Theoretical Approach in Trellis Coded Modulation (TCM) System

H. N. OGBOKE[1], V. C. OGBOH[2], A. O. ANIAGBOSO[3], N. A. NWOYE[4]
[1, 2, 3, 4] *Department of Electrical Engineering, Nnamdi Azikiwe University, Awka*

Abstract- Limited bandwidth, unwanted signal interference in the atmosphere medium during transmission and low data rate transfer among others are the challenges bedeviling communication systems. Hence, the introduction of Trellis Coded Modulation (TCM) system to address these fundamental causes of inefficient and unreliable communication from the transmission point to the consumers has become paramount. The choice of this system, an efficient coding technique improves the coding gain at no extra cost in bandwidth with respect to expansion. Ensures reliable high data rate communication over channels with limited bandwidth at far higher speeds from previously envisaged. The performance of trellis codes on M-ary phase shift keying (MPSK) and M-ary Quadrature Amplitude Modulation (MQAM) shows tremendous coding gain.

Indexed Terms- Bandwidth, Bedeviling, Communication, Modulation, Amplitude, Coding

I. INTRODUCTION

Basically, most commodities are ultimately in high demand but eventually become scarce. Most importantly, a few of them tend to operate with specified range and should be strictly maintained. Furthermore, performance of data transmission over many communication channels has traditionally been marred by unreliable and poor quality of information due to noise interference, low throughput and large bandwidth. The aforementioned challenges in communication led to the design and implementation of a Trellis Coded Modulation (TCM) system. This system improves the coding gain and bandwidth efficiency with respect to the Bit error rate (BER) and high data transfer in a communication system. This is rewarding to digital communication engineers in their various design options and trade-offs for the benefit of the last mile. This article finds its application in many fields such as satellite Communication, digital systems and networks as well as internet for faster transmission of data etc. without bandwidth expansion at a better coding gain.

A. Error Control:

Error control coding is the process of converting source bits into transmitted symbols so as to ensure reliable communication (transmission) even in the presence of noise. The redundancy in the coded sequence is made use of at the receiver to improve the toughness to noise. Coding is used to facilitate detection and correction of errors at the receiver. More importantly, its goal is to prevent errors before they occur by a combination of detection and decoding known as soft decoding. The ultimate aim of error-control coding is to close the gap between the performance of uncoded modulation and the Shannon's limit, allowing a practical system to communicate reliably at a rate close to the Shannon's capacity.

Binary codes and convolutional codes belong to the class of binary codes used for error control. A block code maps blocks of k source bits into blocks of n coded bits where n>k and a code rate of k/n (i.e. a fraction of the total bit rate devoted to information bits). A convolutional coder also produces coded bits at a higher rate (a very long stream of message bits) than the source bits and is not divided into blocks. The coded bits have redundant information about the source bits [1].

B. Coded Modulation Theory

The purpose of channel coding is to address the effects of transmission impairments while aiding the receiver in decision-making process. Specifically, the expectation of the choice of coding and modulation design is to reduce the bit error rate while increasing the reliability of the transmission. The gain achieved due to coding as against the uncoded system is referred to as coding gain. In designing the channel and modulation scheme a number of factors are critically considered. For instance, in a power limited scenario, the system's bandwidth can be traded by accommodating low-rate code which invariably affects the expected throughput as a result of additional parity bits employed. Furthermore, for bandwidth-limited and power limited situations the application of higher coding gain is absolutely necessary. Also, the channel's characteristics determines the type of coding and modulation scheme to be employed to achieve the desired result. In addition, coding gain could be achieved without increasing bandwidth by carefully operating the functions of coding and a type of modulation jointly. Ultimately, the choice of coding and modulation technique requires sacrificing some features to gain other features [2].

C. Coded Modulation Schemes

There are several coded modulation schemes in use for error control measures to ensure efficient transmission of information taking into consideration the characteristics of the channel. Among them is the Bit-Interleaved Coded Modulation (BICM). This scheme utilises bit-based channel interleaving in conjunction with Gray signal labelling. It combines convolutional codes with bit-interleave in order to achieve transmission over fading channels. Another form of coded modulation is Turbo coding that has low coding rate; hence requires some degree of bandwidth expansion. Nonetheless, it is having been standardised for use in third-generation (3G) mobile radio system. BICM Turbo coded modulation (TuCM) came into operation to address the deficiency of Turbo coding- that is, to offer higher spectral (bandwidth) efficiency. Although it performs pretty well in interleaved narrowband Rayleigh fading channels, but worse than Trellis Coded Modulation (TCM) in Gaussian channels due to reduced Euclidean distance (ED) of bit-interleaved scheme. Furthermore, Turbo Trellis Coded Modulation (TTCM) is a new type of modulation scheme that employs TCM scheme as its component code and performs better than TCM and TuCM. Last but not the least is the BICM iteration joint decoding and demodulation (BICM-ID) which uses set partitioning signal labelling. This type was to increase the Euclidean distance of BICM. Also in the list of coded modulation schemes is the Trellis-Coded Modulation (TCM). This is a bandwidth efficient modulation scheme that employs symbol base channel interleaving as well as set partitioning (SP) assisted signal labelling as proposed by Ungerboeck's to obtain a higher Euclidean distance between constellation points. It combines convolutional codes with multi-dimensional signal sets (modulation scheme) for error control technique appropriate for mobile communications [2].

II. TRELLIS-CODED MODULATION

A. Trellis Coded Modulation (TCM)

The performance of data transmission over many communication channels had been marred by unreliable and poor transmission quality due to noise interference. The application of error correcting codes in principle was addition of parity bits known as coding which allows the error pattern to be identified and corrected at the receiver. All error-correction codes when used in real time communication systems provide improvements in error performance at the cost of bandwidth expansion. Hence, one main problem of coding is the reduction in spectral efficiency. In the past, coding generally was not popular for band-limited channels such as telephone channels, where signal bandwidth expansion is not practical [7]. It was observed that treating coding and modulation functions as a separate entity was an impediment in band-limited applications. Hence, the result of Ungerboeck's research in combining the functions of error correction codes and modulation was envisaged and its implementation gave rise to Trellis-Coded Modulation (TCM) which revolutionised telecommunication industry. TCM allows the design of a modem-plus-codec that offers greater resilience to noise than uncoded systems with the same spectral efficiency. Coding gain can be realised with trellis codes because the availability of the Viterbi decoding algorithm makes trellis decoding simple and efficient [13]. TCM offers high coding gain without sacrificing data rate or without increasing either bandwidth or power. However, there is still a trade-off involved. Since TCM achieves coding gain at the expense of decoder complexity the information rate is reduced. In the presence of AWGN, TCM schemes can yield a net coding gain of about 3dB relative to uncoded systems with relative ease, while gains of about 6-dB can be achieved with higher modulation schemes [3].

B. Concept of Trellis Coded Modulation

The idea of TCM emanates from the fact that "not all signal subsets (in a constellation) have equal distance properties". That is, for a non-orthogonal signal set, such as MPSK, antipodal signals have the best distance properties for easily discriminating one signal from the other; while nearest neighbour signals have relatively poor distance properties. The TCM system assigns waveforms to bits with respect to the criterion of better or worse distance properties. Fig 2 is the block diagram of TCM and figure 3 represents the encoder/modulator block.

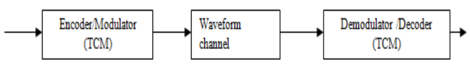


Figure 1: TCM transmitter and receiver block diagram

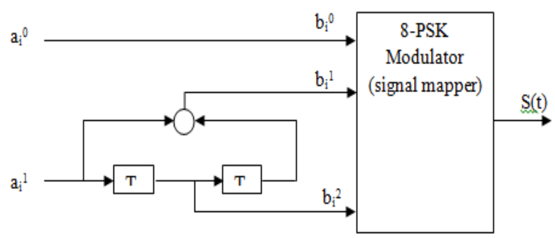


Figure 2: Convolutional encoder with modulator combined (TCM)

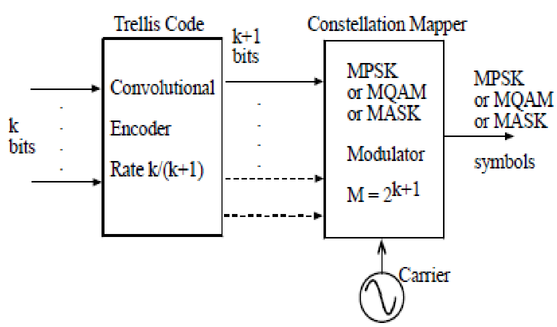


Figure 3: A general form of Trellis-Coded Modulation

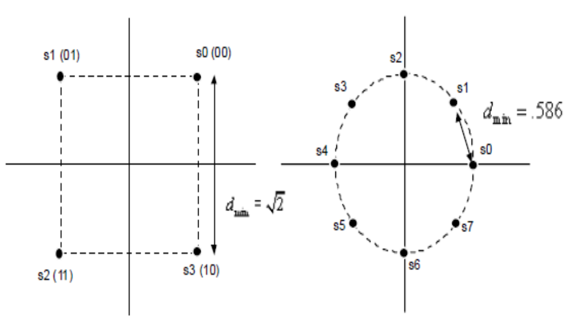


Figure 4: Constellation doubling in TCM, a QPSK signal transmitted using an 8PSK constellation

Figure 4 is a general form of TCM while figure 5 depicts signal mapping. In figure 3, the input, k = 2, n = 3, then R = 2/3 that takes a QPSK signal (M=4) and gives an output 8-PSK signal (M=8). So instead of expanding the bandwidth as the signal goes from QPSK to 8-PSK, it instead doubles the constellation points as shown in figure 5.

C. Features and Principle of TCM

TCM is made up of forward error correction block known as convolutional encoder and a signal mapper (that is a modulator) as shown in the figure 3. The convolutional code is defined by three parameters (n, k, K): the input bits k (a_i^0, a_i^1), the output codeword n (b_i^0, b_i^1, b_i^2), and the constraint length K =k(m-1). The essence of convolutional encoder is that it handles a continuous stream of transmitted data. Unlike the block codes, the convolutional coder, a finite state machine has memory that is influenced by the number of shift registers (constraint factor K). As such the current n-bit output of an (n, k, K) code depends not only on the value of the current k input bits but also on the previous input bit in the shift registers. Hence, the current output of n bits is a function of the last K x k input bits with respect to the modulo-2 adder. The output codeword n is determined by the nature of the connections of the adders. The code rate k/n is the measure of the effectiveness the encoder. The function of a signal mapper is to map the three bits output from the encoder into one of the eight symbols of an 8-PSK signal set of three bits/symbol as against two information bits/symbols of QPSK. The term trellis-coded modulation takes its origin from the fact that the encoded sequences consists of modulated symbols rather than binary digits [3].

The effect of this type of coding is that the Euclidean distance (ED) between the signal elements are closer in 8-PSK constellation than in 4-PSK and it is the task of the decoder to select the best sequence that is closest in Euclidean distance. This criterion for signal selection is called the Maximum Likelihood (ML) criterion. In the decoding process, errors that have the lowest Euclidean distance d_{min} to the correct trellis path can arise which affect the values of the signal-to-noise ratio. So mapping of the binary output of the convolutional encoder onto the constellation points of the modulator plays a crucial role in the overall performance of the trellis code. Hence, mapping rules as given by Ungerboeck are stated below:

a) Parallel signals should be characterised by the highest Euclidean distance between signals.
b) Diverging path signals emerging from a given state be characterised by the second highest Euclidean distance between signals [3].

Set partitioning is used to assign signals in the constellation.

Euclidean distance is the shortest (minimum) distance defined in the I-Q plane between two signals. The distances given in figure 6 are squared and are called squared Euclidean distance (SED) and the smallest of these distances is called minimum squared ED (MSED).

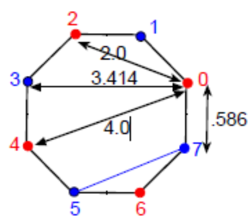


Figure 5: 8 PSK constellation with SED between symbols[18]

Messages are in long sequences and the receiver makes decision based on sequences and not on symbol-by-symbol basis. So the probability of error between sequences is given by:

$$Pe \sim e^{-d2min/2\sigma 2}$$

where, d_{min} is the sequence ED between sequences and $^{\sigma 2}$ is the noise power

D. Signal Redundancy Enlargement

In TCM, with k as the input bits and n output bits is increased by a redundant bit p, n=k+p. Note that coding increases the signal set size from 2^k to 2^{k+p}. So, coding gain is obtained by expanding the uncoded signal set by a factor of 2. Where p=1, the encoder has a rate of k/(k+1) code and subsequently mapping groups of k+1 bits into the set of 2^{k+1} waveforms thereby increasing the signalling set alphabet from $M=2^k$ to $M'=2^{k+1}$. The increase in alphabet size does not affect the bandwidth; however, it results in a reduced distance between adjacent symbol points (for signal sets with a constant average power) which degrades error performance in uncoded system. The free distance determines the error performance. The proximity of signals do not affect performance as long as partitioning rules are followed. The code sequence and distance properties are best determined in the trellis diagram. The main objective of TCM is to assign waveforms to trellis transitions so as to increase the free distance between the waveforms that are the most likely to be confused which ultimately enhances the coding gain [3].

E. BANDWIDTH

The maximum possible symbol rate is determined from bandwidth, B. hence symbol rate, Rs <=2B. We can now determine the size of the alphabet, M to deliver the needed signal BER at the given available power. If we want to transmit a QPSK signal, uncoded with a BER of 10^{-5}. This will require 9.6dB of energy per the ideal Eb/No vs BER relationship. If that much power is not available because the transmitter is small, then an option is to add a code of rate 2/3 to reduce the BER which will give this BER at a smaller Eb/No. So, addition of coding increases the bandwidth by 3/2(i.e. 1/R). If we cannot allow the bandwidth to change, then information rate will have to decrease by the same proportion [8].

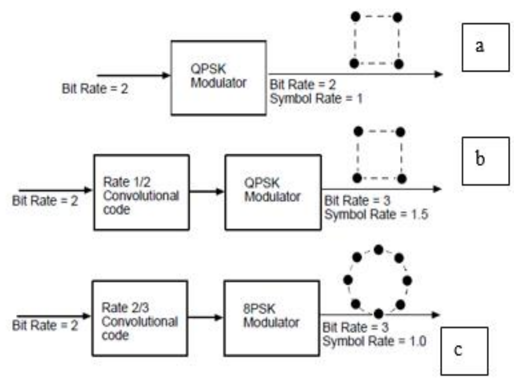


Figure 6: Determination of bandwidth [18]

The figure 7a shows an uncoded signal at bit rate, R_b = 2, output R_b still at 2 bits/s i.e k=1, therefore, going by Rb = kRs, Rs = 1. Then for figure 7b at rate ½ code there is 3 bits/sec at the o/p at k = 2bits/symb, hence Rs = 1.5. While for figure 7c at rate 2/3 code there is 3bits/sec at the o/p at k = 3 bits/symb, hence Rs = 1. As long as the symbol rate is the same, MPSK modulations have the same bandwidth irrespective of the alphabet size.

III. TCM ENCODER

A. Convolutional coder (codes)

This is a finite state machine in which the added redundant bits are generated by modulo-2 convolutions. Convolutional codes are widely used as channel codes in practical communication systems for error correction. The encoded bits depend on the current k input bits and a few past input bits. The main decoding strategy for convolutional codes is based on the widely used Viterbi algorithm.

Convolutional code can be marked by (n, k, K), which means for every k bits, there is an output of n bits and K is called constraint length. Basically, convolutional code is generated by passing the information sequentially through a series of shift registers. Because of the shift registers, convolutional coder has memory, the current n-bit output depends not only on the value of the current block of k input bits but also on the previous K-1 blocks of k input bits. So, the current output of n bits is a function of the last K×k bits.

B. Set Partitioning

Ungerboeck's set-partitioning, a type of signal mapping is devised to improve coding. The basic idea is to map 2^k information bits to 2^{k+1} constellation points such that we can limit the transitions to occur only along the largest SED. The partitioning of the subsets further increases the ED between the signals in that set. The 8-PSK are successively partitioned into disjoint cosets such that the SEDs increases at each level as $\Delta_0 < \Delta_1 < \Delta_2 \ldots$ between the elements of the subsets. Assuming the signal power (i.e. the square of amplitude) is taken to be unity, by calculation the distance Δ_0 (known as Euclidean distance) will be 2 x rsin(π/8) = 0.765. There are two levels of subsets with each level having a different ED such as subsets B0 and B1, C0 to C3 with EDs of $\Delta_1 = 1.414$ and $\Delta_2 = 2$ respectively. In TCM, the trellis transitions are labelled with waveform numbers unlike in convolutional coding where is is labelled with code bits. Because of the good distance property of the 4PSK the decoder can make arbitrary decision for each signal received as against the distance property of uncoded 8PSK signal set. Each subset is called a coset and by the lattice terminology the set partitioning of a coded 8-PSK from uncoded QPSK that shows the ever increasing ED subsets is shown below[3]. See figure 9 and 10.

The number of symbols required in the alphabet can be determined from the bandwidth and bit rate considerations. Consider the uncoded system having 2^m and coded 2^{m+1} number of symbols respectively. It is of great importance to know that the choice of mapping to a large extent affects the performance of the code. From figure 8, it is indicated that the uncoded bits select a signal from the subset and the coded bits select the subset. Hence, there must be 2^{n+1} subsets. A general technique extended to make use of alphabets larger than four [4].

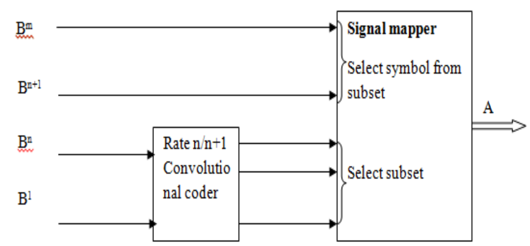


Figure 7: General technique for signal mapping[10]

Mapping has the property that parallel transitions correspond to symbols that are far apart as possible. A systematic way to design such mappings in general is known as set partitioning proposed by Ungerboeck.
Not all the gain is due to the redundancy introduced by the convolutional coder; some of it is due to the constellation chosen for the comparison.

The general rule for mapping by set partitioning are:

- First, maximise the distance between parallel transitions
- Next, maximise the distance between transitions originating or ending in the same state.

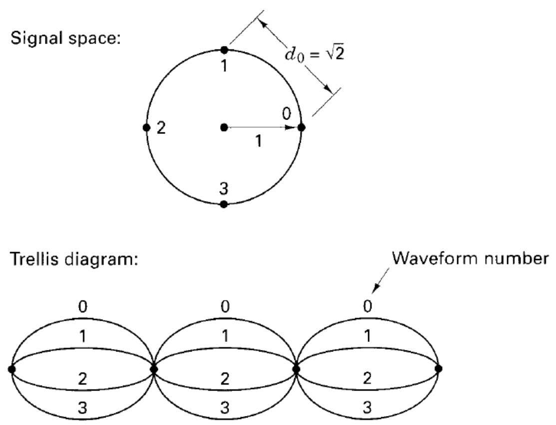


Figure 8: Uncoded 4-PSK and its one-state trellis[3]

With QPSK, k=2

- When coded n,= k+1 = 3
- Code rate r =k/n= 2/3
- No of signal constellations= $2^{k+1} = 2^3 = 8$
- 1st level subset= B_0, B_1
- 2nd level subset= C_0,C_1,C_2,C_3=4 state trellis

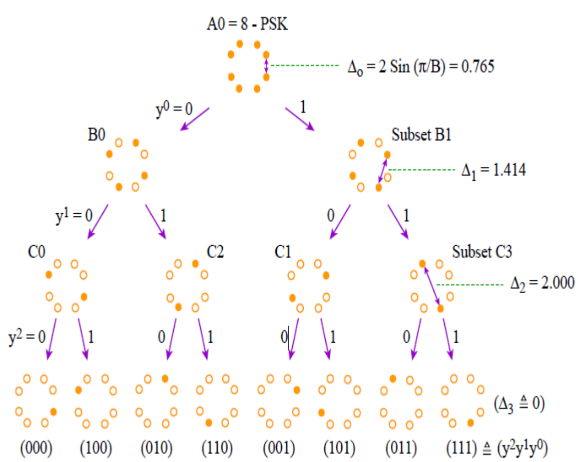


Figure 9:Ungerboeck's 8 – PSK signal set partitioning [3]

B. Waveforms Mapping to Trellis Transitions

The purpose of trellis transition branches to waveforms is to ensure coding gain is achieved. Hence, assigning of waveforms to transitions are guided by rules as devised by Ungerboeck.

The rules ensure that codes have a regular structure and that the ED exceeds the minimum distance of the uncoded reference modulation [3].

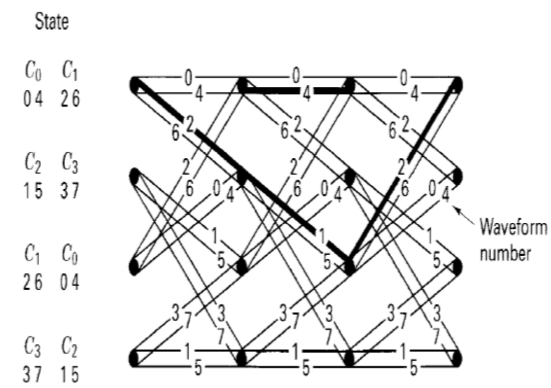


Figure 10: Four-state trellis structure[3]

The four-state trellis figure 11 satisfies the rules stated earlier. With the waveform number assigned to the trellis transitions, the waveform signals can now be assigned in an arbitrary format without the fear of lower ED which could cause misinterpretation of signals leading to error. The trellis structure above will be used in a TCM decoder for detection and decoding of received signals to ensure the ultimate goal – good coding gain is achieved.

IV. TCM DECODER

A decoder is a device used to recover a coded signal transmitted to its original state at the receiver. The two approaches used for decoding convolutional codes are sequential decoding – using Fano algorithm and Maximum Likelihood decoding that uses Viterbi algorithm. The latter is the principle of Viterbi algorithm with soft-decision.

The most important error correction algorithm developed for convolutional codes is the Viterbi decoding algorithm. The algorithm technique compares the received sequence with all possible transmitted sequences and chooses a path through the trellis whose coded sequence differs from the received sequence in a few places. With a valid path established, the decoder can recover the input data bits from the output code bits. Convolutional codes performs excellently well in noisy channels where the possibility of high bit errors is imminent and have been found useful in wireless communication applications [3].

A. DECODING USING THE VITERBI ALGORITHM

In Viterbi algorithm, the principle used to reduce the choices is that errors seldom occur and are distributed randomly. Viterbi algorithm uses a metric and tracks this metric for several trellis paths at once. All paths are followed until two paths converge in one node then the path with larger metric is discarded. This helps to reduce the complexity and gives the Viterbi decoding an advantage over sequential decoding. This form of decoding makes it useful for Trellis Coded Modulation (TCM) as it uses long stream of sequences. In hard-decision Viterbi decoding, this is done using the Hamming distance as a metric. In TCM, the decoding is done with soft decision algorithm and Euclidean distance is used as the metric. The objective is to track “n” possible sequences, keep track of cumulative MSEDs. When paths merge at a state, it follows only the one with the smallest metric [4].

B. TRELLIS CODES

Significant coding gains can be achieved with lower complexity by the use of FSM in the transmitter in conjunction with a multidimensional signal constellation. The lower complexity sterns from the simplicity of the transmitter FSM and the use of Viterbi algorithm for Maximum Likelihood detection at the receiver.

The basic advantage of signal space (trellis) coding is that by going to a higher dimensionality space we can increase the minimum distance in relation to the transmit signal energy. The sequence of data symbols are dependent on one another and this dependence is the essence of achieving coding and shaping gain. The FSM introduces dependence of the successive symbols by its symbol-to-symbol state memory. The coding gain due to the FSM can augment the coding and shaping gain due to constellation design. A signal-space coder based on an FSM is often called a trellis coder.

A penalty is paid in an increase in the number of points per multidimensional symbol, which either reduce the minimum distance or increase the transmitted energy. However, the advantage of working in a multidimensional space more than makes up for this penalty [4].

Furthermore, uncoded input(s) together with coded input are mapped by the signal mapper. However leaving some bits uncoded may affect performance. The extra uncoded bit can be represented in the trellis using parallel branches. A trellis with m-n (m = total no of input bits to mapper and n = total no of input bits to convolutional coder) uncoded bits has 2^{m-n} parallel transitions between every pair of states. The effect of parallel transitions is that a very short error event may occur where a mistake can arise in choosing one of these parallel transitions for the correct one. To minimise its probability, ensure that the ED between the symbols corresponding to the parallel transitions is maximised. To ensure that parallel transitions have symbols as far apart as possible, symbols for parallel transitions are selected from the same subset. The parallel transitions degrade the performance by about 1 dB compared to the system without parallel transitions. The total gain that can be achieved with a trellis code depends on the number of states in the FSM. We can achieve coding gain of 3dB with 4 states, 4.5 dB with 16 states and about 6dB with 128 states or more.

V. TCM PERFORMANCE MEASURE

The performance of TCM could be measured by the amount of coding gain obtained from a modulation scheme used in conjunction with the convolutional encoder as against the uncoded system as explained in this section.

A. Coding Gain

The asymptotic coding gain G, expressed in dB is the comparison with some uncoded system with same average signal power and noise variance is in trellis transition expressed as a ratio of distances squared as shown in the equation below.

$$Coding\ gain, G_{dB} = \frac{d^2_{free/coded}}{d^2_{min/uncoded}}$$

The coding gain is referenced to the baseline signal. With a QPSK at BER of 10^{-5}, requires 9.6 dB of Eb/No is taken from the baseline and for 8-PSK that requires a higher Eb/No at the same BER will be account for from the baseline of coding gain equation.

Although $d_{min/coded}$ is less than $d_{min/uncoded}$, if $d_{free/coded}$ is increased by the summation of branch transitions to be greater than $d_{min/uncoded}$ there will be a positive coding gain. Free distance, d_{free} is obtained from distances between sequences rather than distances between signals. Hence, it is the Euclidean distance of a coded signal in terms of the smallest possible distance between allowed sequences and is measured from all-zero sequence (000) as hamming distance. The Euclidean distance of a coded signal is the smallest between all-zero sequence and one that diverges from it and remerges. The equation above is the major aim of TCM which is to achieve free distance that exceeds the minimum distance of the uncoded modulation signal at the same information rate, bandwidth and power[3].

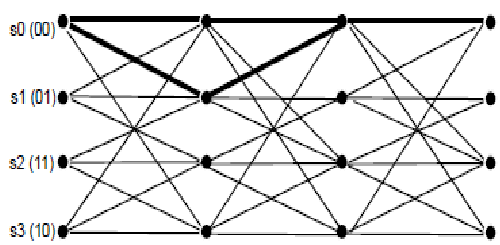


Figure 11: The diverging and remerging of alternate sequences

Note: A sequence is a set of demodulated symbols. When two paths diverge it means there was an error and the decoder made a wrong decision and so further decision is to bring it back to the correct one. A small incorrectly allowed path is a measure of the error correcting capability of the code. This is the concept used by the receiver as it picks neighbouring signal points because they are most likely the correct signal. This is the concept behind Maximum Likelihood decoding (MLD).

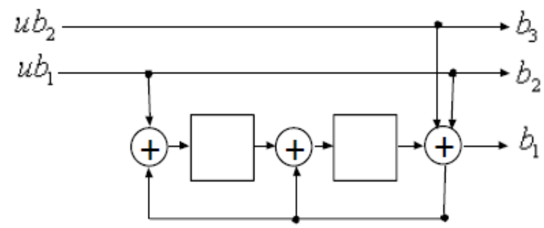
In figure 12, the trellis structure shows that if an error is made at time t=0 and the path diverges and then remerges to the correct sequence in only two segments, the sum of the squared ED for this path is called free distance, d_{free} of the code. The free distance is the sum of the squared Euclidean distance which is

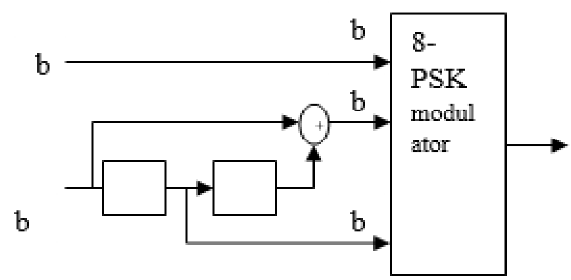
$$d^2_{\underline{free}\atop coded} = d^2_{min} + d^2_{min}$$
$$= (\sqrt{2})^2 + (\sqrt{2})^2 = 4$$

B. CODING GAIN FOR 8-PSK WITH A 4-STATE TRELLIS

Since the top two levels, the main constellation A and subset B in the set partitioning have smaller distances; errors are more likely to occur, hence the use of coded bits to traverse through this part. Considering the 4-state trellis of figure 12, at the receiver, the decoder makes a decision about the coded bits and then the uncoded bit, sending it down one of the four paths at each state. Even if the decoded decision is correct, there would still be a possibility that the uncoded bit will be decoded incorrectly.

Figure 12: a) Rate 2/3 convolutional encoder;





b) Rate ½ convolutional coder

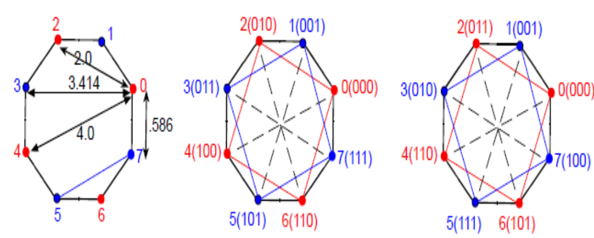


Figure 13: a) 8-PSK symbol mapping;

c) Natural mapping; c) Gray mapping [7]

For instance, using a rate 2/3 non-systematic coder of figure 13a,

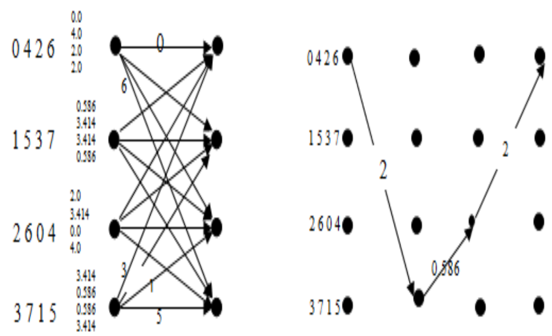


Figure 14: Trellis diagram for 2/3 encoder [3]

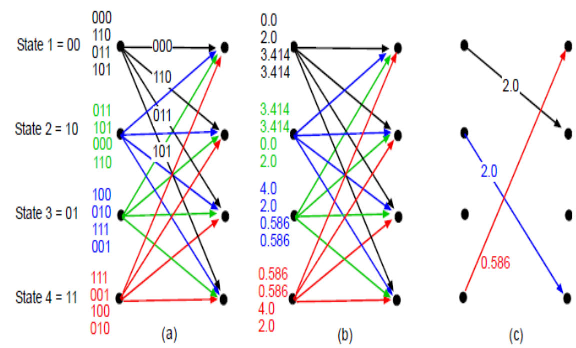


Figure 15: Trellis of a rate 2/3 convolutional code[7]

From the trellis diagram of figure 15, which of the divergent paths will help return the path back to zero state in connection with each other considering the state transitions that give the minimum. The signals at each state are stated side by side to their EDs and the appropriate selections are depicted in figure 14. So the total MSED for this sequence is

2 + 0.586 + 2 = 4.586

$$Coding\ gain, G_{dB} = \frac{d^2_{\frac{free}{coded}}}{d^2_{min/uncoded}}$$

$$= 10\log\left(\frac{4.586}{2}\right) = 3.6dB$$

With natural mapping we have 3.6 dB coding gain. Alternatively, using Gray code and applying the trellis structure above the total MSED resulted to

0.586 + 0.586 + 0.586 = 1.758

Therefore, coding gain =

$10\log\left(\frac{1.758}{2}\right) = -0.56dB$ (a loss).

So mapping is an important consideration. So adding any code of rate 2/3 code in front of an 8-PSK modulator is not beneficial.

For instance, using a rate ½ systematic coder of figure 13b, 010 bits represented by b_1 b_2 b_3 would be mapped to symbol S_2 using this mapping where 0(b1) is the most significant bit(MSB) with the largest distance. Ungerboeck's approach leaves the MSB uncoded to take care of itself as it has a large ED. This turns out to be the key to larger coding gains of this approach. Only the bits that reside at the top levels of set partitioning with smaller SEDs are coded, thus reducing coding rates and increasing bit efficiency.

TCM can therefore be implemented using lower coding rate leaving the MSB uncoded by the use of systematic convolutional encoder as depicted in figure 13b.

Instead of using a rate 2/3 code on both incoming bits, 1 bit is left uncoded and rate ½ code and another bit which still gives 2/3 code rate. The use of ½ code rate has a better coding gain.

TCM basis is to selectively code only some of the bits and take advantage of the increasing ED obtained by set partitioning, then leaving uncoded bits that are naturally protected by their large SED. Starting with a 4-state encoder we have;

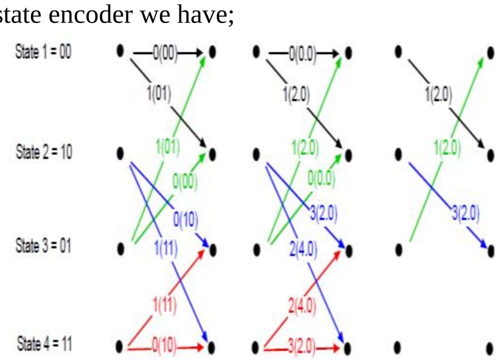


Figure 16: Trellis of rate ½ code

The minimum length paths have the corresponding distances from all zero(00) bit symbol are 2.0, 2.0, 2.0. Therefore, 2+2+2=6; hence the MSED for this code is 6

Therefore, coding gain G(dB) = 10 log(6/2) = 4.77 dB This means that 4.77 dB of rate ½ code is better than 3.6 dB from 2/3 code rate obtained earlier. This is just the coding gain of the rate ½ code with the uncoded bit yet to be considered.

To account for the uncoded bit, consider figure 18, at each state we have two incoming coded bits and one uncoded bit, so each path doubles to account for the two choices for the uncoded bit. At state 00, if the output of the coded bit are 10, then we get 110 if uncoded bit is 1or 010 if it is 0. The doubling of choices is called parallel transitions. (This is of consequence only in a 4-state code)

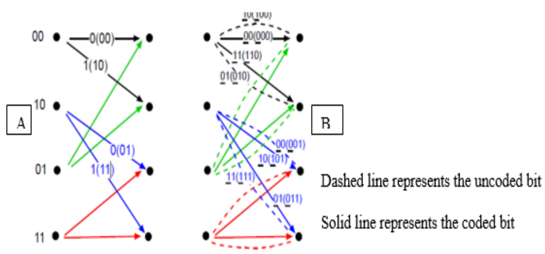


Figure 17: a) rate ½ code trellis, b) rate ½ code trellis with uncoded bit, b1[7]

From the figure above, we have a symbol of 3 bits in parenthesis. The first bit is uncoded while the last two bits are coded from the top trellis where the minimum distance is low and the chances of mistaking one symbol for another is high. Below is the complete trellis of the above 4-state trellis with the chosen MSED using the waveform numbers to calculate the minimum distance from the set partitioning.

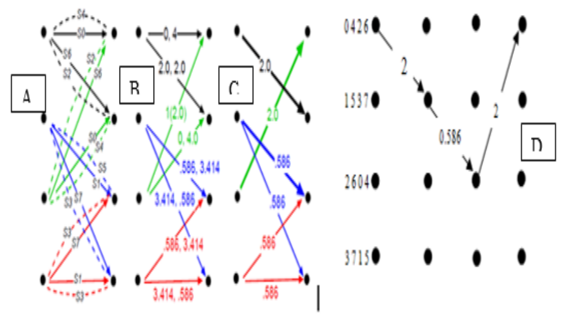


Figure 18: Trellis structure: a) Modified trellis for uncoded bit and replaced by a chosen symbol map(i.e. S0, S4, S2, S6), b)Establish their EDs, c) Choose the path with minimum ED, and d) Represent the path taken in the trellis diagram[18]

Therefore, d_{free} of the code = 2+0.586+2 = 4.586
It will be observed that the parallel pairs (e.g. 0 4 & 2 6) has 180 deg phase shift apart and corresponds to a MSED of 4.0 and is smaller than the sequence staggered MSED (SMSED),(4.586) this error is more likely. In other words, it is more likely that the uncoded bits will be decoded incorrectly than the two coded, simply because they have a larger SMSED. This is called single stage error. When this happens we have:

δ^2_{min} = 4.0 (not d_{min}, but the SED at the bottom level of the partition which contains the uncoded bit).
δ^2_{free} = sum of the MSED (SMSED)
so, the minimum distance of the two determine the overall performance.
Therefore, $d_{free} = \min(\delta^2_{free}, \delta^2_{min}) = \min[4.586, 4.0]$
Coding gain = minimum of the two/ d^2_{min}
Note for QPSK, $d^2_{min/uncoded} = 2.0$
= 10 x log(4/2) = 3 dB
This gain is low compared to 3.6 dB obtained from rate 2/3 code. This is because of the following:
1. The ACG is low because of the single stage errors
2. The major contributor of the low ACG is the existence of parallel transitions – it limits ACG
3. Finally the use of 4-state trellis structure is another limiting factor that encourages parallel transition; hence, the introduction of higher state trellis.

C. Coding Gain for 8-Psk with 8-State Trellis
We can increase the coding gain in TCM by eliminating the parallel transitions by assigning paths so that they do not have parallel transitions and this

can be achieved by increasing the number of states. Increase of states from 4 to 8-states is done by increasing the number of shift registers from 2 to 3.

From the trellis diagram figure 20, which of the divergent paths will help return this path back to all-zero reference in connection of the other at the end point using the waveforms instead of the binary numbers. So at state 1:0426 the smallest distances are 2 and 6 looking at fig 14a while at state 2:1537 the smallest distances are 1 and 7. Also at state 3:2604 the smallest distances are 2 and 6 while at state 4:3715 the smallest distances are 1 and 7. Here we choose the best distance path that will diverge and remerge with links back to zero sequence. Such will be state 1=6 links state 3 at t1, i.e. d_{min}=2; state 2=1 links state 1 at t1 i.e. d_{min}=0.586; state 3=6 links state 2 at t1 i.e. d_{min}=2. So state 4 is not connected due to the choice of the minimum distances that will connect back to zero sequence.

So the total minimum SED(MSED) for this sequence is

$d^2_{free/coded} = 2 + 0.586 + 2 = 4.586$

therefore, coding gain = 10 log(4.586/2) = 3.6dB. Natural mapping thus gives a coding gain of 3.6dB

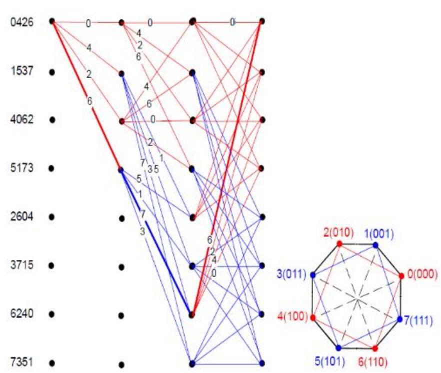


Figure 19: Trellis of rate ½ code with 8 eight states[18]

With more states the four paths are assigned to other states to avoid parallel transitions. From the figure above, S6, S7, S6 appear to have the minimum distances from each branch compared to all zero state. Therefore, the free distance, d_{free} = 2.0 + 0.586 + 2.0 = 4.586

Coding gain = 10 x log (4.586/2) = 3.6 dB

There is an improvement of 0.6 dB as a result of the use of 8-state against 4-state trellis structure. This shows that the use of more states of 8-PSK improves coding gain as depicted in the table below.

No of states	Coding gain (dB)
4	3.0
8	3.6
16	4.1
64	4.8
128	5.0

VI. TRELLIS CODING FOR QAM

A. TCM for QAM

In the same way that set-partitioning is applied to M-ary PSK so it is applicable to M-ary QAM. Consider a coded 16-QAM having three information bits per modulation interval and making reference to an uncoded 8-PSK system. A 16-QAM will need three uncoded incoming bits of which two will be coded to give thee bit output i.e rate 2/3 encoder. See figure 21.

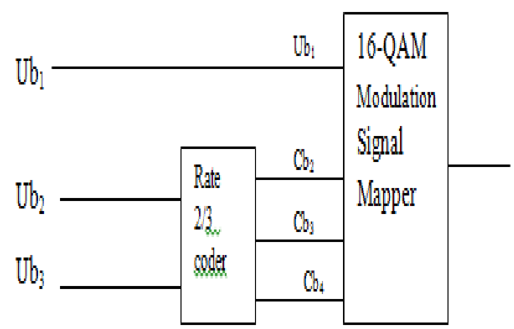


Figure 20: Rate 2/3 convolutional coder for 16-QAM

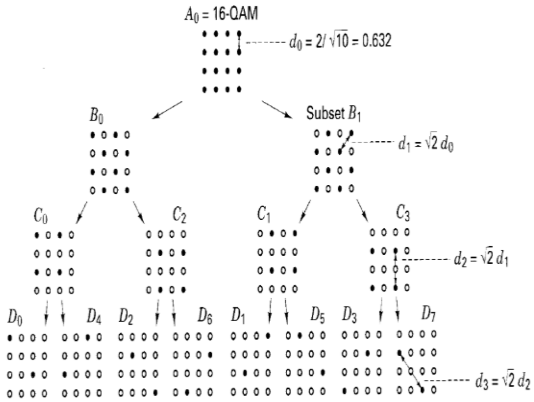


Figure 21: set partitioning of 16-QAM constellation [3]

The table 5 illustrates the coding gain of different states of 16-QAM signal set.

Table 1: Number of states with its Coding gain

No of states	Coding gain (dB)
4	4.4
8	5.3
16	6.1

B. The Pragmatic TCM

This is a type of TCM, a 64 state standard rate ½ convolutional codes for communications as against different rate code for each version of TCM. This was suggested by Viterbi. Its advantage is that with slight modifications it can be used to decode both the coded and uncoded bit. Pragmatic approach is not as effective as the set-partitioning approach.

VII. MULTIDIMENTIONAL TCM (SIGNAL CONSTELLATIONS)

The shape of the signal constellation affects positively or negatively the constellation design. So circular constellation has a lower variance of the data symbols than the square constellation, hence a circular shape is appropriate. The resulting reduction in signal power is called shaping gain.

A second approach to improving a constellation is the use of hexagonal constellation in which points fall on a grid of equilateral triangles, also reduces the variance for the same minimum distance or increase in minimum distance for the same power through changing the relative spacing of the points is called coding gain.

Coding and shaping gain can be combined by changing the points in the circularly shaped constellation to a hexagonal grid while retaining the circular shaping. Neither shaping nor coding gain is feasible in one dimension, but both are available in two dimensions. In the context of trellis code, the benefit of using multidimensional constellation yields additional coding and shaping gain than with 2-dimentional constellation. It is observed that doubling the symbol alphabet is sufficient to achieve almost the available coding gain determined by Shannon limit. However, doubling the constellation size for the same minimum distance will increase the signal energy. The coding gain overcomes this disadvantage. The continuous approximation predicted that doubling the size of the constellation increases the signal energy P, per two dimensions by $2^{2/N}$ for an N-dimensional constellation. Thus the dimension of the constellation increases, the power penalty decreases.

E.g. $P_{(dB)} = 10 \times \log_{10} (2^{2/N})$

Thus it deceases from 3 dB for a two-dimensional constellation to just 1 dB for a six-dimensional constellation. However, multidimensional constellations suffer from a complexity that increases exponentially with dimensionality. This could be checked (mitigated) by combining it with trellis codes. BPSK and PAM are one dimensional (1D) while QPSK is 2-BPSK, hence 2 dimensional modulation and is called 2LD-MPSK-TCM, where dimensionality factor L = 1. So “L” denotes “L” dimensions of 2D MPSK signals. Another way multi-dimensional is referred is by L x MPSK e.g. 2 x MPSK or n x MPSK i.e. n symbols. So 4 x MPSK = 8D-TCM.

The main concept in multi-dimensionality is increasing the number of symbols created in one processing period (i.e. multi- processing) that allows or enhances better performance [4].

Advantages

1. Fractional information rates can be transmitted. Instead of rate 2/3 code in 1 x 8PSK, we have 5/6, 8/9, 11/12. This reduces code overhead by affecting more than one symbol.
2. Better bit efficiency is possible (i.e. ratio of number of input information bits to the no of symbols transmitted per processing period).
3. Smaller peak to average ratio.
4. No additional hardware complexity as multi-D TCM uses pragmatic version of TCM.

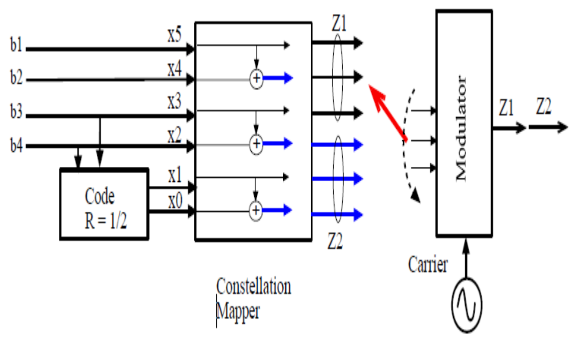


Figure 22: 2x8-PSK, code rate 4/6[7]

There are 4 input bits, 2-coded and uncoded bits respectively to produce 4 bits that produces 6 bits output of the constellation mapper. The constellation mapper uses an algorithm to reorder these bits to output 2 symbols as shown in figure 23.

The 2 inputs through a rate ½ encoder generate 2 parity bits which mean that we have 4 information bits and 2 symbols per processing period.

Therefore, bit efficiency = 4/2 =2 bits/symbol.

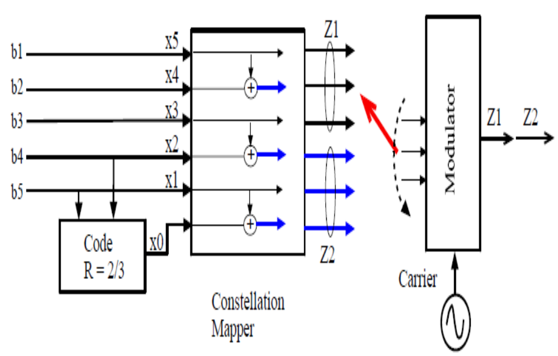


Figure 23: 2 x 8PSK, code rate 5/6 [11]

Bit efficiency = b_1 to b_5/z_1+z_2 = 5/2 = 2.5 bits/symbol

D. COSET Codes:

Comparison of the performance of trellis coded systems against uncoded systems were made with the same transmit energy and spectral efficiency but the sources of improvement – performance were not properly accounted for. In particular, they mix coding gain due to constellation design with that due to the FSM. Furthermore, the methods do not scale well large constellations and to multidimentional constellations. Trellis codes based on coset partition are called coset codes by Forney. Coset codes allows us to separate coding gain due to the lattice, the shaping gain and the additional coding gain due to the FSM. See figure 25.

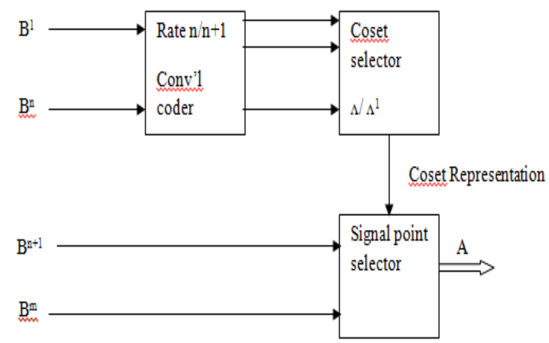


Figure 24: A view of trellis coding that enables deeper understanding of the sources of performance gain.

There are potentially three distinct sources of performance gain:

- the lattice Λ may have coding gain
- the signal point selector will use some shaping region S that may have shaping gain
- the convolutional coder will allow only a particular sequence of cosets to be sent to the signal point selector, and thus introduces its own coding gain by increasing the minimum distance between sequences [4].

E. Coding Gain Due to Redundancy

The convolutional coder allows only a subset of all possible sequences of cosets. Thus, there is redundancy in such sequences. This means that the minimum distance between any two allowable sequences will be larger than the minimum distance between pairs of points in the lattice. Let $d_{min}(C)$ be the minimum distance between any two sequences of cosets allowed by the convolutional coder C where the distance between two cosets is taken to be the minimum distance between any two point is one coset and any point in the other. The minimum distance will dominate the performance of the overall system. Let d_{min} (ʌ) be the minimum distance between any two points in ʌ. Then the convolutional coder has the minimum distance by a factor of $d_{min}(C)/ d_{min}(ʌ)$.

There is a price paid for, however, for this increase in minimum distance. Since there is one more bit emerging from the convolutional coder than going into it, twice as many points in the lattice will be needed within the shaping region to transmit the coded signal. Thus the size of the shaping region will increase. This constellation expansion reduces the gain, since it increases the energy.

Combining the positive and negative effects of the convolutional coder, we get the coding gain due to the convolutional coder which is simply the increase in minimum distance due to the convolutional coder divided by the increases in energy (per two dimensions) due to the constellation expansion [4]. Expressed as

$$Y_C = \frac{d^2\,\min(C)}{d^2\,\min(} (\Lambda) \mathrm{x}\ 2^{P(C)}$$

REFERENCES

[1] Andreas F. M. (2005). Wireless Communications. England: John Wiley & Sons ltd.

[2] John R. B., Edward A. L., Messerschmitt,D. G. (2004). Digital communications. 3rd ed. USA: Kluwer Academic publishers.

[3] Haykin, S. S. (1988). Digital Communications. Canada: John Wiley & Sons Inc.

[4] Otung, I. E. (2001). Communication Engineering Principles. UK: Palgrave MacMillian

[5] http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=31452&tag=1

[6] Andrew J. V., Jack K. W., Ephraim Z., Roberto P ‘’A Pragmatic Approach to Trellis-Coded Modulation’’.

[7] www.complextoreal.com/chapters/tcm.pdf: Accessed 24/08/2012

[8] Telecommunications Instructional Modelling System (TIMS). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1093542&tag=1

[9] Trellis-Coded Modulation with redundant signal sets part 1 introduction by G. Ungerboeck.”

[10] Ungerboeck, G. (1987). Trellis Coded Modulation. IEEE Communications Magazine. 25 (1), 20.

[11] Otung, I E (2010). Short Course In Digital Communication. UK: University of Glamorgan