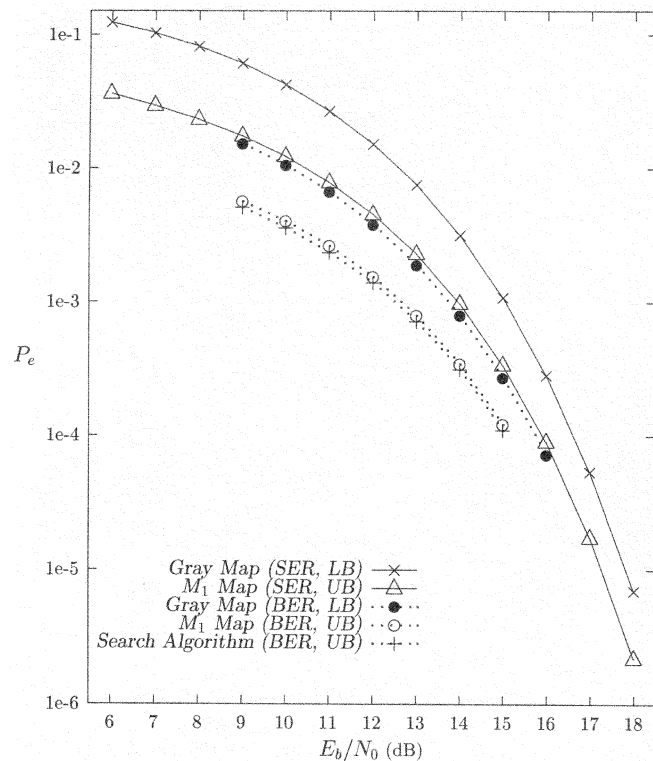

Fig. 3. BER and SER of 16-PSK $p = 0.9$ using the Gray map, the M_1 map, and the search algorithm. Gray (M_1 and search) map curves were obtained using the lower (upper) bound in [19].

are used to obtain the SER and BER results shown in Figs. 2 and 3 and Figs. 5 and 6, respectively. These bounds are tight
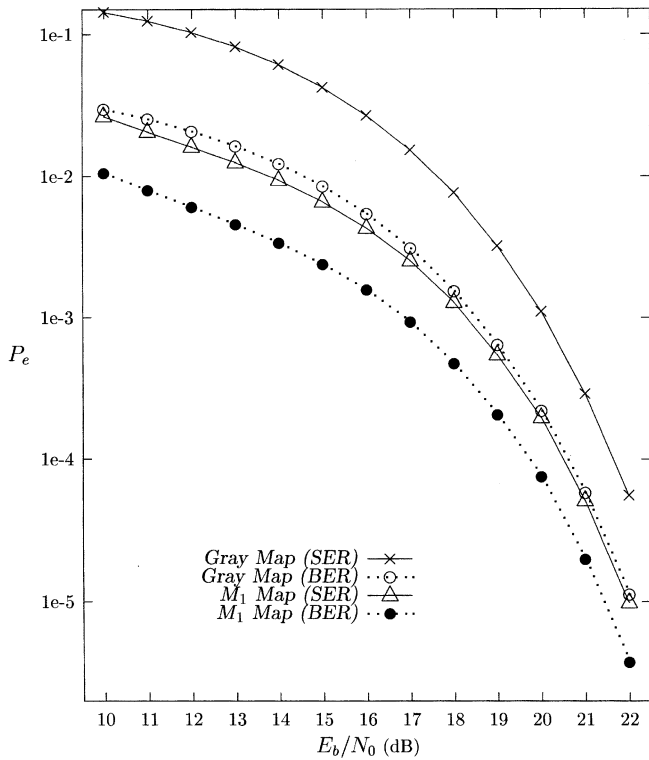
Fig. 4.   BER and SER of 32-PSK at $p = 0.9$ using the Gray and M_1 maps. All curves were obtained by simulation using $25\,000\,000$ symbols.

Fig. 5.   SER of 16-QAM at $p = 0.9$ using the Gray map, the M_1 map, and the search algorithm, and the SER of 16-QAM at $p = 0.1$ using the Gray map. Gray (M_1 and search) map curves were obtained using the lower (upper) bound in [19].

and very easy to evaluate for all values of the source distribution p. In these cases, the upper bound and corresponding lower bound are graphically coincident or nearly graphically coincident and thus provide essentially exact error rates for practical applications; so, for the sake of simplicity, we only show the upper bound for our proposed maps and the lower bound for the Gray map. For $M = 8$ in Fig. 2, we also include simulation results to illustrate the accuracy of the bounds. Moreover, since the bounds are numerically efficient, we are able to provide performance results from exhaustive (globally optimal) and locally optimal mapping searches, in addition to the performance of the heuristic map described in Section III as well as the Gray map. For $M = 32$ and 64, we do not perform a mapping search. Rather, the performance of the heuristic map of Section III is evaluated relative to Gray mapping using simulation in Figs. 4 and 7.

In all the examples considered in this section, we assume that the source distribution p is 0.9 unless otherwise indicated. We also denote the mapping obtained from criteria 1')–2') for QAM or criteria 1)–4) for PSK presented in Section III as mapping M_1.

A. 4-QAM/4-PSK

For 4-QAM/4-PSK, the signals are equidistant from the origin so the average energy per symbol is the same for any mapping. It is not difficult to see that there are only two mappings for a 4-QAM constellation that lead to distinct SER and BER values. These two mappings, starting in the first quadrant and proceeding clockwise, are (a) $\{00, 10, 11, 01\}$ and (b)
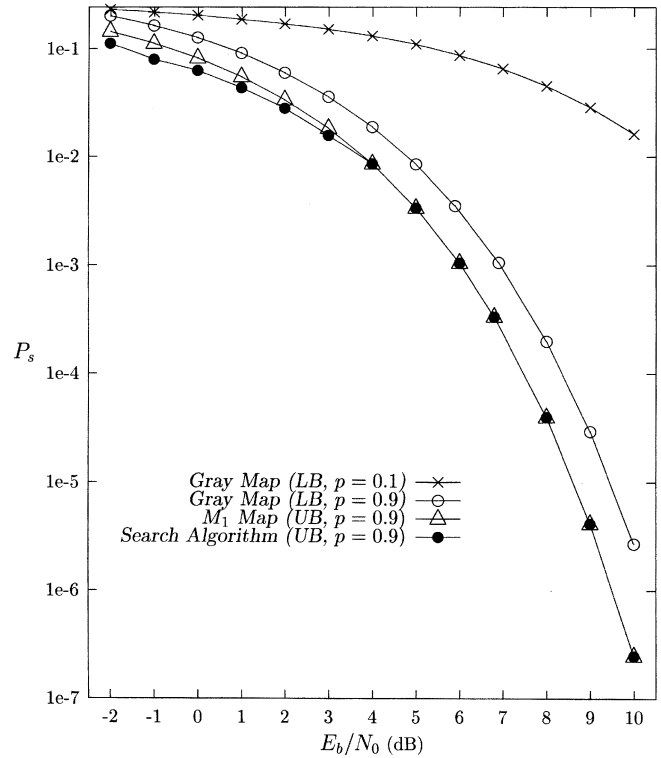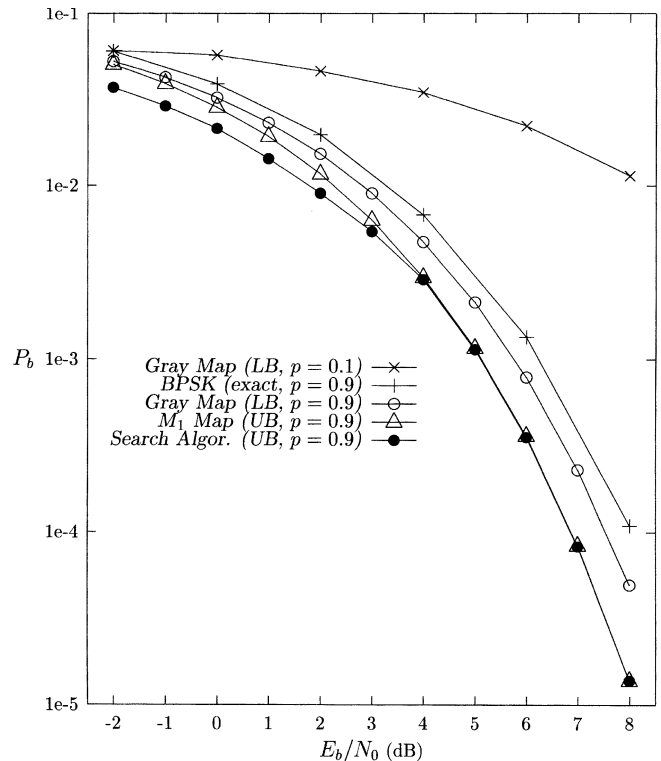

Fig. 6.   BER of 16-QAM at $p = 0.9$ using the Gray map, the M_1 map, and the search algorithm versus the BER of BPSK (2-QAM), and the BER of 16-QAM at $p = 0.1$ using the Gray map. Gray (M_1 and search) map curves were obtained using the lower (upper) bound in [19].
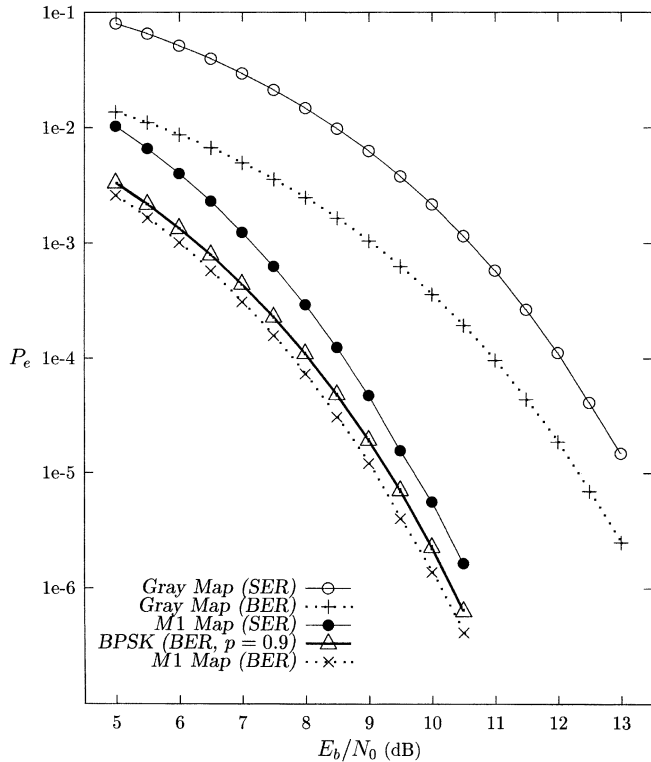
Fig. 7. BER and SER of 64-QAM at $p = 0.9$ using the Gray and M_1 maps obtained by simulation using 25 000 000 symbols, and the exact BER of BPSK at $p = 0.9$.

Fig. 8. Mappings M_1 and Gray (in parentheses) for 16-QAM.

$\{00, 11, 10, 01\}$. Note that mapping (a) is a Gray map. All $4! = 24$ possible mappings for 4-QAM can be obtained from these two mappings by rotations of 90, 180, or 270 degrees, and reflections about the x axis, the y axis, the line $y = x$, and the line $y = -x$. A direct comparison between the two mappings reveals that mapping (b) is optimal for SER at all values of E_b/N_0, while mappings (a) and (b) yield essentially the same BER performance. Note also that mapping (b) coincides with the heuristic M_1 map. For a given SER, the gain of mapping (b) over mapping (a) is on the order of 0.1 dB in E_b/N_0 for SER less than 10^{-3}. Thus, there appears to be no significant loss with the choice of Gray mapping for 4-QAM.

B. 8-, 16-, and 32-PSK

In Fig. 2, we compare the SER and BER performances of a Gray mapped 8-PSK constellation and an optimally mapped 8-PSK constellation. The optimal mapping in this example is determined using an exhaustive search. The optimal SER mapping and the optimal BER mapping are the same. Specifically, proceeding counter-clockwise from the point $(1, 0)$ in the first quadrant of an 8-PSK constellation (illustrated, for example, in [20, p. 225]), they are given by the following map: $\{000\,111\,001\,011\,010\,110\,100\,101\}$. As for the 4-QAM case, the optimal SER/BER mapping is independent of the value of E_b/N_0. Moreover, it is again the case that the optimal mapping can be obtained via the heuristic M_1 map. At an SER and BER equal to 10^{-3}, the gain due to the optimal mapping over the Gray mapping is roughly 1 and 0.5 dB in E_b/N_0, respectively.

For 16-PSK, an exhaustive search is numerically tedious, requiring several days. In this case, we employ the bounds in [19] to evaluate the M_1 and Gray maps as well as, for BER, the search algorithm introduced at the end of Section III. Performance SER and BER results are presented in Fig. 3. At an SER and BER equal to 10^{-3}, the gain due to the M_1 mapping over the Gray mapping is roughly 1 dB in E_b/N_0 for both SER and BER. Observe also that the performance of the M_1 mapping is quite close to that of the locally optimal map obtained by search algorithm.

For 32-PSK, we present simulation SER and BER results for the M_1 and Gray mappings in Fig. 4. At SER and BER equal to 10^{-3}, the gain due to the M_1 mapping over the Gray mapping is more than 1.5 dB in E_b/N_0 for both SER and BER. We can clearly remark that the gain of the M_1/search algorithm maps over the Gray map increases as the number of signals M increases (compare the gains for 4-PSK versus 8-PSK, 16-PSK, and 32-PSK). This can be explained by the fact that as M increases, the likelihood of the most likely symbols substantially increases with respect to the likelihood of the least likely symbols. Hence, as M increases, the error probabilities (under MAP decoding) of the most likely symbols play a more dominant role in the SER/BER expressions than the error probabilities of the least likely symbols. Hence, the M_1 and search algorithm maps which attempt to minimize the error probabilities of the most likely symbols will increasingly outperform the Gray map more prominently as M increases.

C. 16-QAM and 64-QAM

Fig. 8 shows a 16-QAM constellation with a mapping M_1. For $p > 0.5$, the mapping M_1 minimizes the average symbol energy [criterion 1)] and, subject to this, for any E_b/N_0, the mapping M_1 also maximizes the conditional probability that symbol 0000 (the most likely symbol) is decoded, given that 0000 is sent. This is due to the fact that symbol 0000 has the least

111111 110111 110011 110101 | 010101 011101 011111 111110
(110000) (111000) (011000) (010000) | (010010) (011010) (111010) (110010)

010111 010110 011001 100001 | 000101 001011 011010 111100
(110100) (111100) (011100) (010100) | (010110) (011110) (111110) (110110)

011011 010011 010010 010001 | 000011 001001 001110 011110
(100100) (101100) (001100) (000100) | (000110) (001110) (101110) (100110)

111001 011000 010000 000001 | 000000 001010 001100 011100
(100000) (101000) (001000) (000000) | (000010) (001010) (101010) (100010)

110001 110000 010100 000100 | 000010 001000 001101 111000
(100001) (101001) (001001) (000001) | (000011) (001011) (101011) (100011)

110110 110100 100100 100000 | 100010 101000 101010 101110
(100101) (101101) (001101) (000101) | (000111) (001111) (101111) (100111)

101111 101100 100101 100110 | 000110 100011 101001 111010
(110101) (111101) (011101) (010101) | (010111) (011111) (111111) (110111)

111101 101101 100111 110010 | 000111 001111 101011 111011
(110001) (111001) (011001) (010001) | (010011) (011011) (111011) (110011)

Fig. 9.   Mappings M_1 (in bold) and Gray (in parentheses) for 64-QAM.

likely neighbors, subject to criterion 1); thus, the decision region for 0000 is maximized. The remaining symbols are placed in the constellation to successively maximize the decoding regions of 0001, 0100, and 0010, in that order. We note that the principles of Gray encoding have been abandoned in our construction of mapping M_1.

The M_1 map was used as the initial map in our search algorithm described in Section III for a binary source. The results, shown in Figs. 5 and 6, show an improvement over the Gray map of roughly 1 and 0.75 dB in E_b/N_0 (at error rates between 10^{-5} and 10^{-3}) for SER and BER, respectively. For these error rates, the performance of our heuristic map M_1 is very close to that of the map obtained from the search algorithm. We note that these plots illustrate the fact that the optimal map varies with p; using a mapping that is highly mismatched to the source distribution can result in an extremely poor performance (e.g., compare the curves of the Gray map when $p = 0.1$ and $p = 0.9$).

We also observe that 16-QAM with the mapping M_1 achieves around a 1-dB gain over conventional 2-QAM (or BPSK) for $p = 0.9$ and the same BER. This leads us to the interesting observation that, while the traditional wisdom for equally likely signals is that there is a tradeoff between spectral efficiency and BER, with nonuniform signals and fixed constellations there need not be such a tradeoff. Indeed, in this example 16-QAM achieves both four times the spectral efficiency and better BER performance[1] than 2-QAM when $p = 0.9$ and the constellations are fixed.

In Fig. 7, simulation SER and BER results for the M_1 and Gray mappings described in Fig. 9 are shown. We remark from the figures that the M_1 mapping achieves substantial gains over the Gray mapping. At error rates between 10^{-5} and 10^{-3}, improvements of up to 3.5 dB for SER and 3.0 dB for BER are

---

[1]It is important to note, however, that this only holds since both constellations have nonzero mean as they are not allowed to be altered. When appropriately shifted to have zero mean, the shifted 2-QAM constellation performs better than the shifted 16-QAM constellation.

---

obtained. We also note, as in M-PSK signaling, that the gain of the M_1/search algorithm maps over the Gray map increases as M increases (compare the gains for 4-QAM versus 16-QAM and 64-QAM). Furthermore, the mapping gain with increasing M for M-QAM constellations is even larger than for M-PSK. This is due to the fact that, for QAM, the distribution of low-energy constellation points to the most likely symbols provides an additional advantage for the M_1 and search algorithm maps over the Gray map (while for PSK all constellation points have identical energy). Finally, we also observe from Fig. 7 that, as in the case of 16-QAM, the 64-QAM M_1 map outperforms conventional 2-QAM in terms of BER, while achieving six times the spectral efficiency.

V. Comparisons With Zero-Mean Constellations

In this paper, we are investigating the performances of conventional 2-D signaling constellations when used with nonuniform data. It is well known that a minimum-energy constellation is derived from any given constellation by subtracting the centroid of the constellation from each of the signaling points so that the centroid is translated to be at the origin of the signal space [14]. This translation preserves the error rate performance of the original constellation while minimizing the energy required. The PSK and QAM constellations considered here are zero mean when used with uniform data and, hence, are not zero mean when used with nonuniform data. Since it is useful to a system designer to know how much benefit is available, and at what cost by using more complex designs, we determine the improvement that can be achieved by using the minimum-energy constellations relative to the fixed constellations. This information is valuable for determining the cost-performance tradeoff in using, for example, a widely available chipset with conventional modulation constellations as opposed to building a zero mean constellation modulator and demodulator from discrete components. A summary of the performances of the mapped/fixed constellations relative to the minimum energy Gray map constellations is as follows.

- **8-PSK:** When converted to a zero-mean constellation, the average energy of the Gray map constellation is reduced by 4.05 dB; this translates into shifting the BER/SER curves in Fig. 2 by 4.05 dB to the left. The optimal map fixed constellation gains 1 dB at SER $= 10^{-3}$ and 0.5 dB at BER $= 10^{-3}$ over the Gray map fixed constellation. Thus, 3–3.5 dB in SNR is paid in this case for using a fixed constellation.
- **16-PSK:** For zero-mean (shifted) 16-PSK, the energy of the Gray map constellation is reduced by 3.65 dB. Thus, our M_1 mapping scheme recovers about 1 dB in SNR, reducing the fixed constellation loss from 3.65 to about 2.65 dB.
- **32-PSK:** In this case, the energy reduction achieved by the minimum-energy (zero mean) Gray constellation is 3.43 dB. The M_1 map recovers 1.5 dB at BER $= 10^{-3}$, reducing the fixed constellation loss to about 1.9 dB.

These examples for 8-, 16-, and 32-PSK seem to indicate that the fixed constellation loss decreases as the order (number of

signals M) of the modulation increases, and furthermore, that our proposed mappings recover an increasing proportion of the loss as the modulation order increases. Examining QAM constellations, we find the following.

- **16-QAM:** The SNR reduction for the minimum-energy Gray constellation is 3.11 dB. At SER $= 10^{-3}$ in Fig. 5 and BER $= 10^{-3}$ in Fig. 6, the M_1 map gains 1 and 0.75 dB, respectively. Thus, the loss is reduced to 2.11 and 2.36 dB, respectively.
- **64-QAM:** The minimum-energy translation reduces the SNR required by the Gray map by 1.65 dB. Thus, the M_1 map with fixed constellation gains 1.85 dB for SER and 1.35 dB for BER in Fig. 7 over the minimum-energy Gray map constellation. This is a very interesting result, indicating that our proposed mappings used with a fixed constellation can in some cases outperform the minimum-energy Gray map constellation, without the detriments of shifting the mean of the signal constellation to the signal space origin. We observe that the fixed constellation loss decreases as the modulation order increases for QAM, as was also the case for PSK.

An interesting and potentially useful question is how our mappings (optimal or M_1) operating on minimum-energy (zero mean) constellations perform relative to the minimum-energy Gray map constellation. We have investigated this issue and found that, for all the considered cases except for 64-QAM, the Gray map outperforms the designed maps when both constellations are converted to zero mean. However, as the number of signal points M increases, the loss of the M_1 map with respect to the Gray map is reduced, and for 64-QAM the M_1 map minimum-energy constellation has a gain over the Gray map minimum-energy constellation of 2.07 dB for SER and 1.57 dB for BER.

VI. CONCLUSION

In this paper, we proposed and implemented design criteria for 2-D constellation mappings for the transmission and MAP decoding of binary nonuniform memoryless sources over M-ary PSK and QAM modulated AWGN channels. We showed that, with nonuniform signals, constellation mappings which follow the objective of minimizing the average symbol energy and, given this, maximizing the decoding probability of the most likely signals, yield SER and BER performance that is significantly better than Gray encoding maps when the constellations are fixed. We also found that, with an appropriate mapping, 16-QAM and 64-QAM can perform better than conventional 2-QAM, in terms of both spectral efficiency and BER for highly nonuniform sources, although a nonconventional zero-mean minimum-energy 2-QAM (obtained by a mean shift) will perform better. Finally, we briefly considered the performance of our mappings when applied to the (shifted) zero-mean minimum-energy corresponding constellations. While our examples indicate that the minimum-energy Gray map constellations outperform the minimum-energy mapped constellations for binary modulations and low-to-medium-order

PSK and QAM, an example shows that the converse can be true for higher order QAM.

REFERENCES

[1] F. Alajaji, N. Phamdo, N. Farvardin, and T. Fuja, "Detection of binary Markov sources over channels with additive Markov noise," *IEEE Trans. Inform. Theory*, vol. 42, pp. 230–239, Jan. 1996.
[2] F. Alajaji, N. Phamdo, and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 325–336, Sept. 1996.
[3] P. Burlina, F. Alajaji, and R. Chellappa, "Transmission of two-tone images over noisy communication channels with memory," Center for Automation Research, Univ. of Maryland, College Park, Tech. Rep. CAR-TR-814, 1996.
[4] P. Burlina and F. Alajaji, "An error resilient scheme for image transmission over noisy channels with memory," *IEEE Trans. Image Processing*, vol. 7, pp. 593–600, Apr. 1998.
[5] T. Fazel and T. Fuja, "Joint source-channel decoding of block-encoded compressed speech," in *Proc. Conf. Inform. Sciences and Systems*, Princeton, NJ, Mar. 2000, pp. (FA5)1–(FA5)6.
[6] J. Hagenauer, "Joint source and channel coding for broadcast applications," in *Audio and Video Digital Radio Broadcasting Systems and Techniques*. Amsterdam, The Netherlands: Elsevier, 1994.
[7] J. Kroll and N. Phamdo, "Source-channel optimized trellis codes for bitonal image transmission over AWGN channels," *IEEE Trans. Image Processing*, vol. 8, pp. 899–912, July 1999.
[8] W. Xu, J. Hagenauer, and J. Hollmann, "Joint source-channel decoding using the residual redundancy in compressed images," in *Proc. Int. Conf. Communications*, Dallas, TX, June 1996, pp. 142–148.
[9] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sept. 1995.
[10] J. Kroll and N. Phamdo, "Analysis and design of trellis codes optimized for a binary symmetric Markov source with MAP detection," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2977–2987, Nov. 1998.
[11] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. Commun.*, vol. 39, pp. 838–846, June 1991.
[12] B. S. Srinivas, R. E. Ladner, M. Azizoglu, and E. A. Riskin, "Progressive transmission of images using MAP detection over channels with memory," *IEEE Trans. Image Processing*, vol. 8, pp. 462–475, Apr. 1999.
[13] R. E. Van Dyck and D. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding," *Proc. IEEE*, vol. 87, pp. 1734–1750, Oct. 1999.
[14] J. Wozencraft and I. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
[15] I. Korn and J. P. Fonseka, "Optimal receiver for binary signals with nonequal probabilities," in *Proc. Int. Conf. Telecommunications*, Acapulco, Mexico, May 2000.
[16] G. J. Foscini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Trans. Commun.*, vol. COM-22, pp. 28–37, Jan. 1974.
[17] X. Dong, N. C. Beaulieu, and P. H. Wittke, "Signaling constellations for fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 703–714, May 1999.
[18] G. D. Forney, R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureschi, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas. Commun.*, vol. SAC-2, pp. 632–646, Sept. 1984.
[19] H. Kuai, F. Alajaji, and G. Takahara, "Tight error bounds for nonuniform signaling over AWGN channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2712–2718, Nov. 2000.
[20] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.

Glen Takahara (M'96) was born in Hope, BC, Canada, in 1962. He received the B.A. degree (with honors) in mathematics from the University of British Columbia, Vancouver, BC, Canada, in 1988 and the M.S. and Ph.D. degrees in statistics from Carnegie Mellon University, Pittsburgh, PA, in 1990 and 1994, respectively.

In 1993, he joined the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, where he is currently an Associate Professor of Statistics/Mathematics and Engineering. His current research interests include data networking, queueing theory, communication theory, computational statistics, and orientation modeling.

Fady Alajaji (S'90–M'94–SM'00) was born in Beirut, Lebanon, on May 1, 1966. He received the B.E. degree (with Distinction) from the American University of Beirut, Lebanon, and the MS. and Ph.D. degrees from the University of Maryland, College Park, all in electrical engineering, in 1988, 1990, and 1994, respectively.

He held a post-doctoral appointment in 1994 at the Institute for Systems Research, University of Maryland. In 1995, he joined the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, where he is currently an Associate Professor of Mathematics and Engineering. Since 1997, he has also been cross-appointed in the Department of Electrical and Computer Engineering at the same university. He held a sabbatical visit to the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, in the fall of 2001. In 1999, he co-chaired and organized the Sixth Canadian Workshop on Information Theory in Kingston, ON. He also served as chair of the Selection Committee for the 2000 Canadian Award in Telecommunications. His research interests include information theory, digital communications, error control coding, joint source-channel coding, and data compression.
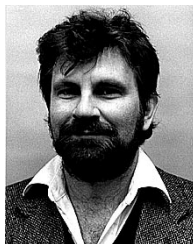
Norman C. Beaulieu (S'82–M'86–SM'89–F'99) received the B.A.Sc. (honors), M.A.Sc., and Ph.D. degrees in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1980, 1983, and 1986, respectively.

He was a Queen's National Scholar Assistant Professor with the Department of Electrical Engineering, Queen's University, Kingston, ON, Canada, from September 1986 to June 1988, Associate Professor from July 1988 to June 1993, and Professor from July 1993 to August 2000. In September 2000, he became the iCORE Research Chair in Broadband Wireless Communications at the University of Alberta, Edmonton, AB, Canada, and, in January 2001, the Canada Research Chair in Broadband Wireless Communications. His current research interests include broadband digital communications systems, fading channel modeling and simulation, pulse shaping, interference prediction and cancellation, and decision-feedback equalization.

Dr. Beaulieu was elected Fellow of The Royal Society of Canada in 2002. He is a Member of the IEEE Communication Theory Committee and served as its Representative to the Technical Program Committee of the 1991 International Conference on Communications, and as Co-Representative to the Technical Program Committee of the 1993 International Conference on Communications and the 1996 International Conference on Communications. He was General Chair of the Sixth Communication Theory MiniConference in association with GLOBECOM'97 and Co-Chair of the Canadian Workshop on Information Theory, 1999. He has been Editor of Wireless Communication Theory of the IEEE TRANSACTIONS ON COMMUNICATIONS since January 1992 and Editor-in-Chief since January 2000, Associate Editor for Wireless Communication Theory of the IEEE COMMUNICATIONS LETTERS since November 1996, and on the Editorial Board of the PROCEEDINGS OF THE IEEE since November 2000. He received the Natural Science and Engineering Research Council (NSERC) E.W.R. Steacie Memorial Fellowship in 1999. He was awarded the University of British Columbia Special University Prize in Applied Science in 1980.

Hongyan Kuai was born in Hubei, China, in March 1972. She received the B.Sc. degree from the University of Science and Technology of China in 1995 and the M.Sc. degree in mathematics from Queen's University, Kingston, ON, Canada, in 1999.

From 1997 to 1999, she held a graduate award from Queen's University, and in 1999 she was the recipient of a graduate fellowship from the Natural Sciences and Engineering Research Council of Canada. Her research interests include information theory, digital communications, coding, and network communications.