



CHALMERS

Chalmers Publication Library

Power efficient modulation schemes [Invited book chapter]

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

Impact of Nonlinearities on Fiber Optic Communication

Citation for the published paper:

Karlsson, M. ; Agrell, E. (2011) "Power efficient modulation schemes [Invited book chapter]". Impact of Nonlinearities on Fiber Optic Communication

Downloaded from: <http://publications.lib.chalmers.se/publication/129303>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Power efficient modulation schemes

Magnus Karlsson and Erik Agrell

1 Introduction

Coherent optical fiber communications had a brief period of popularity in the early 1990s, mainly because the optical links of that day were significantly power-limited. Coherent detection provided a possibility of optically amplifying the signal to a power level that, after photodetection, made the thermal noise negligible. Two things, however, caused those coherent systems to be abandoned. The first was the sheer technical difficulties: a coherent receiver requires a local oscillator laser that is to be phase- and polarization-locked to the received signal. This gave rise to significant technical obstacles, and only a few limited and expensive coherent receiver solutions were demonstrated [17, 26]. The second was the development of the Erbium-doper fiber amplifier (EDFA) that provided an elegant and practical solution to the problem of the thermal noise. By 1995, the EDFA was a commodity in fiber communication systems, simple on-off keying modulation worked well enough, and coherent communication was forgotten.

However, coherent transmission systems got renewed attention around 2005 [12, 34]. This time the motivation was entirely different. A coherent receiver gives access to both the optical phase and amplitude, which provides two important benefits; (i) advanced multilevel modulation formats can be used, that can improve the spectral efficiency; and (ii) electronic distortion mitigation can be used, as the optical field is directly mapped to the electrical signal. Moreover, the practical problems with the coherent detection could now be solved by performing the phase- and polarization tracking by fast digital signal processing. This enabled a third significant

Magnus Karlsson

Photonics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Göteborg, Sweden, e-mail: magnus.karlsson@chalmers.se

Erik Agrell

Communications Systems Group, Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden e-mail: agrell@chalmers.se

benefit: (iii) a practical use of both polarization components for data transmission. By 2008, a landmark development was reported by Sun *et al.* [51]: the first 10 Gbaud coherent transmission system, with a working coherent receiver based on digital signal processing. In this work, we will investigate modulation formats for such links, which have the peculiarity that the signaling space is four-dimensional.

1.1 Optical coherent modulation—background

An electromagnetic carrier wave offers essentially four degrees of freedom (DOFs) in which data can be independently modulated; the I and Q quadratures (or the real and imaginary parts) of each of the x and y polarization components. These four DOFs can also be interpreted as the amplitude, absolute phase, and polarization state of the wave. We will refer to the number of available DOFs in a transmission system as the *dimensionality*, N , of the constellation space. Binary phase-shift keying (BPSK) requires a one-dimensional constellation space and its higher-dimensional generalizations, quaternary phase-shift keying (QPSK) and dual-polarization QPSK (DP-QPSK), have $N = 2$ and $N = 4$, respectively. These constellations form an N -dimensional cube in their respective constellation spaces.

The polarization state is used for information transmission in fixed microwave communication links, e.g., the Ericsson Mini-Link system, and similar methods have also been considered for mobile radio communications, although impairments like fading and polarization interference pose severe difficulties in the latter case [58].

In coherent optical systems, however, all four degrees of freedoms can be readily detected and used for signaling. And indeed, in recent coherent transmission research this is precisely what is done: a binary modulation in each of the four quadratures, enabling four parallel binary data streams that produce a signal with a data rate that is four times the symbol rate [13, 40, 51, 54]. This modulation format is often referred to as DP-QPSK. It is a 16-level modulation format formed by the vertices of a cube in a four-dimensional (4d) constellation space.

Coherent fiber systems using optical amplifiers can, to a good approximation, be modeled as additive white Gaussian noise (AWGN) channels [23–25, 27, 30], which is important since all fundamental theorems and results of AWGN channels will apply [43]. In order to compare the performance of different modulation formats, we will use the receiver *sensitivity*, which is defined as the signal to noise ratio (SNR) required to reach a bit error rate (BER) or symbol error rate (SER) of 10^{-9} , or, which is increasingly common, 10^{-3} . BPSK is often chosen as a reference format, and is (at least in the optical research community) often believed to have the best sensitivity among all possible modulation formats at a given bit rate. Since the DP-QPSK format is four parallel and independent BPSK channels, its sensitivity is the same as that for BPSK. However, as we will show in this chapter, thanks to the geometrical properties of four-dimensional constellation space, there exist modulation formats that have better sensitivities than BPSK [1, 28]. The improvement comes

from jointly optimizing the constellations over all four DOFs, rather than applying independent modulation in each polarization.

In this paper, we will analyze some of those formats, and quantify their sensitivities within the AWGN model. Besides being of fundamental interest, such power-efficient modulation formats may be of practical relevance as they provide means to reduce nonlinear fiber transmission impairments [28], by allowing reduced transmitter power for the same BER. We will here extend previous studies of modulation formats based on average-energy minimization to peak-energy minimization. As will be discussed in Sec. 5, the peak energy may be more critical than the average in systems limited by fiber nonlinearities such as self- and cross-phase modulation (SPM, XPM). We will give several examples of optimized constellations and present their coordinate representations.

Error correction coding is a way of increasing the dimensionality by introducing more DOFs in the transmitted signal space, however at the price of increased system complexity. In this work, we will limit the discussion to the constellation space of the uncoded modulated signal, which is four-dimensional.

Modulation in a four-dimensional constellation space has been investigated previously in the communication theory literature, e.g. [8, 32, 42, 53, 56, 62]. In [56], constellations with more than 12 levels were analyzed in terms of SER. Some simpler formats, including 5-, 8- and 16-level systems, were analyzed in [62]. For reasons that will be apparent later on in this article, the 5-, 8-, 16-, and 24-level schemes are of most interest.

In the optical communication context, 4d modulation was investigated in the early 1990s [5–7, 16], when coherent systems were popular. These papers demonstrated theoretically how optical transmission systems could benefit from 4d modulation techniques, by showing how transmitters and receivers could be realized. Some fundamental sensitivity limits were given in [5, 6]. However, it is not entirely clear from these works under what circumstances the constellations were optimized (for example under an average or maximum symbol energy constraint). Nor do they point out that sensitivity improvements over BPSK could be achieved, which in our opinion is a most important, and not widely known, observation.

We will give a number of examples of modulation formats (e.g., based on 5, 8, 16 and 24 levels) that have *improved* receiver sensitivities over BPSK and DP-QPSK. Two of these (the 8- and 24-level formats) have a reasonable complexity and, contrary to the 5-level system, the transmitter and the bit-to-symbol mapping problem can be solved without too much loss of performance, so we will describe those modulation formats and their implementations in more detail. It should be noted that we are not the first to point out that multilevel formats with sensitivities better than BPSK exist. Rather, their asymptotic sensitivity gains were originally given in [8, 42, 53]. However, that context was different, as they considered increasing the dimensionality of the signal by using two carrier waves, rather than the two polarization components that can be used in fiber communications.

This chapter is structured as follows: In Sec. 2, we lay out the basic definitions and notation, discuss the relation between polarization states and signals in four-dimensional space, and explain the relation between dense sphere packings and

power-efficient constellations. In Sec. 3 we review sphere packing in two and four dimensions, and present two different optimization principles, (minimization of average and maximum symbol energy, respectively) that we use. Then we present optimum constellations and compare them in terms of sensitivity and spectral efficiency. In Sec. 4 we compute and discuss symbol- and bit-error rates for some of the most promising constellations. In Sec 5 we present fundamental sensitivity limits for the coherent (four-dimensional) channel, and discuss the influence of fiber nonlinearities on the results. We also compare and discuss the two families of optimal constellations we have found in more detail. Finally in Sec. 6 we summarize this chapter.

2 Definitions and system model

This section describes the basic properties of the electromagnetic field and how we interpret it as a four-dimensional signal. Then we will go on to describe how this relates to digital signal transmission, and finally show how sphere packings can be used to find power efficient formats. Much of the material in this section is standard textbook material, but as it is scattered over different texts we wish to include it for completeness.

2.1 The four-dimensional optical signal

As mentioned in the introduction, the electromagnetic field has two quadratures in two polarization components, thus in total four degrees of freedom, which span a 4d signal space. The electric field amplitude of the optical wave can be written as a complex, 2-component vector

$$\mathbf{E} = \begin{pmatrix} E_{x,r} + iE_{x,i} \\ E_{y,r} + iE_{y,i} \end{pmatrix} = \begin{pmatrix} |E_x| \exp(i\phi_x) \\ |E_y| \exp(i\phi_y) \end{pmatrix}, \quad (1)$$

where indices x and y denote the polarization components, and r and i the real and imaginary parts, resp., of the field. The coordinate directions x and y are orthogonal to the propagation direction z . The phases ϕ_x and ϕ_y are by definition in the interval $(-\pi, \pi]$.

The electric field may be equivalently described in terms of its phase, amplitude and polarization state (the latter being the *relative* phase and amplitude between the x and y field components) as

$$\mathbf{E} = \|\mathbf{E}\| \exp(i\phi_a) \begin{pmatrix} \cos \theta \exp(i\phi_r) \\ \sin \theta \exp(-i\phi_r) \end{pmatrix}, \quad (2)$$

where $\|\mathbf{E}\|^2 = |E_x|^2 + |E_y|^2$ and $\theta = \sin^{-1}(|E_y|/\|\mathbf{E}\|)$. J denotes the Jones vector, which is usually normalized to unity, i.e., $J^+J = |J|^2 = 1$. Note the distinction between the absolute phase $\varphi_a = (\varphi_x + \varphi_y)/2$ of the field and the relative phase $\varphi_r = (\varphi_x - \varphi_y)/2$ between the field vector components. The relative phase $\varphi_r \in (-\pi, \pi]$ describes the *ellipticity* of the polarization state, with the special cases $\varphi_r = 0, \pm\pi/2, \pi$ for linear polarization and $\varphi_r = \pm\pi/4, \pm3\pi/4$ for circular polarization, and all other cases are called elliptical states of polarization. The angle $\theta \in [0, \pi/2]$ is usually called the *azimuth* as it describes the orientation in the xy plane of the linear polarization states, or, more generally, the major axis of the polarization ellipse.

A final way of expressing the signal is as a four-dimensional vector \mathbf{s} with real components

$$\mathbf{s} = \begin{pmatrix} E_{x,r} \\ E_{x,i} \\ E_{y,r} \\ E_{y,i} \end{pmatrix} = \begin{pmatrix} \|\mathbf{E}\| \cos \varphi_x \sin \theta \\ \|\mathbf{E}\| \sin \varphi_x \sin \theta \\ \|\mathbf{E}\| \cos \varphi_y \cos \theta \\ \|\mathbf{E}\| \sin \varphi_y \cos \theta \end{pmatrix}. \quad (3)$$

The transmitted optical power is $P = \|\mathbf{s}\|^2 = \|\mathbf{E}\|^2 = E_{x,r}^2 + E_{x,i}^2 + E_{y,r}^2 + E_{y,i}^2$. Note that this four-dimensional vector should not be confused with the Stokes vector description of polarization states, which is defined in a completely different way and proportional to the intensity rather than being linear in the field. The three-dimensional Stokes space was used as a signal space for so-called polarization shift keying modulation in the 1990s [4]. However, the lack of an absolute phase description makes constellation points with different absolute phase but same polarization coincide in Stokes space, and it is therefore less useful as a signal space in a coherent communication system with additive noise (see Sec. 2.2). Yet the Stokes space description of the optical field is useful when discussing the polarization properties of the different modulation formats.

As an example, we consider the DP-QPSK modulation format, which uses independent QPSK modulation in both polarization components, i.e., $\varphi_x = m\pi/4$ and $\varphi_y = n\pi/4$ where $m, n \in \{-3, -1, 1, 3\}$, while $|E_x|$ and $|E_y|$ remain the same for all phases. In the notation of (2), the absolute and relative phases φ_a and φ_r are both multiples of $\pi/4$. The 16 possible combinations are schematically shown in Fig. 1, along with the polarization states they correspond to. Thus, the polarization of DP-QPSK varies between four states; linear in the $+45^\circ$ direction for $\varphi_r = 0$, linear in the -45° direction for $\varphi_r = \pm\pi/2$, left-hand circular (LHC) for $\varphi_r = \pi/4$ or $\varphi_r = -3\pi/4$, and right-hand circular (RHC) for $\varphi_r = -\pi/4$ or $\varphi_r = 3\pi/4$.

2.2 Digital transmission over a noisy channel

In general, all entities in (3) vary continuously with time. For the purpose of digital communications, $\mathbf{s}(t)$ is designed to transmit a sequence of information symbols

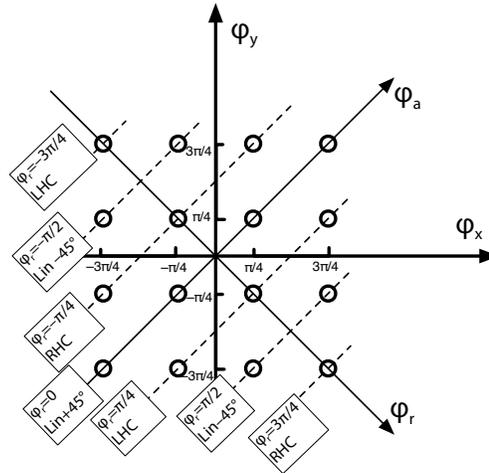


Fig. 1 The phase values used for DP-QPSK modulation. The diagonal axes show the φ_r and φ_a phases. For the φ_r levels, the corresponding states of polarization are denoted as linear $\pm 45^\circ$, LHC, or RHC.

(s_0, s_1, s_2, \dots) , one symbol every T seconds. The symbol s_n is taken from a finite set, or *constellation*, $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_M\}$ of N -dimensional vectors. We assume all constellation vectors to be equally likely. Thus, $\log_2 M$ information bits are transmitted every T seconds, yielding an information bit rate of $R_B = \log_2 M / T$ bits/s.

With linear modulation, $\mathbf{s}(t)$ is generated as

$$\mathbf{s}(t) = \sum_n \mathbf{s}_n p(t - nT), \quad (4)$$

where $p(t)$ is a pulse-shaping function. It may, e.g., be taken as a rectangular pulse of duration T to provide perfect constant-intensity modulation, or a narrower function for RZ pulse shaping. Without loss of generality, we normalize $p(t)$ to unit energy, so that $\int_{-\infty}^{\infty} p^2(t) dt = 1$.

The signal $\mathbf{s}(t)$ is now transmitted over a noisy channel. In the coherent optical systems of today, the dominating noise source is usually either amplified spontaneous emission (ASE) noise from in-line optical amplifiers or shot noise from the local oscillator in the receiver [23, 24, 31]. Both these noise sources are accurately modeled by the additive white Gaussian noise (AWGN) channel, for which the received N -dimensional signal is $\mathbf{r}(t) = \mathbf{s}(t) + \mathbf{z}(t)$, where $\mathbf{z}(t)$ is a vector of N independent, white, and Gaussian noise processes, each with a double-sided spectral density of $N_0/2$ (which is the standard notation in communications literature).

The purpose of the receiver is to recover the sequence (s_0, s_1, \dots) as reliably as possible, given an observation of the signal $\mathbf{r}(t)$. It is well known (see [3, Sec. 2.6] or [39, Sec. 5.1]) that in the absence of inter-symbol interference, the optimal receiver operates by filtering $\mathbf{r}(t)$ and sampling, creating a sequence of so-called *received*

vectors $(\mathbf{r}_0, \mathbf{r}_1, \dots)$ where

$$\mathbf{r}_n = \int_{-\infty}^{\infty} \mathbf{r}(t + nT)p(t)dt. \quad (5)$$

It can be shown that $\mathbf{r}_n = \mathbf{s}_n + \mathbf{z}_n$, where \mathbf{z}_n are independent, Gaussian random vectors with variance $N_0/2$ in each dimension. This equation is a *discrete-time channel model*, which includes modulation, optical transmission, and demodulation. It should not be confused with its continuous-time counterpart $\mathbf{r}(t) = \mathbf{s}(t) + \mathbf{z}(t)$. For instance, the average of the squared field amplitude $\|\mathbf{s}(t)\|^2$ is the optical transmitted power P , while the average of $\|\mathbf{s}_n\|^2$ equals the *average energy per symbol*

$$E_s = \frac{1}{M} \sum_{k=1}^M \|\mathbf{c}_k\|^2 = PT \quad (6)$$

assuming that each symbol in the set is transmitted with the same probability. We also find it useful to define the *maximum energy per symbol* as

$$E_{s,max} = \max\{\|\mathbf{c}_1\|^2, \dots, \|\mathbf{c}_M\|^2\}. \quad (7)$$

Similarly, while the optical noise power $\|\mathbf{z}(t)\|^2$ is (in theory) infinite, the discrete-time noise energy $\|\mathbf{z}_n\|^2$ is finite and equals on average $NN_0/2$, because each of the N components of \mathbf{z}_n has variance $N_0/2$.

The *spectral efficiency*, SE , is generally defined either as the information bitrate per bandwidth (in bits/s/Hz) or as information bits per channel use, where a ‘‘channel use’’ refers to the transmission of two (or sometimes one) real vectors over the discrete-time channel, i.e., to two (or one) dimensions in signal space [3, p. 219]. We follow the latter approach, defining the spectral efficiency as the number of transmitted bits per polarization, where each polarization represents a dimension pair. Formally,

$$SE = \frac{\log_2 M}{N/2} [\text{bits}/(\text{symbol} \cdot \text{polarization})]. \quad (8)$$

With this definition, BPSK, QPSK, and DP-QPSK all have the same spectral efficiency of 2 bits/sym/pol, which actually makes sense, since BPSK uses only one quadrature, i.e., 1/2 polarization.

2.3 Symbol error rates and sphere packing

If the pulse $p(t)$ is suitably chosen, there is no inter-symbol interference and \mathbf{s}_n can be optimally estimated from the single received vector \mathbf{r}_n . The AWGN model means that the received vector \mathbf{r}_n has an isotropic distribution around \mathbf{s}_n in an N -dimensional space, and for a maximum likelihood receiver, the symbol decision is based on which signal in the constellation set is closest (in the Euclidian sense) to the received vector. To put this on more solid mathematical grounds, consider

the constellation $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_M\}$ of M signaling points, or symbols. Each symbol \mathbf{c}_k is surrounded by a decision region, also known as a *Voronoi region*, defined as all points in the N -dimensional Euclidean space that are closer to \mathbf{c}_k than to any $\mathbf{c}_j \neq \mathbf{c}_k$. The probability of receiving symbol \mathbf{c}_k in error is then the probability for a Gaussian variable centered at \mathbf{c}_k to be outside the Voronoi region. For constellations in many dimensions, this probability can in general not be calculated exactly, since the Voronoi regions may have very complex shapes.

However, a simple, yet useful, approximation to the SER is the *union bound*. It builds on the fact that the *pairwise* error probability of confusing the symbols \mathbf{c}_k and \mathbf{c}_j is easy to calculate—it is simply a function of the distance $d_{kj} = \|\mathbf{c}_k - \mathbf{c}_j\|$. The overall SER of a symbol \mathbf{c}_k is then upperbounded by the sum of these pairwise error probabilities over all $j \neq k$. Finally, averaging over all equiprobable symbols \mathbf{c}_k , the union bound on the SER can be expressed as [3, p. 191]

$$SER \leq \frac{1}{M} \sum_{k=1}^M \sum_{\substack{j=1 \\ j \neq k}}^M \frac{1}{2} \operatorname{erfc} \left(\frac{d_{kj}}{2\sqrt{N_0}} \right), \quad (9)$$

where erfc denotes the complementary error function. This bound is in most cases sufficiently accurate at large SNR, and it approaches the true SER asymptotically. We will show numerically later on that it, in our cases, agrees well with exact results for SERs less than 10^{-3} .

We may see directly from (9) that in the limit of high SNR (and low SER), the errors will be dominated by the signals in the set that are closest together, i.e., the term containing $\operatorname{erfc}(d_{\min}/2\sqrt{N_0})$, where $d_{\min} = \min_{j \neq k} \{d_{kj}\}$ is the *minimum distance* of the constellation. Therefore, a judicious selection of signaling levels \mathbf{c}_k that minimizes the average energy per symbol E_s without decreasing d_{\min} is crucial for a modulation format to perform well. This selection is equivalent to the problem of packing M N -dimensional spheres so that E_s (which is equal to the average second moment of \mathbf{c}_k) is minimized. In fact, on a more fundamental level, most coding and modulation problems for AWGN-limited systems may, in the high-SNR regime, be reformulated as sphere-packing problems. Unfortunately, while such sphere packing problems are often easy to formulate, they are notoriously difficult to solve analytically, and one must often resort to numerical optimization techniques to find the best constellations.

We now wish to compare the performance of constellations with different numbers of levels M at a fixed bit rate R_B . We therefore rewrite the dominant term in (9) as

$$\operatorname{erfc} \left(\sqrt{\gamma \frac{P}{R_B N_0}} \right) = \operatorname{erfc} \left(\sqrt{\gamma \frac{E_b}{N_0}} \right), \quad (10)$$

where

$$\gamma = \frac{d_{\min}^2}{4E_b} \quad (11)$$

and $E_b = P/R_B = E_s/\log_2 M$ is the average *energy per bit*. In the following, we will refer to both E_s/N_0 and E_b/N_0 as the SNR, depending on the context. The parameter

γ , which captures the constellation's influence on the SER and is usually given in dB, is called the *asymptotic power efficiency* [3, p. 220], because the power needed for a certain required SER, still at asymptotically high SNR, is proportional to $1/\gamma$. Another interpretation of γ is as the *sensitivity gain* over BPSK to transmit the same data rate, since $\gamma = 0$ dB for BPSK, QPSK, and DP-QPSK.

In fact, most common modulation formats have a penalty with respect to BPSK; for example, M -PSK and M -QAM have [3, pp. 226, 234]

$$\gamma_{M\text{-PSK}} = \sin^2(\pi/M) \log_2 M, \quad (12)$$

$$\gamma_{M\text{-QAM}} = \frac{3 \log_2 M}{2(M-1)}, \quad (13)$$

where (13) is valid for M being a power of 4. We can show from these expressions that both M -PSK and M -QAM have efficiencies $\gamma \leq 0$ dB for all values of M (with the notable exception of 3-PSK, which will be discussed in the next section).

The first general investigation on how the SER depends on the dimensionality N , the constellation size M , and the SNR was done by Shannon in 1959 [44]. By using geometrical sphere-packing arguments, he managed to obtain upper and lower bounds on the SER under rather general conditions. While Shannon's objective was to quantify the performance of capacity-approaching coded systems, our focus in this paper is on uncoded transmission, i.e., low-dimensional constellations, in particular $N = 2$ and 4.

Specifically, we will consider the question: At a given dimension N , and constellation size M , and asymptotic SNR, which modulation format (constellation) has the highest asymptotic power efficiency γ ? Quite surprisingly, this issue was not addressed until recently by us [1, 28] and then only when minimizing the average symbol energy E_s . As noted early [44], minimizing the maximum energy $E_{s,max}$ is also a relevant problem. In the next section, we will therefore present results for both average-energy and maximum-energy minimization.

3 N -dimensional sphere packing results

Before presenting the main results, we will give a brief historical background and introduction to the area of sphere packing.

3.1 *Sphere packings—background*

As we noted in Sec. 2.3, the problem of finding the constellation with maximum asymptotic power efficiency is equivalent to finding the densest packing of M N -dimensional spheres. Here “densest” can be interpreted either as a minimization of the maximum distance from the origin, or as a minimization of the average squared distance from the origin, as mentioned above. In this chapter, we will refer to a sphere-packing constellation designed to minimize the average squared distance as a *cluster* and one designed to minimize the maximum distance as a *ball*¹. It is actually challenging enough to find the best constellations for a fixed number of levels M in a given dimension N . In general, no formal mathematical proof that a certain constellation is the densest is known, and conclusions are rather supported by empirical evidence in the sense that “no better constellations have been found.” In reality, sphere packing optimization often involves the creation of thousands of dense constellations (and various efficient algorithms for this have been proposed), and then selecting the best among these. For high dimensionality and constellation sizes, this can be quite demanding.

For planar clusters, some conjectured optimal constellations were originally presented by Foschini *et al.* [20] for selected values of M up to 16. They are typically hexagonal packings of M circles centered around the origin. This was further demonstrated by Graham *et al.* [22], who numerically computed conjectured optimum packings up to $M \leq 100$ in the plane and even larger constellations ($M \leq 500$) with a suboptimal, greedy technique. In $N = 3$ dimensions, the best known sphere packings, including images of the cases $M \leq 20$, were originally reported by Sloane *et al.* in [46]. Their work has been updated and extended to tables of the best known packings for $N = 3$, $M \leq 99$ and $N = 4$, $M \leq 32$, which are available online [47]. Some early work on ball optimization were reported by Lachs [33], but limited to 10 points in 3 and 4 dimensions. Also, other tables based on numerical optimization have been reported, e.g., in [38], but it is noteworthy that some of the constellations reported there are inferior to those of [47] (one such example is the case $M = 8$, $N = 4$ which is of particular interest to us). We performed our own sphere-packing optimizations for $N = 2, 3, 4$ and $M \leq 16$ that verified the reported values from [47]. For higher dimensions not much is known about good constellations of finite sizes M . Much more is known about the densest infinite-size packings, particularly *lat-*

¹ Mathematically, a “ball” is defined as the set of points in Euclidean space whose distance to a given point is upperbounded by a given constant, i.e., the region bounded by a sphere. “Although physicists often use the term ‘sphere’ to mean the solid ball, mathematicians definitely do not” states Weisstein [55]

tices, for higher dimensions, and most of this work can be found in the extensive review by Conway and Sloane [14].

If the target is to design balls instead of clusters, i.e., to minimize $E_{s,max}$ instead of E_s , the optimization problem can be interpreted as packing M unit-size spheres into a larger sphere, which should be as small as possible. In two dimensions, this problem and its variants have received a lot of attention, as evidenced by Stephenson's extensive bibliography [50]. The best known balls are tabulated by E. Specht for $M \leq 900$ [49]. We are not aware of any published results for $N \geq 3$, but we can derive presumably optimal constellations of moderate sizes based on available results for *spherical codes*.

In a spherical code, all constellation points are required to have the same distance to the origin, and a good spherical code is one where this distance is as low as possible. It is known since the days of Shannon that spherical codes are good for communication over the AWGN channel in very high dimensions [43,44], but this is generally not the case in the low-dimensional applications considered in this chapter. The best known spherical codes are tabulated for $M \leq 130$ and dimensions up to 5 [48]. In this work, we derive balls of size $M \leq K_N + 1$ from spherical codes, where the *kissing number* K_N is the maximum number of nonoverlapping spheres in N -dimensional space that can touch a given sphere with the same size. For two and three dimensions, one has $K_2 = 6$ and $K_3 = 12$, respectively [14], and in four dimensions one has $K_4 = 24$. Like many sphere-packing problems, rigorous proofs of these values are very difficult, and although $K_4 = 24$ was long conjectured [14], it was only recently proven formally [35].

It can be shown that the optimal N -dimensional ball is identical to the optimal spherical code if $M \leq K_N$. Furthermore, if $M = K_N + 1$, we conjecture that the optimal ball is constructed as a spherical code of size K_N with the addition of an extra constellation point at the origin.

As an example of the difference between the maximum and average symbol energy minimization, two-dimensional balls and clusters of size $M = 5$ are shown in Fig. 4.

3.2 Results—sensitivity vs. spectral efficiency

A common way to compare modulation formats [3, 39] is to represent each format as a point in the spectral efficiency vs. sensitivity plane. These sensitivities can be obtained by using the union bound (9) to plot SER vs. SNR as shown e.g. in Fig. 9 in Sec. 4, and then finding the E_b/N_0 required to get a certain SER. This is convenient as it directly shows the SE–sensitivity trade-off, and in addition it can be compared to the Shannon capacity limit which relates the SNR and spectral efficiency as

$$\frac{E_b}{N_0} = \frac{2^{SE} - 1}{SE}. \quad (14)$$

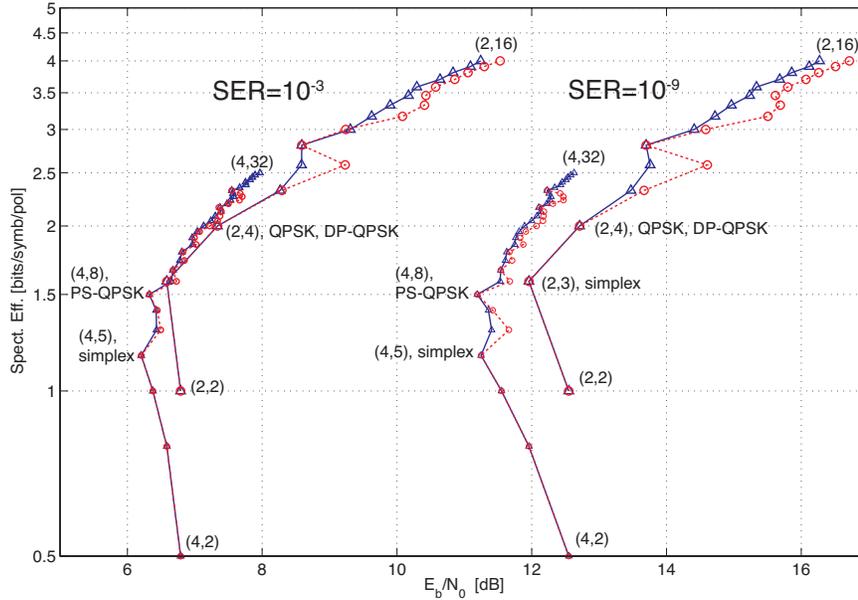


Fig. 2 Spectral efficiency vs. required E_b/N_0 for $SER = 10^{-3}$ and $SER = 10^{-9}$. The optimum constellations are referred to as (N, M) , where N is the number of dimensions and M is the number of points in the constellation. We plot constellations in $N = 2$ up to $M = 16$. In $N = 4$ dimensions, we plot balls (shown as circles connected with dashed lines) up to $M = 25$ as well as clusters (shown as triangles connected with solid lines) up to $M = 32$. Some common modulation formats (QPSK, DP-QPSK) are identical with the optimized $(2,4)$ -constellation. The PS-QPSK format $(4,8)$ is also shown, as are the simplices.

The results are shown in Fig. 2, plotting the optimized constellations for $SER = 10^{-3}$ and $SER = 10^{-9}$. The balls are marked with circles and the clusters with triangles in this graph. One can clearly see the required extra SNR as the SER demand increases to 10^{-9} . Also the difference in sensitivity between the balls and the clusters increases at 10^{-9} , as does the difference between the two- and four-dimensional constellations. It should be noted that the balls will always have a sensitivity penalty relative to the clusters, as we choose to define sensitivity in terms of average energy per bit, E_b . In Sec 5.2 we will show the difference when we use maximum energy per bit, $E_{b,max} = E_{s,max}/\log_2 M$, as a sensitivity measure instead.

Asymptotically, for very low required SERs, the relative difference in sensitivities between the formats approach constant values, although the absolute sensitivity in E_b/N_0 will approach infinity. This situation can be shown by plotting the formats as in Fig. 3 with the (inverse) asymptotic power efficiency on the x-axis. This facilitates a direct comparison between the constellations, as the relative sensitivity differences are approximately the same as in the absolute sensitivity scale of Fig. 2,

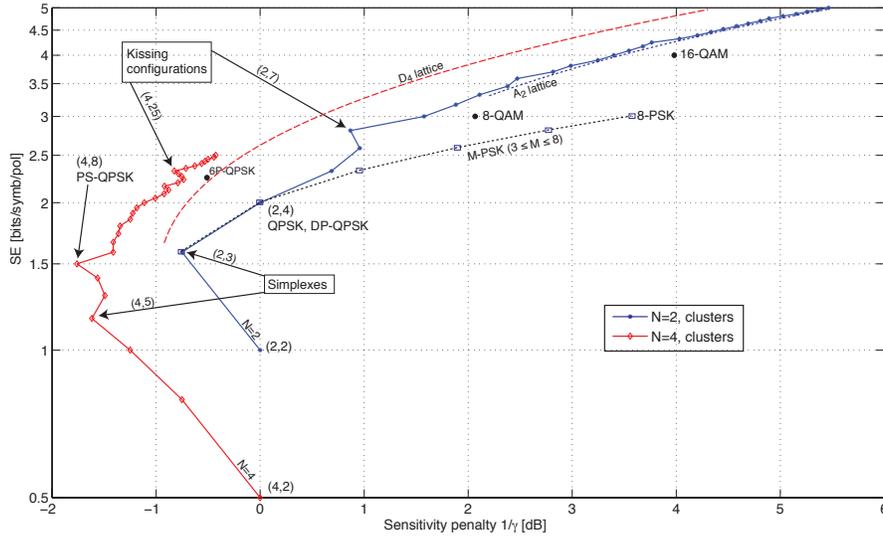


Fig. 3 Spectral efficiency vs. asymptotic power efficiency for $SER = 10^{-3}$. We plot optimized clusters in $N = 2$ and $N = 4$ dimensions. For comparison we also plot the M-PSK, 8- and 16-QAM and 6P-QPSK formats, and the best lattice packings in 2 and 4 dimensions (dashed lines). The optimum constellations have in some cases been marked by (N, M) , indicating dimensionality and number of points.

but the Shannon limit cannot, for example, be included. In this plot we removed the balls from simplicity, but have included some other known formats such as M-PSK, and rectangular 8- and 16-QAM for comparison. We also indicate the *kissing configurations*, i.e., the configurations involving the K_N spheres touching a central sphere, which emerge as local minima for the power efficiency at $M = K_N + 1$ for $N = 2$ and $N = 4$ (but not, e.g., $N = 3$).

As M increases for a given (low) dimension N , the best (densest) packings are known to approach a regular structure called a *lattice*. In two dimensions, the best lattice is generated by placing three circles in a regular triangle (simplex) and extending the pattern indefinitely in all directions. This generates the well-known honeycomb, or hexagonal lattice, usually denoted A_2 . Its density is $\Delta(2) = \pi/(2\sqrt{3}) = 0.91$, which means that the circles cover 91% of the plane. The three-dimensional analogy is the face-centered cubic lattice A_3 , obtained by extending a regular tetrahedron (three-dimensional simplex), with the density $\Delta(3) = \pi/(3\sqrt{2}) = 0.74$. In four dimensions, however, something unexpected happens. Even though a four-dimensional lattice A_4 can be generated from a 4d simplex in perfect analogy with A_2 and A_3 , it is not the densest lattice possible. The densest lattice in four dimensions is denoted D_4 [14], and can be seen as a 4d analogy of the checkerboard pattern. It can be represented by all integer coordinate points such that the coordinates sum to an even integer, and it has the density $\Delta(4) = \pi^2/16 = 0.62$.

The asymptotic power efficiency of a lattice is [14, eq. (32)]

$$\gamma_{\text{lat}} = \log_2(M) \left(1 + \frac{2}{N}\right) \left(\frac{\Delta(N)}{M}\right)^{2/N}, \quad (15)$$

where the densities $\Delta(N)$ are tabulated in [14, Tab. 1.2]. The performance of the densest lattices, A_2 and D_4 , are included as dashed-line asymptotes in Fig. 3.

3.3 Specific formats

In this section we will discuss some of the optimized constellations from Figs. 2–3, and present their coordinates when known. We denote the optimized constellations for M points in N dimensions with $\mathcal{C}_{N,M}$ for clusters and $\mathcal{B}_{N,M}$ for balls. When the coordinates of the constellations are presented, they have been normalized to make the minimum distance between points $d_{\min} = 2$, which corresponds to the packing of unit-radius spheres. We will present both balls and clusters for selected sizes, and emphasize when they are equal, which occurs, we believe, only in a finite number of cases. We will discuss each dimension in turn.

We use the following sources for the best known constellations.

- $\mathcal{C}_{2,M}$ and $\mathcal{B}_{2,M}$ for $N = 2, 4$ and $M = 2, 3, 4$ are M -PSK constellations.
- $\mathcal{C}_{2,M}$ for $M \geq 5$ were designed by Graham and Sloane [22], but the obtained constellations were not reported, only their average second moments. We have reconstructed these constellations based on the conjecture in [22] that they are all subsets of the lattice A_2 .
- $\mathcal{C}_{4,M}$ for $M \geq 5$ were taken from Sloane’s website [47].
- $\mathcal{B}_{2,M}$ for $M \geq 5$ were taken from Specht’s website [49].
- $\mathcal{B}_{4,M}$ for $M \geq 5$ were constructed from the spherical codes in [48] using the methods described in Sec. 3.1.

3.3.1 Two-dimensional constellations, $N = 2$

The two-dimensional clusters are always subsets of the hexagonal lattice, as pointed out in [22]. The two-dimensional balls on the other hand have more irregular structures, and the best known are listed in [49] for $M \leq 900$ (with pictures for $M \leq 804$). The only cases we have found where the balls and clusters are identical are for $M = 2, 3, 4, 7, 31, 55$. We believe these are the only such cases in two dimensions. A property of some balls (but no clusters) is the presence of “loose points,” which are constellations points that are further than the minimum distance from all neighbors and the surrounding circle. Such points can move freely without affecting $E_{s,\max}$, which makes the ball nonunique, and having a continuum of possible average powers E_s . The first loose point arises for $M = 8$ and such points become increasingly common as the constellation size increases. The largest known balls without loose points are $M = 37, 61, 91$. We will below briefly discuss a few two-dimensional balls and clusters of particular interest.

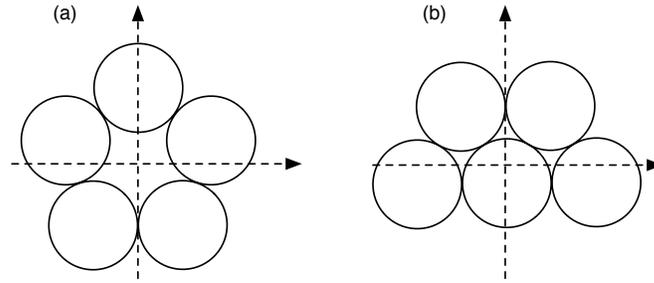


Fig. 4 Optimum five-point constellations in the plane, $(N, M) = (2, 5)$. Minimizing the maximum energy gives the ball $\mathcal{B}_{2,5}$ shown in (a) where all symbols lie on a regular pentagon, and minimizing the average energy gives the cluster $\mathcal{C}_{2,5}$ in (b) which is a subset of the hexagonal packing. This case is further discussed in Sec. 3.3.1.

$M = 2, 3, 4$

These modulation formats are the well-known binary, ternary and quaternary PSK. The clusters and balls coincide for these. The smallest sensitivity over all sizes M is obtained for $M = 3$, and the optimal constellation is the triangle, or *simplex*. It was suggested for modulation in [18, 37] under the name ternary phase-shift keying (3-PSK), and it has a $\gamma = (3/4) \log_2 3 = 0.75$ dB asymptotic sensitivity gain over BPSK. Due to the moderate gain as well as the difficulty of mapping bits to three levels, this format has gained little attention, however. The other constellation points are given by $\mathcal{C}_{2,2} = \mathcal{B}_{2,2} = \{(\pm 1, 0)\}$ for BPSK and $\mathcal{C}_{2,4} = \mathcal{B}_{2,4} = \{(\pm 1, \pm 1)\}$ for QPSK. It is noteworthy that $\mathcal{C}_{2,4}$ is not unique; the constellation points can be continuously deformed to $\mathcal{C}_{2,4} = \{(0, \pm 2/\sqrt{3}), (\pm 1, -1/\sqrt{3})\}$ which is an extension of $\mathcal{C}_{2,3}$ with one point. This constellation is also a cluster, since it has the same E_s [22]. Note also that both BPSK and QPSK have the same power efficiency, 0 dB.

$M = 5$

This is the first case for which the cluster and the ball are not identical. The two cases are shown in Fig. 4. The pentagonal structure, Fig. 4(a), has the same maximum and average energy, $E_s = E_{s,max} = 8/(5 - \sqrt{5}) \approx 2.89$, whereas the hexagonal structure, Fig. 4(b), has average energy $E_s = 68/25 = 2.72$ and maximum energy $E_{s,max} = 112/25 = 4.48$.

$M = 6, 7$

The $M = 7$ constellation is the kissing configuration in two dimensions: six circles touching a unit circle at the origin. The ball and the cluster are identical to this kissing configuration, i.e. $\mathcal{B}_{2,7} = \mathcal{C}_{2,7} = \{(0, 0), (\pm\sqrt{3}, \pm 1), (0, \pm 2)\}$, for all sign

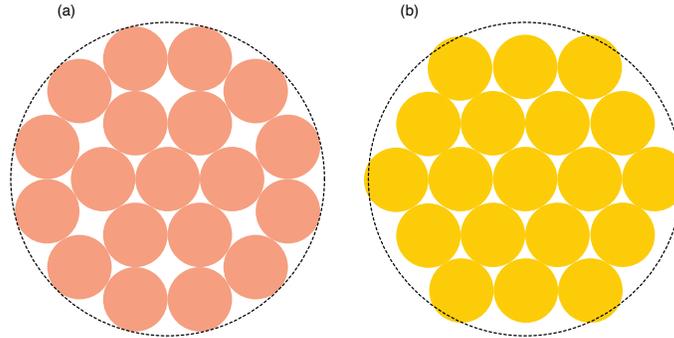


Fig. 5 The ball $\mathcal{B}_{2,19}$ (a) and the cluster $\mathcal{C}_{2,19}$ (b) can be obtained from each other by shifting the outer ring of circles. The smallest circle containing the ball's circles has been drawn around each configuration (dashed), showing that $E_{s,max}$ of the cluster is higher.

combinations. The maximum energy is $E_{s,max} = 4$ and the average energy is $E_s = 24/7 = 3.43$. The asymptotic power efficiency is $\gamma = \log_2(7)/E_s = -0.87$ dB.

The cluster $\mathcal{C}_{2,6}$ is obtained by removing an edge point from $\mathcal{B}_{2,7}$ and recentering the constellation, which gives $E_s = 29/9 = 3.22$. The ball $\mathcal{B}_{2,6}$ is obtained by removing an edge point or a center point, since $E_{s,max} = 4$ irrespective of which point is removed. The average energy will be larger and equals $E_s = 4$ if the center point is removed, which is the choice used in [49] and in the results presented here.

$M = 8, 9$

These balls have both $M - 1$ points in a circle of radius $1/\sin(\pi/(M - 1))$ and a loose point inside this circle.

$M = 15$

This ball consists of a regular structure with 5 inner points in a pentagon and an outer ring of 10 points, arranged so that two outer points touch each inner point.

$M = 19$

The ball and the cluster are different, but very close in structure. Both have hexagonal symmetry, with a $\mathcal{B}_{2,7}$ ball of 7 points in the center, surrounded by 12 outer points. The cluster $\mathcal{C}_{2,19}$ is formed when the outer points form a large hexagon, while in $\mathcal{B}_{2,19}$, the outer points form a circle, as shown in fig 5.

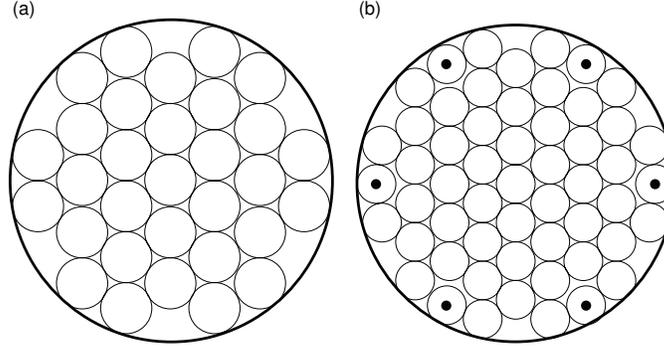


Fig. 6 The constellations $\mathcal{B}_{2,31} = \mathcal{C}_{2,31}$ (a) and $\mathcal{B}_{2,55}$ (b), with coordinates taken from [49]. The cluster $\mathcal{C}_{2,55}$ is obtained by moving the loose points (denoted with black dots) closer to the center, which does not change $E_{s,max}$.

$M = 31, 55$

The two largest known constellations for which the cluster is also a ball can be the same occurs for $M = 31$ and $M = 55$. They are shown in Fig 6. For $M = 55$, the ball has 6 loose points (black) that can be moved without changing $E_{s,max}$. The cluster forces these loose points to lie in the hexagonal lattice.

3.3.2 Four-dimensional constellations, $N = 4$

In four dimensions, the constellations are a bit more difficult to visualize. For $M = 2$ and 4, the clusters and balls are all $(M - 1)$ -dimensional simplices, i.e., 3-PSK and the tetrahedron constellation. We will present some interesting special cases of clusters and balls below, referring to them with the number of points.

$M = 5$

The four-dimensional simplex has 5 points, and is called the *pentachoron*, or *pentatope*, or *5-cell*. It is both cluster and ball. It was discussed in several papers analyzing four-dimensional modulation [6, 8, 32, 53, 56, 62]. Its coordinates can be compactly expressed as

$$\mathcal{C}_{4,5} = \mathcal{B}_{4,5} = \left\{ \sqrt{\frac{2}{5}}(1, 1, 1, 1), -\frac{1}{2\sqrt{10}}(1 - 3\sqrt{5}, 1 + \sqrt{5}, 1 + \sqrt{5}, 1 + \sqrt{5}) \right\} \quad (16)$$

where the second vector should be repeated with all four coordinate permutations. Asymptotically, the pentachoron has a $\gamma = (5/8)\log_2 5 = 1.62$ dB gain over BPSK. As for most constellations in this section, the difficulty of using it for transmission

lies partly in its generation and partly in the difficulty to map bits to 5 constellation levels.

$M = 6$

This is the first instance for which for which the cluster and the ball differ. The cluster, which is the pentachoron plus an extra point, has the coordinates

$$\mathcal{C}_{4,6} = \left\{ \pm \sqrt{\frac{5}{8}}(1, 1, 1, 1), \frac{1}{\sqrt{8}}(-3, 1, 1, 1) \right\} \quad (17)$$

with both signs for the first vector and all four permutations of the second.

The ball is not unique. We use the constellation from [48], whose coordinates can be obtained by rescaling the first vector of (17). After renormalization, this yields

$$\mathcal{B}_{4,6} = \left\{ \pm \frac{1}{\sqrt{2}}(1, 1, 1, 1), \frac{1}{\sqrt{6}}(-3, 1, 1, 1) \right\}. \quad (18)$$

Other, equally good, balls can be obtained by removing any two points from the cross-polytope constellation $\mathcal{B}_{4,8}$ described below.

$M = 7$

Again, the ball is not unique. The constellation in [48] can be identified as

$$\mathcal{B}_{4,7} = \left\{ (\pm 1, \pm 1, 0, 0), (0, 0, \sqrt{2}, 0), (0, 0, -\frac{1}{\sqrt{2}}, \pm \sqrt{\frac{3}{2}}) \right\} \quad (19)$$

with all signs. Thus it consists of four points forming a square in one plane, and three points forming an equilateral triangle in the orthogonal plane. Other versions of the ball can be obtained from $\mathcal{B}_{4,8}$ by removing an arbitrary point.

The cluster $\mathcal{C}_{4,7}$ is obtained from $\mathcal{B}_{4,8}$ by removing any point and shifting the resulting constellation to have zero mean.

$M = 8$

In terms of average bit energy requirements, the cluster $\mathcal{C}_{4,8}$ is the best 4d constellation of any size M , as can be seen from Figs. 2–3. A projection of the constellation is shown in Fig. 7 (a). All its points lie on the 4d sphere, and thus $\mathcal{B}_{4,8} = \mathcal{C}_{4,8}$. Its eight points follow from the biorthogonal representation, which is given by all signs and all permutations of

$$\mathcal{C}_{4,8} = \mathcal{B}_{4,8} = \{(\pm\sqrt{2}, 0, 0, 0)\}. \quad (20)$$

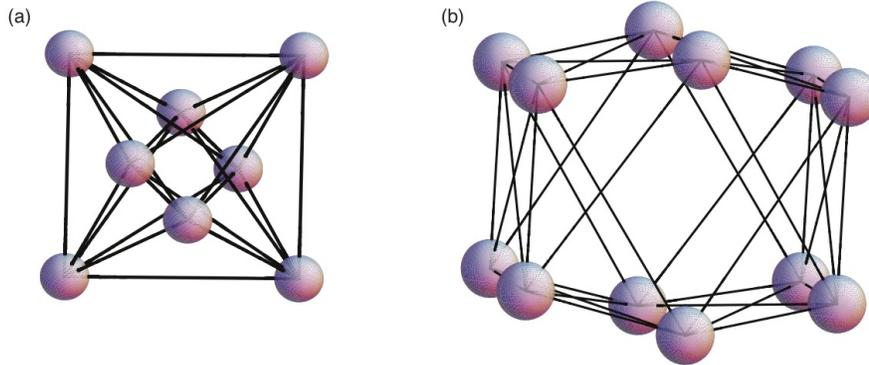


Fig. 7 Projections of the constellations $\mathcal{B}_{4,8} = \mathcal{C}_{4,8}$ (a) and $\mathcal{B}_{4,12}$ (b). The black lines connect nearest neighbors, and they have all the same length in four-dimensional space.

The structure is known as the *cross-polytope*, and it is invariant under a number of symmetries, which simplifies its implementation in a transmission system. A 45 degree absolute phase rotation will bring it into the modified representation

$$\mathcal{C}'_{4,8} = \{(\pm 1, \pm 1, 0, 0), (0, 0, \pm 1, \pm 1)\}. \quad (21)$$

This shows that a modulator based on the cross-polytope can be implemented as QPSK transmission in either x or the y polarization, but *not* both simultaneously as in DP-QPSK [28]. Therefore, we call this modulation format *polarization-switched QPSK* (PS-QPSK).

A third representation is possible as half of the points (e.g., those whose coordinates sum to an even integer) of the cubic (DP-QPSK) constellation. It was described in more detail (including transmitter configurations) in [28]. Since it has only eight levels, its spectral efficiency is reduced to 3 bits per symbol (1.5 bits per polarization), but this is more than compensated for by the minimum distance increasing by a factor of $\sqrt{2}$. Thus the asymptotic power efficiency becomes $\gamma = 3/2 = 1.76$ dB better than DP-QPSK.

$M = 10$

The cluster and ball are identical also here, and this constellation is known as the *rectified 5-cell*, which is formed by the ten points that lie midway between all pairs of points in the 4d simplex. After normalizing, the coordinates can be expressed as

$$\mathcal{C}_{4,10} = \mathcal{B}_{4,10} = \left\{ \frac{1}{2\sqrt{10}}(3 + 3\sqrt{5}, 3 - \sqrt{5}, 3 - \sqrt{5}, 3 - \sqrt{5}), \right. \\ \left. -\frac{1}{\sqrt{10}}(1 - \sqrt{5}, 1 - \sqrt{5}, 1 + \sqrt{5}, 1 + \sqrt{5}) \right\} \quad (22)$$

where the first vector should be taken with its four coordinate permutations and the second vector with its six permutations. This is a rather regular structure, where each point has 6 nearest neighbors at an angular distance of $\cos^{-1}(1/6)$, and the three furthest points all lie at an angular distance of $\cos^{-1}(-2/3)$. The asymptotic power efficiency of this constellation is $\gamma = 1.41$ dB. This structure was originally identified as the optimum by Lachs [33].

$M = 12$

The ball is given by the neat structure

$$\mathcal{B}_{4,12} = \{(\pm a, -b, -b, -b), (\pm a, b, b, -b), (\pm a, b, -b, b), (\pm a, -b, b, b) \\ (0, c, c, c), (0, c, -c, -c), (0, -c, c, -c), (0, -c, -c, c)\} \quad (23)$$

where $a = \sqrt{7/6}$, $b = 1/\sqrt{2}$, and $c = 2\sqrt{2}/3$. As illustrated in Fig. 7 (b), the ball consists of three tetrahedra, uniformly spread along the first coordinate.

The cluster $\mathcal{C}_{4,12}$ is obtained by stretching the middle tetrahedron by about 4 % and then pushing the two outer tetrahedra closer together along the first dimension until all three touch each other. Thus the ball and the cluster have the same symmetries. Graphically, $\mathcal{C}_{4,12}$ looks almost exactly as Fig. 7 (b), with the addition of four more lines representing nearest neighbors. Its coordinates are also given by (23), where in this case $a = 1$, $b = 1/\sqrt{2}$, and $c = (2\sqrt{5} + \sqrt{2})/6$.

$M = 16$

We denote the cubic constellation DP-QPSK with $\mathcal{D}_{4cube} = \{(\pm 1, \pm 1, \pm 1, \pm 1)\}$, with all possible sign selections. This is the most common modulation format in coherent systems, as it is easy to generate and detect. However, it is not a very optimized configuration, neither in an average-energy or maximum-energy sense. The optimum cluster $\mathcal{C}_{4,16}$ is instead a remarkable structure comprising 2 subsets of the D_4 -lattice, with 7 and 9 points, rotated and translated with respect to each other. Its coordinates can be given as

$$\mathcal{C}_{4,16} = \{(a + \sqrt{2}, 0, 0, 0), (a, \pm\sqrt{2}, 0, 0), (a, 0, \pm\sqrt{2}, 0), (a, 0, 0, \pm\sqrt{2}), \\ (a - c, \pm 1, \pm 1, \pm 1), (a - c - 1, 0, 0, 0)\} \quad (24)$$

with all combinations of signs, where $a = (1 - \sqrt{2} + 9c)/16$, and $c = \sqrt{2\sqrt{2} - 1}$. With this representation, which is illustrated in Fig. 8 (a), the cluster can be regarded as four three-dimensional constellations stacked on top of each other along the first dimension: a single point, an octahedron, a cube, and finally another single point. The average symbol energy of this constellation can be expressed as

$$E_s = (279 + 64\sqrt{2} + (7 + 9\sqrt{2})c)/128 = 3.09, \text{ which can be compared to } E_s = 4$$

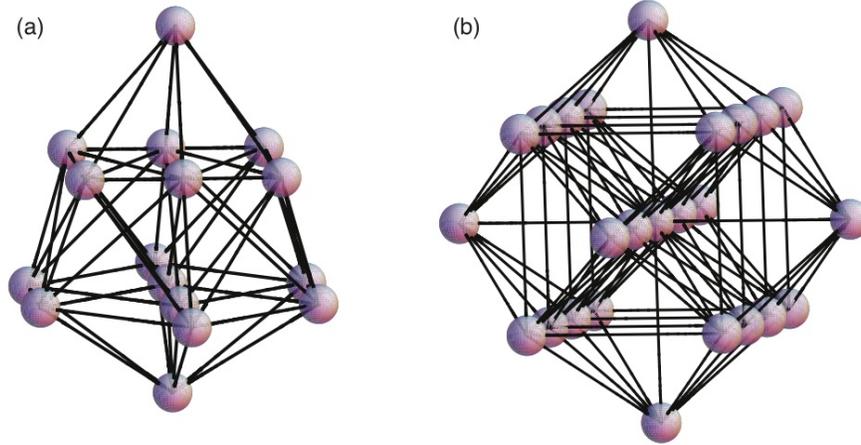


Fig. 8 Projections of the constellations $\mathcal{C}_{4,16}$ (a) and $\mathcal{B}_{4,25} = \mathcal{C}_{4,25}$ (b). The black lines connect nearest neighbors, and they have all the same length in four-dimensional space.

for \mathcal{D}_{4cube} , which in dB units makes the sensitivity of $\mathcal{C}_{4,16}$ 1.11 dB better than DP-QPSK.

The ball $\mathcal{B}_{4,16}$ has no apparent useful symmetries facilitating a nice coordinate representation. Another constant-energy constellation was given in [32] with almost as good performance as $\mathcal{B}_{4,16}$ (having about 0.1 % higher $E_{s,max}$), but the two constellations are geometrically different. This illustrates the occurrence of multiple local minima in numerical constellation optimization.

$M = 23, \dots, 27$

All clusters, and some balls, in the range $M = 23, \dots, 27$ can be derived from the kissing configuration $\mathcal{B}_{4,25} = \mathcal{C}_{4,25}$, which is the four-dimensional analogy of $\mathcal{B}_{2,7} = \mathcal{C}_{2,7}$. It consists of a sphere at the origin and 24 spheres touching this sphere. There is a unique way to arrange 25 spheres in this manner, illustrated in Fig. 8 (b). It forms a subset of the D_4 lattice and is a very symmetrical and dense constellation. It can be formally defined as $\mathcal{B}_{4,25} = \mathcal{C}_{4,25} = \mathcal{B}_{4,24} \cup \{(0,0,0,0)\}$, where $\mathcal{B}_{4,24}$ represents the 24-cell defined below. The constellation $\mathcal{B}_{4,25}$ was discussed in [56] and it has an asymptotic power efficiency of $\gamma = 0.83$ dB.

The ball for $M = 24$ is obtained by removing any point from $\mathcal{B}_{4,25}$. The choice of point to remove does not influence the performance (in perfect analogy with $\mathcal{B}_{4,6}$) and we choose $(0,0,0,0)$, to preserve the symmetry. The ball $\mathcal{B}_{4,24}$ thus defined consists of the 24 vertices of the 4d regular polytope sometimes referred to as the *24-cell*. All five regular Platonic solids in three dimensions (tetrahedron, cube, octahedron, dodecahedron, and icosahedron) have extensions to four dimensions. The 24-cell, however, is the only regular 4d polytope, that, according to Coxeter, is

unique: "...having no analogue [in dimensions] above or below." [15, p. 289]. The 24-cell was considered for communications in [8, 32, 53, 56, 62]. Its coordinates can be expressed in two distinct ways. The first is as the union of the 16 levels of the 4d cube (DP-QPSK) and the 8 levels of a cross-polytope:

$$\mathcal{B}_{4,24} = \mathcal{D}_{4cube} \cup \sqrt{2}\mathcal{B}_{4,8} = \{(\pm 1, \pm 1, \pm 1, \pm 1), (\pm 2, 0, 0, 0)\}, \quad (25)$$

again including all signs and permutations. This demonstrates how the DP-QPSK format can be extended to 24 points without increasing the average symbol energy or reducing the minimum distance. These additional modulation levels were also recently suggested by Bülow [11] to be utilized for forward error correction overhead. The modulation format can be seen as using four absolute phase levels for each of the six polarization states ($x, y, \pm 45^\circ$, LHC, RHC).

The second and more compact description of the 24-cell is

$$\mathcal{B}'_{4,24} = \{\sqrt{2}(\pm 1, \pm 1, 0, 0)\}, \quad (26)$$

again allowing for arbitrary sign choices and coordinate permutations. This is an equally common representation of the 24-cell. A point \mathbf{c}' in $\mathcal{B}'_{4,24}$ can be obtained from a point \mathbf{c} in $\mathcal{B}_{4,24}$ by applying the coordinate transformation [14]

$$\mathbf{c}' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \mathbf{c}. \quad (27)$$

In fiber-optics language, a similar transformation that can be used to transform \mathbf{c} to \mathbf{c}' is $\mathbf{E}' = \mathbf{E} \exp(i\pi/4)$.²

By using the set $\mathcal{B}_{4,24}$, the sensitivity of the DP-QPSK format can be improved by $\log_2(24)/\log_2(16) = 0.59$ dB, but mapping bits to 24 symbols is non-trivial. In [1] we introduced a modulation format called 6P-QPSK by mapping nine information bits to two sequential points in $\mathcal{B}_{4,24}$, which enables 4.5 bits per symbol to be transmitted. This gives an improvement of $\gamma = 9/8 = 0.51$ dB over DP-QPSK.

The cluster $\mathcal{C}_{4,24}$ is obtained by removing an outer point (i.e., not $(0, 0, 0, 0)$) from $\mathcal{B}_{4,25}$ and shifting the resulting constellation to zero mean. It improves on DP-QPSK by 0.79 dB.

For $M = 23$, a ball is obtained by removing two arbitrary points from $\mathcal{B}_{4,25}$. The remaining balls can be shifted around in many ways without changing the maximum energy. The cluster $\mathcal{C}_{4,23}$ is however unique, and it is obtained by removing two adjacent outer points from $\mathcal{B}_{4,25}$ (say, $(1, 1, 1, \pm 1)$) and recentering the constellation.

Clusters for $M = 26$ and $M = 27$ are obtained by adding points from the next layer of D_4 to $\mathcal{B}_{4,25}$. Specifically, $\mathcal{C}_{4,26}$ is obtained by centering $\mathcal{B}_{4,25} \cup \{(2, 2, 0, 0)\}$ and $\mathcal{C}_{4,27}$ is obtained by centering $\mathcal{B}_{4,25} \cup \{(2, 2, 0, 0), (2, 0, 2, 0)\}$. These two clusters

² It was erroneously stated in [1] that the transformation (27) is equivalent to a 45° rotation of the carrier phase of the electric field. It is, if one interchanges row 1 with 2 and row 3 with 4 of the matrix in (27).

are however very weak in terms of maximum power, as will be shown in Sec. 5.2.4. The balls for $M = 26$ and $M = 27$ have no apparent relation to the kissing configuration $\mathcal{B}_{4,25}$ or the D_4 lattice.

$M > 27$

There are several regular 4d constellations with more points. For example, a 48-point constellation can be formed as $\mathcal{B}_{4,24} \cup \mathcal{B}'_{4,24}$, which was discussed in [8, 62]. There are also the regular 600-cell (for $M = 120$) and 120-cell (for $M = 600$) [8, 32, 56, 62], of which the former is good in terms of both average and maximum energy and the second is not good, in analogy with the icosahedron and dodecahedron, resp., in three dimensions [2]. At asymptotically high M , optimal constellations in both senses can be constructed as circular subsets of the D_4 lattice.

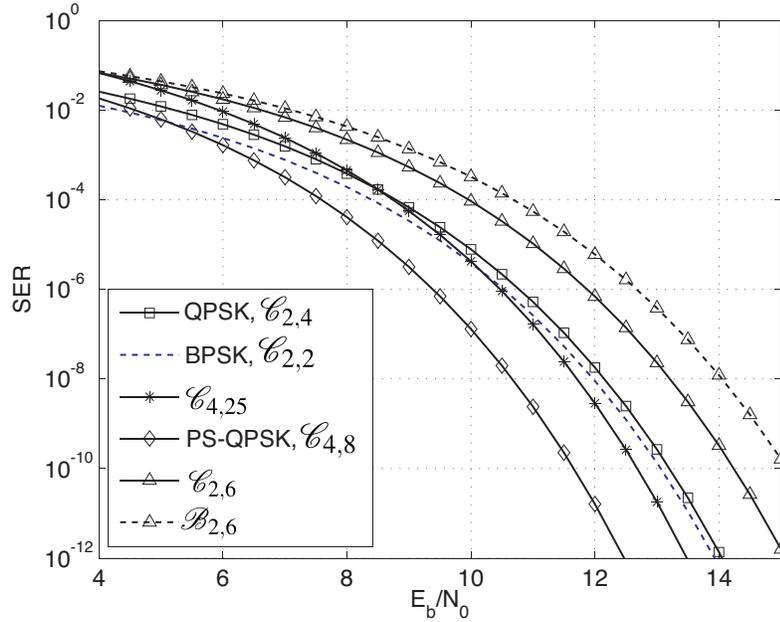


Fig. 9 SER vs. E_b/N_0 (average-energy SNR) for a number of constellations, including QPSK and BPSK.

4 Symbol- and bit-error rates

In this section we will discuss symbol error rates for some of the common modulation formats, and also discuss the difference between maximum-energy and average-energy SNR. We will start with this latter point.

Based on the union bound (9) we can now plot SER vs. SNR for all constellations we known with coordinates. In general, the union bound agrees well with the exact SER for $SER < 10^{-3}$. Note, however, that the SNR can be defined in two different ways: either (which is most common) as E_b/N_0 , i.e. with respect to the average energy per bit, or as $E_{b,max}/N_0$, i.e., with respect to the maximum energy per bit. Figures 9 and 10 shows the SER for the same group of constellations plotted versus these two SNR definitions. For formats where the average and peak symbol energies are the same (e.g., BPSK, QPSK, and PS-QPSK), there will be no difference. However, for formats where the peak and symbol energy differ (as for $\mathcal{C}_{4,25}$) the x-axis will be rescaled when plotting vs. $E_{b,max}$. A more dramatic difference can be seen when comparing clusters and balls that are non-identical. As a simple example of this, we plotted the SER for $\mathcal{C}_{2,6}$ (solid lines, triangles) and $\mathcal{B}_{2,6}$ (dashed lines, triangles) in Figs. 9 and 10. Quite obviously, a constellation that has been optimized with respect to average energy (a cluster) will perform better than a ball when plot-

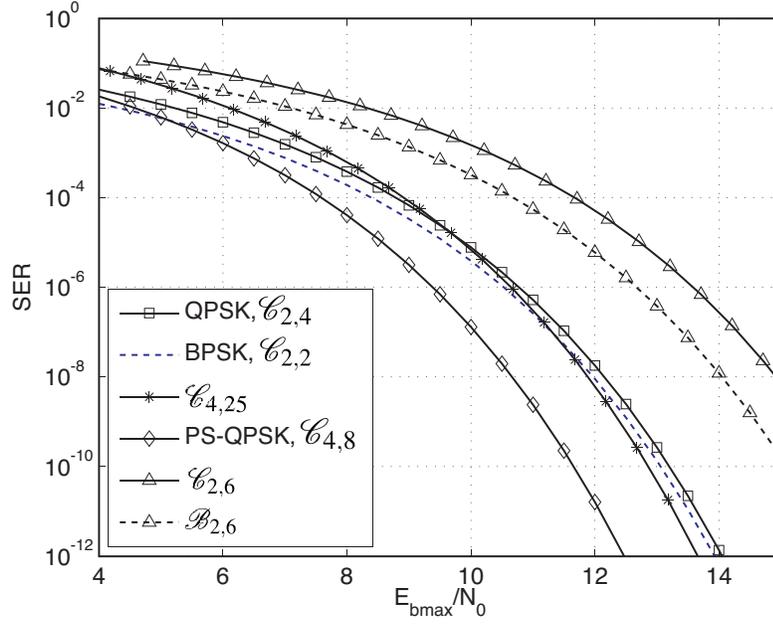


Fig. 10 SER vs. $E_{b,max}/N_0$ (maximum-energy SNR) for a number of constellations, including QPSK and BPSK.

ted vs. average energy (in Fig 9). The situation is reversed when plotting the SER vs. maximum energy (Fig. 10); here the ball performs better than the cluster.

We will now go beyond the union bounds and present exact symbol error rates for three of the most interesting formats, which are:

- The cubic constellation \mathcal{D}_{4cube} , which corresponds to the DP-QPSK format,
- the cross-polytope $\mathcal{C}_{4,8}$, which corresponds to the PS-QPSK format, and
- the 24-cell constellation, $\mathcal{B}_{4,24}$, which is used for the 6P-QPSK format.

The exact SER expressions for these constellations are, resp.,

$$SER_{4cube} = 1 - \left[1 - \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_s}{4N_0}} \right) \right]^4 \quad (28)$$

$$SER_{4,8} = 1 - \frac{1}{\sqrt{\pi}} \int_0^\infty (1 - \operatorname{erfc} x)^3 e^{-(x - \sqrt{\frac{E_s}{N_0}})^2} dx \quad (29)$$

$$SER_{4,24} = 1 - \frac{1}{\sqrt{\pi}} \int_0^\infty (1 - \operatorname{erfc} x)^2 \operatorname{erfc} \left(x - \sqrt{\frac{E_s}{2N_0}} \right) \cdot e^{-(x - \sqrt{\frac{E_s}{2N_0}})^2} dx. \quad (30)$$

Eq. (28) is straightforward to derive due to the simple geometry of the cubic constellations. The $SER_{4,8}$ expression (29) can be found in standard textbooks [3, p.

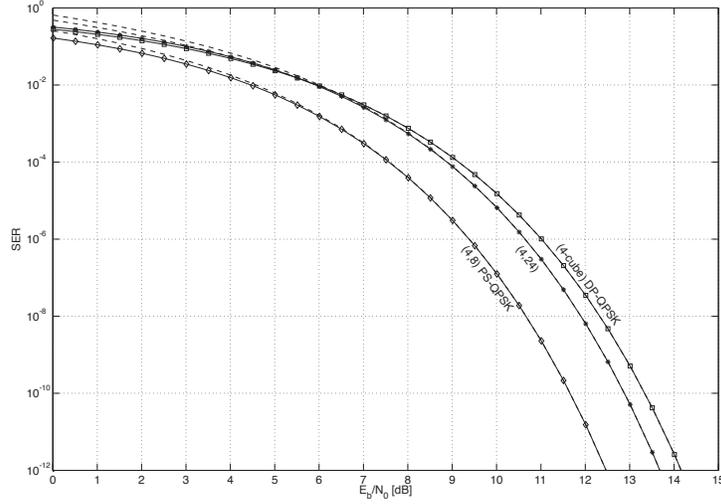


Fig. 11 SER versus E_b/N_0 for $\mathcal{C}_{4,8}$ (PS-QPSK), $\mathcal{B}_{4,24}$, and \mathcal{D}_{4cube} (DP-QPSK). The dashed lines are union bound calculations, whereas the solid lines are exact calculations from (28)–(30). The expected asymptotic improvements are 1.76 dB for PS-QPSK and 0.59 dB for $\mathcal{B}_{4,24}$.

210], [45, p. 201] by recognizing $\mathcal{C}_{4,8}$ as an 8-ary biorthogonal constellation. The derivation of the $SER_{4,24}$ -expression (30) is more cumbersome and reported in [2].

We do not recommend (28)–(30) for numerical evaluation at high E_s/N_0 , as cancellation occurs when subtracting two almost equal numbers. As observed in [59] for the case of $\mathcal{C}_{4,8}$, expanding the polynomials in $\operatorname{erfc}x$ and integrating out the constant term yields

$$SER_{4cube} = \frac{1}{16} \operatorname{erfc} \left(\sqrt{\frac{E_s}{4N_0}} \right) \left[4 - \operatorname{erfc} \left(\sqrt{\frac{E_s}{4N_0}} \right) \right] \cdot \left[8 - 4 \operatorname{erfc} \left(\sqrt{\frac{E_s}{4N_0}} \right) + \operatorname{erfc}^2 \left(\sqrt{\frac{E_s}{4N_0}} \right) \right] \quad (31)$$

$$SER_{4,8} = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_s}{N_0}} \right) + \frac{1}{\sqrt{\pi}} \int_0^\infty \operatorname{erfc} x \cdot (3 - 3 \operatorname{erfc} x + \operatorname{erfc}^2 x) e^{-\left(x - \sqrt{\frac{E_s}{N_0}}\right)^2} dx \quad (32)$$

$$SER_{4,24} = \operatorname{erfc} \left(\sqrt{\frac{E_s}{2N_0}} \right) \left[1 - \frac{1}{4} \operatorname{erfc} \left(\sqrt{\frac{E_s}{2N_0}} \right) \right] + \frac{1}{\sqrt{\pi}} \int_0^\infty \operatorname{erfc} x (2 - \operatorname{erfc} x) \operatorname{erfc} \left(x - \sqrt{\frac{E_s}{2N_0}} \right) \cdot e^{-\left(x - \sqrt{\frac{E_s}{2N_0}}\right)^2} dx. \quad (33)$$

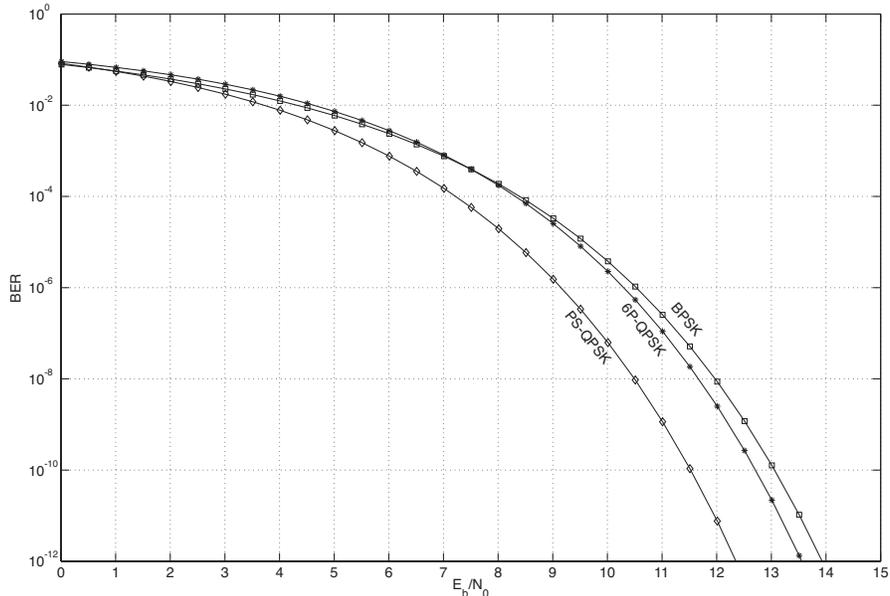


Fig. 12 BER versus E_b/N_0 for PS-QPSK, 6P-QPSK, and BPSK. QPSK and DP-QPSK have the same BER performance as BPSK. The improvement of PS-QPSK over BPSK is 0.97 dB at a BER of 10^{-3} and 1.51 dB at 10^{-9} . The asymptotic gains are again 1.76 dB for PS-QPSK but only 0.51 dB for 6P-QPSK.

In Fig. 11, we plot the SER as a function of E_b/N_0 by using these expressions. Union bounds from (9) are also shown. It is noteworthy that the union bound becomes indistinguishable from the exact values when the SER is less than 10^{-3} .

The BER performance depends on the mapping from information bits to symbols, which in turn depends on the modulator (and demodulator) implementation. If M is not a power of two, all constellation points cannot be used for binary data transmission, but the excess points can be used for framing and control purposes, as in, e.g., Fast Ethernet and Gigabit Ethernet, where 3- and 5-level modulation formats are standardized [52, pp. 285–289]. The amount of excess points can be controlled by mapping bits to a block of symbols rather than to independent symbols. The BER performance of DP-QPSK (or, equivalently, BPSK), PS-QPSK (exact), and 6P-QPSK (approximation) are compared in Fig. 12. We omit these details, which are discussed in [1], and give the results only.

For the DP-QPSK format, the BER performance is equivalent to that of the BPSK channel, which is $(1/2) \operatorname{erfc}(\sqrt{E_b/N_0})$. This property holds for any N -dimensional cubic modulation format, such as BPSK, QPSK, or DP-QPSK. For the PS-QPSK format we map the bits so that opposite points in the constellation have opposite bit patterns, and find that $BER_{PS-QPSK} \approx SER_{4,8}/2$. For the 6P-QPSK format, we map nine bits to two consecutive symbols, and then it is possible to obtain $BER_{6P-QPSK} \approx (5/18)SER_{4,24}$.

5 Sensitivities and nonlinearities

We will now discuss how these power-efficient modulation formats will improve the fundamental quantum-limited sensitivities of optical systems, and also discuss the role of fiber nonlinearities.

5.1 Fundamental sensitivity limits

Under the reasonable assumption that coherent links will use optical amplifiers, the main limiting noise source will be ASE noise from the amplifiers. It has been shown [21] that ASE noise is additive and Gaussian in nature, i.e., that the AWGN model applies to such a system. The optical noise at the receiver has a power spectral density of

$$N_0 = N_a n_{sp} h\nu \frac{G-1}{G} \approx N_a n_{sp} h\nu \quad (34)$$

per polarization [24, 30]. Here N_a denotes the number of in-line amplifiers, G the gain, n_{sp} the spontaneous emission factor of the amplifiers, and $h\nu$ the photon energy. In a polarization diversity homodyne coherent receiver, the optical amplitude is directly mapped to the electrical signal, so our AWGN results can be interpreted by using $E_b/N_0 = n_b/N_a n_{sp}$, where n_b is the average number of photons per bit. In the limit of a single amplifier with 3 dB noise figure ($N_a = n_{sp} = 1$), this implies that E_b/N_0 has a physically appealing interpretation as the number of photons per bit of the received signal. This can be used to translate the results from Fig. 12 to sensitivities (i.e., the number of photons per bit required to get $BER = 10^{-9}$). For BPSK we get the well-known result $E_b/N_0 = 12.5 \text{ dB} = 18$ photons per bit [27, 30]. The most sensitive format, PS-QPSK, improves this with 1.5 dB to 13 photons per bit [28]. The 6P-QPSK format is with 17 photons per bit slightly better than BPSK. All sensitivities (including some other formats discussed in this paper and those at $BER = 10^{-3}$) are found in Table 1.

Table 1 The properties of some common modulation formats, including the ones presented by us. The QAM formats are square grids; the 8-QAM being a 3×3 grid with the center point removed.

Name	Nbr. of pts. M	Nbr. of dims. N	Pow. Eff. γ [dB]	Spectral Eff. [bits/symb/pol]	Sens. at $BER = 10^{-3}$ E_b/N_0 [dB]
BPSK	2	1	0	2	6.8
QPSK	4	2	0	2	6.8
8-PSK	8	2	-3.57	3	10.0
8-QAM	8	2	-3.01	3	9.0
16-QAM	16	2	-3.98	4	10.5
DP-QPSK = \mathcal{C}_{cube}	16	4	0	2	6.8
PS-QPSK = $\mathcal{C}_{4,8}$	8	4	1.76	1.5	5.8
6P-QPSK	$2^{9/2} = 22.6$	4	0.51	2.25	6.9

We believe that these relative improvements of PS-QPSK and 6P-QPSK over BPSK will translate also to other coherent optical channels where the AWGN model applies, such as, e.g., the shot-noise limit [23, 24]. Neglecting pulse position modulation (which has been shown to provide unbounded capacity but is impractical in high-speed links [36]), we can thus conclude that the PS-QPSK modulation format gives the best sensitivity in uncoded optical links [28].

To get some real numbers into these sensitivities, we may note that at a bit rate of $1/T = 10$ Gbit/s, one photon per bit equals a received optical power of -59 dBm, and the sensitivity for BPSK in the ASE limit is then 12.5 dB above this, at -46.5 dBm. Recent experiments, based on offline synchronization algorithms, have succeeded in reaching remarkably close, within 4 dB, of this limit [31]. At higher rates, e.g., 100 Gbit/s, the sensitivity power levels become 10 dB higher in absolute power terms. Eventually, at this and higher rates, the nonlinear distortions of optical fibers will limit the BER, and power-efficient modulation formats such as those outlined in this paper may play an important role in improving the performance.

5.2 Nonlinear effects

The widespread deployment of EDFAs, and the development of high-power optical amplifiers have made the available optical power less of a problem than in the pre-EDFA days. Instead, fiber nonlinearities such as SPM and XPM is becoming increasingly important as limiting factors of fiber capacity [9, 10, 19, 60, 61]. The influence of nonlinearities is complicated by the fact that they are more or less impossible to discuss without also considering the dispersion. Different dispersion management schemes will lead to different impacts of the nonlinearities. For example, links with dispersion compensating fiber inserted periodically will not influence the signal in the same way as links that compensate all accumulated dispersion in the receiver (which is becoming more and more common in coherent systems) [41, 61]. The latter situation is significantly more difficult to analyze; to our knowledge, no analytic approaches are available and one usually has to resort to tedious simulations [10, 61].

The case when the accumulated dispersion is not allowed to grow significantly (by e.g. in-line compensation) is easier to analyze. The simplest approach is to just neglect dispersion, or only account for the walk-off effects in WDM systems. Then it is simpler to investigate how the SPM or XPM alone, or together with ASE noise, distorts the signal. Such links are mainly penalized by, to first order, the SPM/XPM-induced nonlinear phase shift, and to second order, nonlinear phase noise (NLPN). SPM is usually less relevant for equal-amplitude formats, since all constellation points will get the same nonlinear phase shift. On the other hand, it acts over all high-power sections in the system. In absence of dispersion and noise, SPM can be completely cancelled in the receiver by rotating the phase back in proportion to the detected amplitude.

XPM, in contrast, induces phase shifts in proportion to the instantaneous power in all WDM channels, but acts mainly over the walk-off-length between the two

WDM channels considered. It cannot be compensated, unless all WDM channels are simultaneously received and post-processed, which seems very challenging in today's systems. In general, XPM it acts in two ways, one is direct phase modulation and the other is polarization changes, sometimes referred to as *cross-polarization modulation*, XPolM [29, 57].

NLPN comes from the simultaneous action of ASE-induced intensity noise and SPM (or XPM). It will make the channel differ from the AWGN model by causing the phase noise to be larger than the amplitude noise.

There are three different aspects of the nonlinear influence on modulation formats that we shall briefly discuss here. They are (i) the role of the format's power efficiency, (ii) the format's robustness against nonlinear impairments and (iii) the format's influence on other wavelengths via XPM. In general, all these three items will be relevant, but which one is most limiting may likely vary between different system configurations, and would require full WDM system simulations to analyze, which is beyond the scope of this paper.

5.2.1 Power efficiency

Obviously, power efficient formats allow the transmitted power to be reduced, and as a result, the induced nonlinearities will decrease. Thus, for example, we can expect the PS-QPSK format to have 1.76 dB less power than DP-QPSK when transmitting at the same data rate, and naturally, this will be beneficial in links that are affected by nonlinearities.

5.2.2 Nonlinear robustness

The power efficiency is not the whole truth when it comes to nonlinear robustness. We must also consider the robustness to SPM/XPM of the formats. For example, the multilevel pulse-amplitude modulation (PAM) format may tolerate more nonlinear phase noise than QPSK, since the NLPN will move the points in the phase rather than amplitude direction, and hence not closer to a decision boundary. Thus, from this point of view, amplitude modulation might be beneficial in NLPN-limited links. On the other hand, amplitude-modulated formats will get more distorted from SPM, so it may not necessarily be a benefit.

Only scattered work has been done on comparing the nonlinear robustness of different formats in coherent links, so this is a rather open field for research.

5.2.3 XPM-induced crosstalk

Even if, as we saw above, a PAM format may be more robust to nonlinear phase rotation in itself, amplitude-modulated formats are much worse when it comes to their influence on other WDM channels via XPM. This means that the amount of

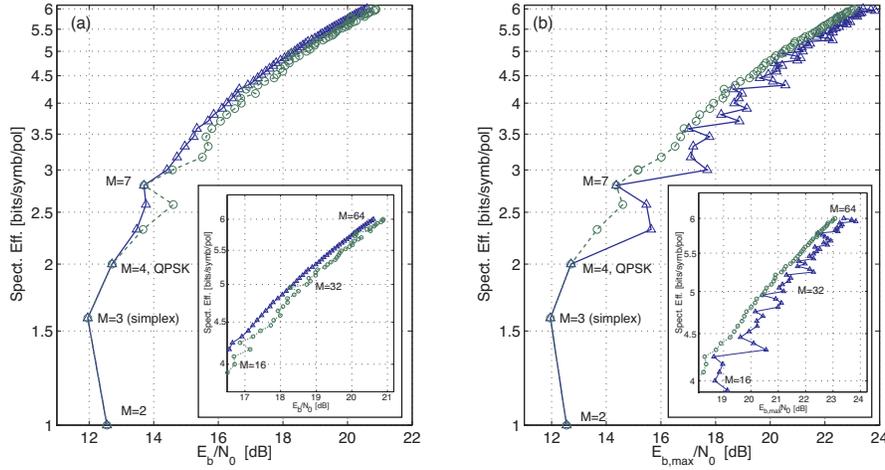


Fig. 13 SE vs. sensitivity for two-dimensional balls (circles, dashed lines) $\mathcal{B}_{2,M}$ and clusters (triangles, solid lines) $\mathcal{C}_{2,M}$, at a sensitivity defined at $SER = 10^{-9}$. The two plots show average (a) and maximum (b) SNR, and the insets are magnifications of the last points up to $M = 64$.

XPM-induced phase shift will depend on which symbols in the WDM channels overlap at a specific instance of time. Therefore, from this point of view, one would prefer equal-amplitude formats. For example, it has been shown that coherent DP-QPSK channels are more severely affected by OOK WDM channels than other DP-QPSK channels [10, 41].

However, in the presence of dispersion, also initially equal-amplitude formats will become amplitude-varying, so how large this effect is will depend on the details of the link and its dispersion management. There is for example work indicating that no optical dispersion compensation reduces the XPM influence [41, 61].

5.2.4 Relevance of maximum energy optimization

In general, all these three items will be relevant, but which one is most limiting may likely vary between different system configurations, and would require full WDM system simulations to analyze, which is beyond the scope of this paper.

It should thus be evident from the above discussion that nonlinear limitations are complex, and depend strongly on link design parameters such as dispersion map, amplifier spacing, WDM channel powers and separation, and, last but not least, modulation formats. As we know that SPM and XPM are determined by instantaneous rather than average power levels, we believe that minimization of maximum symbol energy power is preferred over average energy minimization in situations where nonlinearities are significant. There is thus reason to compare the two optimization schemes in more detail, and it would be interesting to show the formats also on a maximum-energy scale rather than the average bit-energy scale that is

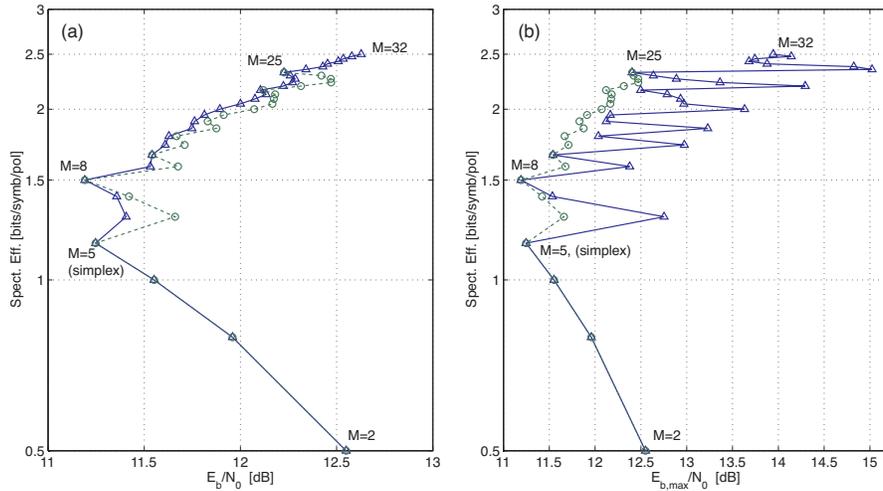


Fig. 14 SE vs. sensitivity for four-dimensional balls (circles, dashed lines) $\mathcal{B}_{4,M}$ and clusters (triangles, solid lines) $\mathcal{C}_{4,M}$, at a sensitivity defined as $SER = 10^{-9}$. The two plots show the same constellations vs. average (a) and maximum (b) SNR, for clusters up to $M = 32$ and balls up to $M = 25$.

usually chosen. This is done in Figs. 13–14, which shows the performance of the clusters and balls of Sec. 3 in terms of average bit energy E_b and maximum bit energy $E_{b,max} = E_{s,max}/\log_2 M$. Obviously, the clusters outperform the balls in terms of average energy, and the balls are better in terms of maximum energy. It is however interesting to see that many clusters are very bad in terms of maximum energy (the (b)-plots), whereas the balls perform fairly well for both measures. The cases in which the cluster and the ball coincide seem, however to be very good constellations in general. In two dimensions this occurs for $M = 2, 3, 4, 7, 31, 55$, which we believe are the only cases. In four dimensions it occurs for $M = 2, 3, 4, 5, 8, 10, 25$, and although this list may not be conclusive as we have not analyzed balls beyond $M = 25$, we believe there is only a finite coinciding cases.

A next step in the research of these optimized constellations will be to make full simulations, including nonlinearities and thereby judging the nonlinear robustness of these formats. Their practical realization may in some cases be complicated by the number of symbols in a constellation not being a power of 2. The transmitters and receivers for non-rectangular constellations are more complex as well, and those are also problems to look into. Nevertheless, a format such as PS-QPSK has none of these problems [28], and to investigate its nonlinear robustness and performance relative to, e.g., DP-QPSK appears to be quite interesting.

6 Summary and outlook

By using numerically optimized sphere constellations, we computed the best sensitivities of four-dimensional modulation formats up to 32 levels, which resulted in the conclusion that PS-QPSK is the format with the overall best sensitivity, 1.76 dB better than BPSK. We have shown that this is the most power-efficient modulation format when using four-dimensional constellations, unless the dimension is somehow increased. This can be done for example by using error-correcting codes, wavelength/space/time division multiplexing, or different modes in multimode fibers.

We also studied constellations that were optimized with respect to peak power, which we believe are relevant in nonlinearly limited systems. Our comparisons show that the mismatch penalty when using a format optimized for peak power in a scenario where the average power is critical is much less than vice versa. Hence, formats optimized for peak power are more robust and should be preferred in applications where both average and peak power are relevant, which is the case for most nonlinear impairments. Analyzing the performance of these modulation formats in nonlinear situations is an open area for future research.

Acknowledgements We wish to acknowledge funding from Vinnova within the IKT grant, and the Swedish strategic research foundation (SSF). We also acknowledge numerous stimulating discussions with all the researchers within the Chalmers fiber-optic communication research center FORCE. Dr. Seb Savory is gratefully acknowledged for a useful discussion, help with the $\mathcal{C}_{4,16}$ cluster, and for providing a few previously overlooked references.

References

1. Agrell, E., Karlsson, M.: Power-efficient modulation formats in coherent transmission systems. *Journal of Lightwave Technology* **27**(22), 5115–5126 (2009)
2. Agrell, E., Karlsson, M.: On the symbol error rate of regular polyhedra (2010). URL <http://arxiv.org/abs/1003.1257v1>
3. Benedetto, S., Biglieri, E.: *Principles of Digital Transmission: With Wireless Applications*. Kluwer Academic / Plenum Publishers, New York (1999)
4. Benedetto, S., Poggiolini, P.: Theory of polarization shift keying modulation. *IEEE Transactions on Communications* **40**(4), 708–721 (1992)
5. Betti, S., Curti, F., De Marchis, G., Iannone, E.: Exploiting fibre optics transmission capacity: 4-quadrature multilevel signalling. *Electronics Letters* **26**(14), 992–993 (1990). DOI 10.1049/el:19900644
6. Betti, S., Curti, F., De Marchis, G., Iannone, E.: A novel multilevel coherent optical system: 4-quadrature signaling. *Journal of Lightwave Technology* **9**(4), 514–523 (1991). DOI 10.1109/50.76666
7. Betti, S., De Marchis, G., Iannone, E., Lazzaro, P.: Homodyne optical coherent systems based on polarization modulation. *Journal of Lightwave Technology* **9**(10), 1314–1320 (1991). DOI 10.1109/50.90929
8. Biglieri, E.: Advanced modulation formats for satellite communications. 61–80, in *Advanced Methods for Satellite and Deep Space Communications*, ed. J. Hagenauer, Springer-Verlag, Berlin (1992)

9. Bononi, A., Bertolini, M., Serena, P., Bellotti, G.: Cross-Phase Modulation Induced by OOK Channels on Higher-Rate DQPSK and Coherent QPSK Channels. *Journal of Lightwave Technology* **27**(18), 3974–3983 (2009). DOI 10.1109/JLT.2009.2021537
10. Bononi, A., Serena, P., Rossi, N.: Nonlinear signal–noise interactions in dispersion-managed links with various modulation formats. *Optical Fiber Technology* **16**, 73–85 (2010)
11. Bülow, H.: Polarization QAM modulation (POL-QAM) for coherent detection schemes. In: *Proc. Optical Fiber Communication and National Fiber Optic Engineers Conference, OFC/NFOEC'09*. Paper OWG2 (2009)
12. Charlet, G., Maaref, N., Renaudier, J., Mardoyan, H., Tran, P., Bigo, S.: Transmission of 40 Gb/s QPSK with coherent detection over ultra-long distance improved by nonlinearity mitigation. In: *Proceedings of European Conference on Optical Communications, ECOC'06*, PDP Th.4.3.6 (2006)
13. Charlet, G., Salsi, M., Renaudier, J., Pardo, O., Mardoyan, H., Bigo, S.: Performance comparison of singly-polarised and polarisation-multiplexed coherent transmission at 10 Gbauds under linear impairments. *Electronics Letters* **43**(20), 1109–1111 (2007). DOI 10.1049/el:20072246
14. Conway, J., Sloane, N.: *Sphere Packings, Lattices and Groups*. Third Edition. Springer-Verlag, New York (1999)
15. Coxeter, H.S.M.: *Regular Polytopes*. Dover Publications, New York (1973)
16. Cusani, R., Iannone, E., Salonic, A., Todaro, M.: An efficient multilevel coherent optical system: M-4Q-QAM. *Journal of Lightwave Technology* **10**(6), 777–786 (1992)
17. Derr, F.: Optical QPSK homodyne transmission of 280 Mbit/s. *Electronics Letters* **26**(6), 401–403 (1990)
18. Ekanayake, N., Tjhung, T.: On ternary phase-shift keyed signaling. *IEEE Transactions on Information Theory* **IT-28**(4), 658–660 (1982)
19. Essiambre, R., Kramer, G., Winzer, P., Foschini, G., Goebel, B.: Capacity limits of optical fiber networks. *Journal of Lightwave Technology* **28**(4), 662–701 (2010)
20. Foschini, G., Gitlin, R., Weinstein, S.: Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise. *IEEE Transactions on Communications* **22**(1), 28–38 (1974)
21. Gordon, J.P., Walker, L.R., Louisell, W.H.: Quantum statistics of masers and attenuators. *Physical Review* **130**(2), 806–812 (1963). DOI 10.1103/PhysRev.130.806
22. Graham, R.L., Sloane, N.J.A.: Penny-packing and two-dimensional codes. *Discrete and Computational Geometry* **5**(1), 1–11 (1990)
23. Ho, K.P.: *Phase-Modulated Optical Communication Systems*. Springer, New York (2005)
24. Ip, E., Lau, A.P.T., Barros, D.J.F., Kahn, J.M.: Coherent detection in optical fiber systems. *Optics Express* **16**(2), 753–791 (2008). Erratum, **16**(26), 21943, (2008)
25. Jacobsen, G.: *Noise in Digital Optical Transmission Systems*. Artech House Publishers, Boston (1994)
26. Kahn, J., Gnauck, A., Veselka, J., Korotky, S., Kasper, B.: 4-Gb/s PSK homodyne transmission system using phase-locked semiconductor lasers. *IEEE Photonics Technology Letters* **2**(4), 285–287 (1990). DOI 10.1109/68.53264
27. Kahn, J., Ho, K.P.: Spectral efficiency limits and modulation/detection techniques for DWDM systems. *IEEE Journal of Selected Topics in Quantum Electronics* **10**(2), 259–272 (2004). DOI 10.1109/JSTQE.2004.826575
28. Karlsson, M., Agrell, E.: Which is the most power-efficient modulation format in optical links? *Optics Express* **17**(13), 10814–10819 (2009)
29. Karlsson, M., Sunnerud, H.: Effects of nonlinearities on PMD-induced system impairments. *Journal of Lightwave Technology* **24**(11), 4127–4137 (2006)
30. Kazovsky, L., Benedetto, S., Willner, A.: *Optical Fiber Communication Systems*. Artech House Publishers, Boston (1996)
31. Kikuchi, K., Tsukamoto, S.: Evaluation of sensitivity of the digital coherent receiver. *Journal of Lightwave Technology* **26**(13), 1817–1822 (2008)
32. Kim, H.G.: 4-dimensional modulation for a bandlimited channel using Q²PSK. *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 3, 1144–1147 (1999)

33. Lachs, G.: Optimization of signal waveforms. *IEEE Transactions on Information Theory* **9**(2), 95–97 (1963)
34. Ly-Gagnon, D., Katoh, K., Kikuchi, K.: Unrepeated optical transmission of 20 Gbit/s quadrature phase-shift keying signals over 210 km using homodyne phase-diversity receiver and digital signal processing. *Electronics Letters* **41**(4), 206–207 (2005)
35. Musin, O.: The kissing number in four dimensions. *Annals of Mathematics* **168**, 1–32 (2008)
36. Pierce, J.R.: Optical channels: Practical limits with photon counting. *IEEE Transactions on Communications* **26**(12), 1819–1821 (1978)
37. Pierce, J.R.: Comparison of three-phase modulation with two-phase and four-phase modulation. *IEEE Transactions on Communications* **COM-28**(7), 1098–1099 (1980)
38. Porath, J.E., Aulin, T.: Design of multidimensional signal constellations. *IEEE Proceedings-Communications* **150**(5), 317–323 (2003). DOI 10.1049/ip-com:20030652
39. Proakis, J.: *Digital Communications*, 4th ed. McGraw-Hill, Boston (2001)
40. Renaudier, J., Charlet, G., Salsi, M., Pardo, O., Mardoyan, H., Tran, P., Bigo, S.: Linear fiber impairments mitigation of 40-Gbit/s polarization-multiplexed QPSK by digital processing in a coherent receiver. *Journal of Lightwave Technology* **26**(1), 36–42 (2008)
41. Roberts, K., O’Sullivan, M., Wu, K.T., Sun, H., Awadalla, A., Krause, D.J., Laperle, C.: Performance of Dual-Polarization QPSK for Optical Transport Systems. *Journal of Lightwave Technology* **27**(16), 3546–3559 (2009). DOI 10.1109/JLT.2009.2022484
42. Saha, D., Birdsall, T.: Quadrature-quadrature phase-shift keying. *IEEE Transactions on Communications* **37**(5), 437–448 (1989). DOI 10.1109/26.24595
43. Shannon, C.E.: Communication in the presence of noise. *Proceedings of the IRE* **37**(1), 10–21 (1949)
44. Shannon, C.E.: Probability of error for optimal codes in a Gaussian channel. *Bell Systems Technical Journal* **38**(3), 611–656 (1959)
45. Simon, M., Hinedi, S., Lindsey, W.: *Digital Communication Techniques: Signal Design and Detection*. PTR Prentice Hall (1995)
46. Sloane, N.J.A., Hardin, R.H., Duff, T.S., Conway, J.H.: Minimal-energy clusters of hard spheres. *Discrete and Computational Geometry* **14**(3), 237–259 (1995)
47. Sloane, N.J.A., Hardin, R.H., Duff, T.S., Conway, J.H.: Minimal-energy clusters, library of 3-d clusters, library of 4-d clusters (1997). URL <http://www.research.att.com/~njas/cluster/>
48. Sloane, N.J.A., Hardin, R.H., Duff, T.S., Conway, J.H.: Spherical codes, part 1 (2000). URL <http://www.research.att.com/~njas/-packings/>
49. Specht, E.: The best known packings of equal circles in the unit circle (2009). URL <http://hydra.nat.uni-magdeburg.de/packing/cci/cci.html>
50. Stephenson, K.: Circle packing bibliography as of september 2005 (2005). URL <http://www.math.utk.edu/~kens/CP-bib.pdf>
51. Sun, H., Wu, K., Roberts, K.: Real-time measurements of a 40 Gb/s coherent system. *Optics Express* **16**(2), 873–879 (2008)
52. Tanenbaum, A.S.: *Computer Networks*, 4th ed. Pearson, Upper Saddle River, NJ (2003)
53. Taricco, G., Biglieri, E., Castellani, V.: Applicability of four-dimensional modulations to digital satellites: A simulation study. In: *Proc. IEEE Global Telecommunications Conference*, vol. 4, 28–34 (1993)
54. Tsukamoto, S., Ly-Gagnon, D., Katoh, K., Kikuchi, K.: Coherent demodulation of 40-Gbit/s polarization-multiplexed QPSK signals with 16-GHz spacing after 200-km transmission. In: *Proceedings of Optical Fiber Communication and National Fiber Optic Engineers Conference, OFC/NFOEC*, vol. 6, PDP29 (2005)
55. Weisstein, E.W.: "Ball.", From Mathworld — a Wolfram Web Resource (2010). URL <http://mathworld.wolfram.com/Ball.html>
56. Welti, G., Lee, J.: Digital transmission with coherent four-dimensional modulation. *IEEE Transactions on Information Theory* **20**(4), 497–502 (1974)
57. Winter, M., Bunge, C.A., Setti, D., Petermann, K.: A Statistical Treatment of Cross-Polarization Modulation in DWDM Systems. *Journal of Lightwave Technology* **27**(17), 3739–3751 (2009)

58. Wu, J., Wu, M.C.: Dual-polarization frequency reuse with frequency band shifting allocation. *IEEE Transactions on Vehicular Technology* **49**(6), 2244–2256 (2000)
59. Xiao, L., Dong, X.: New analytical expressions for orthogonal, biorthogonal, and transorthogonal signaling in Nakagami fading channels with diversity reception. *IEEE Transaction on Wireless Communications* **4**(4), 1418–1424 (2005)
60. Xie, C.: Interchannel nonlinearities in coherent polarization-division-multiplexed quadrature-phase-shift-keying systems. *IEEE Photonics Technology Letters* **21**(5), 274 (2009)
61. Xie, C.: WDM coherent PDM-QPSK systems with and without inline optical dispersion compensation. *Optics Express* **17**(6), 4815–4823 (2009)
62. Zetterberg, L., Brändström, H.: Codes for combined phase and amplitude modulated signals in a four-dimensional space. *IEEE Transactions on Communications* **25**(9), 943–950 (1977)