TCM SCHEMES WITH PARTIALLY OVERLAPPED SIGNAL CONSTELLATION

Ling Kang

Department of Electrical Engineering

McGill University

Montéal, Québec

A thesis submitted to the Faculty

of Graduate Studies and Research in partial fulfillment

of the requirement for the degree of Master of Engineering

March 1991

ABSTRACT

This thesis describes the evolution of Trellis Coded Modulation(TCM) technique and then developes the Analytically Described Trellis Codes with Partially Overlapped Signal Constellation for AWGN channels. This development is shown to obtain a performance gain over the TCM with conventional signal constellation. The new signal constellation is first introduced. A few examples (for one-, two-, and high-dimensional signal sets) are then given to show the performance gain due to the overlap of signal constellation. Bit error probability analysis is carried out for these examples.

SOMMAIRE

Cette thèse décrit d'abord l'évolution de la technique de Modulation de Codage en Treillis et puis développe les Codages en Treillis Décrits Analytiquement en utilisant une constellation de signaux partiellement superposés pour des canaux sous l'influence de bruit gaussien uniforme. Ce développement est démontré d'obtenir des gains en performance par rapport à la Modulation de Codage en Treillis avec constellation conventionnelle des signaux. Premièrement, la nouvelle constellation des signaux sera introduite. En suite, nous présenterons quelque exemples (pour des sets de signaux en un-, deux-, et haute-dimension) démontrant l'accroissement en performance dû à la superposition de constellation des signaux. L'analyse de probabilité d'erreur par unité transmise("bit") sera exécutée sur les exemples présentés.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. M.R. Soleymani, for his invaluable and instructive guidance and generally doing more than he had to throughout this research work.

I would like to thank Salvatore Torrente for the French translation of the abstract and helpful suggestions regarding the computer system. Also, I wish to acknowledge the help of Godfrey Cheng for the English correction.

This work was supported by Canada National Science and Engineering Research Council Grant OG PIN 011.

Finally, special thanks to my husband, Dr. Hua Qin Liu, for his encouragement and unfailing support.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Ungerboeck has developed Trellis Coded Modulation (TCM) as a combined coding and modulation technique for digital band-limited channels The main attraction of TCM is that it allows the achievement of significant coding gains over conventional uncoded multilevel modulation without compromising bandwidth efficiency

In the past, coding and modulation were treated as separate operations with regard to overall system design. In particular, most earlier works on coded digital communication systems independently optimized: 1) conventional (block or convolutional) coding with maximized minimum Hamming distance, and 2) conventional modulation with maximally separated signals, in an Additive White Gaussian Noise (AWGN) environment.

Massey [1] was the first person to show that considerable performance improvement could be obtained by treating coding and modulation as a single entity Following this line of thinking, the first TCM schemes were proposed in 1976 [2] by Ungerboeck and Csajka. Then a more detailed publication [3] in 1982 was shown by 1987 Ungerboeck described further in detail the features of TCM and design method of Trellis Coded Modulation schemes [4, 5].

Ungerboeck's TCM seems to cover the range of possible coding gains (in the 3-6 dB range) with complexity of the order of what we might expect Can they be

further improved? It seems unlikely that further major improvements are possible. But, the upper bound on the gain of trellis code was obtained in [6]. So they still can be improved. Also within the spectrum of performance of already known schemes there will likely be some further embellishments, that will reduce implementation complexity or have other desirable properties, even in the application of TCM in the channels other than the AWGN. Based on the Ungerboeck's TCM many such works have been done. Is this enough? The answer is: no. We still have a lot of works to do. Since 1982 an explosion of research and actual implementations of TCM has taken place [7-11]. Recently, Divsalar has introduced Trellis Coded Modulation into the fading mobile satellite channels [12-14], then expanded the applications of TCM.

In this thesis, we will show that using constellations with partially overlapped signal points, i. e., constellations with less than 2^{n+1}-points, one can achieve a considerable coding gain in comparison to the traditional 2^{n+1}-point constellation. The increase in coding gain is due to a reduction in transmitted signal energy as a result of deleting some high power signals from the original 2^{n+1}-point constellation and replacing them with some of the low power signals.

In chapter 2, the principles and general structure of trellis-coded modulation techniques will be outlined. The main significance of TCM is the mapping method called "mapping by set partitioning". Design criteria for TCM codes will be described first. Then several important works which improve or modify the TCM technique will be briefly reviewed.

Chapter 3 will introduce the new analytically described trellis codes with partially overlapped constellation. The partially overlapped signal constellation will be discussed in both of analytical and coset representations. Examples for one-, two- and four-dimensional signal constellation will be given to describe the use of partially

overlapped constellations for analytically described TCM schemes. Further analysis of new codes will also be discussed.

In chapter 4, the bit error probability performance of these new codes will be analysed. Conclusions and suggestions for further modifications are given in chapter 5.

Chapter 2 Evolution of TCM Techniques

2.1 The General Structure and Principles of TCM

TCM schemes employ redundant nonbinary modulation in combination with a finite-state encoder which governs the selection of modulation signals to generate coded signal sequences. In the receiver, the noisy signals are decoded by a soft-decision maximum-likelihood sequence decoder. The essential new concept of TCM that led to the aforementioned gains was to use signal-set expansion to provide redundancy for coding, and to design coding and signal-mapping functions jointly so as to maximize directly the "free distance" (minimum Euclidean distance) between coded signal sequences. This is because signal waveforms representing information sequences are most impervious to noise-induced detection errors if they are significantly different from each other. Mathematically, this translates into the requirement of signal sequences with large distance in Euclidean signal space. This new concept allowed the construction of modulation codes whose free distance significantly exceeded the minimum distance between uncoded modulation signals, at the same information rate, bandwidth, and signal power.

Following the above arguments, one encoder structure is shown in Fig. 2.1. Here, in order to improve error performance, m bits of data over a period T must be

transmitted in redundantly coded form by a set of 2^{m+1} channel signals. We can easily conclude this as expanding the binary data sequence by suitable convolutional encoding with rate R=m/(m+1), and subsequent mapping of groups of m+1 bits into the larger set of channel signals.

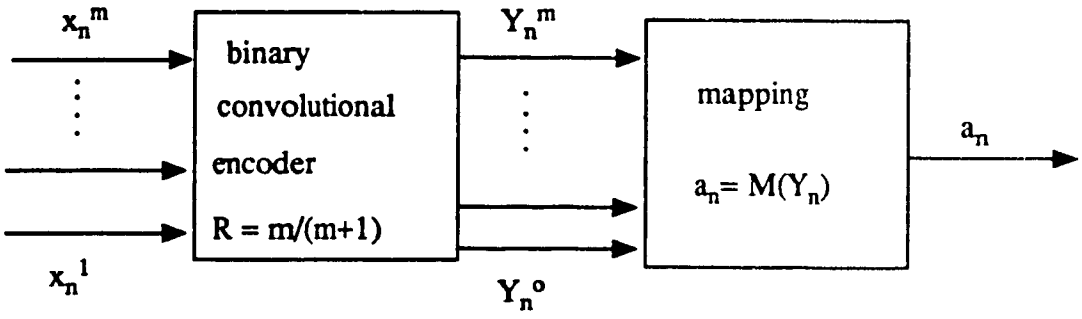

Fig. 2.1 Multilevel encoder structure.

If m bits are not fully encoded and only m′(m′<m) bits are encoded, then another general structure of TCM encoders/decoder is depicted in Fig. 2.2 [5].

According to this figure, TCM signals are generated as follows: When m bits are transmitted per encoder/modulator operation, m′< m bits are expanded by a rate m′/(m′+1) binary convolutional encoder into m′+1 coded bits. These bits are used to select one of $2^{m'+1}$ subsets of a redundant 2^{m+1}–ary signal set. The remaining m–m′ uncoded bits determine which of the $2^{m-m'}$ signals in this subset is to be transmitted With $d(a_n, a_n')$ denoting the ED between channel signals a_n and a_n', the encoder should be designed to achieve maximum free ED:

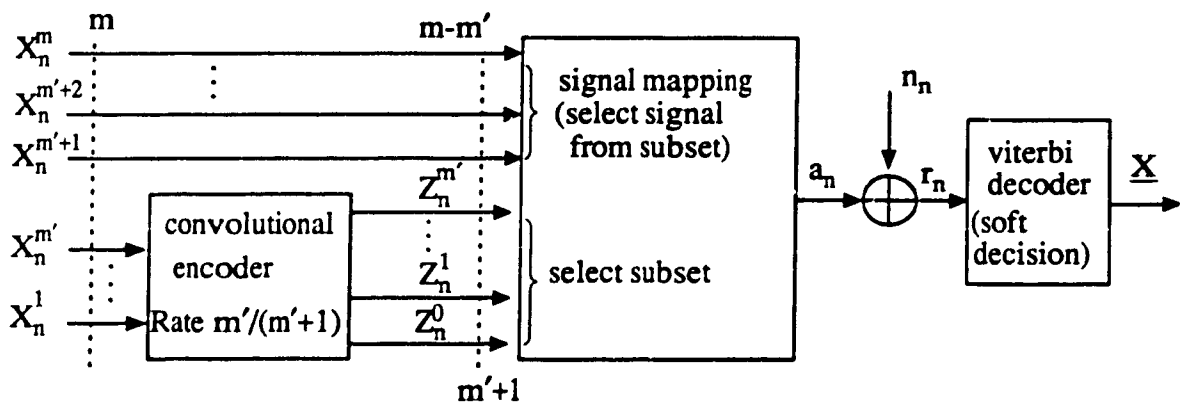

Figure 2.2 General structure of encoder/decoder for trellis-coded modulation.

$$d_{free} = \underset{\{a_n\}\neq\{a_n'\}}{\text{Min}} \left(\sum_n d^2(a_n, a_n') \right)^{1/2} \qquad (2.1)$$

between all pairs of channel–signal sequences $\{a_n\}$ and $\{a_n'\}$ which the encoder can produce.

In order to compare trellis coded modulation with uncoded ones in a convenient way, we consider uncoded 2^m channel signal modulation as a specific one-state trellis coding as shown in Fig. 2.3. The 2^m "parallel" transitions in the one-state trellis diagram of Fig. 2.3 [5] for uncoded 2^m-modulation do not restrict to the

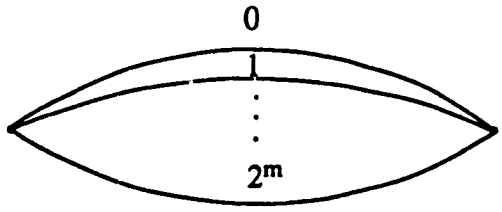

Fig. 2.3 One-state trellis diagram.

sequences of 2^m signals that can be transmitted, that is, there is no sequence coding.

Hence, the optimum decoder can make independent nearest-signal decisions for each noisy 2^m-modulation signal received. The smallest distance between the 2^m signals is denoted as Δ_o. We call it the "free distance" of uncoded 2^m-modulation to use common terminology with sequence-coded systems.

Now, let us discuss some important rules for designing optimum TCM.

2.1.1 Set Partitioning

The main significance of TCM technique is the mapping method between the convolutional binary codes and channel signals, called "mapping by set partitioning" [4, 5]. Through Fig. 2.4 we illustrate this concept for a 16-QAM channel signal set, a signal set of lattice type "Z_2". Generally, the notation "Z_k" is used to denote an infinite "lattice" of points in k-dimensional space with integer coordinates. Lattice-type signal sets are finite subsets of lattice points, which are centered around the origin and have a minimum spacing of Δ_o.

First, set partitioning divides a signal set successively into smaller subsets with maximally increasing smallest intra-set distances Δ_i, i = 0, 1, ... Each partition is two-way. The partitioning is repeated m′+1 times until $\Delta_{m'+1}$ is equal to or greater than the desired free distance of the TCM scheme to be designed. The finally obtained subsets, labeled as C_o, C_1, C_2, C_3 in the case of Fig. 2.4, will henceforth be referred to as the "subsets".

Second, the labeling of branches in the partition tree by the m′+1 coded bits $Z_n^{m'}$, , Z_n^o, in the order as shown in Fig. 2.4, results in a label $Z = \{Z_n^{m'}, ..., Z_n^o\}$ for each subset. The label reflects the position of the subset in the tree. This labeling leads to an important property. If the labels of two subsets agree in the last q positions, but not in

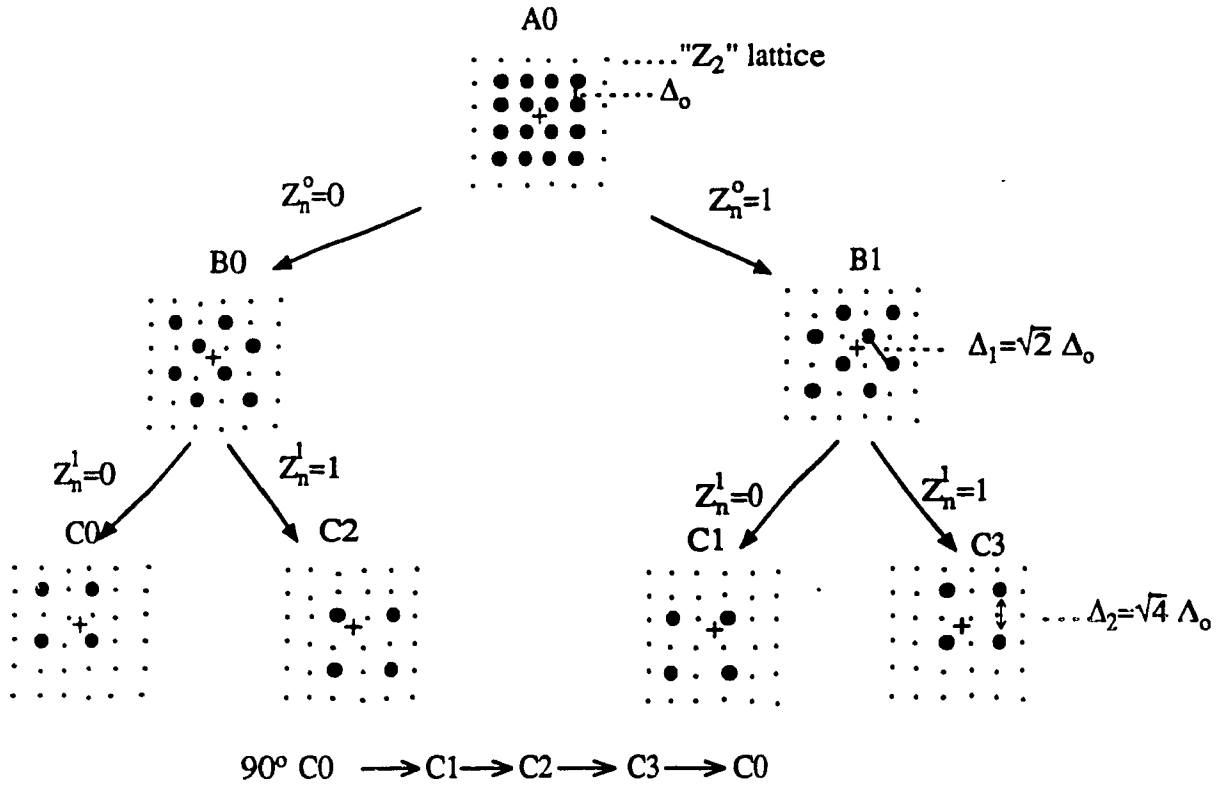

Fig. 2.4 Set partitioning of the 16-QAM signal set (of lattice type "Z_2").

the Z_n^q bit, then the signals of the two subsets are elements of the same subset at level q in the partition tree; thus they have at least distance Δ_q.

Finally, the m−m′ uncoded bits $X_n^{m'}$, ..., $X_n^{m'+1}$ are used to choose a signal from the selected subset. The specific labeling of subset signals by these bits is not particularly important. In the code trellis, the signals of the subsets become associated with $2^{m-m'}$ parallel transitions.

Then, the free Euclidean distance of a TCM code can be expressed as

$$d_{free} = \text{Min}\,[\Delta_{m'+1}, d_{free}(m')] \qquad (2.2)$$

where $\Delta_{m'+1}$ is the minimum distance between parallel transitions and $d_{free}(m')$ denotes the minimum distance between nonparallel paths in the TCM trellis diagram. In the special case of m'=m, the subsets contain only one signal, and hence there are no parallel transitions.

2.1.2 Convolutional Codes for Trellis-Coded Modulation

The rate-m'/(m'+1) convolutional encoder depicted in Fig 2.2 receives m' input bits, and generates m'+1 coded bits which serve as the subset labels $\underline{Z}_n = [Z_n^{m'}, ..., Z_n^0]$ at time n. The set of all possible sequences $\{\underline{Z}_n\}$, which the encoder can generate, forms a convolutional code. Valid code sequences must satisfy the following parity-check equation at all times n,

$$\sum_{i=0}^{m'} (h_v^i Z^i_{n-v} \oplus h^i_{v-1} Z^i_{n-v+1} \oplus \ldots h_o^i Z_n^i) = 0 . \qquad (2.3)$$

In this equation, $\oplus$ denotes modulo–2 addition. The quantity v is called the constraint length, or we can say that encoder realizations with v binary storage elements, which is equivalent to saying that the code has 2^v trellis states. The quantities h_l^i, $v \geq l \geq 0$; $0 \leq i \leq m'$, are the binary parity-check coefficients of the code.

To search optimum TCM codes, $d^2_{free}(m')$ must be as large as possible. Let now $\{\underline{Z}_n\}$ and $\{\underline{Z}_n'\} = \{\underline{Z}_n \oplus \underline{e}_n\}$ be two code sequences, where $\{\underline{e}_n\}$ denotes the error sequence by which these sequences differ. Since the convolutional code is linear, $\{e_n\}$ is also a code sequence. The squared free distance between non-parallel paths in the TCM trellis is bounded by [2]

$$d_{free}^2(m') \geq \underset{\{\underline{e}_n\} \neq \{0\}}{\text{Min}} \sum \Delta^2_{q(\underline{e}_n)} . \qquad (2.4)$$

Here $q(\underline{e}_n)$ is the number of trailing zero in $\underline{e}_n$, that is, the number of trailing positions in which two subset labels $\underline{Z}_n$ and $\underline{Z}_n' = \{\underline{Z}_n \oplus e_n\}$ agree. This means that the distance between signals in the subsets selected by $\underline{Z}_n$ and $\underline{Z}_n'$ is lower-bounded by $\Delta_{q(\underline{e}_n)}$. Minimization has to be carried out over all non-zero code (error) sequences $\{\underline{e}_n\}$ that deviate at, say, time 0 from the all-zero sequence $\{\underline{0}\}$ and remerge with it at a later time. For any given sequence $\{e_n\}$ there exist two coded signal sequences whose signals have at any time n the smallest possible distance between the signals of subsets whose labels differ by $\underline{e}_n$. Usually, this smallest distance equals $\Delta_{q(\underline{e}_n)}$ for all $\underline{e}_n$. If this is the case, the above bound on $d_{free}(m')$ becomes an equation.

This equation is of key importance in the search for optimum TCM codes. It states that free Euclidean distance can be determined in much the same way as free Hamming distance is found in linear binary codes, even though linearity does not hold for TCM signal sequence. It is only necessary to replace the Hamming weights of the e_n (number of 1's in $\underline{e}_n$) by the Euclidean weights $\Delta^2_{q(\underline{e}_n)}$.

2.1.3 Search for Optimum TCM Codes

For a given sequence of minimum intra-set distances $\Delta_0 \leq \Delta_1 \leq \ldots \Delta_{m'}$, and a chosen value of v, a convolutional code with the largest possible value of $d_{free}(m')$ can be found by a code-search program described in [2]. The program performs the search for the $(v + 1).(m' + 1)$ binary parity-check coefficients in a particular order and with a

set of code-rejection rules such that explicit checks on the value of $d_{free}(m')$ are very frequently avoided. Finally, another important aspect of TCM will be described.

2.1.4 Soft-decision Decoding

The Viterbi algorithm is employed as decoder (soft decision). The Viterbi decoding algorithm was discovered and analyzed by Viterbi [15] in 1967. The Viterbi algorithm essentially performs maximum likelihood decoding. However, it reduces the computational load by taking advantage of the special structure in the code trellis. The advantage of the Viterbi decoding, compared with brute-force decoding, is that the complexity of a Viterbi decoder is not a function of the number of symbols in the codeword sequence. The algorithm involves calculating a measure of similarity, or distance, between the received signal, at time t_i, and all the trellis paths entering each state at time t_i. The Viterbi algorithm removes from consideration those trellis paths that could not possibly be candidates for the maximum likelihood choice. When two paths enter the same state, the one having the best metric is chosen; this path is called the surviving path. This selection of surviving paths is performed for all the states. The decoder continues in this way to advance deeper into the trellis, making decisions by eliminating the least likely paths. The early rejection of the unlikely paths reduces the decoding complexity. Note that the goal of selecting the optimum path can be expressed, equivalently, as choosing the codeword with the maximum likelihood metric, or as choosing the codeword with the minimum distance metric.

2.2 TCM with Multidimensional Constellations

Because of the effects of carrier-phase offset, recently, there have been a number of investigations into trellis coding with signal sets defined in more than two dimensions [8, 9, 16-20]. When a carrier-modulated two-dimensional TCM signal is demodulated with a phase offset Δ_ϕ, the soft-decision decoder then operates on a sequence of complex-valued signals $\{r_n\} = \{a_n \exp(j\Delta_\phi) + W_n\}$, where the a_n are transmitted TCM signals and the W_n denote additive Gaussian noise. The phase offset Δ_ϕ could be caused, for instance, by disturbances of the carrier phase of the received signal which the phase-tracking scheme of the receiver cannot track instantly.

The effect of this phase offset can be explained as follows. In the trellis diagrams of TCM schemes, there exist long distinct paths with low growth of signal distance between them, that is, paths which have either the same signals or signals with smallest distance Δ_0 assigned to concurrent transitions. In the absence of phase offset, the non-zero squared distances Δ_0^2 and the squared larger distances of diverging or merging transitions add up to at least the squared free distance. However, if phase offset rotates the received signals such that received signals become located halfway between the signals of the original signal set, the difference in distance between received signals and the signals on distinct transitions that are Δ_0 apart may be reduced to zero. There may then be no difference in distance between a long segment of received signals and two distinct trellis paths, just as though the code were catastrophic. At this point, the decoder begins to fail.

In general, it is desirable that TCM codes have as many phase symmetries as possible to ensure rapid carrier-phase resynchronization after temporary loss of synchronization. In practical systems, multi-dimensional signals can be transmitted as

sequences of constituent one- or two-dimensional (1-D or 2-D) signals. In this section, 2k-D TCM schemes are considered which transmit m bits per constituent 2-D signal, and hence mk bits per 2k-D signal. The principle of using a redundant signal set of twice the size needed for uncoded modulation is maintained. Thus, 2k-D TCM schemes result in less signal redundancy in the constituent 2-D signal sets

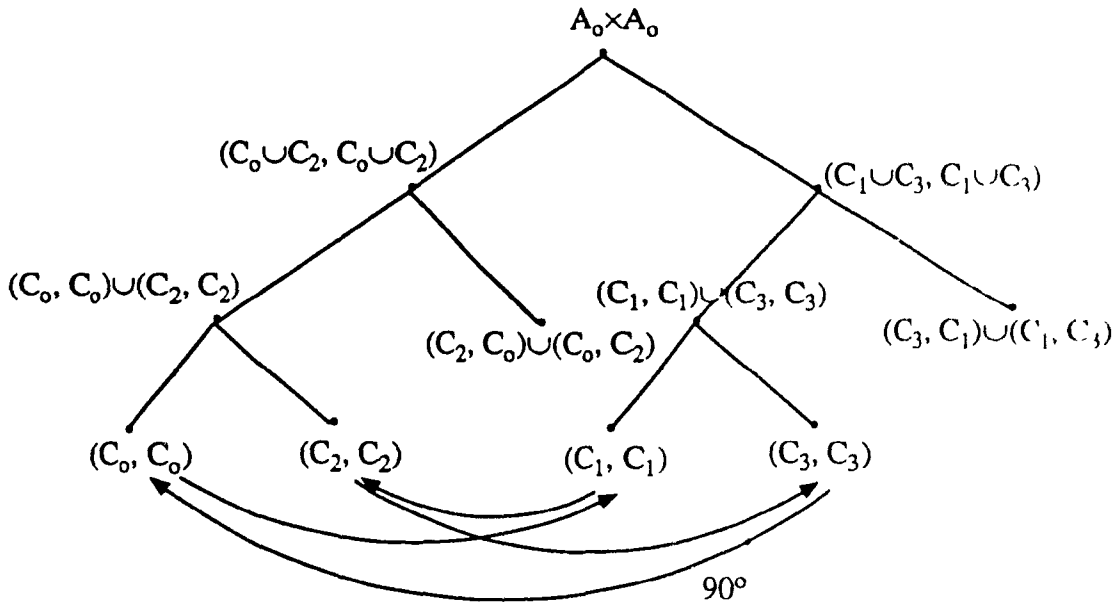

Fig. 2.5 Partition of four-dimensional signal sets.

It has been found that multi-dimensional TCM schemes have more 90° phase invariance [20]. This is an important advantage which we desire. Also the larger signal spacing should make multi-dimensional TCM systems less sensitive to phase offset [5].

Now, let us analyse the multi-dimensional signal sets. For 2-D TCM schemes with "Z_2"-type signal sets, the minimum signal spacing Δ_0 must be reduced by approximately the factor $\sqrt{2}$ (-3dB) to have the same average signal power as for uncoded modulation. But this loss in signal spacing can be compensated by coding to

obtain an overall improvement in free distance. Then the lower signal redundancy of multi-dimensional TCM schemes with "Z_{2k}"-type signal sets results only in a reduction of the minimum signal spacing by the 2k-th root of 2, so coding has to contribute less than in the case of 2-D TCM to detain the same gain in free distance.

One example of set partitioning of four-dimensional signal sets is shown in Fig 2.5 [20]. The corresponding "Z_2"-type signal sets is in Fig. 2.4. Note that sets (C_o, C_o)$\cup$(C_2, C_2), (C_1, C_o)$\cup$(C_o, C_2), (C_1, C_1)$\cup$(C_3, C_3), (C_3, C_1)$\cup$(C_1, C_3) of Fig. 2.5 have the largest distance between them, and the effect of a 90° rotation is shown in Fig. 2.5.

2.3 TCM with Asymmetric Signal Constellation

Symmetric signal constellations, i.e., those with uniformly spaced signal points, have been used traditionally for TCM schemes. But Divsalar and Simon use asymmetric signal constellations instead of Ungerboeck's symmetric ones for TCM to improve the performance [11, 21]. The results in [21] show that for 2-state TCM, significant performance improvement is achievable in some case relative to the equivalent symmetric design. The gain in free Euclinean distance of the two-state trellis-coded asymmetric M level signal constellation over the uncoded M/2-point one is much better than the gain in symmetric M level signal constellation over the uncoded one. But in high coding complexity (the number of states is more than two in the trellis diagram), the amount to be gained by asymmetry diminishes; although the overall of the asymmetric coded system is improved, relative to the equivalent bandwidth uncoded M/2-point system.

By designing asymmetric signal constellations and combining them with

optimized trellis coding, one can further improve the performance of coded systems without increasing power requirements or changing the bandwidth constraints imposed on the system. How to design asymmetric signal constellations? For signal sets with one degree of freedom, e.g., MPSK and M-AM, the optimum asymmetric 2^{n+1}-point constellation to be used with rate n/(n+1) trellis coding is composed by a symmetric 2^n-point constellation and a phase rotated (for MPSK) or amplitude translated (for M-AM) version of itself. For some signal sets with two degrees of freedom, for example, QAM, it appears that the optimum 2^{n+1}-point asymmetric structure is achieved by the optimum 2^n two-dimensional AM-PM structure with a rotated version of itself. In general, for arbitrary two-dimensional structures, the optimum asymmetric set is undoubtedly achieved by combining the optimum symmetric set containing half the number of points with a translated and rotated version of itself.

Another way using Ungerboeck's "set partitioning" concept to describe the M-point asymmetric construction is to imagine partitioning the symmetric M-point constellation into two M/2-point constellations with maximally separated signals and then perform an appropriate rotation (MPSK), translation (M-AM), or combination of rotation and translation (QAM) of one subset with respect to the other. Upon optimization of the amount of translation, rotation, or the combination of the two, the resulting two subsets can be used as the first level of set partitioning in Ungerboeck's method. The example of asymmetric signal sets is shown in Fig. 2.6. The approach of assigning signals to transitions of the trellis code is still based on a mapping rule called "mapping by set partitioning" [3].

Finally, the procedure for designing good trellis codes, combined with optimum asymmetric signal constellation, can be summarized by the following;

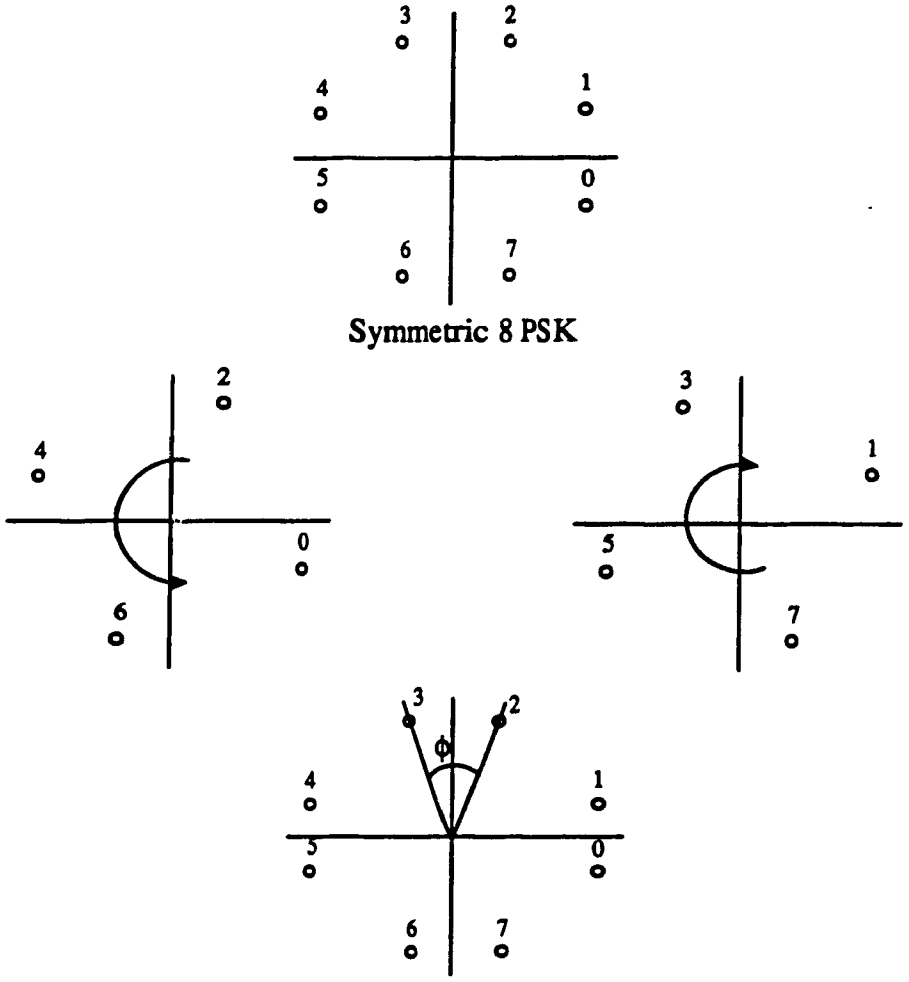

Fig. 2.6. The procedure of designing asymmetric 8-PSK signal sets.

(a) Use the Ungerboeck's mapping rule called "mapping by the set partitioning" method to partition the signal constellation.

(b) Assign signals from either of the two partitions (each containing 2^n signals) generated at the first level of partitioning in (a) to transitions

diverging from a given state. Similarly, assign signals from the other of these two partitions to transitions remerging into a given state. These assignments should be made such that the minimum distance between diverged transitions and the minimum distance between remerged transitions are as large as possible.

(c) Find the free Euclidean distance of the TCM.

(d) Maximize the free Euclidean distance of (c) with respect to the rotation angle ϕ, or the translation Δ, or both. These values of ϕ and Δ define the optimum asymmetric signal constellation.

2.4 A New Description of TCM

Ungerboeck's technique for TCM consists of finding an underlying convolutional code and then mapping the output coded bits into channel signals according to rules referred to as "mapping by set partitioning". As an alternate approach, an analytic or algebraic, description of trellis codes has been introduced by Calderbank and Mazo [10]. They have shown how to realize these two operations in a single-step procedure.

They directly consider the relationship of output channel signals and input binary bits in TCM. A TCM code is then described as a "sliding window" method of encoding a binary data stream $\{a_i\}$, $a_i = 0, 1$, into a sequence of real numbers $\{x_i\}$ which are transmitted into a noisy transmission channel. For the rate of k bits per channel symbol, each channel input x_i will depend not only on the most recent block of

k data bits that enter the encoder but will also depend on the v bits preceding this block. Formally,

$$x_i = x\left(a_{ik}, a_{ik-1}, \ldots, a_{ik-(k-1)} ; a_{(i-1)k}, \ldots, a_{(i-1)k-(v-1)} \right), \qquad (2.5)$$

most recent k-bit block, v bits preceding this block.

Here, the states of encoder are determined by the preceding bits, and for each certain state 2^k possible output symbols are associated for each possible input block of k bits. The state for the next channel symbol is determined by shifting every a_i appearing in (2.5) k places to the right, dropping the right-most k bits, and inserting the new k-bit block at the beginning.

Drawing this encoding procedure sequentially in time results in a trellis structure; hence the name trellis code. The procedure of analytical description trellis codes can be shown as in Fig. 2.7.

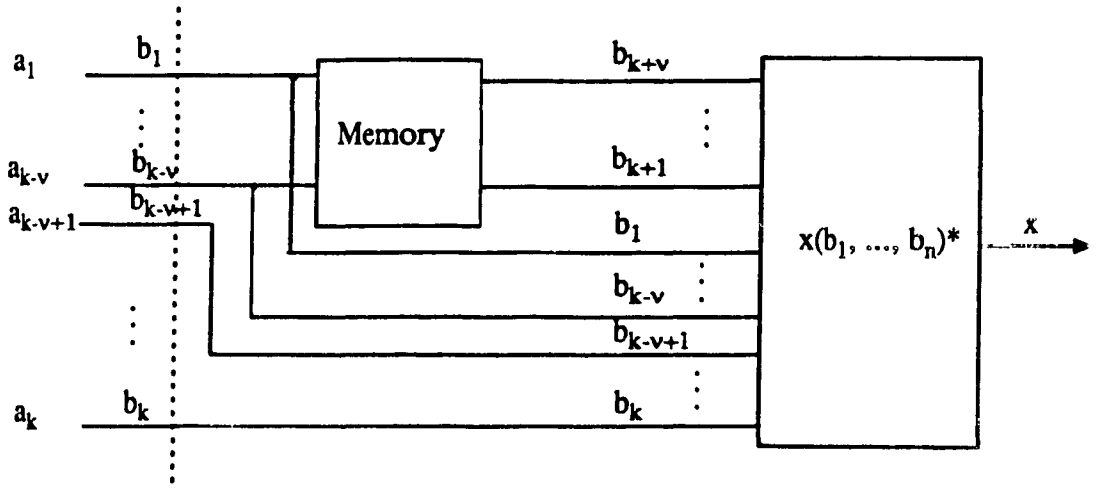

Fig. 2.7 Analytic description of trellis codes.

The channel input signal x_i is considered as a real-valued function of (k+v) binary variables (AM). Then the simplification of (2.5) is shown as follow,

$$x_i = x\,(a_1, a_2, \ldots, a_n)\,, \qquad (2.6)$$

where $n = k + v$. In mathematics we note that any real-valued function x may be written as a sum of products of the a_i,

$$x(a_1, \ldots, a_n) = x_o + \sum_{i=1}^{n} x_i a_i + \sum_{\substack{i,j=1 \\ j>i}}^{n} x_{ij} a_i a_j + \ldots + x_{1\ldots n} a_1 a_2 \ldots a_n\,. \qquad (2.7)$$

There are 2^n constants on the right side of (2.7), one for each distinct product of the variables a_i. They are determined by the 2^n values that $x(a_1, \ldots, a_n)$ can take as pointed out in [22]. Through (2.7), they realize the purpose to express directly the relationship between output channel signal and input binary data.

For trellis codes, in order to take advantage of the signal symmetry, both in the signal set and more generally in the trellis, a conversion from binary to ± 1 input bits is performed. Then (2.7) is changed to the desired form,

$$x \triangleq x(b_1, \ldots, b_n) = x_o + \sum_{i} x_i b_i + \sum_{i,j} x_{ij} b_i b_j + \ldots + x_{1\ldots n} b_1 \ldots b_n\,. \qquad (2.8)$$

where $b_i = \pm 1$ and the constants in (2.8) are not the same as in (2.7). Calderbank and Mazo suggested that a method to determine the constants on the right side of (2.8) is to regard the 2^n values of $x(b_1, \ldots, b_n)$ as a vector x of dimension 2^n. Similarly, regard the 2^n values taken by each product of the variables b_i as a vector of dimension 2^n. Then, the vector of coefficients on the right side of (2.8) can be viewed as the Hadamard

transform of the vector x [23].

The alternative description of trellis codes described in above is analytic rather than graphical, allowing some statistical questions to be easily treated, and still the transmitted power spectrum is not affected by the coding. Furthermore, many practical codes are simple to describe.

Calderbank and Mazo restricted their analysis to real-valued functions. That is, they only looked at the one-dimensional(1-D) case or Amplitude Modulation. Turgeon and Mclane extend their ideas to multi-dimensions and to rotationally invariant trellis codes [24]. We can consider respectively each component of complex signal as a real-valued function of input bits.

From Calderbank and Mazo's method the channel signals is expressed as a function of input binary bits in terms of 2^n constants. The resulting function may be viewed with an equivalent matrix equation [25] as follows:

$$\underline{X} = \underline{B}\,\underline{D} \tag{2.9}$$

where $\underline{X}$ = channel signal matrix

$\underline{B}$ = Hadamard matrix with elements ± 1

and $\underline{D}$ = matrix of constants to be determined.

But here, in equation (2.9), we have,

$$\underline{X}^T = \{x(b_1, b_2, ..., b_n)\}, \quad b_i = \pm 1$$

$$(D \times 2^n \text{ matrix})$$

$$\underline{B}_k = \{1, b_1, b_2, ..., b_1 b_2 ... b_n\} \tag{2.10}$$

$$(\text{kth row of } 2^n \times 2^n \text{ matrix } \underline{B})$$

$$\underline{D}^T = \{d_o, d_1, ..., d_{12...n}\}$$

($D \times 2^n$ matrix of unknown)

where D is the number of signal dimensions and superscript T denotes matrix transpose.

Then, we can make the following conclusion: although Calderbank and Mazo only looked at the one-dimensional trellis-coded schemes, however, no matter what the number of signal dimensions is, the input bits can only take two values, that is ± 1 Hence, the Hadamard matrix $\underline{B}$ remains unaffected by the choice of dimension. The only change is that the matrix of constants, $\underline{D}$, has the same number of dimensions as the channel signals x. Clearly, equation (2.9) could always be used for any trellis code.

2.5 Multiple TCM for Gaussian and Fading Channels

A trellis coded modulation technique referred to as multiple trellis coded modulation (MTCM) is developed by Divsalar and Simon [26, 27], wherein more than one channel symbol per trellis branch is transmitted. The principle behind their discovery is to design a rate nk/(n+1)k (k = 2, 3, 4 ...) encoder and combine it with a 2^{n+1}-point signal constellation outputting k of these signal points (one for each group of n+1 encoder output symbols) in each transmission interval. In each transmission interval, kn bits enter the encoder and k symbols leave the modulator, we still have a unity bandwidth expansion relative to an uncoded 2^n-point uncoded system When values of k are greater than 1 (k = 1 corresponds to the conventional TCM system), the values of d_{free} are increased with symmetric modulations for certain cases. They have found simple two-state trellis codes for symmetric MPSK and AM modulations which can achieve 3dB gain over uncoded modulation at very high signal-to-noise ratios

without bandwidth expansion and without reduction in information bit rate.

Indeed, MTCM may also be thought of as a coding onto a multidimensional signal set constructed from successive channel symbols [7, 16, 18, 28-30]. But, there are differences between the two. The encoder used in Multidimensiona! Trellis Codes is rate nk/(nk+1) [5] which is different from a rate nk/(n+1)k encoder used in MTCM. In other words, MTCM differs from the other multi-dimensional codes in the sense that the size of the channel symbol set is always twice the size of the uncoded set used for comparison. MTCM is also different from lattice coding [28]. The theory of set construction used in lattice coding cannot be directly applied to MPSK modulation.

For k=2, the partition of signal sets is similar to that of high-dimensional sets. But, for k=3 or more, we cannot directly use the partition of high-dimension signal sets. Since the simple two-state MTCM shows a significant improvement over the uncoded case, we describe the general mapping rule for 2-state MTCM. Suppose that the original signal set is first divided into two subsets B_0 and B_1 as maximally increasing intra-set minimum distance. The rules used here are,

(1) the transitions emanating from state "0" will be assigned with signals in partition B_0; partition B_1 will be used for transitions emanating from state "1";

(2) for 2^{nk-1} parallel paths between like states (or unlike states), assign to each parallel path a sequence of k symbols from a partition (B_0 or B_1), such that the minimum squared distance between any two of these parallel paths is equal to twice the minimum squared distance between points in the partition;

(3) the remaining 2^{nk-1} k-tuples formed from symbols in the same partition are assigned to the parallel paths corresponding to a transition to an unlike state. The minimum squared distance among all pairs of parallel paths between unlike states will also be twice the minimum squared distance between points in the partition.

According to above mapping rule, the minimum squared distance among all pairs of paths consisting of a path between like states and a path between unlike states both originating from the same state is only equal to the minimum squared distance between the points in the partition. The place where the trellis multiplicity k has its influence is in regard to the minimum squared distance among all pairs of paths consisting of a path between like states and one between unlike states where the two paths originate from two different states. With the above k-tuple assignments, this minimum squared distance is k times the minimum squared distance between points in one partition and points in the other. d_{free} performance is improved when the minimum distance associated with the error event path of length 2 increase with the increase of k.

Recently, Divsalar and Simon have introduced the technique of Multiple Trellis Coded PSK for fading mobile satellite channels. In previous publications [12, 31-33], they have considered the performance of conventional and multiple trellis codes in a Rician fading environment characteristic of the mobile satellite channel. Results were reported for both the case of coherent detection and deferentially coherent detection with and without the use of channel state information (CSI). The primary emphasis in these previous works was the degradation in performance produced by the fading for trellis codes designed to be optimum on the additive Gaussian noise channel

(AWGN).

In recent papers [13, 14], they look more carefully into the properties of the average bit error probability of trellis coded modulation (TCM) and then proceed to use these as design criteria for conventional and multiple trellis codes operating over a fading channel. It is shown that, over Rician fading channels with interleaving/ deinterleaving, the asymptotic performance of TCM at high signal-to-noise ratio (SNR) is dominated by several other factors depending on the value of the Rician parameter **k**, i.e., the ratio of direct plus specular power (coherent components) to diffuse power (noncoherent component). In particular, for small values of **k** (the channel tends toward Rayleigh), the primary design criteria for high SNR become:

1) the length of shortest error event path.

2) the product of branch distances along that path with d_{free} a secondary consideration.

Thus, the longer is the shortest error event path and the larger is the product of the branch distances along that path, the better the code will perform even though d_{free} does not achieve its optimum value over the AWGN!

The set partitioning method used for optimally designing trellis codes to be transmitted over the Rician fading channel is different from Ungerboeck's method, which is used for optimally designing trellis codes to be transmitted over the additive white Gaussian noise (AWGN) channel. The procedure will be based on the optimum performance criteria developed for this channel. In fact, it may not lead to a tree structure.

Why have they used multiple TCM for fading channel and not conventional trellis codes (i.e., those with one channel signal per branch)? Considering the criteria for designing TCM to achieve minimum error probability performance over the fading channel described above, indeed, it was shown that the analogy to maximizing diversity was to design the code such that the length (as measured by the number of channel signals) of the shortest error event path is maximized. It is observed that for trellises with parallel paths, conventional trellis codes are limited to a diversity of one. Furthermore, for trellises with no parallel paths, the diversity achievable with conventional trellis codes is still limited to the number of branches along the shortest error event path. But allowing for multiple symbols per trellis branch, i.e., multiple trellis coded modulation (MTCM), provides an additional degree of freedom for designing a code to meet the optimization criteria on the fading channel. In particular, we are able to achieve diversities larger than those achievable with conventional trellis codes having the same number of trellis states. It is here where the MTCM technique exploits its full potential.

2.6 Other Works to Improve TCM

Some interesting new TCM schemes that exhibit modest improvements are shown by Forney, JR. et al. in [7]: a 2-state code that has a nominal coding gain of almost 3 dB, and an 8-state trellis code with a coding gain of 4.5 dB. The idea in the 2-state scheme is to use subsets that are partially overlapping and partially distinct. The other one in the 8-state scheme is to use a 4-dimensional constellation as the basic constellation(4-dimensional trellis codes have also been studied by Wilson [34] and Fang et al. [35].). Two simple examples will be described to illustrate these ideas

For the 2-state scheme they use a 80-point signal constellation, which is divided to four subset A′, A", B′, B". Each of four subsets has 32-points. A′ and A" both include 24 same inner A points, but each includes a different 8 outer points; and similarly with B′ and B". Thus A′ and A" are partially overlapping and partially disjoint, and so are B′ and B". For a 2-state trellis diagram, the squared distance between two paths beginning and ending at common nodes remains $2d_o^2$ since the basic distance properties between A subsets and B subsets remain [4]. And there also is a slight reduction in power due to the increased constellation size.

In another 8-state trellis code they first divide a two-dimensional 2^n-point rectangular grid into four subsets . Then the binary convolutional encoder for this scheme, operates on pairs of symbols (4-dimensional signals). During each pair of symbol intervals, three bits enter the encoder and four coded bits are produced. The first two coded bits select the subset for the first symbol and the second two bits select the subset for the second symbol. The minimum free Hamming distance of this convolutional code is 4. The squared distance between any two sequences corresponding to different encoded outputs is at least $4d_o^2$. Over two symbol intervals, 2n-1 bits enter the modem and one parity check is generated, giving 2n bits to select the two signal points. Since (n-1/2) bits/symbol enter the modem, there is a loss of 1.5 dB due to the larger signal constellation. And then a gain of 6 dB in distance is reduced to a net nominal coding gain of 4.5 dB.

Traditionally, for a rate of n bits per channel symbol, a 2^{n+1} point constellation is used for TCM schemes, in both of Ungerboeck's methods and Calderbank and Mazo's schemes. Calderbank and Mazo's analysis description of TCM has an important advantage than Ungerboeck's which is to simplify the procedure of TCM design. Even though Forney, et al. [7] have introduced a new channel signal

constellation which has less than 2^{n+1} points for trellis coded QAM modulation which has the advantage of reduction of average power of channel signals. But they restricted to 2-D dimensional one, and also did not find its analysis description. In this thesis, we study the partially overlapped signal constellations which have less than 2^{n+1} points for 1-D, 2-D and multi-dimension. To keep the advantages of both of the partially overlapped constellation and the analytic description of TCM, we present the analytical description of TCM's with the partially overlapped constellation in one and two dimensions. Furthermore, analytically described trellis coded modulation with high-dimensional partially overlapped constellation is studied, because the larger signal spacing of multi-dimensional constellation can make TCM systems less sensitive to phase offset and also compared to low dimensional schemes, this results in less signal redundancy in the constituent low dimensional signal sets [36, 37].

Chapter 3 Analytically Described Trellis Codes with Partially Overlapped Constellation

There is a growing need for reliable transmission of high quality voice and digital data in digital communication systems. These systems are both power and bandwidth limited. By integrating both coding and modulation into a single entity, the technology of TCM has provided significant coding gain without bandwidth expansion when comparing to uncoded conventional uncoded signal sets. This is the work that has laid the foundation for the design and development of all power and bandwidth efficient digital modems in practice today [3-5]. Recently, many people work on modification and improvement of TCM. These works have developed and perfected TCM to adapt the need of high quality transmission of digital signal. In this chapter, we will show our new analytically described TCM with partially overlapped constellation[36, 37]. It reduces implementation complexity and also possesses other desirable properties.

3.1 Introduction

In Ungerboeck's TCM representation, first n input bits enter a rate n/(n+1) convolutional encoder, and then the n+1 output bits of the encoder are assigned to the channel signals, from a 2^{n+1}-point constellation, according to the "mapping by set

partitioning rule". The added redundancy obtained by doubling the number of channel signals (compared to a corresponding uncoded scheme) allows Ungerboeck to achieve significant gain over uncoded modulation.

Calderbank and Mazo presented an analytical description of the trellis codes which combine Ungerboeck's two steps into one [10]. In Calderbank and Mazo's representation, they express the channel input signal, as a function of the input bits (2.8). To search the optimum codes, they use the computer to find d^2_{min}. The method is to use the graph of possible transitions, where edges are labeled with the contribution of the transition to d^2. More specifically, in this graph the state to which the transition is made is used to determine which terms of the code are excited when evaluating. A state (0 ... 0) has been introduced as an initial state, and the length l of an error event corresponds in this graph to a path of length l from (0 ... 0) to (0 ... 01).

Turgeon and Mclane have used the same method for two- and four-dimensional cases[24].

Ungerboeck, Calderbank and Mazo [10], Turgeon and Mclane [24] still use the traditional 2^{n+1}-point signal constellation for input n bits codes. In this chapter, we investigate the use of constellations having less than 2^{n+1}-points called the partially overlapped signal constellations [7]. Partially overlapped signal constellations are formed by omitting some of the high power signal points of a conventional 2^{n+1}-point constellation. This results in a constellation with subsets having common signal points. Since the number of points in such a constellation is less than what n+1 bits at the output of convolutional encoder can address, some of the signals, preferably signals having lowest power, should be used more than once. This in turn results in a saving in transmitted energy, and therefore, in a coding gain as compared to the TCM schemes using conventional constellation.

3.2 Partially Overlapped Signal Constellation

3.2.1 Analytical Representation

In representation of (2.8), before we set x_o equal to zero since it does not contribute to the distance. Also, in order to have 2^{n+1} signal points, exactly n+1 terms of (2.8) should have non-zero coefficients. For example, for a 16AM/8AM TCM code, i.e., for n=3, only four term expressions of the form,

$$x = \alpha_o f_o(b_1, ..., b_m) + \alpha_1 f_1(b_1, ..., b_m) \cdot \alpha_2 f_2(b_1, ..., b_m) + \alpha_3 f_3(b_1, ..., b_m) \quad (3.1)$$

should be considered, where $f_i(.)$, $i = 0,...,3$ are products of different subsets of $\{b_1, ..., b_m\}$, and, therefore take values ± 1.

It is obvious that in order to have 2^{n+1} distinct signal points, i.e., for TCM schemes with conventional constellation, there should not be any linear dependence between subsets of $\{\alpha_o, ..., \alpha_n\}$ with coefficients ± 1. To see this we assume that $\sum_{j=0}^{k-1} \alpha_j c_j = 0$ where, $c_j = \pm 1$, $j = 0, ..., k-1$, and $k \le n+1$. Then, these k coefficients result in 2^k-1 (and not 2^k) values, and, therefore, if there is no other such linear dependence between remaining n+1-k coefficients, the number of distinct signal points will be $(2^k-1)2^{n+1-k}=2^{n+1}-2^{n+1-k}$. For example, in the case of n=3, if we let $\alpha_o=1$, $\alpha_1=2$, $\alpha_2=\alpha_3=4$ in (2.3), then we have a linear dependency relationship of the form $\alpha_2-\alpha_3=0$, and, therefore, k=2, and the number of points will be $2^4-2^{4-2}=12$. If there are more linear dependency relationships between the coefficients, then more points will overlap

resulting in a constellation with even fewer points. Let's consider the situation where there are two linear dependency relationships involving k_1 and k_2 terms. Furthermore, assume that k_o of the terms are common in two relationships. Then, it is easy to show that the number of points in the resulting constellation is,

$$N = 2^{n+1} - 2^{n+1-k_o} - 2^{n+1-(k_1+k_2-k_o)} . \qquad (3.2)$$

As an example, take the expression (2.3) again, but this time with $\alpha_o = 1$, $\alpha_1 = \alpha_2 = 2$, and $\alpha_3 = 4$. Then, we have two linear combinations, $\alpha_1 - \alpha_2 = 0$ and $\alpha_1 + \alpha_2 \ \alpha_3 - 0$. so, $k_1 = k_o = 2$, $k_2 = 3$ and, therefore, the resulting constellation has only 10 points, namely, ±1, ±3, ±7 and ±9. Deriving expression similar to (3.2) for other cases of interest is a straight forward practice.

3.2.2 Coset Representation

Calderbank and Sloane [28] made the observation that signal points can be considered as finite subsets of infinite lattices. This observation resulted into representing the TCM schemes in terms of coset codes [19]. A lattice Λ is an infinite regular set of points in an N-dimensional space. Signal points are taken from a finite subset of points lying inside a coset of Λ the form $\Lambda+c$ which is the set of N-tuples of the form $\lambda+c$, where $\lambda \in \Lambda$, and therefore $\Lambda+c$ is a translate of Λ. A sublattice Λ' of Λ is a subset of elements of Λ which is itself a lattice. A sublattice Λ' induces a partition of Λ, denoted by Λ/Λ', into equivalence class, we obtain a set of coset representatives for the partition Λ/Λ', denoted by $[\Lambda/\Lambda']$. Then we can write the following coset decomposition of Λ,

$$\Lambda = \Lambda' + [\Lambda/\Lambda']. \qquad (3.3)$$

That is, every element of Λ can be written as $\lambda = \lambda' + c$ where $\lambda' \in \Lambda'$ and $c \in [\Lambda/\Lambda']$. For example, take $\Lambda = Z$, i.e., the set of all integers and $\Lambda' = 2Z$, the set of even integers. Then 0, and 1 from a system of coset representatives for the partition $Z/2Z$, i.e., Z can be regarded as the vision of $2Z$, and $2Z+1$.

Using the idea of coset decomposition, a TCM code can be represented using the structure of Fig. 3.1 [19].

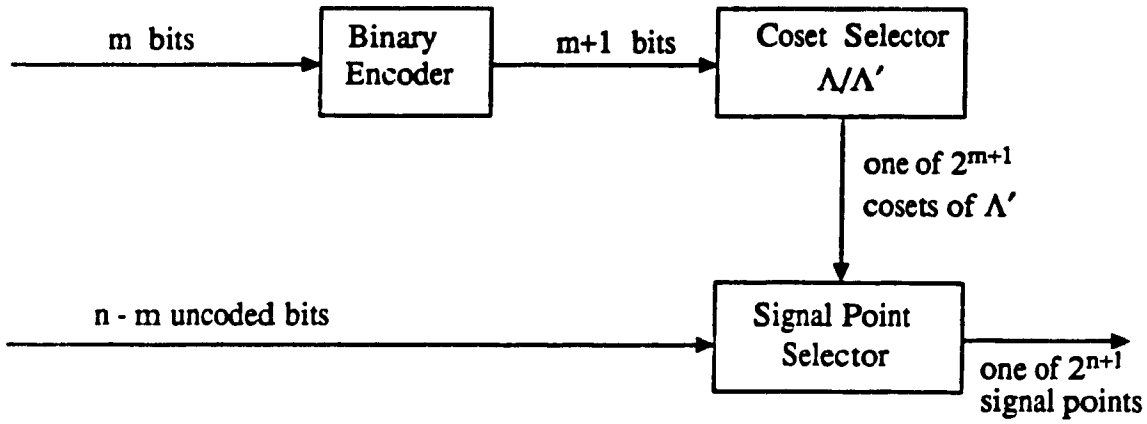

Fig. 3.1 General structure of a coset code [19].

It is evident that when the number of points in the constellation is less than 2^{n+1}, the points cannot be partitioned as explained above. However, there is still possibility of decomposition into overlapping subsets. In particular, consider the 10-point constellation discussed above. Points of this constellation can be written as,

$$x = f_0 + 2f_1 + 2f_2 + 4f_3 = (f_0 + 2f_1) + 2(f_2 + 2f_3). \qquad (3.4)$$

From (3.4), one can immediately see that the constellation can be decomposed using finite subsets of lattices $4Z$ and $2Z$ (more accurately $4Z + 2$ and $2Z + 1$). That is, each point in the constellation can be written as $\lambda = \lambda' + c$ where $\lambda' \in \{-6, -2, 2, 6\}$ and $c \in \{-3, -1, 1, 3\}$. Now, the structure of Fig. 3.1 can be modified, for an $n = 3$, $m = 1$ TCM code, as follows. Two bits at the output of the encoder choose one of four "coset leaders," ± 1, ± 3, and uncoded bits b_2 and b_3 choose a point from the set $c + \{\pm 2, \pm 6\}$ where c is the "coset leader" selected.

3.3 Examples

In this section, we present two examples demonstrating the application of partially overlapped signal constellations for the design of TCM schemes. Both examples are simple two state codes. The first example is a one-dimensional scheme using the 10-point constellation described above. The second example uses a 12 point QAM constellation.

3.3.1 AM Examples

In this sub-section, we consider a two-state, $n = 3$ code using the 10-point constellation discussed above. First bit, b_1, specifies the state. Therefore, there are four parallel transitions between any two states specified by the values of b_2 and b_3. It is obvious that there are four groups of such parallel transitions between states Each group can be specified by b_4 and b_1. where, b_4 and b_1 determine the old and new states, respectively. In order to minimize the distance between the parallel branches, we assign to each set of parallel paths one of the cosets of $\{\pm 2, \pm 6\}$. In other words, b_1 and

b_4 determine $c \in \{\pm 1, \pm 2\}$ and $b_2\, b_3$ choose one of four points in $c + \{\pm 2, \pm 6\}$. That is, the channel signal x can be written as $x = x_1 + x_2$, where x_1 is a two term expression with coefficients 1 and 2 and involving only b_1 and b_4 while x_2 has coefficients 2 and 4, involves b_2 and b_3 only. In 2.1 the TCM design criteria has been discussed, the optimum TCM is which with the largest minimum squared Euclidean distance called the squared "free distance". It is easy to write a program to search for the codes with the largest squared "free distance" among the codes satisfying the above mentioned assignment rules. However, simple codes such as present example, can be designed by hand. In particular, the following additional mapping rule can be helpful. Fig. 3.2 shows a two stage decomposition of the 10-point constellation into four previously mentioned subsets. Note that while sets A′ and A" (also B′ and B") overlap, none of the pairs (A′, B′), (A", B"), and (A′, B"), (A", B′) has points in common.

Our additional mapping rule of assigning the signals to the transitions in the trellis diagram is as the following,

1. parallel transitions are associated with signals with maximum distance between them, in other words, the signals in the four previously mentioned subsets.

2. non-overlapping signal sets are assigned to transition sets leaving or merging to a given state.

3. signals are used in the trellis diagram with unequal frequency, some signals with lowest power are used once more than that other high-power signal be used.

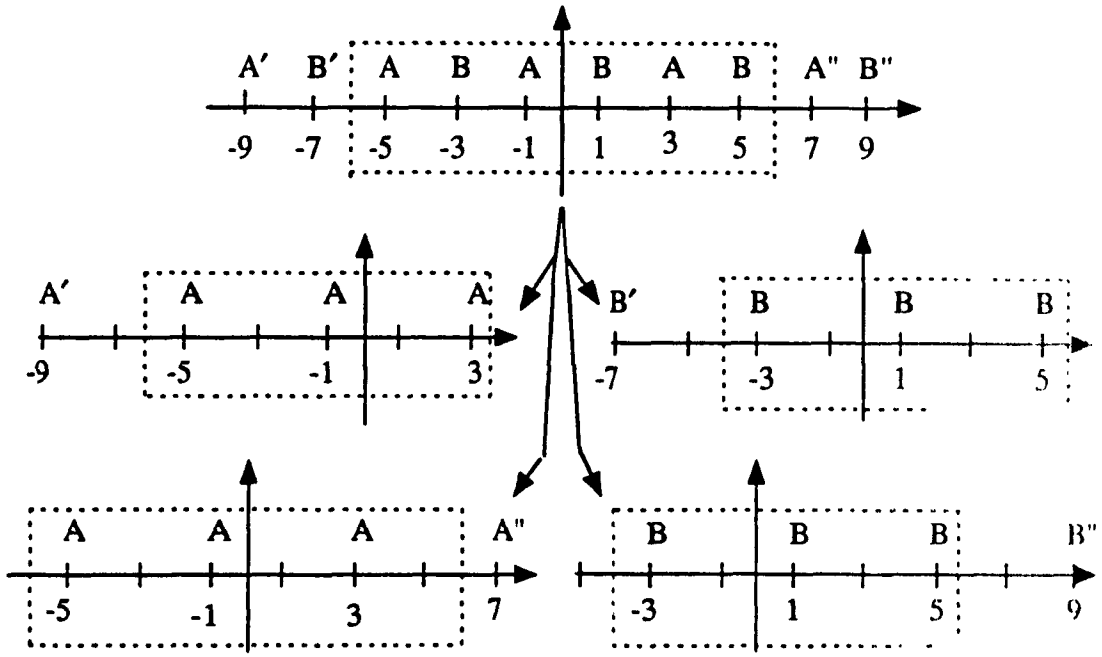

Fig. 3.2 Decomposition of 10-AM constellation.

A trellis diagram for a two-state, n = 3 trellis code using the 10 point constellation is shown in Fig. 3.3.

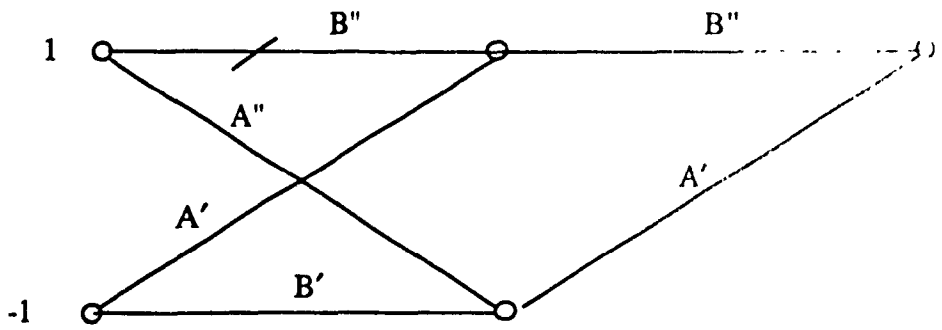

Fig. 3.3 A trellis diagram for 2-state, n = 3 trellis code.

Note that this subset assignment is not unique. We can simply replace each A′(B") with A"(B") and vice versa and obtain another code (Fig. 3.4).

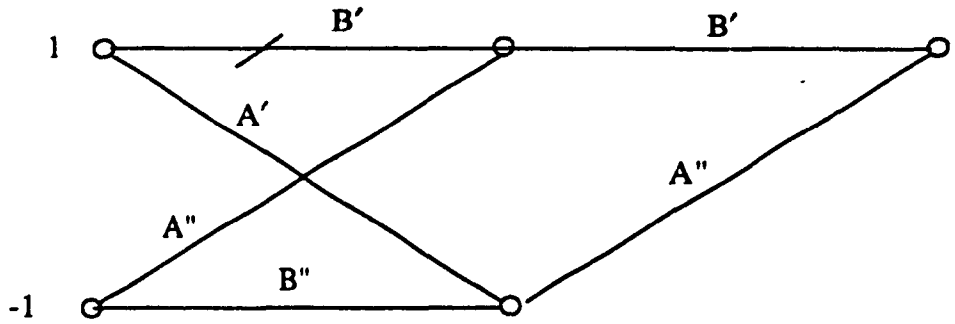

Fig. 3.4 Another trellis diagram for 2-state, n=3 trellis code.

Also assignment of uncoded bits, b_2 and b_3, to points of each subset is not specified in Fig. 3.3. If we assign the bits to points of the subset using Gray encoding, we have the TCM scheme with the following analytical representation,

$$x = x_1 + x_2 + = (b_1 b_4 + 2b_4) + (2b_2 b_3 + 4b_2). \qquad (3.5)$$

The values of x_1, and x_2 as a function of b_1, b_4, and b_2, b_3 respectively are given in Fig. 3.5.

Another possible code for a trellis diagram in Fig. 3.4 is,

$$x = x_1 + x_2 = (b_1 b_4 - 2b_4) + (2b_3 + 4b_2). \qquad (3.6)$$

Now we calculate the coding gain of this scheme over uncoded 8-AM modulation. Firstly, we discuss uncoded 8-AM modulation. As it has been discussed in 2.1, we consider uncoded 8-AM modulation as a specific one-state trellis coding as shown in Fig. 2.3. The smallest distance between 8 "parallel" transitions in the

x_2	b_2	b_3
6	1	1
2	1	-1
-6	-1	1
-2	-1	-1

x_1	b_1	b_4
3	1	1
-3	1	-1
1	-1	1
-1	-1	-1

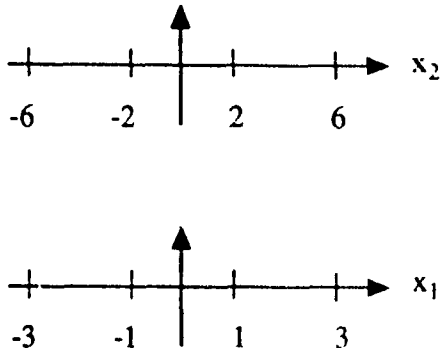

Fig. 3.5 Value of x_2 and x_1 depending on b_2, b_3 and b_1, b_4 respectively.

one-state trellis diagram is called the "free distance" of uncoded 8-AM modulation. So that $d^2_{free}=4$ and,

$$P_{av} = \frac{2(1^2+3^2+5^2+7^2)}{8} = 21 .$$

The free squared distance for the TCM scheme represented in (3.6) with trellis diagram shown in Fig. 3.3 is

$$d^2_{free} = d^2_{min}(A'', B'') + d^2_{min}(A', B'') = 2d_o^2 = 8,$$

and the average power is,

$$P_{av} = \frac{2[9^2+7^2+2(5^2+3^2+1^2)]}{16} = 25 .$$

So, the coding gain is,

$$G = 10 \log \frac{8/25}{4/21} = 2.253 \text{ dB}$$

which is 1.335 dB more than 0.918 dB coding gain of the conventional TCM scheme

using 16-AM constellation. More coding gain can be obtained by increasing the overlap between the subsets of the constellation. For example, a similar TCM scheme using an 8-AM constellation, i.e., a constellation with totally overlapped subsets gives a coding gain of 3 dB over conventional uncoded scheme. However, the problem with this code, which can also be considered as a limiting case of an asymmetric TCM scheme [21], is that it is catastrophic, it has infinite length error events with infinite Euclidean distance and nonzero probability. Note that, our code also has infinite length error events with distances not exceeding 8. However, in this case the probability of these error events is zero. This is due to the fact that with a probability of $p = 0.25$ a point chosen at random form A′(B′) is not a member of A"(B"). That is, two paths diverging from a given state are forced to reconverge to a common state on average p symbols [7], e.g., every 4 symbols in our scheme.

3.3.2 QAM Examples

In these examples, we use a 12-point 2-dimensional constellation. Again, the code has two states and n=3. The constellation together with corresponding set decomposition is shown in Fig. 3.6. Note that this constellation is formed by removing four outer points of a regular 16-point constellation and using four inner most points twice.

The trellis diagram is as that of the AM example shown in Fig. 3.3. A possible analytical representation is,

$$x = x_1 + x_2 = [b_1 + jb_4] + [(b_2 + b_3) + j(b_2 - b_3)]. \qquad (3.7)$$

Another one is,

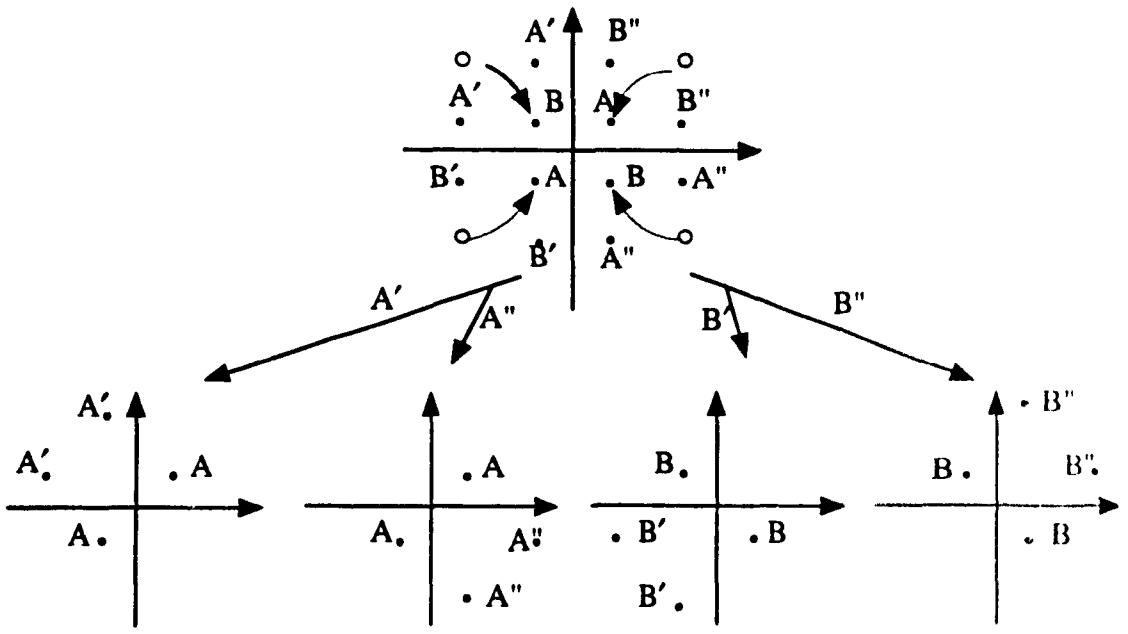

Fig. 3.6 12-QAM constellation and its set decomposition.

$$x = x_1 + x_2 = [-b_1 + jb_4] + [(b_2 - b_3) + j(b_2 + b_3)], \qquad (3.8)$$

and the trellis diagram for this code is same as which is shown in Fig. 3.4. The values of x_1, and x_2 as a function of b_1, b_4, and b_2, b_3, respectively are given in Fig. 3.7 for (3.7). Here, first b_1 and b_4 select $c \in \{\pm 1, \pm j\}$ and then, b_2 and b_3 choose a point from the subset $c + \{\pm 2, \pm j2\}$. The minimum squared distance between parallel path is 8. For the minimum error event (of length two, see Fig. 3.3 and Fig. 3.4), we have,

$$d^2_{free} = 8,$$

and the average power is,

$$P_{av} = \frac{8[1^2+3^2+1^2+1^2]}{16} = 6 .$$

x_2	b_2	b_3	x_1	b_1	b_4
2	1	1	1+j1	1	1
j2	1	-1	1-j1	1	-1
-j2	-1	1	-1+j1	-1	1
-2	-1	-1	-1-j	-1	-1

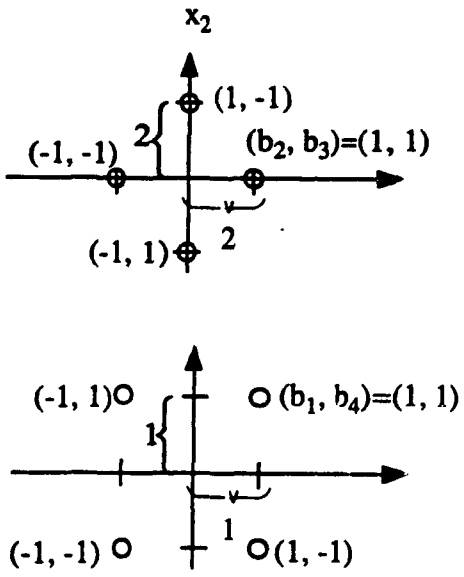

Fig. 3.7 Value of x_2, x_1 depending on b_2, b_3, and b_1, b_4 respectively.

For a corresponding 8-PSK, with average power of 1,

$$d^2_{free} = 0.765^2 = 0.585.$$

And therefore, the coding gain of our scheme over uncoded 8-PSK is

$$G = 10 \log \frac{8/6}{0.585/1} = 3.58 \text{ dB}.$$

The coding gain over an uncoded 8-AMPM, with a minimum distance of 8 and an average power of 10, is

$$G = 10 \log \frac{8/6}{8/10} = 2.22 \text{ dB}.$$

This is 0.46 dB more than the coding gain of a TCM scheme with a 16-point QAM constellation.

For the QAM example, p=0.5, i.e., a point chosen at random from the subset A′(B′) has a fifty percent chance of not being in A"(B"). That is, two paths diverging from a common state are forced to reconverge to a common state on average every two symbols.

An extra 0.79 dB coding gain can be obtained using a constellation with totally overlapped subsets, i.e., using 8-point QAM constellation shown in Fig. 3.8. Then, the coding gain of our scheme over uncoded 8-PSK is achieved to 4.36 dB, and the coding gain over an uncoded 8-AMPM will be 3 dB. That is because of the reduction of the average power P_{av}. However, by doing so, we will have p 0, and therefore, the code will be catastrophic.

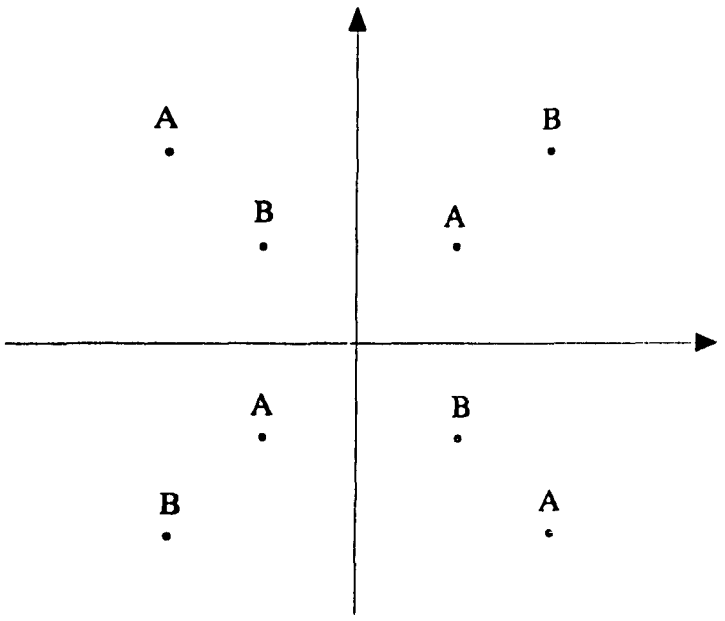

Fig. 3.8 A 8-point QAM constellation.

3.4 Further Analysis of the New Codes

Calderbank and Mazo have developed the analytic description of the trellis coded modulation using only one step instead of Ungerboeck's two steps; finding an underlying convolutional code and then mapping the output coded bits onto channel signals according to rules referred to as "mapping by set partitioning". The advantage of their work is to simplify the procedure of TCM design. Author has analysed the optimum codes which are shown in Calderbank and Mazo's paper [10], and then found that these codes are also suitable to the Ungerboeck's "mapping by set partitioning rule". Now, we can get this conclusion:

Even though that Ungerboeck and Calderbank, Mazo have used different methods to search the optimum TCM codes, the results of searching optimum codes are the same; Certainly, some results in Calderbank and Mazo's codes are slightly better than Ungerboeck's codes. But this is only because of the asymmetry of signal constellation. If they use the same signal constellation with Ungerboeck's codes, the results should be the same as that of Ungerboeck's codes.

The important advantage in Calderbank and Mazo's method and also in our new codes is that optimum TCM codes can be simply searched using only one step. But for further understanding of the new code, it is very helpful to analyse the result of searching optimum codes using the graphical method rather than analytic one.

3.4.1 The Set Decomposition of the Partially Overlapped Signal Constellation

It is helpful if we compare the set decomposition of our new signal

constellation to Ungerboeck's "set partitioning" rule, the set decomposition of the partially overlapped signal constellation is similar to that set partitioning of the traditional constellation. From Fig. 3.2 and Fig.3.6, it is shown that the first step of set decomposition is the same as Ungerboeck's. Subsets at this level are not at all overlapped. The intra-set distance is maximally increased. The second step of set decomposition is different. The subsets at this level which are from the same subsets at level 1 are partially overlapped and partially disjointed. Also the intra-set distance at this level is the same as the one at level 1. But the important thing is that the number of subsets and the number of points in each subset are the same as those in set partitioning at this level of the traditional 2^{n+1}-point constellations if this new signal constellation is for input n bits per symbol TCM.

3.4.2 The Overlapping Property of the New Signal Constellation

More coding gain can be obtained by increasing the overlap between the subsets of the constellation which has been discussed above. If there is no overlap between the subsets, that is the traditional signal constellation set partition which is used to use for TCM before. If a TCM scheme use the signal constellation with totally overlapped subsets, it will give a significant coding gain. However, as previously mentioned in section 3.3, this type of limiting case results in a catastrophic trellis code In practice, one would not use this limiting case. It just be considered as theoretical research. How much extent of overlap is the optimum choice for the new constellation? The main consideration is to reduce the average power of the constellation and at the same time to keep the number of points in each subset and the number of subsets the same as that of traditional constellation set partitioning. So we like to repeat using the

low-power points instead of the high-power points. In other words, we like to have as much extent of overlap of subsets as possible. However, in order to avoid a catastrophic trellis code, we must keep the parts of subsets which are disjointed. For AM constellation, all points have different powers. So, subsets are required only to keep the smallest number of high-power points for the disjointed part to reduce P_{av}. But for QAM constellation, some points of the constellation have the same power. Also for the convenience of analytic representation, the subsets should be needed to keep the number of points for disjointed parts which make the average power the smallest.

3.4.3 The Mapping Rule for Input Bits to Output Channel Signals

The mapping rule for input bits to output channel signals can be considered to be similar to Ungerboeck's codes, but they are different. (a) Even though there is no binary rate $n'/(n'+1)$ convolutional encoder (here $n'= 1$), n' input bits (b_1) and the bits (b_4) for the state are still used to select one of $2^{n'+1}= 4$ subsets. The function is the same with Ungerboeck's codes; (b) the remaining $n - n'$ bits (if total input bits are n bits per symbol) determine one of the $2^{n-n'}$ signals within the selected subset; (c) all signals in the final level subsets should be equiprobable. It should be noted here that because some subsets are partially overlapped with lower-power points, so that some low-power signals are used twice as often as other high-power signals.

Finally, we compare our codes with Calderbank and Mazo's in graphical analysis. According to the above discussion, for input bits $b_1...b_n$ per symbol, we let transmitted channel signal λ is composed of x_1 and x_2. x_2 is the function of the most recent bit b_1 and the previous bit b_{n+1} which is for state and x_1 is the function of $b_2...b_n$. The basic structure of subsets at final level is formed by the different values of x_2. And

the values of x_1 give the several positions to shift the basic structure and then to form the several subsets. For AM TCM the difference between Calderbank and Mazo's codes and our newly designed codes is the values of x_1. They give the values of x_1 to be different enough from each other to make the subsets to be unoverlapped at all. But the values of x_1 according to the new design rule make some of the subsets partially overlapped each other. The principles for QAM are the same and are not necessary to be discussed again.

3.5 Multi-dimensional Trellis Codes

To immune to a carrier phase offset of θ in the receiver, a rotationally invariant coherent modulation(/demodulation) scheme is desirable. In an uncoded coherent modulation scheme, rotational invariance can be easily achieved by using a differential encoding(/decoding) technique, which maps (recovers) input data into (from) the phase difference between the current and previous (estimated) signal points.

For coded coherent modulation schemes, the issue of rotational invariance is more complicated. A general structure for coded coherent modulation schemes is shown in Fig. 3.9 where n is the number of information bits. x_n and y_n are the coordinates of a 2-D signal point. In the receiver, the coherent demodulator outputs a received signal point $(\hat{x}_n, \hat{y}_n)$ which is noisy and possibly rotated by a phase angle equal to a phase ambiguity of the signal constellation. To further make the coded modulation scheme rotationally invariant, a necessary condition is that a rotation of any valid sequence of signal points by any phase ambiguity of the constellation results in another valid sequence. In addition, if a rotation of the signal points results in, at the output of the channel decoder, a sequence of information bits different from what is transmitted, then

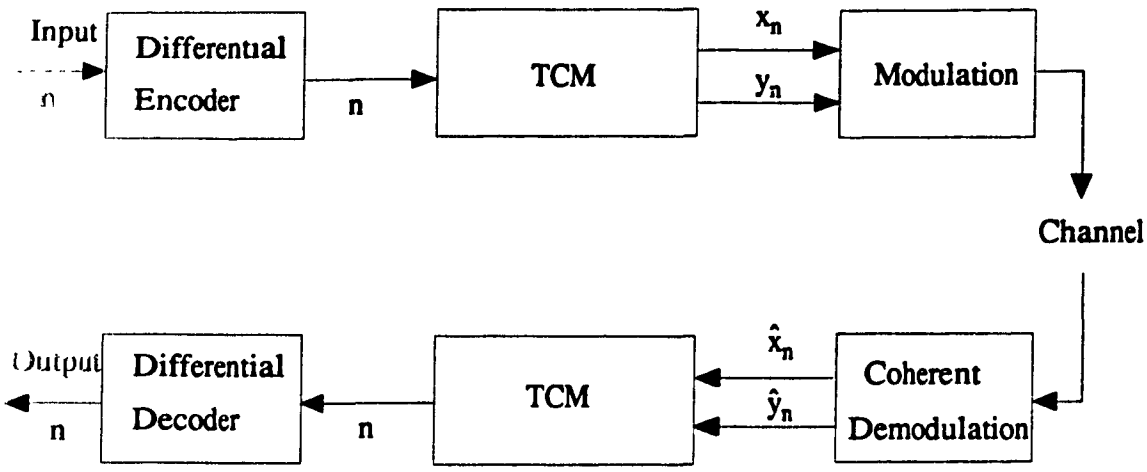

Fig. 3.9 General structure for rotationally invariant coded coherent modulation schemes.

a differential encoding technique may be used to recover the transmitted data irrespective of the rotation.

TCM schemes with multi-dimensional constellations, where the constellation has more than two dimensions and may be formed by concatenating a few (constituent) 2-D constellations, have been shown more advantageous than TCM schemes with 2-D constellations in the case where the 2-D constellation is a QAM [5, 7, 16, 17, 19]. The important advantage on rotational invariance of trellis-coded multi-dimensional QAM schemes is the following; It was easier to make a multidimensional TCM scheme rotationally invariant than a 2-D scheme [17]. The reason is that some or all of the phase ambiguities of a multi-dimensional constellation may be removed by a careful partition of the multi-dimensional constellation without the involvement of the trellis encoder. So that this larger signal spacing can make TCM systems less sensitive to phase offset problem. Another advantage as before we talked in 2.2 is that; Because the multi-dimensional scheme introduces fewer redundant bits than the 2-D scheme, it has

a significantly smaller constituent 2-D constellation and a better trade off between complexity and performance.

In this section, we extend our research to high-dimensional signal constellation. An TCM scheme with 4-D constellation, where the constellation is formed by concatenating a two (constituent) 2-D (QAM) partially overlapped constellations as shown in 3.3.2 (QAM example), is shown. Through this example the procedure of the design of TCM with high-dimensional partially overlapped constellation is illustrated. In keeping with our previous work, in our example, the 12-point partially overlapped QAM constellation shown in Fig. 3.6 is used for 2-D constellation, n=3 bits is transmitted per constituent QAM signal, and hence 2n=6 bits per 4-D signal. A two-state case is considered. People also can write a program to search for the optimum codes. However, this simple example can be designed by hand. The following mapping rule can be very helpful for people to understanding clearly the code design procedure. Because there are 32 parallel transitions between any two states and there are four groups of such parallel transitions between states, so that this 4-D signal constellation should have 32×4=128 points. From Fig. 3.6, we can see that the 2-D (QAM) constellation is divided to four subsets A′, A", B′, B". Then, the partition of 4-D 128-point constellation can be shown in Fig. 3.10.

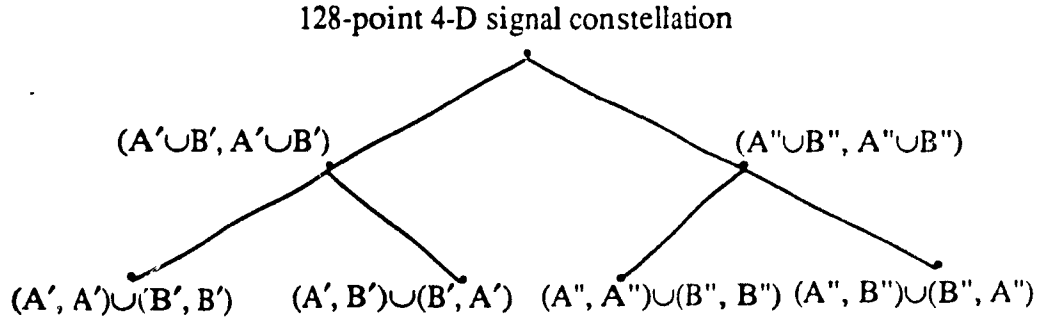

Fig. 3.10 The partition of 4-D 128-point constellation.

A trellis diagram for the two-state, n=3, trellis code with 2k-D (k=2) constellation is shown in Fig. 3.11.

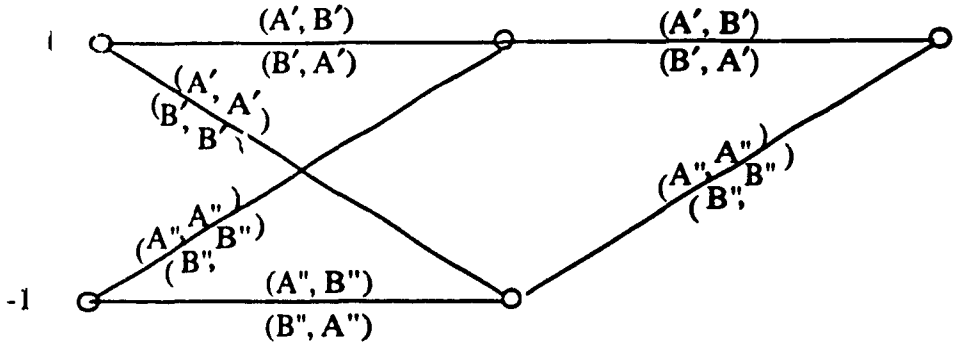

Fig. 3.11 Trellis diagram for a 2-state, n = 3, k = 2 trellis code.

First bit, b_1, specifies the state. Each of four groups of parallels between states can be specified by b_7 and b_1. The 32-parallel transitions between any two states is specified by b_2 - b_6. A possible analytical representation is

$$x = (x^{(1)}, x^{(2)}) = (x_1^{(1)} + x_2^{(1)}, x_1^{(2)} + x_2^{(2)}) \qquad (3.9)$$

$$= \{[-b_7 + jb_2b_7] + [(b_3 + b_4) + j(b_3 - b_4)], [-b_7 - jb_2b_1] + [(b_5 + b_6) + j(b_5 - b_6)]\}.$$

The minimum squared distance between parallel paths is 8. From Fig. 3.11, it is easy to see that $d^2_{free} = 8$. The average power is $P_{av} = 6$. The coding gain is the same as the corresponding TCM with 2-D (QAM) partially overlapped signal constellation. But, as we discussed earlier, trellis-coded modulation schemes using multi-dimensional constellation are more advantageous than the usual two-dimensional schemes because the multi-dimensional schemes can be more easily made rotationally invariant; and comparing with 2-D TCM schemes, this method is in less signal redundancy in the constituent 2-D signal sets.

3.6 Conclusion

In this chapter, we obtain a new analytically described trellis coded modulation scheme using partially overlapped signal sets. By presenting two simple two-state examples, it was shown that using TCM schemes with partially overlapped signal constellations, one can obtain coding gains superior to those of conventional TCM schemes. The coding gain of the proposed schemes was also compared to those of the limiting cases of totally overlapped constellations. The latter, while providing higher coding gains, can not be used in practice, due to their catastrophic behaviour. Then, the further analysis of our new codes comparing with Ungerboeck's method is described. Finally, through an example we extended our work to TCM with high-dimensional signal constellation. The important advantages of TCM with multi-dimensional constellation are discussed.

Chapter 4 Performance Analysis

In this chapter, we first introduce a performance analysis method for TCM which is used in many papers about TCM [21, 26]. Then, this method is used for performance analysis of our new codes. To keep the continuity of our work, two examples shown in chapter 3 are analyzed here according to the bit error probability criterion. The results are superior to those of conventional TCM schemes.

4.1 The Method of Performance Analysis

At time k, for every n information bits per symbol, the rate n/(n+1) trellis coded modulation transmits a signal x_k chosen from a 2^{n+1} signal set. We denote a coded symbol sequence of length N by $\mathbf{x} = (x_1, x_2, ..., x_N)$. Then, x_k is considered as the kth element of $\mathbf{x}$ and is a nonlinear function of the n information input bits denoted by u_k and the v preceding bits for states denoted by s_k, i.e.,

$$x_k = f(s_k, u_k). \tag{4.1}$$

Then, the next state s_{k+1} at time k+1 is a nonlinear function of the present state s_k and the input u_k. In mathematical terms,

$$s_{k+1} = g(s_k, u_k). \tag{4.2}$$

Corresponding to x, the channel outputs the sequence $\mathbf{r} = (r_1, r_2, ..., r_N)$, where the kth element r_k, representing the output at time k, is given by,

$$r_k = x_k + n_k, \tag{4.3}$$

where n_k is a sample of a zero mean Gaussian noise process with variance δ^2.

To find the average bit error probability performance of the Viterbi decoder, we must first find the pairwise error probability $p(\mathbf{x} \to \hat{\mathbf{x}})$ between the coded sequence $\mathbf{x} = \{x_k\}$ and the estimated sequence $\hat{\mathbf{x}} = \{\hat{x}_k\}$. Assume that $|x_k|^2 = 1$. Then, using the Bhattacharyya bound [38], we have,

$$p(\mathbf{x} \to \hat{\mathbf{x}}) \le D^{\Delta}, \ \Delta = \sum_k \delta^2(S_k, U_k) \tag{4.4}$$

where,

$$\delta^2(S_k, U_k) \stackrel{\Delta}{=} \left| f(s_k, u_k) - f(\hat{s}_k, \hat{u}_k) \right|^2 \tag{4.5}$$

with $\hat{s}_k$ and $\hat{u}_k$ are the estimates of the state of the decoder and the information symbol, respectively. Also, D is the Bhattacharyya parameter which in this case is given by,

$$D = \exp(-1/8\delta^2). \tag{4.6}$$

The parameter D of (4.6) can be related to the system bit energy-to-noise ratio E_b/N_o by first recognizing that $\delta^2 = (2E_s/N_o)^{-1}$ where E_s is the M-ary symbol energy. Since, for n/(n+1) trellis coding, n input bits of energy E_b produce n+1 code symbols, which in turn result in a single M-ary symbol of energy E_s, then clearly $E_s = nE_b$. Using these observations in (4.6) gives the desired relation for D in terms of E_b/N_o, namely,

$$D = \exp(-nE_b/4N_o). \tag{4.7}$$

The pair state S_k and the pair-information symbol U_k [38, 39] are defined as,

$$S_k \triangleq (s_k, \hat{s}_k), \quad U_k \triangleq (u_k, \hat{u}_k). \tag{4.8}$$

We are in a correct pair state when $\hat{s}_k = s_k$ and in an incorrect pair state when $\hat{s}_k \neq s_k$.

In terms of the above definitions, it can be shown by analogy with the results in [40] that the average bit error probability P_b is upper bounded by,

$$P_b \leq \frac{1}{n} \frac{d}{dz} T(D, z)\Big|_{z=1} \tag{4.9}$$

or obtain an even tighter bound,

$$P_b \leq \frac{1}{2n} \operatorname{erfc}\left(\sqrt{\frac{nE_b}{N_o} \frac{d_{free}^2}{4}} \right) D^{d_{free}^2} \frac{d}{dz} T(D, z)\Big|_{z=1}, \tag{4.10}$$

where erfc(x) is the complementary error function, d^2_{free} is the squared free distance of the trellis code, and T(D, z) is the transfer function of its pair-state diagram,

$$T(D, z) = \frac{1}{m} V^t [I-A]^{-1} W, \tag{4.11}$$

and,

$$\frac{d}{dz} T(D, z)\Big|_{z=1} = \frac{1}{m} V^t [I - A(1)]^{-1} A'(1) [I - A(1)]^{-1} W, \tag{4.12}$$

m is the number of correct states. The vectors V and W have dimension $m^2 + m$ with elements taking on values 1 and 0. A(z) is an $(m^2 + m) \times (m^2 + m)$ pair-state transition matrix with elements

$$a(S_k, S_{k+1}) = \begin{cases} \sum_{u_k \in U_k} \frac{1}{2^n} z^{w(U_k)} D^{\delta^2(S_k, U_k)}; & \text{if } U_k \text{ is nonempty set} \\ 0; & \text{otherwise} \end{cases} \tag{4.13}$$

where,

$$U_k = \{(u_k, \hat{u}_k) \mid (\hat{s}_k, \hat{u}_k) \neq (s_k, u_k),\ S_k \notin S_d,\ S_{k+1} = G(S_k, U_k) \notin S_t\} \tag{4.14}$$

in which S_t and S_d are sets of all true and dummy correct pair states, respectively, and,

$$G(S_k, U_k) \triangleq (g(s_k, u_k), g(\hat{s}_k, \hat{u}_k)). \tag{4.15}$$

Finally, the free Euclidean distance of the code [39] is,

$$d^2_{free} = \lim_{D \to 0} \log_2 \frac{T(2D, 1)}{T(D, 1)} \quad . \tag{4.16}$$

Note that if $|x_k|^2 \neq 1$, then in (4.10)

$$d^2_{free} = \frac{d^2_{min}}{|x_k|^2} \quad , \tag{4.17}$$

and in (4.13),

$$\delta^2 (S_k, U_k) \triangleq \frac{|f(s_k, u_k) - f(\hat{s}_k, \hat{u}_k)|^2}{|x_k|^2} \quad . \tag{4.18}$$

In previous chapters, our entire discussion has focussed on performance gain as measured by improvement in minimum free distance of the trellis code. In the limit as the system bit error probability becomes arbitrarily small, this measure is equivalent to the improvement in required bit energy-to-noise spectral density ratio for a given average bit error probability. So that maximizing d^2_{free} is synonymous with minimizing the average bit error probability. As we have discussed earlier, in our new codes optimization of our partially overlapped condition produces signal sets wherein the signal subsets are overlapped at all. This results in the smallest possible $|x|^2$ and the largest possible $d^2_{free} = d^2_{min} / |x_k|^2$. However, this results in a catastrophic trellis code.

Our code will also have infinite length error events with distances not exceeding d^2_{free}. However, in this case the probability of these error events is zero. In next two sections, we will discuss the bit error probability performance of our codes.

4.2 Rate 3/4, 2-state Code with 10-AM Partially Overlapped Constellation Comparing with Corresponding TCM with 16-AM Constellation

From a practical standpoint, one is often interested in the reduction of bit energy-to-noise spectral density ratio for a given average bit error probability. Previous results [10, 11, 21, 40, 41] on conventional trellis coding showed that such reductions were possible. Using pair-state diagrams and upper bounds on bit error probability computed from the transfer function of these diagrams, we shall now determine the magnitude of these performance gains for our new codes. Without going into great detail, it has been shown that a tight upper bound on the bit error probability of trellis codes is given by (4.10).

First, consider the rate 3/4, 2-state trellis code with 10-AM partially overlapped constellation (shown in Fig. 3.2) and the following analytical representation,

$$x = x_1 + x_2 = (b_1 b_4 + 2b_4) + (2b_2 b_3 + 4b_2).$$

The relationship between input bits b_1, b_2, b_3 and the output signal x is listed in table (4.1). The trellis diagram is as in Fig. 3.3 and the equivalent pair-state diagram for computing T(D, z) is shown in Fig. 4.1. In Fig. 4.1, the branches are labeled with a gain of the form,

Table 4.1 The relationship between input bits and output signals for 10-AM partial overlapped constellation

b_1 b_2 b_3	1 -1 1	1 -1 -1	1 1 -1	1 1 1
$x \rightarrow (A')$	-9	-5	-1	3
b_1 b_2 b_3	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
$x \rightarrow (A'')$	-5	-1	3	7
b_1 b_2 b_3	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
$x \rightarrow (B')$	-7	-3	1	5
b_1 b_2 b_3	1 -1 1	1 -1 -1	1 1 -1	1 1 1
$x \rightarrow (B'')$	-3	1	5	9

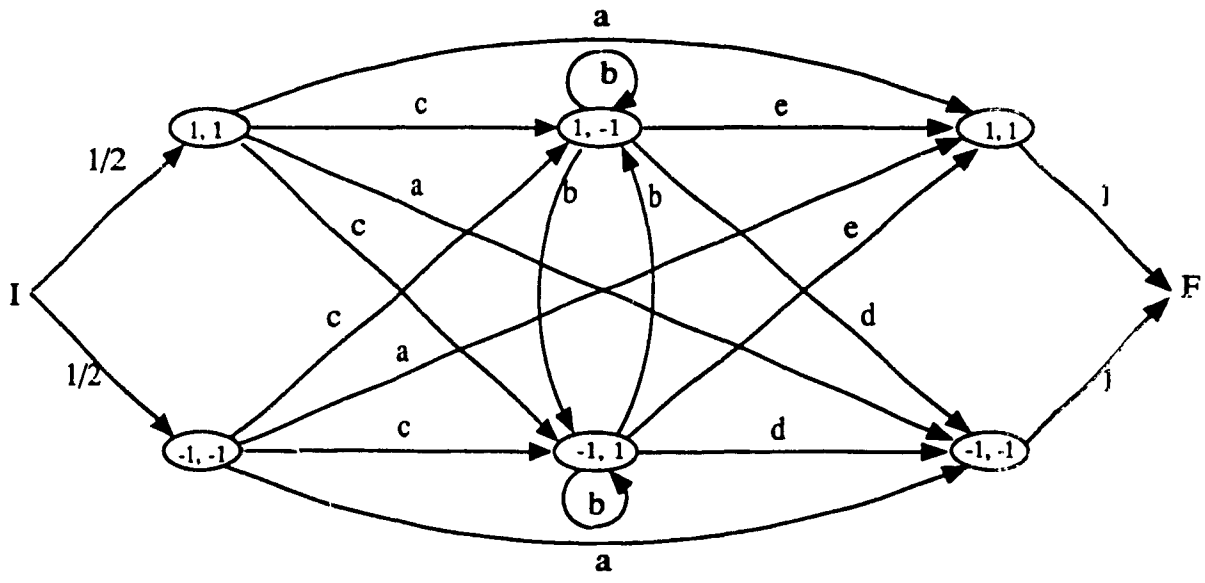

Fig. 4.1 Pair-state diagram corresponding to Fig. 3.3.

$$G = \frac{1}{2^n} z^{\Omega} D^{\delta^2}. \qquad (4.19)$$

Here z is an index, n is input bits per symbol, Ω is the Hamming distance between input bit sequences and δ^2 is the squared Euclidean distance between output signals for the transition between the pair states (if $P_{av} \neq 1$, which should be divided by P_{av}) In accordance with (4.19), then in Fig. 4.1,

$$a = 1/8[(6D^{16/Pav} + 2D^{144/Pav})z + 4D^{64/Pav}z^2]$$

$$b = 1/8[4D^{16/Pav}z + (3 + 4D^{64/Pav} + D^{256/Pav})z^2 + 2(D^{16/Pav} + D^{144/Pav})z^3]$$

$$c = 1/8[4D^{4/Pav}z + (3D^{4/Pav} + 3D^{36/Pav} + D^{100/Pav} + D^{196/Pav})z^2 + 2(D^{36/Pav} + D^{100/Pav})z^3]$$

$$d = z^{-1}c = 1/8[4D^{4/Pav} + (3D^{4/Pav} + 3D^{36/Pav} + D^{100/Pav} + D^{196/Pav})z + 2(D^{36/Pav} + D^{100/Pav})z^2]$$

$$e = 1/8[4D^{36/Pav} + (3D^{4/Pav} + 3D^{100/Pav} + D^{36/Pav} + D^{324/Pav})z + 2(D^{4/Pav} + D^{196/Pav})z^2], \qquad (4.20)$$

where for 10-AM partially overlapped constellation, $P_{av} = 25$. The transfer function of Fig. 4.1 is easily computed as,

$$T(D, z) = 2(a + \frac{c(e + d)}{1 - 2b}). \qquad (4.21)$$

For $n = 3$, $d^2_{free} = 8/25 = 0.32$, $D = e^{-nEb/4No} = e^{-0.75Eb/No}$, the upper bound on P_b is,

$$P_b \le \frac{1}{6}\,\mathrm{erfc}\left(\sqrt{0.24\,\frac{E_b}{N_o}}\right) e^{0.24Eb/No} \cdot H(D)\ , \qquad (4.22)$$

where,

$$H = \left.\frac{dT(D, z)}{dz}\right|_{z=1}\ , \qquad (4.23)$$

so that Fig. 4.4 is an illustration of the upper bound of (4.22). For the purpose of comparison, let us discuss the upper bound on P_b for the corresponding rate 3/4, 2-state, 16-AM TCM scheme. Using Ungerboeck's method, the signal constellation, the set partitioning and the 2-state trellis diagram are shown in Fig. 4.2 and Fig. 4.3,

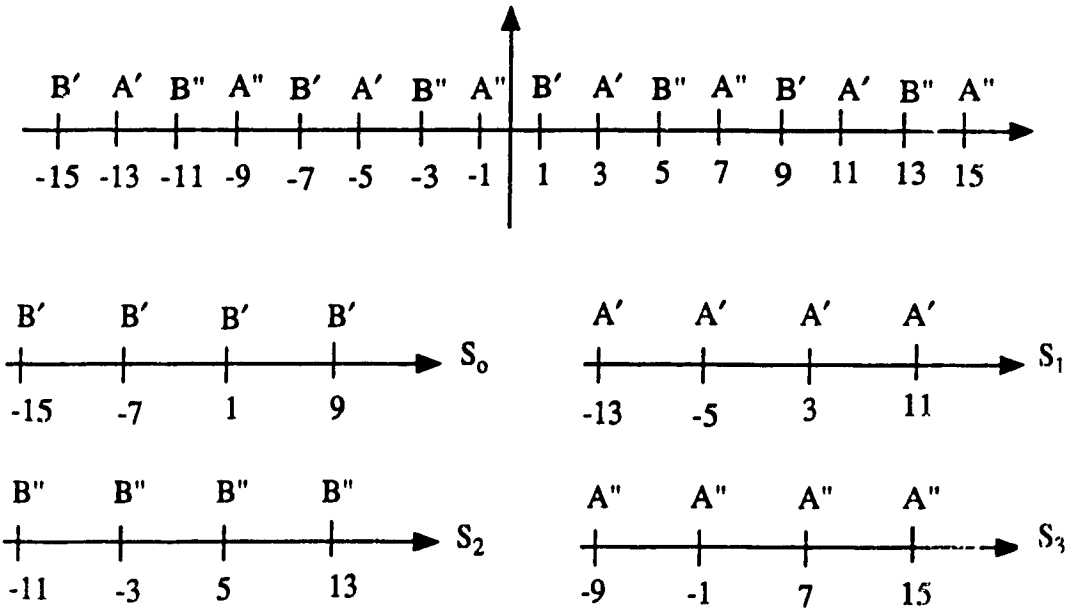

Fig. 4.2 16-AM; signal constellation and its set partitioning.

respectively. The corresponding pair-state diagram for computing T(D, z) is the same

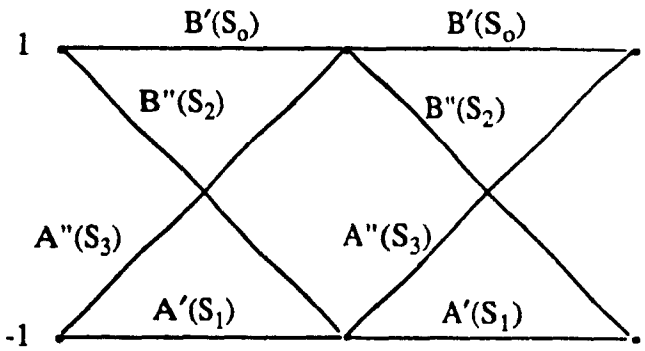

Fig. 4.3 2-state trellis diagram corresponding to Fig. 4.2.

as the one in Fig. 4.1.

If we have the relationship between input bits and output signal as shown in table 4.2, then,

$$a = 1/4[(3D^{64/Pav} + D^{576/Pav})z + 2D^{256/Pav}z^2]$$

$$b = 1/8[4D^{4/Pav}z + (3D^{36/Pav} + 3D^{100/Pav} + D^{484/Pav} + D^{676/Pav})z^2 + 2(D^{196/Pav} + D^{324/Pav})z^3]$$

$$c = 1/8[4D^{16/Pav}z + (3D^{16/Pav} + 3D^{144/Pav} + D^{400/Pav} + D^{784/Pav})z^2 + 2(D^{144/Pav} + D^{400/Pav})z^3]$$

$$d = 1/8[4D^{4/Pav} + (3D^{36/Pav} + 3D^{100/Pav} + D^{484/Pav} + D^{676/Pav})z + 2(D^{196/Pav} + D^{324/Pav})z^2]$$

$$e = 1/8[4D^{36/Pav} + (3D^{4/Pav} + 3D^{196/Pav} + D^{324/Pav} + D^{900/Pav})z + 2(D^{100/Pav} + D^{484/Pav})z^2] \qquad (4.24)$$

where for 16-AM constellation $P_{av} = 85$. For $n = 3$, $d^2_{free} = 20/85 = 4/17$, $D = e^{-nE_b/4N_o} =$

Table 4.2 The relationship between input bits and output signals for 16-AM signal constellation.

b_1 b_2 b_3	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
$x \rightarrow (A')$	-13	-5	3	11
b_1 b_2 b_3	1 -1 1	1 -1 -1	1 1 -1	1 1 1
$x \rightarrow (A'')$	-9	-1	7	15
b_1 b_2 b_3	1 -1 1	1 -1 -1	1 1 -1	1 1 1
$x \rightarrow (B')$	-15	-7	1	9
b_1 b_2 b_3	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
$x \rightarrow (B'')$	-11	-3	5	13

$e^{-0.75Eb/No}$, the upper bound on P_b is,

$$P_b \leq \frac{1}{6}\operatorname{erfc}\left(\sqrt{\frac{3}{17}\frac{E_b}{N_o}}\right) e^{3/17(Eb/No)} \cdot H(D) \ , \qquad (4.25)$$

where H(D) is expressed as in (4.23). So that the upper bounds on P_b of (4.25) are shown in Fig. 4.4 to compare with that of (4.22). Both upper bounds on P_b are also shown in table 4.3. We observe from these results that, over the range of E_b/N_o illustrated, our code with the 10-AM partially overlapped constellation is better than the conventional trellis code with 16-AM constellation. The improvement increases with the increase in SNR.

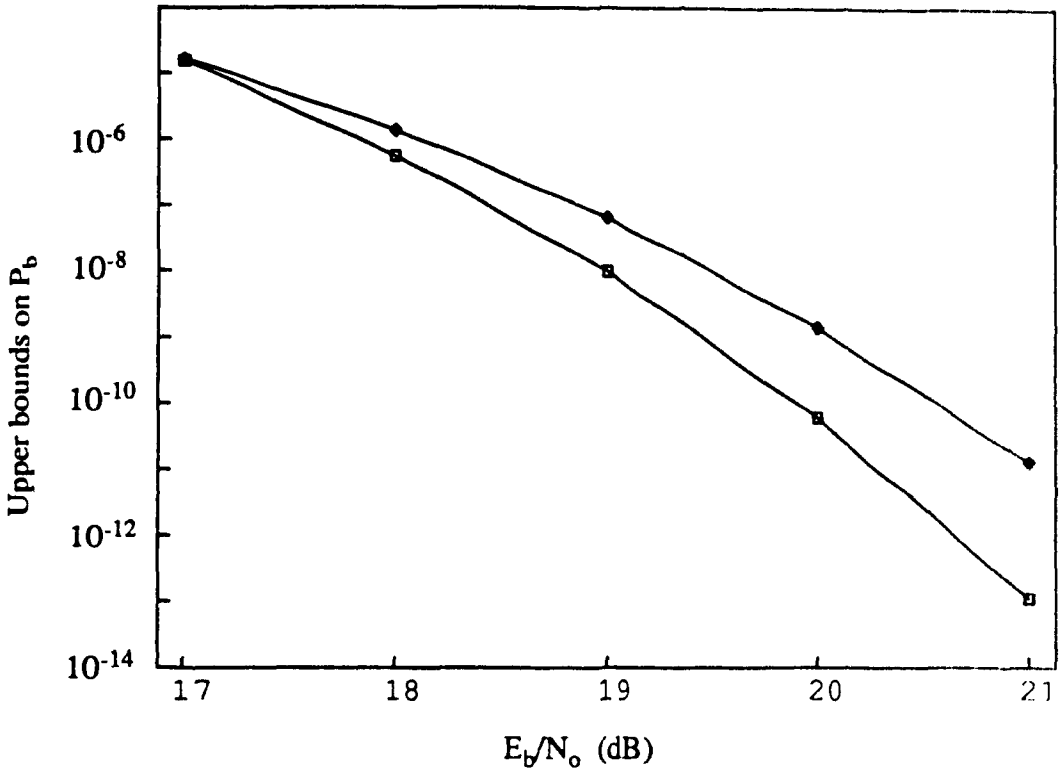

Fig. 4.4 A comparison of the performance of two trellis coded AM modulation (◇), trellis code with 16-AM; (□), trellis code with 10-AM partially overlapped constellation.

Table 4.3 The upper bounds on P_b for two TCM with AM modulation

P_b \ E_b/N_o(dB)	17	18	19	20	21
10-AM	1.35×10^{-5}	5.39×10^{-7}	9.66×10^{-9}	6.21×10^{-11}	1.11×10^{-13}
16-AM	1.67×10^{-5}	1.35×10^{-6}	6.32×10^{-8}	1.44×10^{-9}	1.31×10^{-11}

4.3 Rate 3/4, 2-state Code with 12-QAM Partially overlapped constellation Comparing with Corresponding TCM with 16-QAM constellation

An example of rate 3/4, 2-state trellis code with 12-QAM partially overlapped constellation which is shown in chapter 3.3.2 is discussed here. A 12-QAM signal constellation is shown in Fig. 3.6. Its analytical representation between input bits, bits for state and output channel signal is,

$$x = x_1 + x_2 = [b_1 + jb_4] + [(b_2 + b_3) + j(b_2 - b_3)].$$

The relationship between input bits and output signals is shown in table 4.4. And the

Table 4.4 The relationship between input bits and output signals for 12-QAM partially overlapped constellation.

b_1 b_2 b_3				
$x \rightarrow (A')$	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
	-1-j	-3+j	-1+j3	1+j
b_1 b_2 b_3				
$x \rightarrow (A'')$	1 -1 1	1 -1 -1	1 1 -1	1 1 1
	1-j3	-1-j	1+j	3-j
b_1 b_2 b_3				
$x \rightarrow (B')$	- 1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
	-1-j3	-3-j	-1+j	1-j
b_1 b_2 b_3				
$x \rightarrow (B'')$	1 -1 1	1 -1 -1	1 1 -1	1 1 1
	1-j	-1+j	1+j3	3+j

corresponding 2-state trellis diagram is as in Fig. 4.5. Then the equivalent pair-state

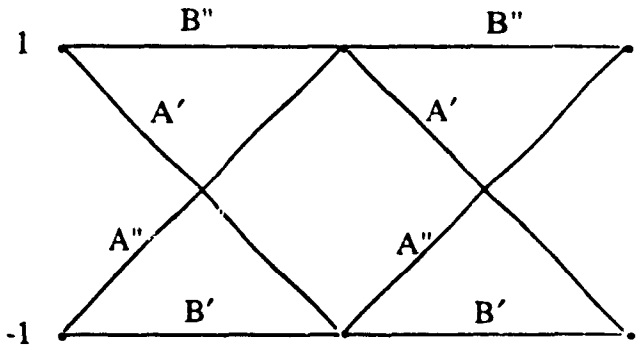

Fig. 4.5 A 2-state trellis diagram.

diagram for computing T(D, z) is shown in Fig. 4.6, where,

$$a = D^{8/Pav}z + (1/2)D^{16/Pav}z^2$$

$$b = 1/4[2D^{8/Pav}z + (1 + 2D^{16/Pav} + D^{32/Pav})z^2 + (D^{8/Pav} + D^{40/Pav})z^3]$$

$$c = 1/8[4D^{4/Pav} + 4(D^{4/Pav} + D^{20/Pav})z + (D^{4/Pav} + 2D^{20/Pav} + D^{36/Pav})z^2]$$

$$d = 1/8[4D^{4/Pav}z + 4(D^{4pav} + D^{20/Pav})z^2 + (D^{4/Pav} + 2D^{20/Pav} + D^{36/Pav})z^3] \quad (4.26)$$

For 12-QAM partially overlapped constellation shown in Fig. 3.6, $P_{av} = 6$ From Fig. 4.6 we can easily calculate the transfer function,

$$T(D, z) = 2(a + \frac{2cd}{1-2b}). \quad (4.27)$$

For $n = 3$, $d^2_{free} = 8/6 = 4/3$, $D = e^{-nEb/4No} = e^{-0.75Eb/No}$, the upper bound on P_b is,

$$P_b \le \frac{1}{6}\,\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_o}}\right)\, e^{(Eb/No)} \cdot H(D)\ , \qquad (4.28)$$

where H(D) is expressed as in (4.23).

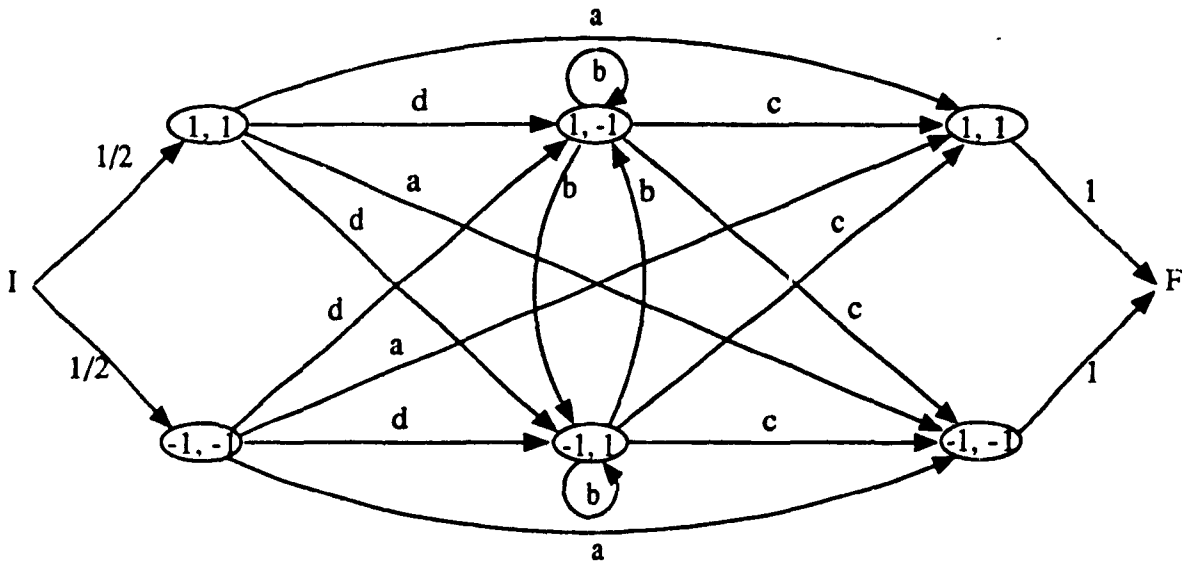

Fig. 4.6 Pair-state diagram corresponding to Fig. 4.5.

The upper bounds of (4.28) are shown in Fig. 4.9. Now we evaluate the performance of the conventional rate 3/4, 2-state, 16-QAM TCM scheme. The signal constellation and its set partitioning is shown in Fig. 4.7 and the 2-state trellis diagram is as in Fig. 4.8 according to Ungerboeck's method. The pair-state diagram is as in Fig. 4.6. Suppose we have the relationship between input bits and output signals as shown in table 4.5, then,

$$a = D^{16/Pav}z + (1/2)D^{32/Pav}z^2$$

$$b = 1/4[2D^{4/Pav}z + (D^{4/Pav} + 2D^{20/Pav} + D^{36/Pav})z^2 + (D^{20/Pav} + D^{52/Pav})z^3]$$

$$c = 1/4[2D^{4/Pav} + (D^{4/Pav} + 2D^{20/Pav} + D^{36/Pav})z + (D^{20/Pav} + D^{52/Pav})z^2]$$

$$d = 1/8[4D^{8/Pav}z + 4(D^{8/Pav} + D^{40/Pav})z^2 + (D^{8/pav} + 2D^{40/Pav} + D^{72/Pav})z^3] , \tag{4.29}$$

where P_{av} = 10 for 16-QAM constellation. The transfer function is the same as in (4.27). Then for n=3, d^2_{free} = 12/10 = 1.2, $D = e^{-nEb/4No} = e^{-0.75Eb/No}$, the upper bound on P_b is,

$$P_b \leq \frac{1}{6}\operatorname{erfc}\left(\sqrt{0.9\frac{E_b}{N_o}}\right) e^{0.9(Eb/No)} \cdot H(D) , \tag{4.30}$$

here, H(D) is expressed as in (4.23).

The upper bound of (4.30) is shown in Fig. 4.9 to compare with that of (4.28). Also the comparison of the upper bounds of the two codes is listed in table 4.6. From these results it is easy to observe that: over the range of E_b/N_0 illustrated, our 12-QAM partially overlapped constellation code is also better than the conventional trellis code with 16-QAM signal constellation.

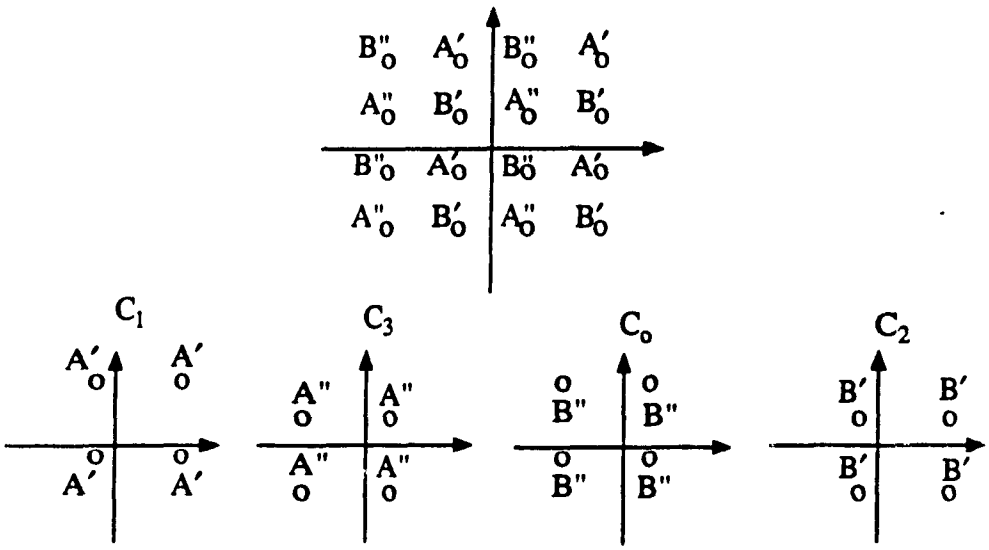

Fig. 4.7 16-QAM constellation and its set partitioning.

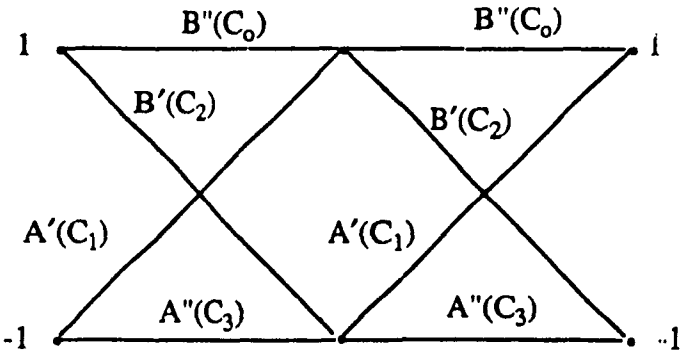

Fig. 4.8 A 2-state trellis diagram.

Table 4.5 The relationship between input bits and output signals for 16-QAM signal constellation.

b_1 b_2 b_3	1 -1 1	1 -1 -1	1 1 -1	1 1 1
$x \rightarrow (A')$	3-j	-1-j	-1+j3	3+j3
b_1 b_2 b_3	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
$x \rightarrow (A'')$	1-j3	-3-j3	-3+j	1+j
b_1 b_2 b_3	-1 -1 1	-1 -1 -1	-1 1 -1	-1 1 1
$x \rightarrow (B')$	3-j3	-1-j3	-1+j	3+j
b_1 b_2 b_3	1 -1 1	1 -1 -1	1 1 -1	1 1 1
$x \rightarrow (B'')$	1-j	-3-j	-3+j3	1+j3

Table 4.6 The upper bounds on P_b for two TCM with QAM modulation

P_b \ E_b/N_o(dB)	13	14	15	16	17	18
12-QAM	1.85×10^{-9}	9.20×10^{-11}	1.20×10^{-14}	2.94×10^{-18}	8.15×10^{-23}	1.64×10^{-28}
16-QAM	2.38×10^{-9}	1.94×10^{-11}	5.04×10^{-14}	2.94×10^{-17}	2.35×10^{-21}	1.86×10^{-26}

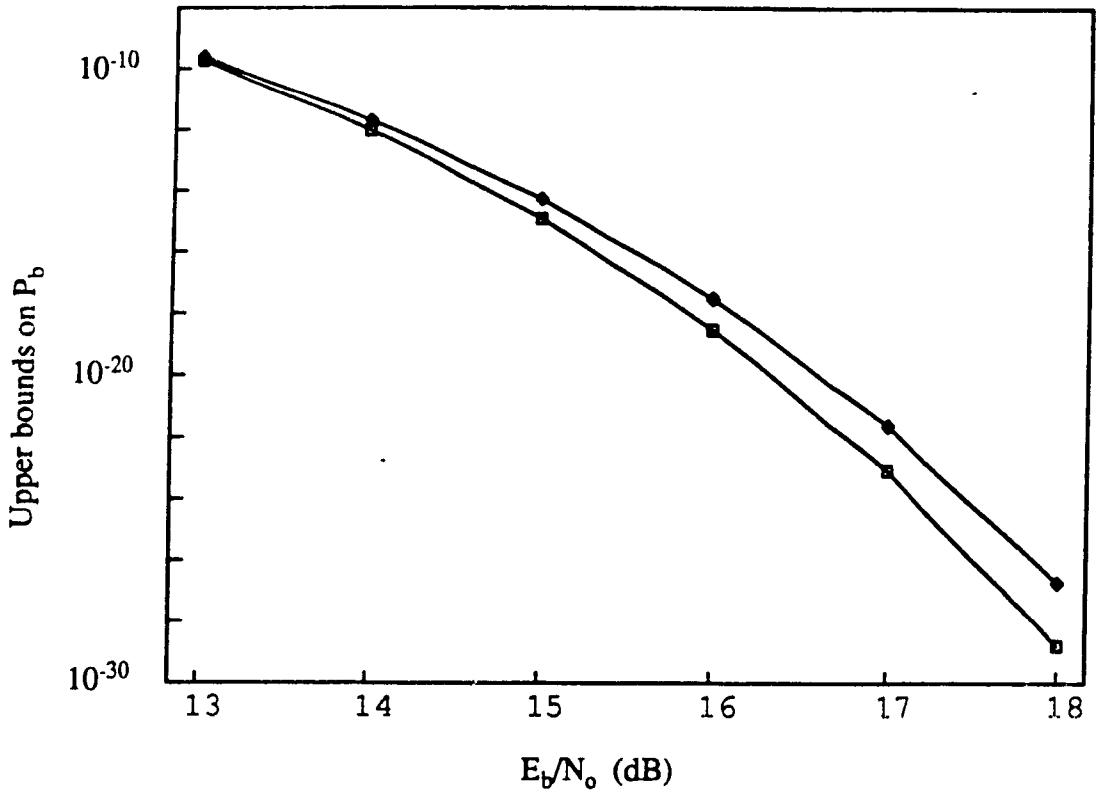

Fig. 4.9 A comparison of the performance of two trellis coded QAM modulation: QAM modulation: (◊), trellis code with 16-QAM; (□), trellis code with 12-QAM partially overlapped constellation.

Chapter 5 Conclusions

5.1 Concluding Remarks

In this thesis, the basic principles and structure of Trellis Coded Modulation (TCM), first developed by Ungerboeck, was described. The key element is set mapping method called "mapping by set partitioning". Some previous research and investigations for improvement and modification of Ungerboeck's TCM were presented. Based on the previous works the possibility of further improvement was discussed. The results showed that it is still possible to improve TCM to reach the upper bound of TCM coding gain.

The important part in this thesis is the newly developed analytically described trellis code with partially overlapped signal constellation, which is also shown in [36, 37]. Instead of using traditional 2^{n+1}-ary signal constellation, these codes use partially overlapped signal constellation which have lesss than 2^{n+1} points. It has the advantage that the average power of the signal constellation is reduced. The analytical representation and coset representation of this partially overlapped signal constellation were presented.

Several examples were given to show the performance gain due to the overlap of the signal set. The performance gains due to the coding and the overlap were

evaluated in terms of the minimum free Euclidean distance d_{free} of the trellis. The results showed that, introducing an appropriate overlap into the constellation design of the analytically described trellis codes, one can easily improve its performance. For M-AM and QAM modulation, it was shown that for low coding complexity significant performance improvement is achievable in comparison to the equivalent conventional design. In particular, the gain in free Euclidean distance of the two-state analytically described trellis coded M level totally overlapped signal constellation over the uncoded M/2-point one approaches 3dB.

The optimum codes can be easily searched by computer. Because these new TCM codes are analytically described.

For comparison of the new codes with Ungerboeck's TCM and Calderbank's method, set partitioning method was used to analyze these examples. The difference was clearly presented.

The comparison between the values of the minimum free Euclidean distance d_{free} divided by P_{av} of new TCM code and the equivalent conventional TCM is an indication of the reduction in required E_b/N_o that can be achieved for arbitrarily small system bit error rates. Bit error probability analysis for these examples was carried out. For sufficiently large values of E_b/N_o, these new codes are better than the equivalent conventional coded schemes.

5.2 Directions for Future Work

In our new coding method described above, the optimum distance was searched with fixed overlap. Although this led to a considerable increase in the performance of bit error probability, but it is still not the optimum values which can be

obtained. The optimum values of upper bound of P_b could be obtained by changing overlap. This requires further investigation.

Since TCM with high-dimensional signal constellation has advantages such as less sensitivity to phase jitter and less redundancy, this work was extended to analytically described trellis codes with high-dimensional partially overlapped signal constellation. The basic principle to form this high-dimensional constellation is to concatenate a few (constituent) low-dimensional partially overlapped signal constellations. These TCM schemes still keep the performance gain achieved by the overlap of low-dimensional partially overlapped constellations, and also have the advantages achieved by high-dimensional signal constellation. In section 2.6 an example of TCM scheme with 4-D signal constellation, where the constellation is formed by concatenating a two (constituent) 2-D (QAM) partially overlapped constellations, is shown. The results obtained clearly demonstrate the above prediction. TCM schemes with partially overlapped high-dimentional signal constellations are also worth further investigation.

REFERENCES

[1] J. L. Massey, "Coding and modulation in digital communications," in Proc. 1974 Int. Zurich Seminar Digital Comm., Zurich, Switzerland, Mar. 1974, pp. E2(1)-E2(4).

[2] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing the channel alphabet and introducing sequence coding," 1976 Int. Symp. Inform. Theory, Ronneby, Sweden, June 1976.

[3] G. Ungerboeck, "Channel coding with multilevel/phase signals," IEEE Trans. Inform. Theory, vol. IT-28, Jan. 1982, pp. 55-67.

[4] G. Ungerboeck, "Trellis coded modulation with redundant signal sets - Part I: Introduction," IEEE Comm. Magazine, vol. 25, No. 2, Feb. 1987, pp. 5-11.

[5] G. Ungerboeck, "Trellis coded modulation with redundant signal sets - Part II: State of the art," IEEE Comm. Magazine, vol. 25, No. 2, Feb. 1987, pp. 12-21.

[6] A. R. Calderbank, J. E. Mazo, and H. M. Shapuo, "Upper bounds on the minimum distance of trellis codes," Bell syst. Tech. J., Vol. 62, No. 8, Oct. 1983, pp. 2617-2646.

[7] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S U Qureshi, "Efficient modulation for band-limited channels," IEEE J. Selected Areas Comm., vol. SAC-2, Sept. 1984, pp. 632-647.

[8] L. F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space - Part I: 180 degrees," IEEE J. Selected Areas Comm., vol. SAC-2, Sept. 1984, pp 659-672.

[9] L. F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space - Part II: nonlinear codes," IEEE J. Selected Areas Comm., vol. SAC-2, Sept. 1984, pp. 672-686.

[10] A. R. Calderbank and J. E. Mazo, "A new description of trellis codes," IEEE Trans. Inform. Theory, vol. IT-30, Nov. 1984, pp. 784-791.

[11] D. Divsalar and J. H. Yuen, "Asymmetric MPSK for trellis codes," presented at GLOBECOM'84, Nov. 26-29, 1984.

[12] D. Divsalar and M. K. Simon, "Trellis coded modulation for 4800 to 9600 bps transmission over a fading satellite channel," IEEE J. Selected Areas Comm , vol. SAC-5, Feb. 1987, pp. 162-175.

[13] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Performance Criteria," IEEE Trans. Comm., vol. 36, No.9, Sept 1988, pp. 1004-1012.

[14] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Set Partitioning for Optimum Code Design," IEEE Trans. Comm., vol. 36, No. 9, Sept. 1988, pp. 1013-1021.

[15] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE Trans. Inform. Theory, vol. IT 13, April 1967, pp. 260-269.

[16] A. R. Calderbank and N. J. A. Sloane, "Four-dimensional modulation with an eight-state trellis code," AT&T Tech. J., vol. 64, May-June 1985, pp. 1005-1017.

[17] L. F. Wei, "Trellis-coded modulation with multidimensional constellations," IEEE Trans. Inform. Theory, vol. IT-33, No. 4, July 1987, pp. 483-501.

[18] A. R. Calderbank and N. J. A. Sloane, "An eight-dimensional trellis code," Proc. of the IEEE, vol. 74, May 1986, pp. 757-759.

[19] G. D. Forney Jr., "Coset codes - Part I: Introduction and Geometrical classification," IEEE Trans. Inform. Theory, Vol. IT-34, Sept. 1988, pp. 1123-1151.

[20] L. F. Wei, "Rotationally invariant trellis-coded modulations with multidimensional M-PSK," IEEE J. Selected Areas in Comm., vol. 7, No 9,

Dec. 1989, pp. 1281-1295.

[21] D. Divsalar and M. K. Simon, "Trellis coding with asymmetric modulations," IEEE Trans. Comm. vol; Com-35, No. 2, Feb. 1987, pp. 130-141

[22] O. Agazzi, D. G. Messerschmitt, and D. A. Hodges, "Nonlinear echo cancellation of data signals," IEEE Trans. Comm. Technol., vol Com 30, No.11, Nov. 1982, pp. 2421-2433.

[23] M. Harwit and N. J. A. Sloane, "Hadamard Transform Optics" New York, Academic, 1979.

[24] J. M. Turgeon and P. J. Mclane, "Minimal complexity design of analytically described trellis codes," in Proc., ICC'87, June 7-10, 1987, pp 23.2.1-23.2.7

[25] M. K. Simon and D. Divsalar, "A new description of combined trellis coding with asymmetric modulation," JPL Publ., 85-45, July 15, 1985.

[26] D. Divsalar and M. K. Simon, "Multiple trellis coded modulation (MTCM)," IEEE Trans. Comm., vol. 36, No. 4, April 1988, pp. 410-419.

[27] D. Divsalar and M. K. Simon, "Generalized multiple trellis coded modulation (MTCM)," in Proc. ICC'87, June 7-10, 1987, pp. 20.3.1-20.3.7

[28] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and

cosets," IEEE Trans. Inform. Theory, vol. IT-33, Mar. 1987, pp. 177-195.

[29] A. LaFanechere and D. J. Costello, Jr., "Multidimensional coded PSK systems using unit-memory trellis codes," presented at 23rd Allerton Conf. Comm., Oct. 2-4, 1985.

[30] S. G. Wilson, "Rate 5/6 trellis coded 8-PSK," IEEE Trans. Comm., vol. Com-34, Oct. 1986, pp. 1045-1049.

[31] D. Divsalar and M. K. Simon, "Multiple trellis coded modulation (MTCM)," JPL Publ. 86-44, MSAT-X Rep. 14, Nov. 15, 1986.

[32] D. Divsalar and M. K. Simon, "Multiple trellis coded modulation (MTCM)," JPL Publ. 86-44, MSAT-X Rep. 141, Nov. 15, 1986.

[33] M. K. Simon and D. Divsalar, "The performance of trellis coded multilevel DPSK on a fading mobil satellite channel," JPL Publ. 87-8, MSAT-X Rep. 144, June 1, 1987.

[34] S. G. Wilson and H. A. Sleeper, "Four-dimensional modulation and coding: an alternate to frequency reuse," Univ. Virginia, Rep. UVA/528200/EE83/107, Sept. 1983.

[35] R. Fang and W. Lee, "Four-dimensionally coded PSK systems for combatting effects of severe ISI and CCI," in Proc. 1983 IEEE GLOBECOM. Conv. Rec.,

1983, pp. 30.4.1-30.4.7.

[36] M. R. Soleymani and L. Kang, "TCM schemes with Partially overlapped signal constellation," accepted for publication, IEEE Trans. Comm..

[37] M. R. Soleymani and L. Kang, "Trellis coding with partially overlapped signal sets," accepted for presentation at ICC'91, June 23-26, 1991.

[38] A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding. New York: McGraw-Hill, 1979.

[39] D. Divsalar, "Performance of mismatched receiverson bandlimited channels," Ph.D. dissertation, Univ. California, Los Angeles, 1978.

[40] M. K. Simon and D. Divsalar, Combined Trellis Coding with Asymmetric MPSK Modulation. Pasadena, CA: JPL Publication 85-24, May 1, 1985.

[41] D. Divsalar and M. K. Simon, "Combined trellis coding with asymmetric modulations" in Proc. GLOBECOM'85, Dec., 1985, pp. 21.2.1-21.2.7