PART I

FOUR-DIMENSIONAL Q^2PSK: Theory and Application to mobile communication

CHAPTER 2

FOUR-DIMENSIONAL Q^{2}PSK SIGNALLING: THEORETICAL FOUNDATION

In this chapter the concepts of four-dimensional Quadrature Quadrature Phase-Shift Keying (Q^{2}PSK) are considered, as a means of improving communication efficiency on the bandlimited Gaussian channel. Investigations into the constraints imposed by finite bandwidth on the dimensionality of the spectrally and power efficient Q^{2}PSK modulation strategy are carried out. In particular, attention is focused on the derivation of a suitable four-dimensional orthonormal set of basis functions for, and channel capacity of Q^{2}PSK. The capacity is then compared to that of two-dimensional Multiple Phase-Shift Keying (M-PSK) signals.

2.1 FOUR-DIMENSIONAL SIGNALLING

The term signal constellation is defined as the geometric arrangement of symbols within a given signal space. The power efficiency of a signal constellation is defined as

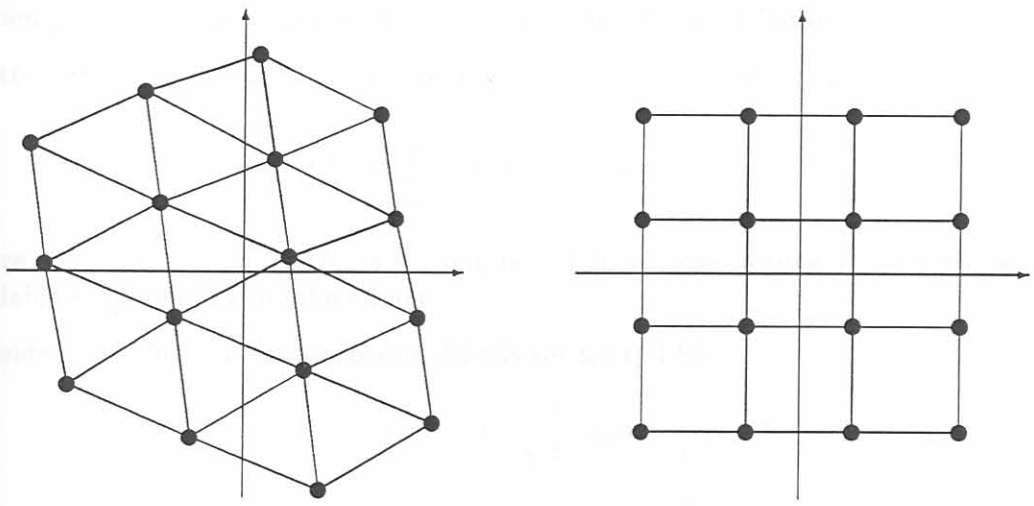
$$\Lambda = \frac{d^2_{free} \cdot R_s}{E_s} \tag{2.1}$$

where d^2_{free} is the minimum squared Euclidean distance between any pair of signals, E_s is the average energy per symbol, and R_s is the symbol rate. The energy efficiency largely depends on the signal space geometry. The main objective of a digital communication system design is to minimise the average energy requirement. Specifically, the Four-Dimensional (4D) design problem is

to locate M signal points in 4D space, $\mathbf{R}^4$, so that for a given minimum Euclidean distance between signals, d_{free}, the average (or peak) energy is minimised. Letting $\overline{s_i}$ denote signal locations and $|\cdot|$ the norm, the problem to be solved is

$$\text{minimise } \frac{1}{M}\sum_{i=1}^{M} |\,\overline{s_i}\,|^2 \tag{2.2}$$

subject to $|\,s_i - s_j\,| \geq d_{free}$, $i \neq j$. This is the classical sphere packing problem for which ample previous work has been done [16, 50]. For illustrative purposes, a Two-Dimensional (2D) arrangement is considered. For large M the best arrangement in 2D places signal points on vertices of equilateral triangles, which tessellate the plane. This is sometimes referred to as a hexagonal lattice, as the decoding regions are hexagons centered at each signal point. For finite M in 2D, the paper by Foschini *et al.* [50] provides optimal constellations as well as symmetrical constellations. For example, the optimum $M = 16$ constellation in 2D has the arrangement shown in Figure 2.1a, while Figure 2.1b illustrates the standard symmetric 16-point Quadrature Amplitude-Shift Keying (16-QASK) design, which may be visualised as a Cartesian product of two One-Dimensional (1D) 4-level Amplitude Modulation (AM) systems.



(a) (b)

Figure 2.1: 16-ary constellations in Two-Dimensions (2D): (a) Optimal 16-ary design. (b) Standard 16-QASK design.

The optimal constellation is more complicated to implement, especially as far as the receiver is concerned. The concepts of the foregoing example may be extended to three and four dimensions, as illustrated in [51]. Related work may be found on Four-Dimensional (4D) modulation in the papers by Welti and Lee [17] and of Zetterberg and Brändström [18]. The Zetterberg and Brändström codes have the property that signal vectors lie on a 4D sphere (equal-energy), whereas Welti and Lee codes are allowed to consume all of the 4D space within the sphere. The equal-energy constraint constitutes a significant penalty as M becomes large, in the same way as M-ary PSK become less efficient than M-ary amplitude/phase modulation in 2D signal space. For the rest of the chapter specific 4D designs are considered, where the symbols coincide with the vertices of a hypercube. A detailed analysis and discussion of these signalling schemes can be found in the book by Wozencraft and Jacobs [22].

2.2 Q²PSK SIGNALLING

In general, the available number of dimensions per second, D, is larger or equal to the symbol rate R_s, $D \geq R_s$, since the number of vertices on a hypercube of $N = DT_s$ dimensions is 2^{DT_s} and the number of signals required is $M = 2^{R_s T_s}$. It is clear that not all the vertices need be used.

The Q²PSK signal space provides a 16-ary constellation in 4D with signals of the normalised form $(\pm 1, \pm 1, \pm 1, \pm 1)$, i.e. the vertices of a hypercube centered at the origin of the signal space [12, 11]. Since the usual association of each of the four bits with ± 1 modulation on a particular dimension applies for detection, minimum bit error probability demodulation can be achieved by means of sign detection in each coordinate direction. As stated in [16], it is difficult to find a 16-ary design that outperforms 4D hypercube signalling, when evaluated in terms of spectral and energy efficiency, as well as implementation complexity.

The double-sided Nyquist-sense definition of bandwidth is adopted throughout this dissertation [22]. It states that theoretically a 4D modulation can transmit $\log_2 M$ *bits per symbol* with a pass band signal bandlimited to a total bandwidth of $W = 2/T_s$, where T_s is the 4D symbol duration. The theoretical spectral efficiency for Q²PSK with $M = 16$, is therefore $\eta_f = \log_2 M = 4$ *bits/s/Hz*. The spectral efficiency depends only on M and not upon the constellation, whereas the energy efficiency does depend on a specific geometry within the constellation.

Restricting the signals $s_i(t)$ to the vertices of a hypercube implies that each signal has the form

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\psi_j(t) \ , \ \text{for } i = 0, 1, \ldots, M-1 \tag{2.3}$$

where $s_{ij} = \pm\sqrt{E_N}$, $N = DT_s$, is the number of dimensions in time T_s, and E_N is defined as the available signal energy per dimension.

Consider the following 4D orthonormal basis set for Q²PSK:

$$\begin{aligned}
\psi_1(t) &= \frac{2}{\sqrt{T_s}} q_1(t) \cos 2\pi f_c t \\
\psi_2(t) &= \frac{2}{\sqrt{T_s}} q_2(t) \cos 2\pi f_c t \\
\psi_3(t) &= \frac{2}{\sqrt{T_s}} q_1(t) \sin 2\pi f_c t \\
\psi_4(t) &= \frac{2}{\sqrt{T_s}} q_2(t) \sin 2\pi f_c t
\end{aligned} \tag{2.4}$$

defined over $| t | < T_s/2$, and zero outside this interval, where

$$\begin{aligned}
q_1(t) &= \cos 2\pi f_d t \\
q_2(t) &= \sin 2\pi f_d t
\end{aligned} \tag{2.5}$$

are the quadrature shaping pulses. Here f_c and f_d are the carrier and deviation frequencies, respectively. In general the deviation frequency is related to T_s as:

$$f_d = \frac{h}{T_s} \tag{2.6}$$

where h is the so-called *deviation ratio*, and $T_s = 1/f_s$ the symbol duration. The basis set $\{\psi_i(t)\}$ forms an orthonormal basis under the restriction

$$f_c = nf_d = \frac{nh}{T_s}, \quad n\epsilon \mathbf{I} \geq 2. \tag{2.7}$$

The minimum value of h is 0.5, corresponding to the minimum frequency deviation necessary for signals to remain orthogonal. In general, any other four mutually orthonormal waveforms could be utilised, provided they conveniently represent the $M = 16$ signals in the 4D signal space.

A difficulty in transmitting sequences of orthonormal pulses is that most physical channels introduce distortion, for instance when undistorted pulses that do not overlap tend to be "smeared", i.e., spread over time greater than T_s. The result, called Intersymbol Interference (ISI), causes loss of orthogonality, leading to a smaller value of D attainable in practice. In practice, the maximum number of essentially orthogonal waveforms that can be transmitted in time T_s through a channel with bandwidth W is limited to between T_sW and $3/2\ T_sW$, where T_sW is the so-called *time-bandwidth product.* With our previous definition of bandwidth ($W = 2/T_s$), the number of orthogonal waveforms is limited between 2 and 3 waveforms.

The orthogonality of the basis set remains invariant under the translation of the origin by multiples of T_s. In other words, if the basis set defined in (2.4) is translated, then orthogonality will be maintained over every interval of T_s centered around $t = mT_s$, with m an integer [12].

2.2.1 Q²PSK Modulation

The orthogonality of $\{\psi_i(t - mT_s)\}$ suggests the modulation scheme which is shown in Figure 2.2 [11, 12]. Data from a binary source at rate $1/T_b$ is demultiplexed into four streams $\{a_i(t)\}, i = 1, \ldots, 4$, with the duration of each data pulse equal to $T_s = 4T_b$. Each data stream is multiplied by the outputs of the basis signal generator $\{\psi_i(t)\}$, producing four mutually orthogonal data streams which are summed to form the modulated Q²PSK signal.

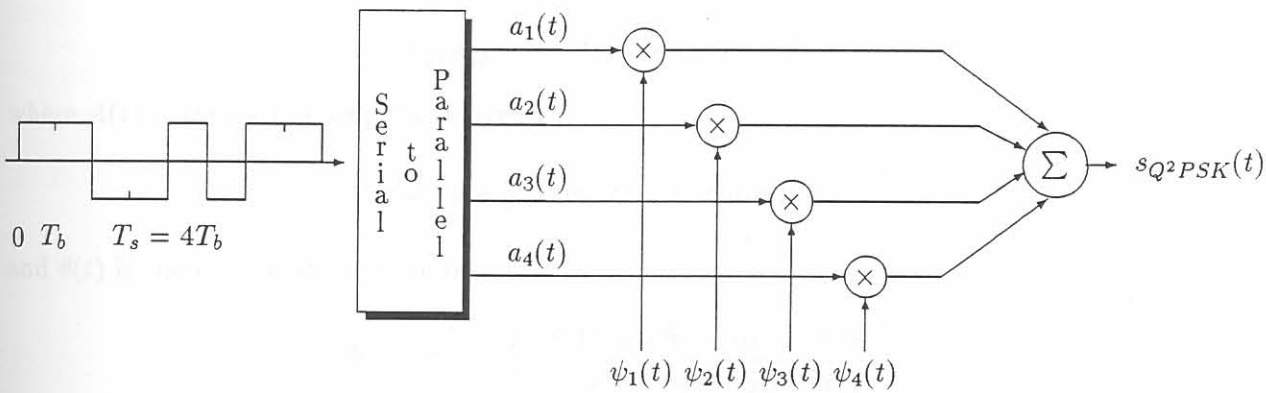


Figure 2.2: Quadrature-Quadrature Phase-Shift Keying (Q²PSK) modulator.

From Figure 2.2 and (2.4), one can represent the Q²PSK signal as

$$s_{Q^2PSK}(t) = \frac{\sqrt{T_s}}{2}\{a_1(t)\psi_1(t) + a_2(t)\psi_2(t) + a_3(t)\psi_3(t) + a_4(t)\psi_4(t)\} \tag{2.8}$$

$$= a_1(t)q_1(t)\cos 2\pi f_c t + a_2(t)q_2(t)\cos 2\pi f_c t$$
$$+ a_3(t)q_1(t)\sin 2\pi f_c t + a_4(t)q_2(t)\sin 2\pi f_c t \quad (2.9)$$
$$= a_c(t)\cos\left[2\pi f_c + b_c(t)2\pi f_d\right]t + a_s(t)\sin\left[2\pi f_c + b_s(t)2\pi f_d\right]t \quad (2.10)$$

where the additional $\sqrt{T_s}/2$ is a normalising factor. Recall, that the deviation ratio is defined as $h = f_d T_s = 1/2$. The relationships between binary quantities $a_c(t), a_s(t), b_c(t), b_s(t)$ and the information data streams, $\{a_i(t)\}_1^4$, are given by

$$\begin{aligned} a_c(t) &= a_1(t), \quad b_c(t) = -a_4(t)/a_1(t) \\ a_s(t) &= a_3(t), \quad b_s(t) = +a_2(t)/a_3(t) \end{aligned} \quad (2.11)$$

The modulating signals $\{\psi_i(t)\}$ affects the bit streams $\{a_i(t)\}$ in two ways. Firstly shaping of the symbol pulse. The second effect is to translate the baseband spectrum to a pass band region. It should be noted that the two pulse trains associated with either of the two carrier are orthogonal over any interval of duration T_s centered around mT_s. Two of the dimensions come from the orthogonality of the carriers, while the other two come from the orthogonality of the data shaping pulses, $q_1(t)$ and $q_2(t)$ (defined in (2.5)). In other words, two carriers and two data shaping pulses are pairwise quadrature in phase. Hence, the name Quadrature-Quadrature Phase-Shift Keying.

At any instant the Q²PSK signal can be analysed as consisting of two signals: the one cosinusoidal with frequencies $(f_c \pm f_d)$, and the other sinusoidal with frequencies $(f_c \pm f_d)$. The separation between the two frequencies associated with either of the two signals is $1/T_s$. This is the minimum spacing that one needs for coherent orthogonality of two FSK signals, as in MSK. The Q²PSK signalling can be thought of as consisting of two minimum shift keying signalling schemes in parallel, which are in quadrature with respect to each other. Since the two schemes are in quadrature, it follows that the bandwidth efficiency will be twice that of conventional MSK.

2.2.1.1 Constant Envelope Q²PSK

One can write the Q²PSK signal, given in (2.8) as

$$s_{Q^2PSK}(t) = A(t)\cos\left(2\pi f_c t + \theta(t)\right) \quad (2.12)$$

where $A(t)$ is the carrier amplitude given by,

$$A(t) = \left(2 + \left[a_1(t)a_2(t) + a_3(t)a_4(t)\right]\sin\frac{2\pi t}{T_s}\right) \quad (2.13)$$

and $\theta(t)$ is the carrier phase given by,

$$\theta(t) = \tan^{-1}\left(-\frac{a_3(t)\cos\frac{2\pi t}{T_s} + a_4(t)\sin\frac{2\pi t}{T_s}}{a_1(t)\cos\frac{2\pi t}{T_s} + a_2(t)\sin\frac{2\pi t}{T_s}}\right) \quad (2.14)$$

The Q²PSK signal, in the absence of any additional constraint, does not maintain a constant envelope. It is well known that constant envelope modulation techniques are desirable, as it makes the modulated signal relatively immune to channel non-linearities. In other words, *hard-limiting* should not degrade phase information. In addition, constant envelope techniques enable the use of automatic gain controllers at the demodulators, and can even tolerate hard-limiting without spectral degradation.

A simple block coding prior to modulation, proposed by Saha [11, 12], provides a constant envelope. The block coding scheme can be described as follows: The coder accepts serial input data and for every three information bits $\{a_1, a_2, a_3\}$, it generates a codeword $\{a_1, a_2, a_3, a_4\}$, such that the fourth bit is an odd parity check for the three information bits, represented by

$$a_4(t) = -\frac{a_1(t) \cdot a_2(t)}{a_3(t)} \tag{2.15}$$

The constant envelope feature is, however, achieved at the expense of a 25% decrease in bandwidth efficiency. This Constant Envelope (CE) Q²PSK scheme will be considered in more detail in Chapter 4.

2.2.1.2 Continuous Phase Q²PSK

In designing a modulation scheme, continuity of phase of the Radio Frequency (RF) carrier may be an additional desirable feature in certain situations apart from the minimum energy and bandwidth constraints. With continuity in phase, high frequency content , i.e., secondary sidelobes of the power spectral density may be significantly reduced. This will bring about a sharper spectrum fall-off, and relaxation of the restrictions on subsequent bandlimiting filters. This is desirable in certain situations where the cost of filtering after modulation is prohibitive and out of band radiation needs to be restricted at a low level. Also, in a bandlimited situation, faster spectral fall-off of the signal itself may result in less ISI and hence, a lower average bit energy requirement for a specified bit error rate.

Considering the expression for the carrier phase of Q²PSK in (2.14), it is noted that the carrier phase does not maintain continuity in phase. The foregoing continuous phase benefits motivates an investigation into the possibilities of achieving phase continuity in the Q²PSK signal. A wide variety of continuous phase modulations are found in the open literature [52, 53, 54]. Many of the better-known techniques, employ some sort of correlative coding which introduces finite memory into the modulated signal. In order to achieve phase continuity, it is necessary only to modify one of the two data shaping pulses, $q_1(t)$ or $q_2(t)$. In other words phase continuity can be achieved without the use of any modulation with memory.

The conventional Q²PSK scheme uses two data shaping pulses. One of them is a half cosinusoid, $q_1(T)$, and the other is a half sinusoid, $q_2(t)$. If one replaces the half sinusoid by a full sinusoid over the same signalling interval, $\mid t \mid \leq T_s/2$, the RF carrier will display phase continuity [12], but at the expense of more bandwidth.

2.2.2 Q²PSK Demodulation

2.2.2.1 Analogue implementation

Implementation of a Q²PSK analogue demodulator was proposed by Saha and Birdsall [11] and later on by De Gaudenzi and Luise [55], where traditional analogue techniques were considered. At the demodulator, four identical coherent generators are available, and the orthogonality of the orthonormal basis set is used to separate the four information bit streams, $\{a_i(t)\}$. In the presence of AWGN, a correlation receiver will perform the process of demodulation in the optimum sense of minimum probability of error sense.

However, the tendency in communication systems is for the utilisation of digital techniques and algorithms for the implementation of more efficient modems; direct translation of analogue techniques to a digital implementation are far from being an optimal approach. Realisation of a fully digital implementation of the modem requires the utilisation of computationally efficient modulation, demodulation, synchronisation and also coding algorithms.

2.2.2.2 Digital implementation

Feiz and Soliman followed DSP techniques to derived a Maximum Likelihood (ML) demodulator for four dimensional modulation schemes. The work of Feiz and Soliman was developed having a Q²PSK modem in mind, and resulted in a DSP solution with reasonable complexity. To this end a conventional DSP based implementation of the Q²PSK modulator and demodulator is considered in this study, proposed and implemented by Acha [23]. In Chapter 4 more detail concerning the realisation of the digital Q²PSK modem will be given.

2.2.3 Spectral efficiency analysis of Q²PSK

The data streams $a_i(t)$ $i = 1, \ldots, 4$ used in (2.8) are assumed to be independent and at any instant each stream can take on either the value $+1$ or -1 with a probability of one half. This implies that in each $T_s = 4T_b$ (second) interval the Q²PSK signal can be one of $M = 16$ possible equally probable waveforms. Let $s_i(t), i = 0, \ldots, M-1$ represent these waveforms. Probability of occurrence of $s_i(t)$ is $\rho_i = 1/M$ for all i. The signal set constituted by $\{s_i(t)\}$ has the following characteristics:

- for each signal waveform $s_i(t)$ of the set, there is also a negative waveform $-s_i(t)$
- the stationary probabilities of $s_i(t)$ and $-s_i(t)$ are equal, and
- the transition probability between any two waveforms is the same.

Such a signalling source is set to be Negative Equally Probable (NEP) [56]. Its overall spectrum is characterised by the absence of a line spectrum and furthermore is independent of the transition probabilities themselves. The overall spectral density is given by [56]

$$S_{Q^2PSK}(f) = \sum_{i=0}^{M-1} \rho_i \mid S_i(f) \mid^2 \tag{2.16}$$

where $S_i(f)$ is the Fourier transform of $s_i(t)$ and is given by

$$S_i(f) = \int_{-\infty}^{\infty} s_i(t)\, e^{-j2\pi ft} dt \tag{2.17}$$

2.2.3.1 Conventional Q²PSK

The time limited pulse shaping waveforms can be written in the following equivalent signal forms:

$$q_1(t) = \sqrt{\frac{2}{T_s}} \cos\left(\frac{\pi t}{T_s}\right) \quad ; 0 \le t \le T_s$$

$$= \sqrt{\frac{2}{T_s}} \text{ rect}\left(\frac{t}{T_s}\right) \cos\left(\frac{\pi t}{T_s}\right) \tag{2.18}$$

(2.19)

$$\begin{aligned} q_2(t) &= \sqrt{\frac{2}{T_s}} \sin\left(\frac{\pi t}{T_s}\right) \quad ; 0 \le t \le T_s \\ &= \sqrt{\frac{2}{T_s}} \text{ rect}\left(\frac{t}{T_s}\right) \sin\left(\frac{\pi t}{T_s}\right) \end{aligned} \tag{2.20}$$

Suppose $Q_1(f)$ and $Q_2(f)$ are the Fourier transforms of the normalised timelimited pulse shaping waveforms, $q_1(t)$ and $q_2(t)$, respectively. Then, it can be shown, using results in [56, 57, 12], that the equivalent baseband version of the Power Spectral Density (PSD) is given by

$$S_{Q^2PSK} = \frac{1}{2}\left[|\, Q_1(f)\,|^2 + |\, Q_2(f)\,|^2\right] \tag{2.21}$$

where

$$Q_1(f) = \frac{\sqrt{8T_s}}{\pi}\left(\frac{\cos \pi f T_s}{4f^2T_s^2 - 1}\right) \tag{2.22}$$

$$Q_2(f) = -j\frac{\sqrt{32T_s^3 f}}{\pi}\left(\frac{\cos \pi f T_s}{4f^2T_s^2 - 1}\right) \tag{2.23}$$

Substituting (2.22) and (2.23) into (2.21), the baseband power spectral density, $S_{Q^2PSK}(f)$, and accordingly its Constant Envelope (CE) version, is given by

$$S_{Q^2PSK}(f) = \left(\frac{8T_s}{\pi^2}\right)(1 + 4f^2T_s^2)\left(\frac{\cos \pi f T_s}{4f^2T_s^2 - 1}\right)^2 \tag{2.24}$$

Expressed in terms of bit duration, $T_b = 1/f_b$

$$S_{Q^2PSK}(f) = \left(\frac{32T_b}{\pi^2}\right)(1 + 64f^2T_b^2)\left(\frac{\cos 4\pi f T_b}{64f^2T_b^2 - 1}\right)^2 \tag{2.25}$$

To provide a means of comparison, the PSDs of MSK and QPSK (with rectangular symbols) signalling schemes, as functions of T_b, are given by [57]:

$$S_{QPSK}(f) = \frac{2T_b}{\pi^2}\left(\frac{\sin 2\pi f T_b}{2f T_b}\right)^2 \tag{2.26}$$

$$S_{MSK}(f) = \frac{16T_b}{\pi^2}\left(\frac{\cos 2\pi f T_b}{16f^2T_b^2 - 1}\right)^2 \tag{2.27}$$

The power spectral densities of MSK, QPSK and Q²PSK are shown in Figure 2.3, as functions of normalised frequency, fT_b.

From these graphs, it is observed that MSK has a wider main lobe (the first null is at $0.75/T_b$), than QPSK (the first null is at $0.5/T_b$), and also wider than Q²PSK (the first null is at $0.375/T_b$). However, the PSD of MSK has lower sidelobes than QPSK and Q²PSK at frequencies removed from the main spectral lobe.

In order to obtain quantitive information about the spectral compactness, a measure of the percentage of total power captured in a specified bandwidth has to be performed. This is plotted in Figure 2.4. For a small specified bandwidth, the percentage power captured in Q²PSK is larger than that in QPSK and MSK, when operating at the same bit rate. Beyond a bandwidth of $1.2/T_b$, the spectral behavior of QPSK and Q²PSK becames almost identical.

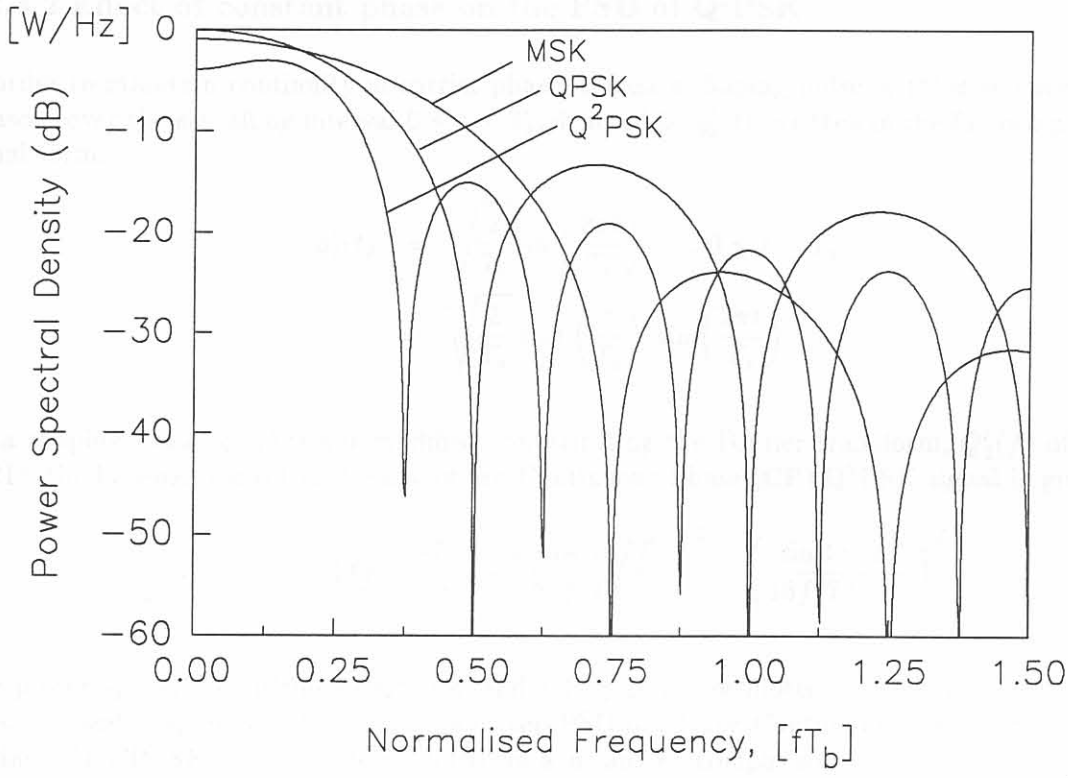


Figure 2.3: Power spectral densities of MSK, QPSK and Q²PSK modulated signals.

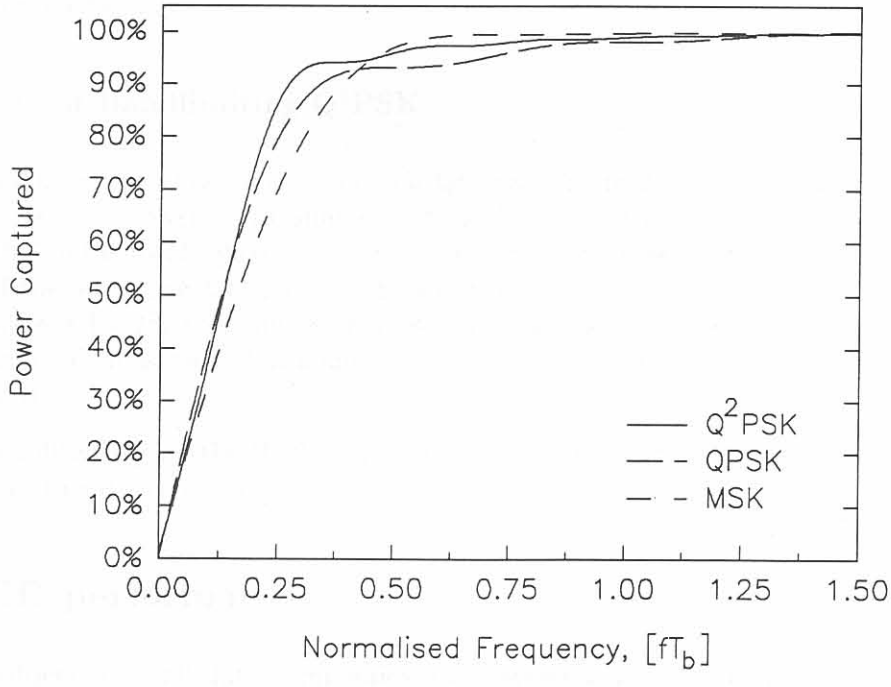


Figure 2.4: Power captured as function of bandwidth of MSK, QPSK and Q²PSK signals.

2.2.3.2 Effect of constant phase on the PSD of Q²PSK

In order to maintain continuity in carrier phase, the data shaping pulse, $q_2(t)$ is replaced by a full sinusoid over the signalling interval $0 \leq t \leq T_s$, denoted by $q_2'(t)$, written in the following equivalent signal form:

$$\begin{aligned} q_2'(t) &= \sqrt{\frac{2}{T_s}} \sin\left(\frac{2\pi t}{T_s}\right) \quad ; 0 \leq t \leq T_s \\ &= \sqrt{\frac{2}{T_s}} \operatorname{rect}\left(\frac{t}{T_s}\right) \sin\left(\frac{2\pi t}{T_s}\right) \end{aligned} \tag{2.28}$$

Data shaping pulse, $q_1(t)$ is not modified. Substituting the Fourier transform, $Q_2'(f)$ of $q_2'(t)$ into (2.21), the baseband spectral density of the Continuous Phase (CP) Q²PSK signal is given by

$$S_{CP-Q^2PSK}(f) = \frac{4T_b}{\pi^2}\left\{4\left(\frac{\cos 4\pi f T_b}{64f^2T_b^2-1}\right)^2 + \left(\frac{\sin 4\pi f T_b}{16f^2T_b^2-1}\right)^2\right\} \tag{2.29}$$

The power spectral densities of Q²PSK and CP-Q²PSK are plotted in Figure 2.5, as functions of normalised frequency, fT_b. In addition, the PSD of M-ary Continuous Phase Frequency Shift Keying (M-CPFSK, $M = 4$) is included as a means of comparison. Figure 2.6 illustrates the percentage of total power captured in a specified bandwidth for Q²PSK, CP-Q²PSK and 4-CPFSK.

In spite of the sharper asymptotic spectrum fall-off, the continuous phase version Q²PSK signal, for a finite bandwidth, captures almost the same power as the original one. The continuous phase version, therefore, does not seem to render any improvement with respect to conventional Q²PSK in terms of energy efficiency.

2.2.3.3 Effects of Bandlimiting Q²PSK

When the spectral compactness of conventional Q²PSK (Figure 2.4) is considered, it is noted that only 90.7% of the total power is transmitted within the Nyquist bandwidth, $W = 2/T_s = 1/(2T_b)$. Thus, nearly 10% of the total signal power will be lost in the process of band pass filtering, resulting in spreading of the baseband data pulses, which in turn causes ISI. For the same finite bandwidth, Continuous Phase (CP) Q²PSK, captures only 88.3% of the radiated power. When a MSK signalling scheme is considered, it is seen that almost the entire signal power (99.05%) is contained within the bandwidth W.

The effect of bandlimiting on the BEP of Q²PSK will be evaluated by means of simulation in Part III, Chapter 7 of this dissertation.

2.2.4 BER performance

The ultimate objective of all data communication systems is to achieve the minimum Bit Error Rate (BER) with a minimum amount of average energy per bit, E_b. In practice, BER performance is usually evaluated under the assumption of a bandlimited channel corrupted by AWGN. The

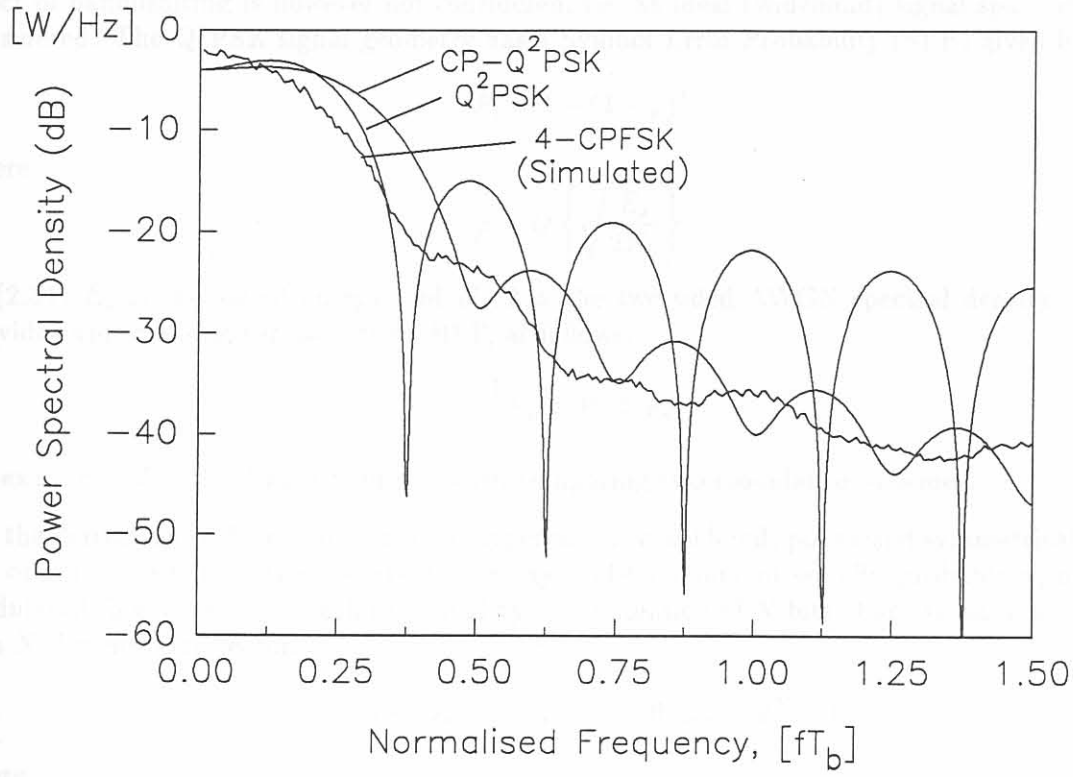


Figure 2.5: Power spectral densities of Q²PSK, CP-Q²PSK and 4-CPFSK modulated signals.

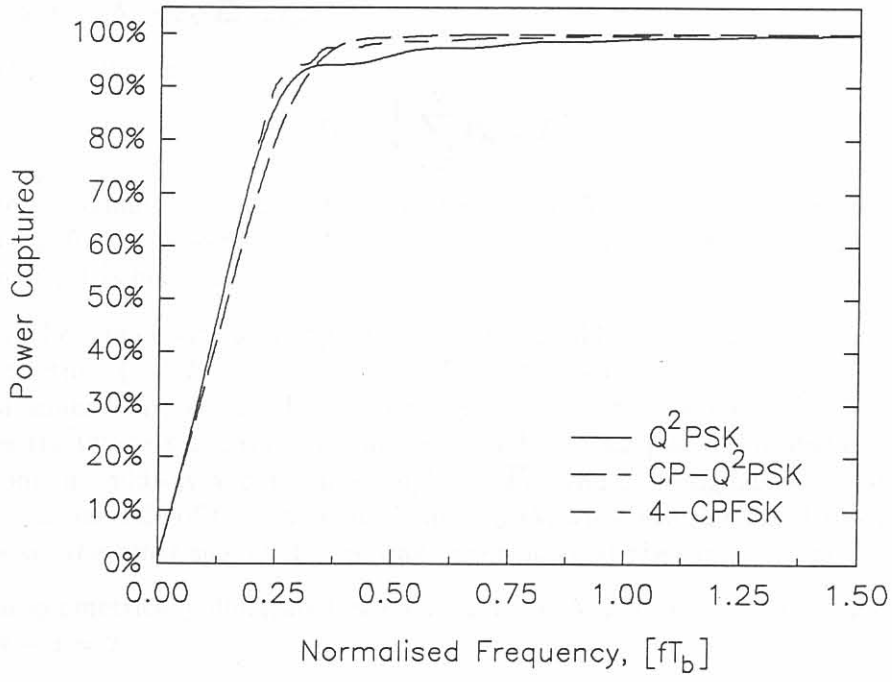


Figure 2.6: Power captured as function of bandwidth for Q²PSK and CP-Q²PSK signals.

effect of bandlimiting is however not considered, i.e. an ideal (wideband) signal space geometry is considered. The Q²PSK signal geometry has a Symbol Error Probability (SEP) given by [22]:

$$P_s = 1 - (1 - \rho)^4 \tag{2.30}$$

where

$$\rho = Q\left\{\sqrt{\frac{E_s}{2N_o}}\right\} \tag{2.31}$$

In (2.31) E_s is the signal energy and $N_o/2$ is the two-sided AWGN spectral density. The SEP provides upper and lower bounds on BEP, as follows:

$$\frac{1}{4}P_s \leq P_b \leq P_s \tag{2.32}$$

An exact calculation of P_b is required when comparing two modulation schemes.

For the derivation of P_b a N-dimensional hypercube is considered, positioned symmetrically around the origin to minimise the average bit energy. The number of equally probable signals in the modulated signal set is 2^N, each presented by a combination of N bits. Each signal may be written as a N-dimensional vector:

$$s_i = (s_{i1}, s_{i2}, \ldots, s_{iN}),\ i = 0, 1, \ldots, 2^N - 1 \tag{2.33}$$

where

$$s_{ij} = \begin{cases} +d/2 \\ \text{or} \\ -d/2 \end{cases} \quad \text{for all } i, j \tag{2.34}$$

represent the projections of the i-th symbol s_i onto the N basis vectors of the N-dimensional signal space, and $d = 2\sqrt{E_s/N} = 2\sqrt{E_b}$ [22].

The average BEP is given by

$$P_b = \frac{1}{N}\sum_{i=1}^{N} P_{bi} = P_{bi} \tag{2.35}$$

where P_{bi} is the probability of error in the i-th bit position. The last equality in (2.35) comes from the equality of P_{bi} for all i, because of the symmetry in signal space geometry. In order to derive an equation for P_b, it is necessary to calculate P_{bi}.

To calculate P_{bi}, the signals within the signal space are divided into two sets: $\{+d/2, s_{i2}, s_{i3}, \ldots, s_{iN}\}$ and its image partner $\{-d/2, s_{i2}, s_{i3}, \ldots, s_{iN}\}$. These two sets of signals will lie on two parallel hyper planes of dimension $(N-1)$. The midway hyper plane of the same dimension is considered, which separates the two sets and is equidistant from each original plane. The distance of any signal in either set from the midway hyper plane is $d/2 = \sqrt{E_b}$. Thus, the signals with $+d/2$ in the first bit position are on one side of this plane at distance $\sqrt{E_b}$, while the signals with $-d/2$ in the first bit position are on the other side of the midway hyper plane at the same distance.

The foregoing is geometrically illustrated in Figure 2.7, for $N = 3$, with midway hyper plane being of dimension $N - 1 = 2$.

An error in the first bit position occurs when the noise component n_0, associated with the first bit position, displaces a signal to the other side of the midway hyper plane. The probability of such

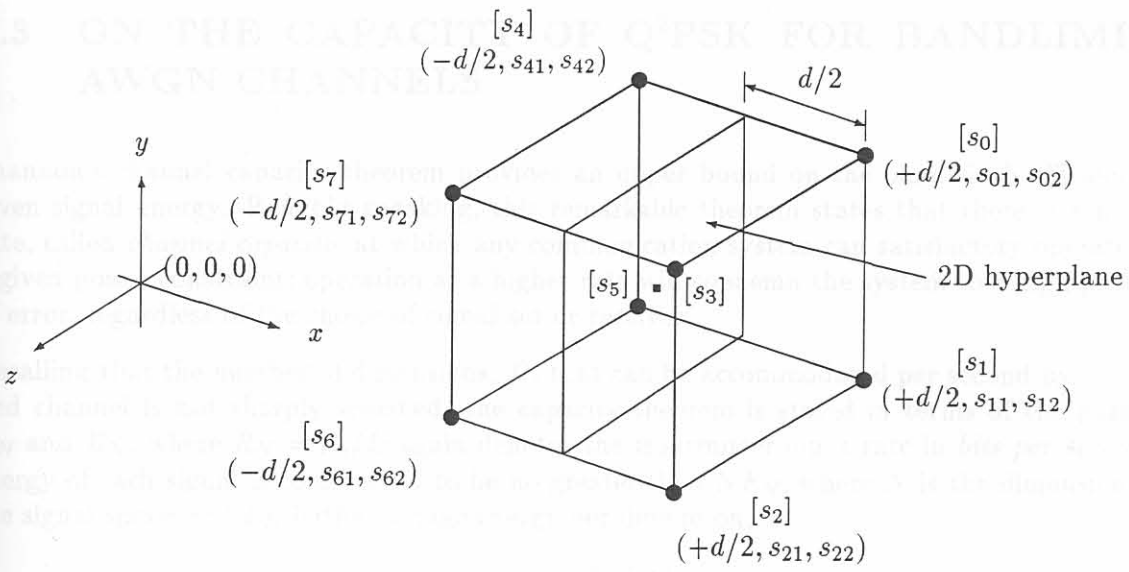


Figure 2.7: Signals on vertices of $N = 3$ dimensional hypercube.

an event is

$$\begin{aligned} P_{b1} &= \int_{\sqrt{E_b}}^{\infty} \rho_{n0}(x) \cdot dx \\ &= Q\left\{\sqrt{\frac{2E_b}{N_o}}\right\} \end{aligned} \tag{2.36}$$

where $p_{n0}(x)$ is the probability density function of Gaussian noise with variance per dimension $N_o/2$, the overall BEP simplifies to

$$\begin{aligned} P_b &= P_{b1} \\ &= Q\left\{\sqrt{\frac{2E_b}{N_o}}\right\} \end{aligned} \tag{2.37}$$

since the conditional probability P_{bi} is equal for all i. This probability of error holds for a hypercube of any dimension N.

Recall, that the BEP of Binary Phase-Shift Keying (BPSK) is exactly the same as the BEP for Q²PSK, given by (2.37). BPSK uses two antipodal signals that can be considered as the vertices of a hypercube of dimension one. Similarly, QPSK and MSK, which use a set of four biorthogonal signals, can be considered as using the vertices of a hypercube of dimension two. Further, it was shown that Q²PSK uses the vertices of a 4D hypercube. Thus, BPSK, QPSK, MSK and Q²PSK belong to the same class of signalling schemes using the vertices of 1D, 2D and 4D hypercubes, respectively. Also, each of them requires the same energy per bit, E_b, to maintain a specific level of communication realiability. The latter is true when the channel is not bandlimited, and corrupted only by AWGN.

In a practical situation channel bandlimiting causes ISI, which has a different effect on each of theses schemes. Since the signal space of the hypercube no longer remains ideal, the energy efficiency is degraded in the case of Q²PSK.

2.3 ON THE CAPACITY OF Q²PSK FOR BANDLIMITED AWGN CHANNELS

Shannon's channel capacity theorem provides an upper bound on the bandwidth efficiency for a given signal energy. Roughly speaking, this remarkable theorem states that there is a maximum rate, called *channel capacity*, at which any communication system can satisfactory operate within a given power constraint; operation at a higher rate will condemn the system to a high probability of error, regardless of the choice of signal set or receiver.

Recalling that the number of dimensions, D, that can be accommodated per second by a bandlimited channel is not sharply specified, the capacity theorem is stated in terms of the parameters E_N and R_N, where $R_N = R_b/D$ again denotes the transmitter input rate in *bits per second.* The energy of each signal is constrained to be no greater than NE_N, where N is the dimensionality of the signal space and E_N is the average energy per dimension,

$$E_N = \frac{\text{joules/s}}{\text{dimensions/s}} \tag{2.38}$$

In the particular case of transmission over an AWGN channel, the capacity theorem may be expressed in the form

$$C_N = \frac{1}{2}\log_2\left(1 + 2\frac{E_N}{N_o}\right) \tag{2.39}$$

C_N is called the *Gaussian channel capacity*, measured in *bits/dimension.*

Let us examine the limits in performance gains that may be achieved when Q²PSK is considered, excluding the effects of ISI (i.e., wideband transmission in AWGN). With perfect timing and carrier-phase synchronisation, samples are taken at time instants $iT_s+\tau_s$, where τ_s the appropriate sampling phase. The output of the modulation channel becomes

$$\overline{z_i} = \overline{a_i} + \overline{w_i} \tag{2.40}$$

where $\overline{a}_i$ denotes a N-dimensional discrete channel signal vector transmitted at modulation time iT, and $\overline{w}_i$ is an independent normally distributed noise sample with zero mean and variance σ^2 along each dimension. The average SNR is defined as

$$\text{SNR} = \frac{E\{|\,\overline{a}_i^2\,|\}}{E\{|\,\overline{w}_i^2\,|\}} = \frac{E\{|\,\overline{a}_i^2\,|\}}{N\sigma^2} \tag{2.41}$$

where $N = 4$ for Q²PSK. When normalised average signal power is assumed ($E\{|\,\overline{a}_i^2\,|\} = 1$), the SNR is simply given by $1/4\sigma^2$.

The capacity, C^* of a Discrete Memoryless Channel (DMC) in the case of a continuous-valued output, assuming AWGN and equiprobable code occurrence, can be written as [27]:

$$C^* = log_2(M) - \frac{1}{M}\sum_{k=0}^{M-1} E\left\{\log_2 \sum_{i=0}^{M-1} \exp(A_{ik})\right\} \tag{2.42}$$

where

$$A_{ik} = -\frac{|\,\overline{a}^k + \overline{w} - \overline{a}^i\,|^2 - |\,\overline{w}\,|^2}{N\sigma^2} \tag{2.43}$$

Using a Gaussian random generator, C^* has been evaluated by Monte Carlo averaging of (2.42) for Q²PSK and Constant Envelope (CE) Q²PSK. Also, the results presented by Ungerboeck [27] for 2D M-PSK modulations are repeated here for comparison purposes. In Figure 2.8, C^* is plotted as a function of SNR. The value at which an uncoded symbol-error probability $P_s = 10^{-5}$ is achieved is also indicated for 4-PSK (QPSK), Q²PSK and its constant envelope version, CE-Q²PSK. These results are summarised in Table 2.1.

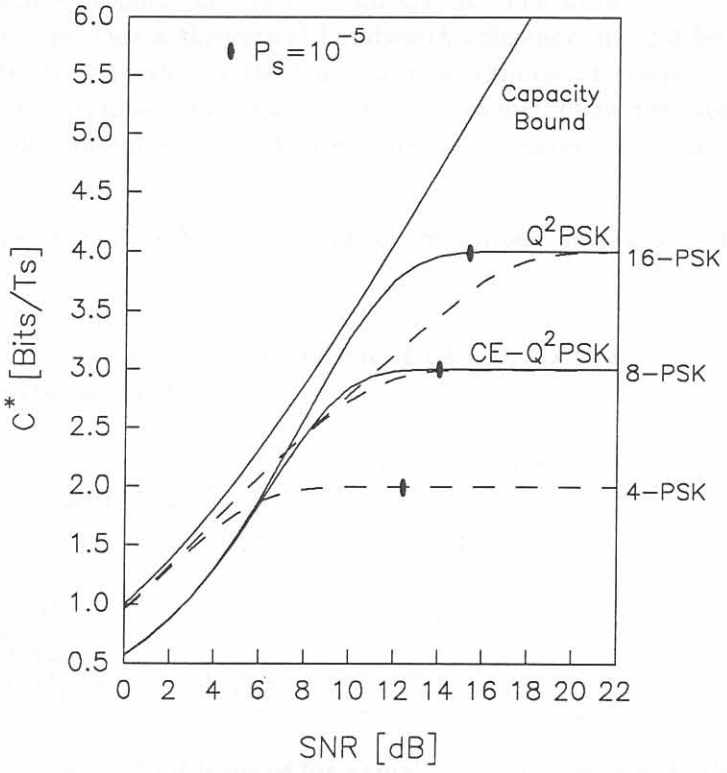


Figure 2.8: Channel capacity C^* of 2D M-PSK ($M = 4$, 8 and 16, denoted by dashed lines) and 4D Q²PSK modulation for bandlimited AWGN channels.

Table 2.1: Comparison of required SNR to achieve a symbol-error probability, $P_s = 10^{-5}$.

Modulation	Required SNR
QPSK	12.9 *dB*
Q²PSK	14.5 *dB*
CE-Q²PSK	13.9 *dB*

To interpret Figure 2.8, consider uncoded 4-PSK operating at 2.0 $bits/T_s$ where $P_s = 10^{-5}$ occurs at $SNR = 12.9$ *dB*. If the modulation is changed from 2D to 4D, it is clear that by considering either Q²PSK or CE-Q²PSK, error-free transmission at 2.0 $bits/T_s$ is theoretically possible at $SNR = 6.9$ *dB*, constituting a gain of 6.0 *dB* compared to QPSK.

The latter observation indicates that if the modulation is changed from 2D to 4D, exploitation of the extra dimensions for coding, can lead to a theoretical gain in the order of 4.8 *dB*. The net result is a coding gain, achieved without alteration of the data throughput. The preceding discussion provides the motivation for 4D modulation.

2.4 CONCLUDING REMARKS: CHAPTER 2

The search for more effective modulation schemes, both from the energy and bandwidth saving viewpoint, is unquestionably one of the relevant trends in modern communication system research. As mentioned before, the key novelty of Q^{2}PSK is the efficient use of degrees of freedom in an expanded multidimensional (4D) signal space. The modulated signal bears a four-fold dimensionality as opposed to the bidimensionality of conventional QPSK. The utilisation of a four-dimensional modulation scheme may provide a theoretical bandwidth efficiency, η_f of 4 *bits/s/Hz*. In terms of spectral efficiency Q^{2}PSK outperforms the traditional workhorse of efficient modulations, MSK, within similar energy constraints. This gain is attained even though the well-known features of phase-continuity and constant envelope of MSK-signals are no longer shared (at least in the uncoded case) by Q^{2}PSK.

These observations are supported in the comparison of various measures of bandwidth listed in Table 2.2.

Table 2.2: Bandwidth comparison of MSK, QPSK, 4-CPFSK, Q^{2}PSK and CP-Q^{2}PSK signalling as a function of power containment.

	Power Containment		
Modulation Type	77%	90%	99%
MSK	$0.29/T_b$	$0.39/T_b$	$0.59/T_b$
QPSK	$0.23/T_b$	$0.34/T_b$	$1.24/T_b$
4-CPFSK	$0.2/T_b$	$0.25/T_b$	$0.45/T_b$
Q^{2}PSK	$0.2/T_b$	$0.27/T_b$	$1.0/T_b$
CP-Q^{2}PSK	$0.32/T_b$	$0.338/T_b$	$1.25/T_b$

Comparing the curves of Figure 2.4 it is noted for a small specified bandwidth, the percentage power captured in Q^{2}PSK is larger than that in QPSK and MSK, when operating at the same bit rate. Beyond a bandwidth of $1.2/T_b$, the asymptotic behavior of QPSK and Q^{2}PSK (also CE-Q^{2}PSK) became almost identical. The continuous phase version CP-Q^{2}PSK signal (see Figure 2.6), in spite of the sharper asymptotic spectrum fall-off, captures almost the same power as the conventional one for a finite bandwidth. For the CP-Q^{2}PSK version the latter is achieved at a 25% decrease in throughput. For this reason it does not seem feasible to pursue this technique further, with respect to conventional Q^{2}PSK in terms of energy efficiency.

When the bandwidth utilisation of conventional Q^{2}PSK is compared to that of 4-CPFSK (depicted in Figures 2.5 and 2.6), it is noted that these techniques exhibit very similar spectrum utilisation within a specified bandwidth. Considering the asymptotic behavior, it is clear that the continuous phase 4-CPFSK definitely provides a better solution in terms of spectral fall-off. However, the inherent phase trellis as a result of the memory employed, restricts the degrees of freedom provided in 4-CPFSK. For this reason, the code structure of Q^{2}PSK is better suited for the application of sophisticated error correction techniques, without having to cope with restrictions on the availability of the extra dimensions.

Considerations of channel capacity of Q^{2}PSK for bandlimited AWGN channels have shown that a theoretical gain of 6.0 *dB* in SNR can be expected when 4D modulation is used in the place of 2D modulation. The latter observation is in effect an indication of the expected theoretical gain when coding is employed. That is, changing the modulation from 2D to 4D and exploiting the extra

dimensions to add coding, a theoretical gain in the order of 6.0 *dB* will result, without altering the data throughput. The 4D Q^2PSK signalling scheme can thus be used as a basis to implement coded transmission.

In the following chapter the Q^2PSK digital communication system is introduced and discussed in detail, including the fading channel model and primary system specifications.