



CHALMERS

Chalmers Publication Library

Cn. 2. Multidimensional Optimized Optical Modulation Formats

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:

Enabling Technologies for High Spectral-efficiency Coherent Optical Communication Networks, Ch. 2

Citation for the published paper:

Karlsson, M. ; Agrell, E. (2016) "Cn. 2. Multidimensional Optimized Optical Modulation Formats". Enabling Technologies for High Spectral-efficiency Coherent Optical Communication Networks, Ch. 2 pp. 13-64.

Downloaded from: <http://publications.lib.chalmers.se/publication/236628>

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)

CHAPTER 2

MULTIDIMENSIONAL OPTIMIZED OPTICAL MODULATION FORMATS

M. KARLSSON¹ AND E. AGRELL²

¹Photonics Laboratory, Department of Microtechnology and Nanoscience
Chalmers University of Technology, Gothenburg, Sweden

²The Communication Systems Group, Department of Signals and Systems
Chalmers University of Technology, Gothenburg, Sweden

2.1 Introduction

The development of advanced digital signal processing (DSP) to enable intradyne coherent optical receivers [1–3] caused a paradigm shift within optical communications, and there is little doubt that the future of optical transport will be coherent. Coherent receivers ideally map the optical signal to the electrical domain, which enables a lot of novel advanced communication algorithms to be implemented in optical links, e.g., digital equalization and advanced modulation. One of the most profound developments was that intradyne receivers enabled all four quadratures of the optical signal (or in optical terms amplitude, phase, and polarization states) to be modulated and detected. This was realized already in the early 90s when Betti et al. investigated the modulation of all four quadratures in optical links [4–7]. Even if coherent detection was demonstrated already in 1990 by Derr [8], it was too complicated to be commercially interesting and the research faded.

title, edition.

By author Copyright © 2015 John Wiley & Sons, Inc.

As optical transmission systems had traditionally used rudimentary modulation (typically on-off-keying (OOK) or differential phase-shift-keying (DPSK), [9]), the coherent receivers meant great opportunity to study novel modulation formats, tailored for the emerging coherent optical links. The first such format used was the polarization-multiplexed quadrature shift keying (PM-QPSK) [2,3], which in its simplest form is binary phase shift keying (BPSK) in all four quadratures in parallel. As coding and modulation are key building blocks in the design of any communication link, it is a natural first approach to separate them and study the performance of each block separately. Most of the research reviewed and presented in this chapter will be dealing with the modulation format only, but we emphasize that that is only part of the problem in designing a good optical transmission link. The second part is to add forward error correcting (FEC) codes, preferably tailored and co-optimized with the modulation formats, an area often referred to as *coded modulation*. However, a discussion on that topic is beyond the scope of this chapter, and we refer the interested reader to, e.g., [10, 11] for a recent overview and introduction.

The choice of modulation format in a link is crucial in that it sets an upper limit on the achievable *spectral efficiency*, which loosely speaking measures how well the channel real estate (bandwidth and signaling dimensions) are utilized. The addition of FEC will always reduce the spectral efficiency (but with the crucial benefit of increasing the noise tolerance). Nevertheless there is a deep relation between coding and modulation. Specifically, all FEC codes can be interpreted as a multidimensional modulation format by considering a sequence of time slots as dimensions. The converse does not necessarily hold. Although many multidimensional modulation formats, in particular those with a regular structure, can be interpreted as a low-dimensional modulation format in combination with a FEC code, this is not always the case. The relation between modulation and coding will be discussed further in sections 2.2.1 and 2.4.

In the choice of modulation format, there is an inherent threefold trade-off between the spectral efficiency, the noise tolerance and complexity of the format. In this chapter, we aim to shed some light on these trade-offs, by investigating relatively simple, low-dimensional formats in 4 dimensions. Such research is not new; 4d formats were investigated already in the 70s by Welty and Lee [12] and by Zetterberg and Brändström [13]. Also, the work by Biglieri [14] contains some of the 4d formats we will discuss in this chapter, as well as discussions on lattices and lattice cuts, which we will also cover. The novelty is the application to the optical channel with its specifics and trade-offs when it comes to signal generation, transmission, and detection. Therefore we will devote quite some effort to review and describe implementations and experiments.

This chapter is organized as follows. In the next section, we will give basic definitions and performance metrics for modulation formats that are common in the literature. In section 2.3, the most interesting formats and their performances are theoretically described and characterized. Next, in section 2.4, we study how low-dimensional codes can be used to extend the known formats to higher dimensions and spectral efficiencies. In section 2.5, the relatively large body of experimental

work done on multidimensional modulation in coherent links that has been done in the last few years is reviewed, and finally section 2.6 concludes.

2.2 Fundamentals of digital modulation

An optical communication channel, like any other physical propagation or storage medium, is what in communication theory is called a *waveform channel*, which communicates a time-varying voltage (or electric field) from one point to another. If the channel is used to transmit digital data, then there are only a finite number of possible waveforms of a given length, and every such waveform corresponds to a certain sequence of bits. The process of mapping bits into waveforms and vice versa is called *digital modulation*. This can be done in a multitude of ways, depending on the type of channel. Some common optical system models, and their preferred modulation techniques, are reviewed in section 2.2.1. In section 2.2.2, we discuss several optical channel models that have been proposed to account for the fiber propagation effects.

In order to compare modulation formats and select a suitable one for implementation in a particular communication system, a *performance metric* is needed. There is a multitude of such metrics, for a variety of purposes. A modulation format that is superior in one sense may very well be inferior in another. This is the topic of section 2.2.3.

2.2.1 System models

A *multidimensional channel* is one that offers the possibility of transmitting multiple waveforms simultaneously. These waveforms could consist of the two quadratures of an amplitude- and phase-modulated lightwave, the two polarizations, multiple wavelengths in a wavelength-division multiplexed (WDM) system, multiple modes, or multiple cores. Each of the parallel waveforms can be thought of residing in one dimension. The traditional paradigm, and the least complex solution, is to transmit independent data on all of these dimensions. However, improved performance can be obtained by encoding data jointly on several dimensions, i.e., by *multidimensional modulation*. This improvement is most prominent if the waveforms interfere with each other during transmission, but significant gains can be achieved even if the waveforms are transmitted independently. The topic of multidimensional modulation will be revisited in section 2.5.5.

The mapping of bits into waveforms can be thought of as a three-step process. First, redundant bits are added to the payload. This overhead serves several purposes: to indicate a frame structure, which allows the interpretation of the received bit stream as a sequence of data packets; to provide address information for proper routing; and to provide error resilience via FEC. These functions, albeit crucial for the operation of an optical communication network, are all outside the scope of the present chapter.

Second, m bits at a time are mapped into a *symbol*, which is a vector in an N -dimensional space. The set \mathcal{X} of all $M = 2^m$ symbols are called a *constellation*.

This is the single most important entity in the definition of a modulation format; indeed, it is so important that the term “modulation format” is sometimes used as a synonym for constellation.

Third, the sequence of symbols is mapped into a set of waveforms. The standard way to do this is via a *linear modulator*. Denoting the sequence of N -dimensional symbols with $\mathbf{x}[k]$, for $k = \dots, 0, 1, 2, \dots$, the vector of N waveforms is computed as

$$\mathbf{x}(t) = \sum_k \mathbf{x}[k] \phi(t - kT), \quad (2.1)$$

where T is the symbol time and $\phi(t)$ is a given pulse shape.

At this point, it should be emphasized that the discrete-time sequence $\mathbf{x}[k]$ is fundamentally different from its continuous-time counterpart $\mathbf{x}(t)$ and they should not be confused with each other. The waveforms $\mathbf{x}(t)$ needs to be considered in order to analyze signal spectra as well as propagation effects such as distortions, filtering, added noise and other hardware limitations. On the other hand, the sequence $\mathbf{x}[k]$ is the quantity of interest to analyze bit and symbol error rates (BER and SER), mutual information, channel capacity, etc.

The vector $\mathbf{x}(t)$ represents N baseband waveforms. Each of these waveforms are now multiplied with a carrier, for transmission over an N -dimensional channel, which, as explained in the beginning of this subsection, consists of multiple quadratures, polarizations, wavelengths, modes, and/or cores.

At the receiver side, the reverse operations are performed using a coherent receiver. First, the symbol clock, carrier phase, and polarization are recovered using either blind or pilot-aided estimation algorithms [15–17]. A balanced detector now outputs the N received baseband waveforms, represented by the vector $\mathbf{y}(t)$, which should hopefully resemble $\mathbf{x}(t)$.

Second, the waveforms are filtered and sampled. The obtained sequence of N -dimensional vectors is

$$\mathbf{y}[k] = \int_{-\infty}^{\infty} \mathbf{y}(t) h(kT - t) dt \quad (2.2)$$

for $k = \dots, 0, 1, 2, \dots$, where $h(t)$ is the impulse response of the receiver filter. The received symbol sequence $\hat{\mathbf{x}}[k]$ is now determined by identifying, independently for each k , the point in \mathcal{X} closest to $\mathbf{y}[k]$, in some well-defined sense that depends on the channel model. Ideally, the receiver filter is chosen as a *matched filter* $h(t) \sim \phi(T_d - t)$, where T_d is the processing delay. Furthermore, the pulse $\phi(t)$ is chosen to satisfy the *T-orthogonality criterion*

$$\int_{-\infty}^{\infty} \phi(t) \phi(t - kT) dt = 0, \text{ for all integers } k \neq 0, \quad (2.3)$$

which avoids intersymbol interference for linear channels, i.e., $\mathbf{y}[k]$ depends on $\mathbf{x}[k]$ but not on $\mathbf{x}[k \pm 1], \mathbf{x}[k \pm 2], \dots$

Third and last, the received bit sequence is obtained by concatenating the bits corresponding to each symbol. Then the digital overhead is removed, which includes the operations of FEC decoding and frame synchronization.

It is also possible to consider blocks of K symbols $\mathbf{x}[k], \mathbf{x}[k+1], \dots, \mathbf{x}[k+K-1]$ as a supersymbol, taken from a constellation of NK dimensions. In general, this technique improves the performance at the cost of a higher transmitter and receiver complexity. A similar effect can be achieved at a more manageable complexity by applying a FEC code before modulation. Specifically, if a block code with codeword length $n = mK$ is applied to the bit stream before modulation, the resulting symbol sequence can be regarded either as a sequence of dependent N -dimensional symbols or as a sequence of independent NK -dimensional supersymbols. We will see examples of such NK -dimensional constellations designed from standard FEC codes in section 2.4.

2.2.2 Channel models

A complication for optical links is that the fiber propagation of the signal waveform is conventionally modeled with a nonlinear partial differential equation, the nonlinear Schrödinger equation (NLSE), where fiber dispersion, nonlinearities and amplifier noise distorts the signal. This is not the desired discrete-time model that a communication engineer would like to have when designing the coding and modulation algorithms. There are generally three problems associated with taking the fiber propagation to a usable discrete-time model. (i) To correctly model the transitions between symbols and waveforms (discrete and continuous time). Usually the transmitter is modeled as a continuous pulse source multiplied with discrete data in each symbol time, ignoring the sum in (2.1). This works fairly well, but one may have unwanted intersymbol interference in the symbol borders that is often neglected. The receiver side, going from the continuous waveform to a discrete data sample, is often modeled as an integrate-and-dump filter, i.e., restricting the integral in (2.2) to an interval of length T . This is not penalty-free, and it is theoretically complicated when the signal spectrum is distorted or broadened so one cannot guarantee matched filtering or sampling without aliasing. (ii) The NLSE and fiber transmission is *nonlinear* in the general case, and often operated in a regime where the nonlinearity cannot be neglected. In this case, the received signal is generally affected by intersymbol interference even if (2.3) is satisfied and linear ISI in the channel is removed. (iii) The coherent receiver should have negligible distortions, i.e., operate in a regime (strong local oscillator with low phase noise) where it linearly maps the optical field to the electrical domain for sampling and detection. In addition, perfect timing synchronization and compensation for channel impairments are assumed.

Often these problems are neglected, which leads to the standard additive white Gaussian noise (AWGN) model for coherent links, where the signal is only distorted by additive amplifier spontaneous emission (ASE) noise [18, 19]. Good agreement between simulations and experiments are evidence that this approach works reasonably well for many systems.

Of the above problems, the nonlinearity is the most serious one, but thanks to the recent developments of the *Gaussian noise (GN) model* [20–22], it can be dealt with by a simple extension of the AWGN model. The GN model applies to links with strong dispersive broadening during propagation and electronic dispersion compen-

sation in the receiver. Then the impact of the nonlinearity can be accurately modeled as additive white Gaussian noise with a variance proportional to the average signal power cubed, which was first observed by Splett et al. already in 1993 [23]. In such links, the presented format optimizations (which rely on the noise being uniform in all dimensions) will still work well. The GN model is known to agree well with experiments and to be a useful system design tool, but the usefulness for, e.g., capacity estimates in nonlinear links can be questioned [24].

A second model accounting for fiber nonlinearities is the *nonlinear phase-noise model* [25, pp. 157, 225]. This applies to links where the dispersion is negligible, e.g., with optical in-line compensation and/or low baudrates. Then the nonlinear self-phase modulation will, together with the ASE noise, lead to constellations with a spiraling shape. The model has also been extended to dual polarizations by Beygi et al. [26].

2.2.3 Constellations and their performance metrics

The starting point for digital modulation theory is, since long before the invention of fiber-optic communications, the scenario consisting of an additive white Gaussian noise channel, no coding, optimal detection (maximum likelihood, ML), and asymptotically low error probability. In this scenario, the BER and SER are both proportional to $Q(d/\sqrt{2N_0})$ [27, 28], where $Q = (2\pi)^{-1/2} \int_x^\infty \exp(-z^2/2) dz$ is the Gaussian Q function, d is the minimum Euclidean distance between points in the constellation, and N_0 is the noise power spectral density. Modulation formats are therefore traditionally designed in order to maximize (a normalized version of) the minimum distance d . Nevertheless, such modulation formats are often applied even in scenarios where the minimum distance does not govern the performance, such as for non-Gaussian or nonlinear channels, in coded systems, with suboptimal receivers, or at nonasymptotic error probabilities.

The following performance metrics are often used to quantify the performance of modulation formats [28].

Spectral efficiency The spectral efficiency or normalized bit rate is defined as [29, 30]

$$\beta = \frac{\log_2 M}{N/2},$$

where N and M , as defined in section 2.2.1, give the number of dimensions and constellation points, resp. The spectral efficiency gives the number of bits per channel use, where every (complex) channel use involves two dimensions. It also gives the bitrate per bandwidth, in bit/s/Hz, if Nyquist signaling is applied (sinc pulse shaping). A related quantity is $\beta/2$, which gives the number of bits per dimension, and can be interpreted as the data rate per bandwidth in bit/s/Hz, if rectangular pulse shaping is applied and bandwidth is defined as the width of the spectral main lobe.

Average and peak symbol energy The average symbol energy, also called the second moment or the mean squared Euclidean norm, is

$$E = \frac{1}{M} \sum_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x}\|^2$$

and the peak symbol energy is

$$E_{\max} = \max_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x}\|^2.$$

If the pulse $\phi(t)$ in (2.1) satisfies (2.3), then

$$\lim_{n \rightarrow \infty} \frac{1}{2nT} \int_{-nT}^{nT} \|\mathbf{x}(t)\|^2 dt \sim \frac{E}{T} \int_{-\infty}^{\infty} \phi^2(t) dt,$$

i.e., the continuous-time average energy is proportional to the discrete-time average energy E . Unfortunately, there exists no analogous relation between the continuous-time and discrete-time peak energies. Constellation designs based on E_{\max} tend nevertheless to be relatively good also in terms of the continuous-time peak energy, but not necessarily optimal [31].

Average bit energy $E_b = E / \log_2 M$ gives the average energy needed to transit one bit of information.

Constellation figure of merit The constellation figure of merit CFM is defined as [29, 30]

$$CFM = \frac{d^2 N}{2E}.$$

This is, assuming AWGN, no coding, optimal detection (maximum likelihood), and asymptotically high signal-to-noise ratio, SNR, (low error probability), the relevant power metric if modulation formats are compared at the same *bandwidth*.

Power efficiency The (asymptotic) power efficiency is [32, eq. (5.8)], [27]

$$\gamma = \frac{d^2}{4E_b} = \frac{\beta CFM}{4}. \quad (2.4)$$

This is, under the same conditions as for the CFM , the relevant power metric if modulation formats are compared at the same *bit rate*.

Gain The gain is quantified with respect to a baseline modulation format at the same spectral efficiency β , commonly chosen as *pulse-amplitude modulation* (PAM) [29, 30]. A PAM constellation has

$$CFM_{\text{PAM}} = \frac{6}{2^\beta - 1}$$

and

$$\gamma_{\text{PAM}} = \frac{3\beta}{2(2^\beta - 1)}. \quad (2.5)$$

Multidimensional extensions of PAM such as quadrature-amplitude modulation (QAM) and polarization-multiplexed (PM) QAM have the same CFM and γ . Geometrically, the baseline constellations represent cubic subsets of the cubic lattice. The gain is defined as

$$G = \frac{CFM}{CFM_{\text{PAM}}} = \frac{\gamma}{\gamma_{\text{PAM}}},$$

also for spectral efficiencies β for which no PAM constellation exists.

Mutual information, MI The mutual information is defined as

$$I(\mathbf{X}; \mathbf{Y}) = \iint f(\mathbf{x}, \mathbf{y}) \log_2 \frac{f(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})f(\mathbf{y})} d\mathbf{x}d\mathbf{y}, \quad (2.6)$$

where \mathbf{X} and \mathbf{Y} are the channel inputs and outputs, respectively, and f denotes the distribution of the stochastic variables indicated by its arguments.

Complexity Finally, some words should be said about complexity. It is one of the most important figures of merit, and it should be considered in any implementation, in order to keep the latency, energy consumption, and cost within reasonable levels. Nevertheless it is one of the hardest parameters to quantify numerically, depending not only on the modulation format but also on the transmitter and receiver algorithms as well as the hardware platform. As a crude rule of thumb, the complexity increases with the dimension, number of points, and irregularity of the constellation.

2.3 Modulation formats and their ideal performance

In this section, we will briefly review the various modulation formats and format optimizations that have been presented in the literature. Without doubt, the most commonly used formats are the PAM formats, based on the cubic lattice, possibly in N dimensions. Their performance is well known and stated in section 2.2.3. Their popularity is mostly due to their simplicity of generation and detection, but if some of that simplicity is sacrificed, much better performance (in terms of noise tolerance or spectral efficiency) can be achieved. The formats presented in this section will be examples of that.

We will extensively discuss *format optimization* in the below. It is important to emphasize that the outcome of such an optimization is heavily dependent on *what* is optimized and *which constraints* are assumed under the optimization. The simplest and most common scenario is to assume additive white Gaussian noise, no coding, optimal detection (ML), and asymptotically high SNR (low error probability). This ideal scenario will be studied in this section. Modulation optimization for some specific nonlinear and non-Gaussian channel models will be summarized in 2.3.2.

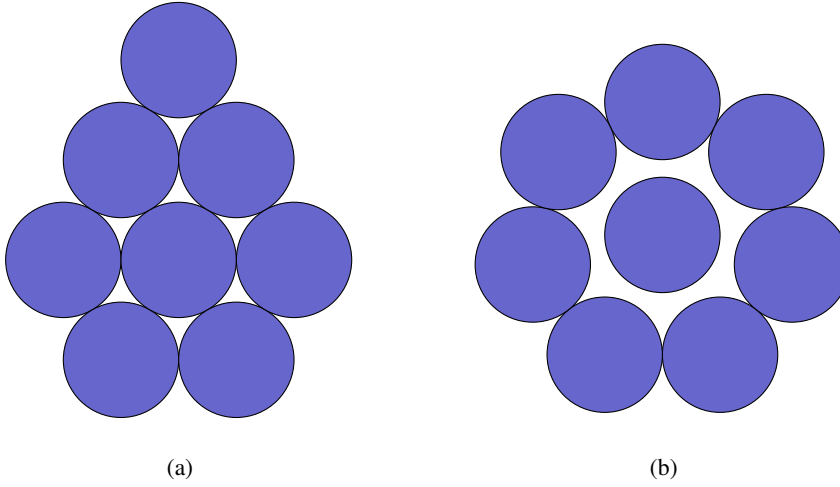


Figure 2.1 The cluster $\mathcal{C}_{2,8}$ (a) and the ball $\mathcal{B}_{2,8}$ (b).

In the limit of high SNR, the formats with the lowest symbol error rate (SER) can be found from optimized packings of solid spheres [27, 31, 33]. For a constant dimensionality and number of spheres, such packing optimization can be done by either minimizing the average distance of the spheres from the origin (the average second moment E) or by minimizing the maximum distance (the maximum symbol energy E_{\max}). To emphasize this difference, the constellation of M spheres with minimum E in dimension N is called the *cluster* $\mathcal{C}_{N,M}$, and the constellation with lowest E_{\max} is called the *ball* $\mathcal{B}_{N,M}$. Sometimes the clusters and balls coincide, but in general they do not. A simple example of the latter arises for 8 points in 2d, as shown in figure 2.1. This example also shows that the balls may be nonunique, as the center point is loose, and can be freely moved without affecting E_{\max} .

In addition to the balls and clusters, one can also compare different formats at the same bit rate (where γ is the relevant metric), or at the same bandwidth (where CFM is used). Both cases will be discussed below, in section 2.3.1. The balls and their relevance were discussed in [31] and will be briefly touched upon in section 2.3.2.

2.3.1 Format optimizations and comparisons

This and the next few subsections will focus mainly on the *clusters*, i.e., the N -dimensional, M -point constellations that minimize the average symbol energy (second moment) E . Tables with coordinates of those constellations are given in, e.g., [34] for 2d clusters and [35] for 3d and 4d clusters. These and other constellations are available online [28]. All these are numerically optimized results, presented as tables of coordinates. Some of the most interesting constellations are presented in

Table 2.1 Known densest lattices, their number of nearest neighbors and densities.

Dim.	Densest lattice	Neighbors (kissing number)	Density Δ
2	A_2	6	$\frac{\pi}{2\sqrt{3}} = 0.91$
3	A_3	12	$\frac{\pi}{3\sqrt{2}} = 0.74$
4	D_4	24	$\frac{\pi^2}{16} = 0.62$
8	E_8	240	$\frac{\pi^4}{384} = 0.25$
24	Λ_{24}	196560	$\frac{\pi^{12}}{12!} = 0.0019$

exact analytic form in [27, 31]. Quite often, the clusters possess some symmetry that facilitates a nice coordinate description.

In the limit of many points, the clusters will be spherical cuts from the regular *lattices* that are known to be the best packings in the given dimension. The best packing lattices are only known exactly in dimensions 2, 3, 4, 8 and 24, and they are listed in Table 2.1, together with their *densities*, Δ , which denotes the fraction of N d space that is filled by packing nonoverlapping spheres at the lattice points. The power efficiency for a spherical cut of M lattice points in N -d space can, if M is sufficiently large, be well approximated as [36, eq. (32)]

$$\gamma_{\text{lat}} = \log_2(M) \left(1 + \frac{2}{N}\right) \left(\frac{\Delta}{M}\right)^{\frac{2}{N}}. \quad (2.7)$$

This expression is derived by assuming a uniform point density in the spherical cut. This approach can be expected to be better with increasing M , significantly exceeding the nearest neighbor number, so that many lattice cells are enclosed in the cut. If an N -d hypercubic cut is carried out rather than a spherical cut, a penalty of $\pi e/6 = 1.53$ dB (the so-called *shaping gain*) is sacrificed for large N . In a similar manner we have the *CFM* and gain for the lattices as

$$CFM_{\text{lat}} = 2(N+2) \left(\frac{\Delta}{M}\right)^{\frac{2}{N}} \quad (2.8)$$

$$G_{\text{lat}} = \frac{N+2}{3} \left(M^{\frac{2}{N}} - 1\right) \left(\frac{\Delta}{M}\right)^{\frac{2}{N}}. \quad (2.9)$$

2.3.1.1 General properties of the metrics Properties of the best known clusters, for $N = 2, 4$, and 8 and selected values of M , are shown in figure 2.2. The coordi-

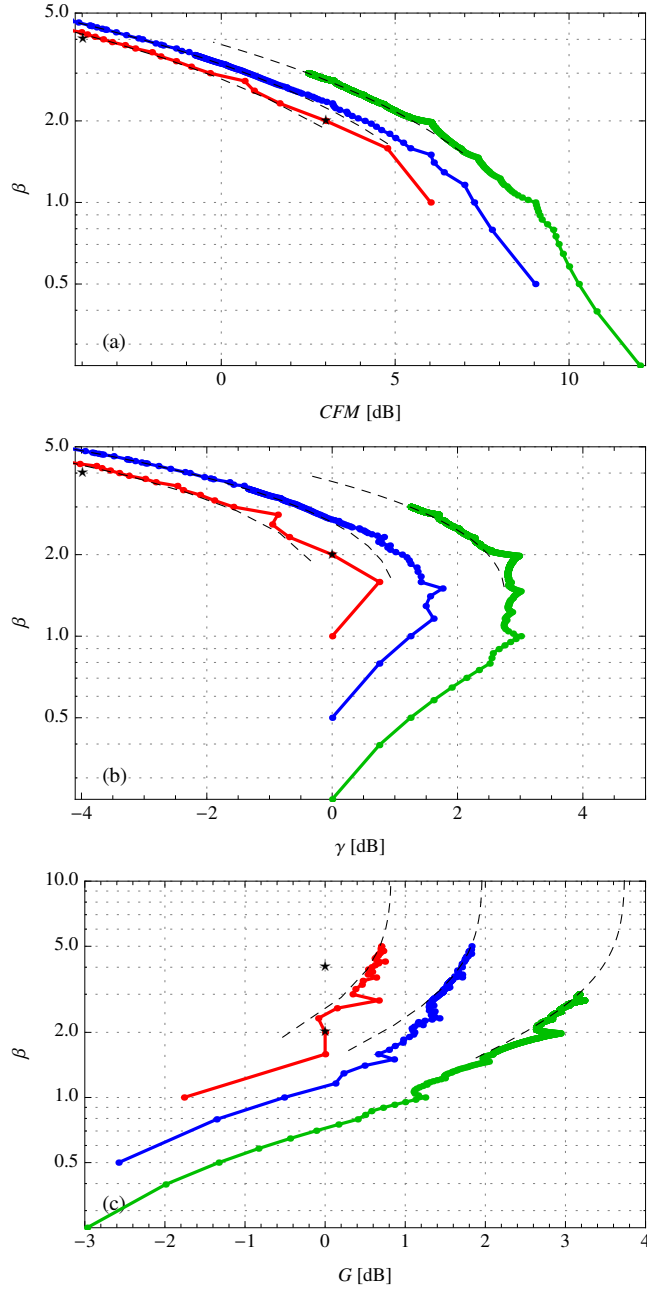


Figure 2.2 Spectral efficiency of the best known clusters, plotted vs. CFM , γ , and G , resp. Dimension $N = 2$ (red), 4 (blue), and 8 (green). The performance of the A_2 , D_4 and E_8 lattices using (2.7)–(2.9) is shown with dashed lines. The stars show BPSK (corresponding to QPSK in 2d and PM-QPSK in 4d) and 4-PAM (corresponding to 16-QAM in 2d and PM-16-QAM in 4d).

nates of the clusters are available online [28]. We conjecture that these clusters are all optimal for their values of N and M .¹

The spectral efficiency β is shown versus the three power measures: CFM , power efficiency γ , and gain G . This also shows the qualitatively different behavior of the three metrics (CFM , γ , and G). We will now discuss the general behavior of these metrics with spectral efficiency β (or M , since $\beta \sim \log_2(M)$).

The $CFM \sim 1/E(M)$ decreases monotonically with spectral efficiency β , as it compares formats at the same bandwidth (same baudrate), thus showing essentially how the second moment $E(M)$ increases with M . For large M , one can expect the clusters to behave as lattice packings, and the CFM to decrease as $\sim M^{-2/N}$ according to (2.7).

The γ on the other hand weighs in the data rate by multiplying CFM with $\log_2(M)$, giving it a dependence $\gamma \sim \log_2(M)/E(M)$. It can be shown that γ always increases up to at least the simplex ($M = N + 1$). However, for large M , the dependence is the lattice's $\sim \log_2(M)M^{-2/N}$, which will eventually decrease with M , and we conclude that for every dimension $N > 1$, γ has a maximum γ_{\max} at some value M_{opt} . The values of γ_{\max} and M_{opt} are only known, or conjectured, for $N = 2, 3, 4, 8$ and listed in Table 2.2. Not much is known about the general dependence of γ_{\max} and M_{opt} on the dimension N . However, a crude approximation can be obtained from the lattice expression, and maximizing $\log_2(M)M^{-2/N}$ for real M . This optimum is

$$M_{\text{opt,lat}} = \exp(N/2) \quad (2.10)$$

$$\gamma_{\max,\text{lat}} = \frac{N+2}{2e \log_2(2)} \Delta^{\frac{2}{N}}. \quad (2.11)$$

These values are compared to the exact known values² in Table 2.2, and the agreement is surprisingly accurate, given the rough approximation involved by approximating the discrete points with the homogeneous lattice distributions. It is also interesting to note that $M_{\text{opt,lat}}$ corresponds to $\beta_{\text{opt,lat}} = 1/\log_2(2) = 1.44$ bits per symbol per dimension pair, independently of N .

The gain G is defined as the performance relative to the cubic-lattice PAM constellations (QPSK, 16-QAM, PM-QPSK, PM-16QAM, etc.), which all have $G = 0$. The clusters show a rapid improvement over the cubic lattice as β increase, as is clear from figure 2.2 (c). At high spectral efficiencies, the gain will approach the asymptote given by

$$G_{\max} = \frac{N+2}{3} \Delta^{\frac{2}{N}} \quad (2.12)$$

which is 0.84, 1.97, and 3.72 dB in the respective 2d, 4d, and 8d cases.

¹The conjecture does not extend to $N = 8$ and $M > 128$. These clusters, which are also included in figure 2.2, are unpublished and not as well optimized as those on [28].

²With “exact known values” we mean, like for most other sphere packing problems, that they are found from extensive numerical optimization. No formal proofs of optimality are known.

Table 2.2 Known maxima for γ and their optimum number of points M_{opt} . For comparison are shown the optima based on the asymptotic lattice expression 2.7.

Dim. N	γ_{max} [dB]	M_{opt}	$\gamma_{\text{max,lat}}$ [dB]	$M_{\text{opt,lat}}$
2	0.75	3	-0.15	2.8
3	1.25	4	0.36	4.5
4	1.76	8	0.98	7.3
8	3.01	16	2.73	54

2.3.1.2 2-dimensional formats The 2d clusters are in almost all cases part of the hexagonal lattice A_2 , which is the densest packing of many spheres in 2d space. The only exception is $M = 4$, for which every rhombic constellation with vertex angle between 60° and 120° , including the square constellation (QPSK), have the same average symbol energy E as a four-point subset of A_2 . Second moments of clusters for M up to 500 are listed in [34].

Foschini et al. [33] found the optimum 2d clusters in the cases of practical interest ($M = 8$, $M = 16$) by numerical optimization already in 1974, but clearly these results have not taken on in the community, and there are at least three reasons for this: (i) the noninteger coordinates makes a practical implementation more difficult, (ii) the gains G over square QAM constellations are never more than 0.84 dB according to (2.12), and (iii) (less important) the hexagonal constellations do not lend themselves to a straightforward bit-to-symbol mapping. QAM constellations are therefore dominating in practical 2d systems.

The full set of 2d clusters up to $M = 32$ are shown in red for the three metrics (γ , CFM and G) in figure 2.2. The most common formats QPSK and 16-QAM are shown as stars in figure 2.2, and in figure 2.2 (c) they are references at $G = 0$. In the limit of many points, the 2d clusters have performance close to the A_2 lattice (shown with a dashed line), which is not surprising since they *are* cuts from this lattice, as shown by Graham et al. [34].

The highest γ is seen to arise for $M = 3$ (3-PSK), at $\beta = 3/2$. However, as for all the other 2d clusters (except for QPSK), it has seen limited use, although being discussed in the literature [37, 38].

2.3.1.3 4-dimensional formats The 4d clusters $\mathcal{C}_{4,M}$ are shown in figure 2.2 in blue. Similar optimizations are found in, e.g., [39] as well as in online listings of coordinates by, e.g., Sloane et al. in [35]. A more detailed description, including exact coordinate descriptions of some interesting clusters, were provided in [27, 31]. For communication purposes, the powers of two, $M = 8, 16, 32 \dots$, are of particular interest, and they will be discussed separately below.

Generally, the optimum, or nearly optimum, 4d constellations that are subsets of the D_4 lattice are easier to implement than the corresponding 2d clusters, since the D_4 lattice is a subset of the regular cubic (integer) lattice \mathbb{Z}^4 . They will thus have a better opportunity to find wide use than the 2d clusters. Also, higher gains G are attainable in 4d than 2d.

A few specific cases have caused interest in the research community, and will be discussed separately below, namely $M = 4, 8, 16$, and 24 , as well as the higher powers of 2.

3d simplex The best packing of 4 points in 4d, the cluster $\mathcal{C}_{4,4}$, is to put them in a regular tetrahedron, also known as the 3-dimensional *simplex* [40, p. 178]. Obviously, this is not a 4d object at all, since at least 5 points are required to span a 4d object, but it is the best packing of 4 points in all dimensions $N \geq 3$. Moreover, numerical evidence indicates that all clusters $\mathcal{C}_{N,M}$ where $M \leq N + 1$ are the M -ary simplices. In optical communications, this format was proposed and evaluated by Dochhan et al. [41] as an alternative to PM-BPSK, over which it has a 1.25 dB asymptotic sensitivity gain.

PS-QPSK The maximum γ in 4d occurs for $M = 8$, as originally pointed out in [42]. Geometrically, the format is the 4d *cross-polytope*, and also known in the communications community as 8-ary *biorthogonal modulation*. The biorthogonal (or cross-polytope) format consists of all permutations and signs of signal vectors with zeroes at all coordinates except one [40, p. 178]. Exact SER expressions for all biorthogonal formats are given in [43, eq. (4.102)], [44]. Gray mapping is not possible for biorthogonal formats, but assuming the “obvious” bit to symbol mapping that flips all bits between opposing symbol pairs $\pm 1, 0, 0, 0 \dots$, an exact expression for the BER was given in [43, p. 203].

The 8-ary biorthogonal format was originally proposed for optical coherent systems by Betti et al. [5], although it had been considered for communications much earlier [12, 13, 45]. It can even be considered as a special case of permutation modulation, introduced already by Slepian in 1965 [46].

In 4d, the cross-polytope can take on many representations [42]; in addition to the permutations of $\pm(1, 0, 0, 0)$ it can be regarded as the odd (or even) parity subset of the 4d cube (PM-QPSK). It can thus also be seen as resulting from a parity-check code applied to the standard PM-QPSK [47, 48], as will be discussed in section 2.4.1. The strength lies in that it loses less in spectral efficiency than it gains in sensitivity over PM-QPSK, so compared at the same bit rate, it gains $\gamma = 3/2$ or 1.76 dB in power efficiency. At a finite BER of 10^{-3} , its gain is around 1 dB.

Transmission simulations of PS-QPSK in nonlinearly-limited fiber links were presented in, e.g., [49–51]. The general result is that the power efficiency improvement over conventional PM-QPSK can be translated into a reach extension or increased amplifier span losses, which has also been seen in experiments as discussed in section 2.5 below.

6PolSK-QPSK/24-cell The 24-cell is a four-dimensional polytope that is, according to Coxeter [52], "... a peculiarity of four-dimensional space... having no analogue [in dimensions] above or below." The constellation consists of 24 vertices equally spaced from the origin and each other, and plays an important geometric role of being the Voronoi cell of the D_4 -lattice, as well as the the 4d kissing constellation, the latter being proved relatively recently [53]. The kissing constellation, consisting of the 24-cell and a point at the origin, is also the cluster $\mathcal{C}_{4,25}$, notably a local maximum in the γ vs. β plot, figure 2.2 (b). The cluster $\mathcal{C}_{4,24}$ is *not* the 24-cell, but $\mathcal{C}_{4,25}$ with an outer point removed and centered at the center of gravity.

Nevertheless, the 24-cell $\mathcal{C}_{24\text{-cell}}$ performs quite well as a format in its own right. The points can be given as the hypercube in union with the cross-polytope, i.e.,

$$\mathcal{C}_{24\text{-cell}} \in \{(\pm 1, \pm 1, \pm 1, \pm 1), (\pm 2, 0, 0, 0)\}$$

taken with all permutations and sign selections. An alternative, rotated and rescaled, representation is all permutations and sign selections of $(\pm 1, \pm 1, 0, 0)$. The 24-cell has a spectral efficiency $\beta = \log_2(24)/2 = 2.29$ and power efficiency $\gamma = \log_2(24)/4 = 1.15$ (or 0.59 dB). An exact expression of its SER was given in [27] and derived in [44]. It was suggested for communications in [12, 13, 54]. In coherent optical communications, it was first proposed by Bülow et al. in 2009 [55], and later identified as the 24-cell in [27]. In optical links, it is realized by transmitting QPSK in one of 6 different polarization states (x,y, $\pm 45^\circ$, right/left-hand circular) [55], and hence referred to as 6-polarization shift-keying (6PolSK)-QPSK.

The 24-ary nature of 6PolSK-QPSK makes the bit-to symbol mapping nontrivial, although a scheme was proposed in [27] based on mapping 9 bits to two subsequent 6PolSK-QPSK symbols. The resulting format has a slightly reduced γ of 0.51 dB and spectral efficiency of 2.25 bit/symbol/polarization [27].

M -SP-QAM and the D_4 -lattice As was mentioned above, the D_4 lattice is the densest packing of many points in 4d space [36]. It is therefore of importance when finding useful modulation formats for 4d transmission lines. There are systematic and low-complexity ways of doing this, rather than resorting to sphere packing optimizations and the above-mentioned clusters. The idea is to cut finite portions from the D_4 lattice by using a cubic or spherical cut. The former is easiest in implementations, but the latter is better from a theoretical perspective, which is ultimately, for many points, 1.5 dB better, as was discussed above.

Generally, the D_4 lattice can be defined, e.g., as all points with integer coordinates that sum to an even number. It can be obtained from the cubic (integer) lattice \mathbb{Z}^4 in two ways, either by *reduction*, or *extension* [56, 57]. The reduction scheme is similar to Ungerboeck's *set partitioning*, (SP), introduced for trellis-coded modulation [58]. The idea is to remove half of the points in \mathbb{Z}^4 points, e.g., those with odd parity. Thus one has

$$D_4 = \left\{ (k_1, k_2, k_3, k_4) \in \mathbb{Z}^4 \mid \sum k_i = \text{even} \right\}. \quad (2.13)$$

The extension scheme is instead to start from \mathbb{Z}^4 add a shifted variant of \mathbb{Z}^4 half an integer in every dimension, i.e.,

$$\mathbb{Z}^4 \cup \mathbb{Z}^4 + \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right).$$

However if the center-of-mass should remain at the origin (which is most efficient from a power efficiency point of view), it is better to shift the two cubic lattices an equal amount in opposite directions, i.e.

$$D_4^* = \mathbb{Z}^4 - \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \cup \mathbb{Z}^4 + \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \quad (2.14)$$

Both methods give the D_4 lattice (apart from a rescaling) and both are useful when obtaining power efficient modulation formats, especially with high spectral efficiency.

The reduction scheme was originally suggested for optical communications by Coelho and Hanik [59] who called the resulting formats M -ary set-partitioned QAM, or M -SP-QAM. Later, Karlsson and Agrell [57] extended the concept to the whole hierarchy of formats obtainable from extension or reduction of the standard rectangular QAM formats.

In the reduction process, the minimum distance squared is increased by a factor of two, at the expense of losing 1 bit per symbol. Applying this to PM-QPSK leads to the PS-QPSK format with a gain of $2 \times 3/4 = 3/2 = 1.76$ dB over PM-QPSK. Applying the same technique to PM-16-QAM is more attractive, leading to a gain of $2 \times 7/8 = 7/4 = 2.43$ dB over PM-16-QAM. This format is called 128-SP-QAM, and after being introduced by Coelho and Hanik [59] it was studied in simulations of nonlinear transmission by Renaudier et al. [60] and Sjödin et al. [61]. The latter paper also discussed the problem of bit-to symbol mapping and maximum-likelihood decoding for the format.

By using the extension scheme on PM-QPSK, one obtains the 32-SP-QAM format [57, 60] which has $\gamma = 0$ dB, i.e., the same power efficiency as PM-QPSK, but transmitting 5 bits per symbol rather than 4. By using extension and reduction for known QAM formats, an M-SP-QAM hierarchy with $M = 8, 32, 128, 512, 2048 \dots$ can be realized, and in the recent review article by Fischer et al. [62], more properties of these formats are given, including, e.g., mutual information. The following relations between power efficiency and spectral efficiency, corresponding to (2.5), can be derived for the SP-QAM hierarchy [57]

$$\gamma_{\text{SP-QAM}} = \frac{3\beta}{2^{\beta+\frac{1}{2}} - 1}, \quad (2.15)$$

$$\gamma_{\text{SP-QAM}} = \frac{3\beta}{2^{\beta+\frac{1}{2}} - \frac{1}{2}}, \quad (2.16)$$

where (2.15) holds for SP-QAM formats obtained by reduction of a rectangular PAM format and (2.16) for extension.

Other 4d formats of interest The 16-ary cluster $C_{4,16}$ is 1.11 dB better (in the γ sense) than the hypercube (PM-QPSK), but the coordinates are not very nice. Layered along one coordinate axis, it consists of a 3d octahedron and a 3d cube, sandwiched between two single points. Explicit forms of the coordinates were given in [31, 63]. The mutual information reveals only a marginal improvement over PM-QPSK, see section 2.4.3 and [63], although it has received some experimental interest as discussed in section 2.5.

Another improvement over the PM-QPSK format was proposed by Sjödin et al. [64]. Referred to as *subset-optimized PM-QPSK* (SO-PM-QPSK), the idea was to improve PM-QPSK by rescaling one (e.g., the even-parity) subset and leaving the other unchanged. By optimizing the rescaling to 1.618 (the golden ratio), a 0.44 dB γ improvement over PM-QPSK can be obtained. ML decoding schemes, SER, and BER for this format are given in [64].

It is possible to obtain a nice symmetric 256-point format by cutting the D_4 lattice with a spherical cut around a deep hole, as pointed out, e.g. in [12, 65] and recently described and experimentally implemented by Eriksson et al. [66]. The levels comprise all 4d vectors that lie within a radius of 6, whose coordinates are odd integers and where the coordinate sum is a multiple of 4. Remarkably this is exactly 256 vectors, and the $\gamma = 16/27 = -2.27$ dB, and it is quite likely the most power-efficient 256-ary constellation in 4d.

2.3.1.4 8- and higher-dimensional formats The 8d clusters $C_{4,M}$ are shown in green in figure 2.2. The maximum power efficiency, γ_{\max} , is 3.01 dB and occurs for $C_{8,16}$, which is the 8d cross-polytope, or biorthogonal 16-ary modulation. Interestingly, almost the same γ is obtained for $M = 58$ and 241 [28]. Expressions for the SER and BER of this format can be found in [43], and it was recently experimentally realized, as will be discussed in section 2.5.

For higher spectral-efficiency constellations, the best 8d lattice packing is given by the E_8 lattice [36], which can be obtained from the D_8 -lattice in union with a shifted D_8 , i.e.,

$$E_8 = D_8 \cup D_8 + \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right),$$

where D_8 is defined in analogy with (2.14).

Koike-Akino et al. [67] and Millar et al. [68] discussed two ways of obtaining 8d constellations, namely by cutting (spherical) parts from the E_8 lattice and using known block codes. The latter will be more discussed in section 2.4. In [67], they classified a few promising 8d modulation formats ($M = 128, 256$) in terms of γ as well as in terms of nonlinear transmission reach. In [68], they went deeper and generalized the study to even higher dimensions as well, e.g., 6d, 16d and 24d. In 16d, the Barnes–Wall lattice is known to be the densest, and a promising 16d constellation with $M = 2^{11}$ points was found from cuts of this lattice which might be the γ_{\max} of 16d, even if further studies are required before this can be settled. It would, interestingly, be in agreement with the approximate expression (2.10). In 24d, the

Leech lattice (and the associated Golay code) was used. More details on the latter, including experiments, were given in [69, 70].

2.3.1.5 PPM-based formats Pulse-position modulation (PPM) is a well-known technique to increase power efficiency at the expense of spectral efficiency. The idea is to frame 2^p symbols in time to a K -ary “supersymbol”, of which one slot is selected for the transmission of a single pulse. One can in this way transmit $\log_2(K)$ bits per supersymbol. It has been suggested to combine PPM with higher-order modulation formats to a hybrid K -PPM- M -QAM format by transmitting modulated data in the selected PPM slot [71, 72]. In [47], the efficiency γ of various such formats were theoretically calculated and compared. It was pointed out that PS-QPSK is equivalent to 2-PPM-QPSK—it is just another set of four dimensions. To use PPM (i.e., subsequent symbol slots) instead of polarization is often an easier way of realizing more dimensions, and in fact such formats come closer to being an FEC code. Recently, the PPM idea was further generalized to allow supersymbols with an arbitrary number of nonzero slots (instead of just one) [73]. For example, with inverse PPM, iPPM, the idea is to transmit in all symbols but one in the PPM frame. In this way, formats could be realized that have both higher spectral efficiency and higher sensitivity than PM-QPSK. An example is 8iPPM-QPSK, which has $\beta = 2.13$ and $\gamma = 0.84$ dB.

2.3.2 Optimized formats in nonlinear channels

Most formats discussed above have been optimized for the linear AWGN channel. However, as pointed out, the fiber is nonlinear, and often systems are operated in a weakly nonlinear regime, where the signal power is optimized as a trade-off between SNR and nonlinear distortions. What can then be done for the nonlinear channel in term of format optimization?

Within the GN model, the format optimization can be essentially the same as for the linear AWGN model assumed above, since the noise will be uniform and approximately Gaussian in all dimensions. Obviously, the formats based on optimizing minimum distance is reasonable for very high SNRs, and in a model with limited SNR (as the GN model) one would need to optimize at a constant SNR. This can be done, but is more computationally demanding. Not much has been published in that area, apart from the work by Foschini et al. [33].

One could also argue (see the discussion in [31]) that the balls would be better than the clusters, since they suffer less penalties for average power-limitations than clusters do for maximum power-limited channels [31]. Then again, in the GN model the average signal power is relevant, which speaks in favor of the clusters. Moreover, as discussed in section 2.2.3, there is no simple mapping between peak-power limits in discrete and continuous time. The former is easier to analyze, the latter makes more physical sense. Therefore a more rigorous comparison between balls and clusters in nonlinear links remains to be done.

Format optimizations have also been done for non-Gaussian channels, such as the phase-noise channel model used in, e.g., [21, 74–76]. Lau and Kahn [74] compared

various 4-point constellations, and managed to improve the nonlinear tolerance significantly by going from QPSK to a constellation with 3-PAM plus a fourth point further out. In [21, 76], constellations in the nonlinear regime were considered. A comparison was made between constellations with points on 2 and up to 5 different radii. By optimizing 16-point constellations, a few dB of increased nonlinear tolerance was seen.

In this context, the recent work by Kayhan and Montorsi [77] on constellation optimization should be mentioned, although they considered a linear phase-noise channel model. As Foschini et al. [33], they considered finite SNR, but used a different target function for the optimization process (approximations and variants of mutual information).

In [75], *satellite constellations* were introduced, in order to show that the channel capacity of any channel (linear or nonlinear) may not decrease with signal power. They are formed by taking a standard format, e.g., 8-PSK, and moving one point far out from the rest. This yields a constellation whose minimum SER, as well as maximum MI, occurs at a high average power, which can be made arbitrarily high by moving the lone point (satellite) further from the rest. A similar trick was used by Steiner [78] when optimizing formats in the low-SNR regime.

2.4 Combinations of coding and modulation

So far, the comparisons of modulation formats in this chapter have concerned uncoded transmission. Modern optical communication systems, however, often include some kind of FEC coding [79–81]. For best system performance, the code should influence the choice of modulation format. For example, a modulation format with high spectral efficiency may require a lower-rate code (better error protection capability) than a modulation format with lower spectral efficiency.

This section will discuss optimization of modulation formats in coded systems. We will distinguish between three cases, depending on the type of decoder employed, which pose quite different requirements on the choice of modulation format. The three cases are soft-decision decoding, hard-decision decoding, and iterative decoding, which loosely correspond to weak, medium, and strong coding, respectively. Most of this section will be devoted to the first case, which is more intimately connected to the problem of constellation design.

2.4.1 Soft-decision decoding

We here consider the application of a relatively short-length, well-structured block code, such as a single-parity check code, Hamming code, or Reed–Muller code. The employment of such simple codes imply low latency, simple encoding and decoding hardware, and hence low energy consumption. Nevertheless, significant improvements over uncoded transmission can be obtained.

If the decoder is an optimal soft-decision decoder, in the sense that it finds the codeword that is closest to the received word in Euclidean distance, then the com-

bination of code and modulation can be regarded as a single higher-dimensional modulation operation. This approach has been developed extensively in the communications literature [29, 82] [43, Sec. 4.1] and more recently in an optical context [68].

To be precise, let \mathcal{C} be a binary linear block code with parameters (n, k, d_H) , where n is the total number of bits per codeword, k is the number of information bits per codewords, and d_H is the minimum Hamming distance. The code rate is k/n . If this code is used in combination with a (low-dimensional) constellation $\tilde{\mathcal{X}}$ with dimension \tilde{N} and size \tilde{M} , then $\log_2 \tilde{M}$ bits are needed to index each point in the constellation, and the k information bits in a codeword suffice to index a block of $k/\log_2 \tilde{M}$ constellation points (assuming that this is an integer). This block can be regarded as a point in a larger constellation \mathcal{X} with parameters

$$N = \frac{n\tilde{N}}{\log_2 \tilde{M}},$$

$$M = \tilde{M}^{k/\log_2 \tilde{M}} = 2^k.$$

There exists no general expression for calculating the minimum Euclidean distance d from the parameters of \mathcal{C} and $\tilde{\mathcal{X}}$. It depends heavily on the mapping from bits to symbols, which needs to be done with some care.

The simplest, and most common, special case is to let $\tilde{\mathcal{X}}$ be a BPSK constellation $\tilde{\mathcal{X}} = \{\pm\sqrt{\tilde{E}}\}$, where \tilde{E} is the symbol energy, with parameters $\tilde{N} = 1$ and $\tilde{M} = 2$. This yields, for every code \mathcal{C} , a constellation \mathcal{X} with parameters

$$\begin{aligned} N &= n, \\ M &= 2^k, \\ d^2 &= 4\tilde{E}d_H, \\ E &= n\tilde{E}. \end{aligned} \tag{2.17}$$

Geometrically, the constellation \mathcal{X} resides on the vertices of an n -dimensional hypercube. The binary code is used to select a subset of these vertices.

Consider, for example, the 4-bit single-parity check code with parameters $(n, k, d_H) = (4, 3, 2)$. This yields, by (2.17) and (2.4), a 4-dimensional constellation with 8 points and power efficiency $\gamma = d^2 \log_2 M / (4E) = d_H k / n = 3/2 = 1.76$ dB. This constellation is identical to the PS-QPSK constellation, described in section 2.3.1.3 [27].

Using this approach, a large number of high-dimensional constellations can be designed from standard binary block codes. An attractive family of codes for this purpose is the Reed–Muller (RM) codes. These codes have rather good performance (albeit not optimal) at short block lengths n , and furthermore, there exist fast encoding and decoding algorithms, alleviating the need for table look-up [83, Sec. 4.3]. An RM code is specified by two integer parameters, u and r , chosen such that $u \geq 1$ and $0 \leq r \leq u$. The parameters of the $\text{RM}(r, u)$ code are $(n, k, d_H) = (2^u, \sum_{i=0}^r \binom{u}{i}, 2^{u-r})$. Special cases are repetition codes ($r = 0$),

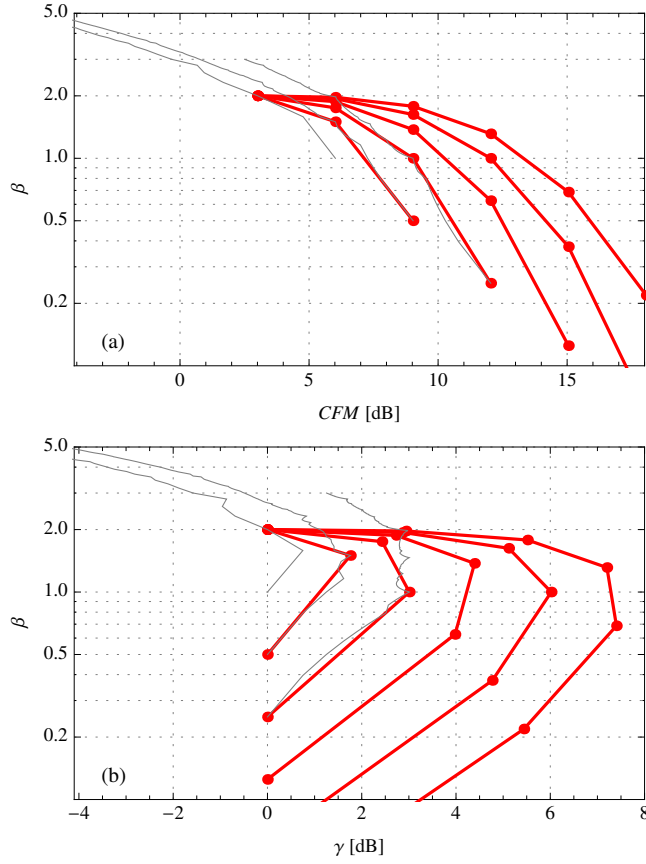


Figure 2.3 Reed–Muller codes with BPSK modulation. Codeword lengths $n = 4, 8, 16, 32, 64$ (red curves, from left to right). Minimum distance $d_H = 1, 2, 4, \dots, n$ (dots on each curve). Dimension $N = n$. The best known clusters from figure 2.2 are shown for reference (gray).

single-parity check codes ($r = u - 1$), and the universe code (i.e., uncoded transmission, $r = u$).

The parameters of some constellations obtained from Reed–Muller (RM) codes are illustrated in figure 2.3. The obtained constellations are apparently quite competitive, compared with the best known constellations at the same dimensions and spectral efficiencies, and in several instances the RM codes actually yield the best known constellations. For any $\beta < 2$, arbitrarily high CFM and γ can be obtained by choosing suitable RM code parameters. This makes RM codes attractive instruments for constellation design, especially since low-complexity coding and decoding algorithms are known.

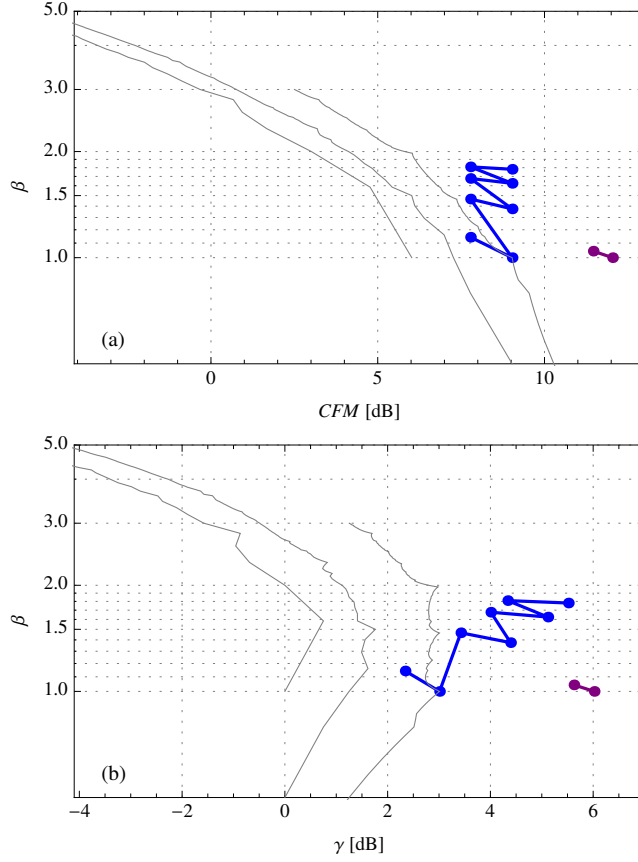


Figure 2.4 Hamming codes and their extended versions with BPSK modulation, with codeword lengths $n = 7, 8, 15, 16, 31, 32, 63, 64$ (blue). Golay ($n = 23$) and extended Golay code ($n = 24$), also with BPSK modulation (purple). Dimension $N = n$.

Analogous curves are presented in figure 2.4 for Hamming codes, the Golay code, and their extended versions. It turns out that the Hamming codes yields a constant CFM of 7.78 dB, regardless of their size, and similarly, the extended Hamming codes yield $CFM = 9.03$ dB. The power efficiency increases with the codeword length n and reaches asymptotically 4.77 dB for Hamming codes and 6.02 for extended Hamming codes. The extended Golay code also reaches $\gamma = 6.02$ dB, at a lower spectral efficiency [69].

Theoretically, nothing prevents us from designing extremely high-dimensional constellations by applying the methods of the previous section to codes with long codewords. Consider, for example, the ubiquitous (255,239) Reed–Solomon (RS) code, which was standardized by the ITU-T in 2000 [84]. This code encodes 239 information bytes into blocks of 255 transmitted bytes, and it has an error-correction

capability of 8 bytes. This error-correcting capability corresponds to 8 bit errors in the worst case, but if the bit errors come in burst, many more than 8 bit errors can be corrected, as long as the errors do not affect more than 8 bytes in total.

The parameters of the (255,239) RS code, converted from bytes to bits, are $(n, k, d_H) = (2040, 1912, 17)$. This code can, if combined with BPSK modulation as in (2.17), be regarded as a 2040-dimensional constellation with parameters $\beta = 1.87$, $CFM = 15.31$ dB, $\gamma = 12.02$ dB, and $G = 11.79$ dB. Compared with the curves in figures 2.2–2.4, this constellation has an impressive performance, falling far to the right of any of the curves. This exemplifies the essence of coding, which is to improve the power efficiency by increasing the dimensionality.

One could ask whether it makes sense to consider a 2040-dimensional constellation with $M = 3.7 \cdot 10^{575}$ points. Clearly it is not possible to enumerate or store all the points. However, if soft-decision decoding is used in the receiver, the performance predicted by the constellation analysis above is indeed achievable, for well-structured codes such as RM and RS. Generalizing, we conclude that any low-dimensional modulation scheme with soft-decision FEC is equivalent to a high-dimensional modulation scheme without FEC.

In all these cases, the spectral efficiency β never goes above 2, which is the spectral efficiency for uncoded BPSK. This is a serious limitation in practical optical system implementations, where higher and higher spectral efficiencies are being targeted nowadays. To circumvent this limitation, one must employ a multilevel constellation instead of BPSK. This leads to a simple type of coded modulation [11]. Below we will give some simple examples, based on single-parity check codes and RM codes.

A natural extension to BPSK is to let $\tilde{\mathcal{X}}$ be a regular PAM constellations with $\tilde{M} = 2^{\tilde{m}}$ points. If the distance between two neighboring PAM points is \tilde{d} , then the average symbol energy of $\tilde{\mathcal{X}}$ is $\tilde{E} = (\tilde{M}^2 - 1)\tilde{d}^2/12$. As explained above, the k information bits in a codeword are divided into groups of \tilde{m} bits, thus indexing a block of k/\tilde{m} PAM symbols. This block of PAM symbols constitutes an N -dimensional constellation \mathcal{X} with parameters

$$\begin{aligned} N &= \frac{n}{\tilde{m}}, \\ M &= 2^k, \\ d^2 &\geq \tilde{d}^2 d_H, \\ E &\leq \frac{n\tilde{E}}{\tilde{m}} \end{aligned} \quad (2.18)$$

for a suitably chosen (Gray-coded) mapping from bits to PAM symbols.

An important special case arises by applying a single-parity check code with parameters $(n, k, d_H) = (\tilde{m}N, \tilde{m}N - 1, 2)$. The codeword length n is chosen so that the obtained constellation \mathcal{X} is N -dimensional. The obtained constellations are plotted in figure 2.5 for various values of \tilde{m} and N . The 4d case has been studied in optical communications under the name SP-QAM (see section 2.3.1.3). If, for example, the code with parameters (8,7,2) is mapped to a Gray-coded 4-PAM constellations, then

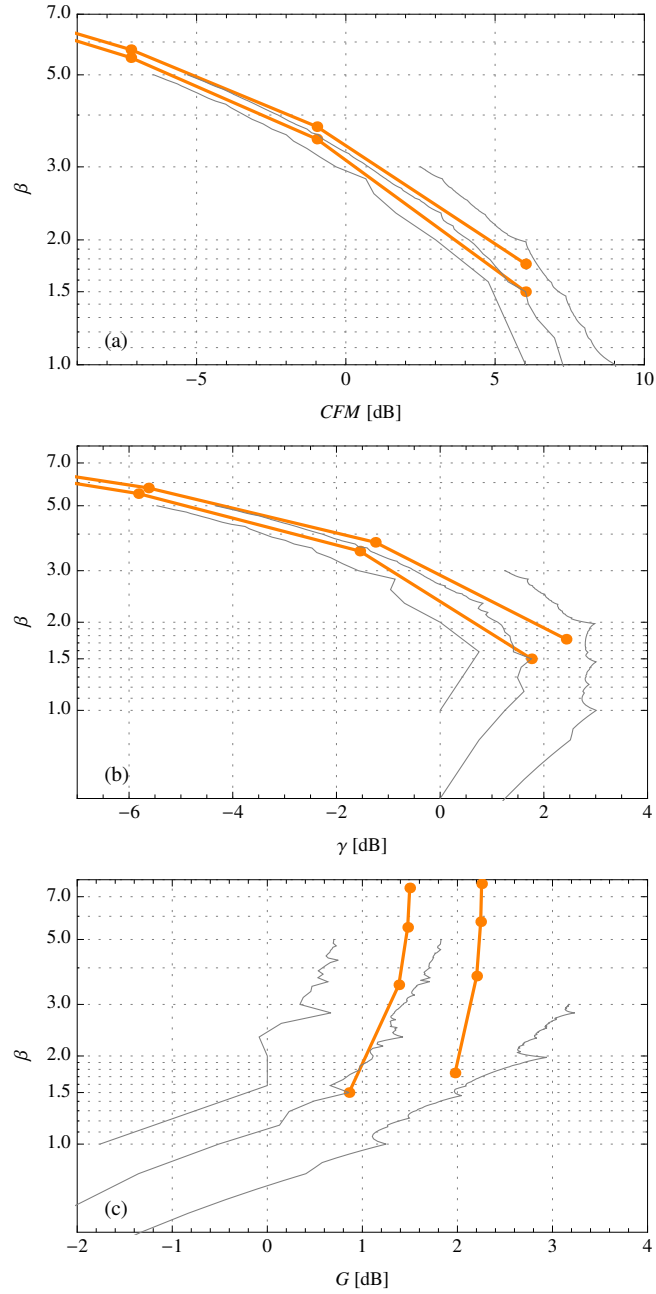


Figure 2.5 Single-parity check codes mapped to Gray-coded \tilde{M} -PAM, where the code lengths n is chosen so that the dimension $N = n/\log_2(\tilde{M})$ is an integer. Dimensions $N = 4, 8$ (from left to right). PAM alphabet size $\tilde{M} = 2, 4, 8, 16$ (points on each curve). The left curve ($N = 4$) represents SP-QAM, see section 2.3.1.3.

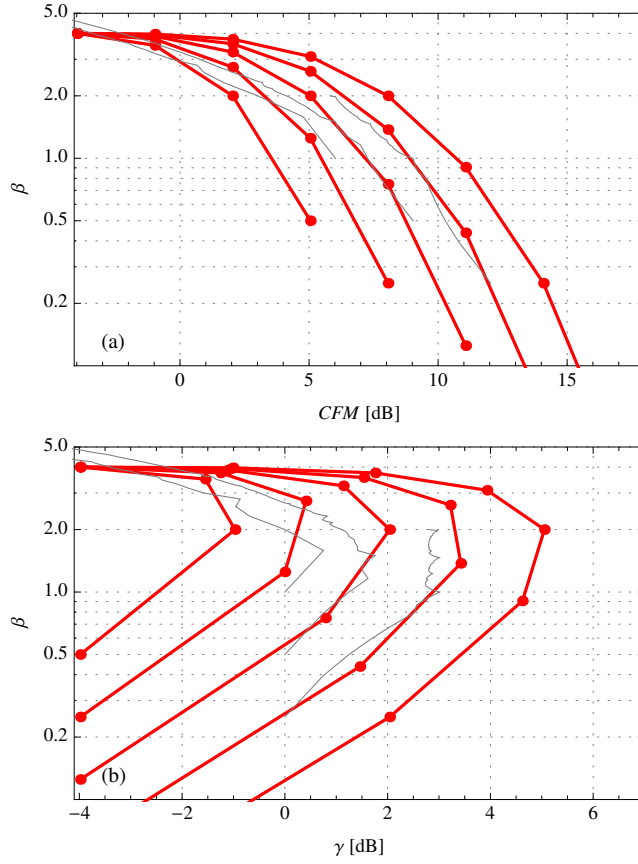


Figure 2.6 RM codes with Gray-coded 4-PAM. Codewords lengths $n = 8, 16, 32, 64, 128$ (from left to right). Minimum distance $d_H = 1, 2, 4, \dots, n$ (dots on each curve). Dimension $N = n/2$.

the parameters of the resulting 128-point constellations are, by (2.18), $d^2 \geq 2\tilde{d}^2$ and $E \leq 4\tilde{E} = 5\tilde{d}^2$, which yields $\gamma \geq 7/10 = -1.55$ dB, at a spectral efficiency of $7/2$. This modulation format is represented by one of the orange dots in figure 2.5 (b). As n increases, the gain G of SP-QAM converges to 1.51 and 2.26 for $N = 4$ and 8, resp. [57], which should be compared with the maximum possible gains 1.97 and 3.72 dB, resp., in (2.12). The asymptotic gain as N and M both approach infinity is 3 dB. There is no gain to be harvested in 2d by this method.

The same types of codes as in figures 2.3 and 2.4 were applied to 4-PAM, which yielded the results in figures 2.6 and 2.7. The obtained constellations are relatively weak, compared with the best known constellations at the same dimensions and spectral efficiencies. Nevertheless, the results show that it is in principle to achieve arbitrarily high CFM and γ at any $\beta < 4$, if the codeword length is increased sufficiently.

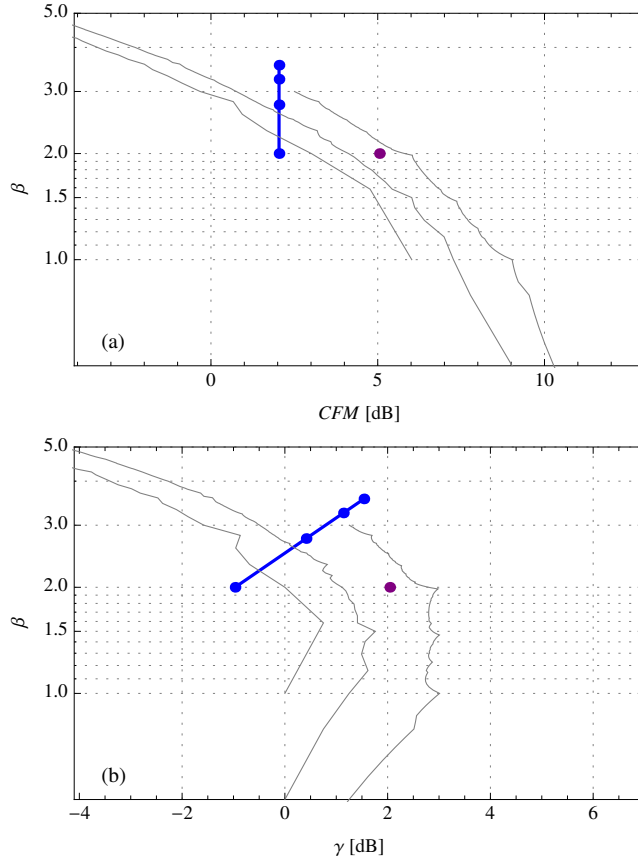


Figure 2.7 Extended Hamming codes ($n = 8, 16, 32, 64$) and extended Golay code ($n = 24$) with Gray-coded 4-PAM. Dimension $N = n/2$.

2.4.2 Hard-decision decoding

Most commercially deployed long-haul fiber-optical communication systems use hard-decision decoding, which can be realized at a significantly lower hardware complexity than soft-decision decoding. The decoder can be implemented using binary logic, with no need for analog-to-digital conversion. This in turn admits the use of stronger codes (longer codeword lengths). Reed–Solomon codes are the most popular codes in this context, but BCH (Bose–Chaudhuri–Hocquenghem) codes, Hamming codes, and convolutional codes have also been considered. The interested reader is referred to [79] for a survey of the field, which is outside the scope of this chapter.

In a system with hard-decision decoding, the geometric framework in the previous section makes less sense. In this case, performance metrics based on the minimum

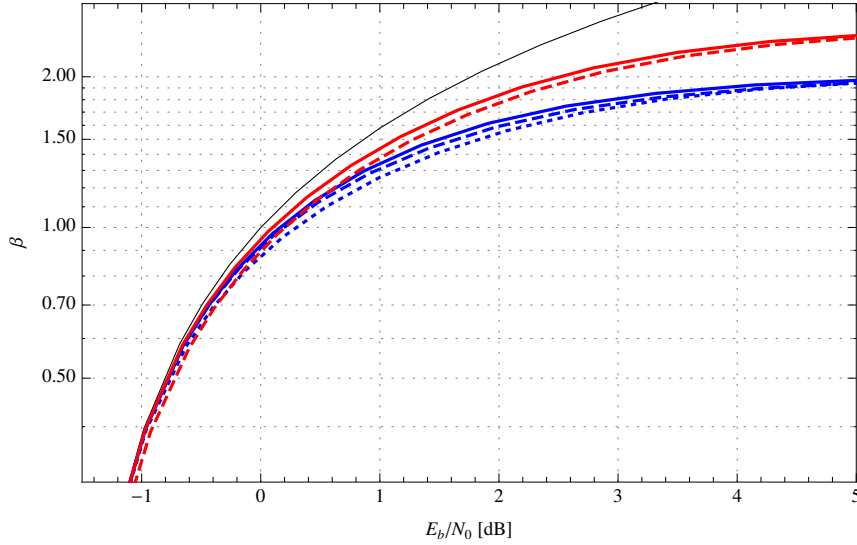


Figure 2.8 Spectral efficiency β , calculated as mutual information per dimension pair, as a function of the SNR per bit for selected 4d formats. Blue: 16 points, red: 32 points. Solid: best known constellations $\mathcal{C}_{4,M}$. Blue dashed: PM-QPSK, blue dotted: SO-PM-QPSK, red dashed: 32-SP-QAM. Black: AWGN channel capacity.

Euclidean distance are misleading, and modulation and coding should be kept separate in the analysis. The standard system design method is to choose a modulation format that guarantees a certain BER, the so-called FEC limit, which is typically in the range of 10^{-3} to 10^{-4} [84], and trust the FEC to bring down the BER to a negligible level. This design principle is very popular in practice, since it decouples the FEC from the rest of the system, which facilitates experimental work. Two main weaknesses with this standard approach is that it offers no simple mechanism to optimize the code rate (varying the FEC limit) and that it does not account for the bursty nature of errors.

2.4.3 Iterative decoding

Modern codes such as low-density parity check (LDPC) codes and turbo codes have revolutionized wireless communications, and an equally promising potential is envisioned in optical communications [80, 81]. These codes are typically very long (in the order of 10000 bits) and have a pseudorandom structure. Algebraic decoding would be far too complex for such codes, but there have been devised efficient iterative decoding algorithms, which gradually improves an estimate of the transmitted codeword, using either soft or hard decisions. Such decoders are not guaranteed to find the optimal codeword, but nevertheless these codes have excellent performance.

In some cases, they even approach the maximum spectral efficiencies predicted by Shannon in 1948 [85].

Shannon proved that the achievable data rate of a given modulation format is upper-bounded by the MI, defined in (2.6). Furthermore, recent research has shown that a performance very close to the MI is achievable using long LDPC codes and soft-decision iterative decoding. The spectral efficiency in bits per dimension pair is plotted in figure 2.8 for some common 4d constellations, as a function of the SNR per bit. The input \mathbf{X} is drawn uniformly from a constellation \mathcal{X} , given by the modulation format, and a memoryless AWGN channel is assumed. The spectral efficiency is here calculated as $\beta = I(\mathbf{X}; \mathbf{Y})/(N/2)$ and the SNR per bit as E_b/N_0 , where N_0 is the power spectral density of the Gaussian noise.

Although the plot only includes a small number of constellations, several interesting conclusions can be drawn. At a given target β , it is practically always beneficial in terms of power efficiency to increase the number of points in the constellation. The fact that the minimum distance decreases is fully compensated for by using a lower code rate (higher overhead). The gain obtained by increasing the number of points is however negligible at low β , where practically any constellation performs close to capacity.

If the number of points is kept constant, then the constellations designed for optimum uncoded performance tend to be good also in terms of MI. The gains in dB are however less than the corresponding gains in terms of γ or *CFM*. Interestingly, PM-QPSK is better than SO-PM-QPSK for all β values, in contrast to the uncoded performance where SO-PM-QPSK is 0.44 dB better (see section 2.3.1.3). This effect is even more pronounced in systems with bit-wise receivers, where PM-QPSK is the best known constellation, despite its simple structure, gaining approximately 1 dB over the cluster $\mathcal{C}_{4,16}$ [86].

2.5 Experimental work

To experimentally demonstrate 4- and higher-dimension modulation formats, one needs to be able to simultaneously access all dimensions in the transmitter and receiver. Depending on how the dimensions are physically realized in the channel (e.g., time, frequency, or spatial dimensions), this can be more or less complicated, as the used dimensions must be synchronized and not drift between symbols. This often requires tailored DSP algorithms for the considered modulation formats.

In this section, we review the experimental work done on mainly 4d formats, where the four dimensions are the conventional four quadratures (I/Q in each of the x and y polarizations). We will divide the discussion into first realizations of the transmitter and transmission link properties (section 2.5.1) and second the receiver algorithms (section 2.5.2), including DSP and decoding, with a summary table in section 2.5.3. Then, in section 2.5.4, we will discuss format detection, i.e., how to simply determine the transmitted symbol from the received 4d vector, without resorting to a full search of the Euclidean distances to all points in the whole constellation.

We will finally discuss alternative ways of extending dimensions in signal space in section 2.5.5 from a complexity and implementation perspective.

2.5.1 Transmitter realizations and transmission experiments

We will below describe the experimental work in similar order as in the above theory section 2.3.1.3.

3d simplex Dochhan et al. [41] proposed and demonstrated the 3d simplex (tetrahedron) transmission, by using a 4-channel digital-to-analog converter (DAC) driving a conventional PM-QPSK modulator. The format transmits QPSK in one polarization and BPSK in the other, thus leaving one quadrature unmodulated. Of the resulting 8 levels (forming a cube), the 4 with odd parity were selected, giving the desired tetrahedron. The symbol rate was 16 Gbaud, corresponding to a data rate of 32 Gbit/s, which was transmitted over 300 km of single-mode fiber. The back-to-back sensitivity was approximately 1 dB better than PM-BPSK, in agreement with theory, and the nonlinear robustness was similar to PM-BPSK. In [87], Yamazaki et al. developed a simple integrated modulator structure for this format.

PS-QPSK The first experimental realization of a 4d format in fiber optic transmission, was probably the demonstration by Sjödin et al. of PS-QPSK at 30 Gbit/s in 2011 [88]. In this experiment, a conventional I/Q-modulator for QPSK was used, and then the data was split into two arms, driven by a pair of Mach–Zender amplitude modulators in a push–pull constellation, meaning that either one or the other arm was blocked. Then the two arms were multiplexed together by a polarization combiner. In this way, two bits were encoded in the QPSK symbol and the third in the choice of polarization. Similar transmitter structures were used also by other groups around this time, e.g., Millar et al. [89] and Nelson et al. [90].

An alternative transmitter setup was used by Fischer et al. [91] who used a PM-QPSK transmitter with a programmable bit-pattern generator, driving the four bits with a preprogrammed pattern. Three of the bit streams were driven by uncorrelated (delayed) pseudo-random sequences, and the fourth was formed as a parity (exclusive OR, XOR) bit from these three sequences. This transmitter was also used by Renaudier et al. [92], who also introduced a timing offset between the two polarizations to facilitate the receiver DSP, as will be discussed in section 2.5.2 below.

In [93], Yamazaki et al. presented an integrated modulator optimized for PS-QPSK, which could directly generate PS-QPSK in a single device driven by three binary drive signals. It has the additional benefit of avoiding the inherent 3 dB coupling loss in the I/Q modulators when the I and Q quadratures are mixed.

Following the initial demonstration of single-channel transmission [88] came a stream of experimental demonstrations of PS-QPSK; first single-channel demonstrations at higher data rates, e.g., 42 Gbit/s [89, 94] and 112 Gbit/s [91, 95], and then WDM experiments over ultralong distances (thousands of kilometers) [89, 90, 95–97]. The general conclusion was that the improvement in transmission distance predicted in simulations [49, 50] was experimentally verified. Typically, PS-QPSK

achieved 10-25 % longer transmission reach (at a $\text{BER}=10^{-3}$) than PM-QPSK at the same data rate [89,97,98], both in single-channel systems and with WDM. Masalkina et al. [99] used PS-QPSK in a 20 Gbit/s orthogonal frequency division multiplexed (OFDM) transmission experiment. Lavery et al. [95] demonstrated that digital back-propagation could extend the reach for 112 Gbit/s PS-QPSK from 4600 km to 5600 km.

6PolSK-QPSK Experimental generation of 6PolSK-QPSK was first demonstrated in 2012 independently by Buchali and Bülow [100] and by Fischer et al. [101]. The transmitter structure used in these experiments was based on a 4-channel 28-Gbaud DAC driving a dual-polarization I/Q modulator at the three levels $\{-1, 0, 1\}$. By forming all 24 permutations and sign selections of the vector $(0, 0, \pm 1, \pm 1)$, the 24-cell is obtained. The symbol-to-bit mapping followed the suggestion in [27], by mapping 9 bits to 2 subsequent symbols. In back-to-back measurement, Fischer et al. [101] demonstrated 2.1 and 3.4 dB implementation penalties for PM-QPSK and 6PolSK-QPSK, respectively, for 28 Gbaud and a BER of around 10^{-3} . This corresponds to an extra implementation penalty for 6PolSK-QPSK of 1.3 dB. In [102] Ding et al. generated 6PolSK-QPSK by using a single dual-drive (i.e. not I/Q) Mach-Zehnder modulator in each polarization.

In transmission, Fischer et al. [101] demonstrated transmission of 19 WDM channels over 3400 km, which is less than the 4800 km of PM-QPSK. However, by applying a rate 455/511 Reed–Solomon code to 6PolSK-QPSK, making the bit rates the same for both formats, the transmission distance gap was partly bridged. A later extension of this work by Tanimura et al. [103] used a more advanced setup, including pre-emphasis to compensate for DAC imperfections, investigated the nonlinear tolerance in some detail. It was found that an inner FEC could be beneficial in removing bursty errors due to nonlinearities.

One conclusion from these works was that the lack of Gray mapping for the 6PolSK-QPSK induces an extra penalty relative to PM-QPSK, which is troublesome in the nonasymptotic regime.

M-SP-QAM According to predictions and simulations in, e.g., [57,59,61,92], the SP-QAM formats emerged as an interesting 4d format generalization with increased spectral efficiency over PM-QPSK. In 2013, 128-SP-QAM was demonstrated by three independent groups [104–106]. The transmitter structures in these experiments were similar; based on pre-programmable DAC:s using 8 bit streams of which one is a parity check bit.

Eriksson et al. [104, 107] used 12 Gbaud for 128-SP-QAM and compared with 10.5 Gbaud PM-16-QAM to obtain the same data rate of 84 Gbit/s. Both single channel and 9 WDM channel (25 GHz separation) transmission were compared. PM-16-QAM had a slightly higher implementation penalty (2.1 dB) relative to 128-SP-QAM (1.5 dB), attributed to the improved Euclidean distance of 128-SP-QAM. The back-to-back sensitivity improvement of 128-SP-QAM was 1.9 dB at the same bit rate, and 2.9 dB at the same symbol rate, in close agreement with theoretical expectations. In transmission, the reach was 1300 km (for PM-16QAM in a WDM

system) and 2000 km (for 128 SP-QAM, also in WDM), i.e., a 54 % improvement. It was also concluded that PM-16-QAM has a larger penalty when going from single channel to WDM than 128-SP-QAM.

Zhang et al. [106, 108] used 128 SP-QAM (denoted “half-4D-16QAM”) over 294 channels, (16.64 Gbaud at 17 GHz separation) covering the full C-band, at 104 Gbit/s each to achieve 30.58 Tb/s. The transmitter used bit-interleaved coded modulation (BICM) together with a 20 % overhead LDPC code. They also used DACs for Nyquist channel shaping to reduce interchannel crosstalk. This data was then transmitted over 7230 km, making it (to that date) the experiment with the highest bit rate-times-distance product of 221 Pb/s \times km. In a later experiment [108], digital back-propagation was used to increase the transmission distance to 10300 km, but at a reduced data rate of 21 Tb/s.

Renaudier et al. [105] demonstrated both 32-SP-QAM and 128-SP-QAM at 28 Gbaud. In back-to-back experiments, they achieved 0.8 and 1.5 dB implementation penalties, resp., relative to the AWGN theory. Then they propagated 16 channels in a circulating loop constellation, at optimized power (which was the same for all formats), and measured transmission distance. At a BER of $4 \cdot 10^{-3}$ they could propagate PM-QPSK, 32-SP-QAM, 128-SP-QAM, and PM-16-QAM, respectively, over 18000, 14000, 7000, and 4000 km.

At ECOC 2013, two independent groups [109, 110] compared 32-SP-QAM with another spectral efficiency $\beta = 2.5$ -bits/symbol/pol scheme, namely hybrid PM-QPSK/PM-8-QAM. The hybrid scheme means that half of the transmitted symbols are PM-QPSK ($\beta = 2$) and half are PM-8-QAM ($\beta = 3$). In [109], 65 symbols of PM-QPSK were followed by 65 symbols of PM-8-QAM. In [110], the interleaving was every second symbol, but done in both polarizations in a staggered manner, so that each symbol slot contained one polarization of QPSK and one polarization of 8-QAM, and then the next symbol swapped formats in the polarizations. The 8-QAM format used in both papers was the star-shaped constellation consisting of two QPSK constellation at different radii, rotated 45° relative to each other.³ Both studies concluded (in line with theoretical predictions) that 32-SP-QAM had better performance in terms of transmission distance.

32- and 128-SP-QAM were also demonstrated in few-mode fiber transmission over 42 km by van Uden et al. [111], but then using pairwise time slots, rather than polarizations, to set up the 4d space, and subsequently propagating 4d symbols independently in two polarizations and three spatial modes. Also, the PS-QPSK counterpart, denoted time-switched (TS)-QPSK, was implemented. The transmitter setup consisted of a 4-channel programmable DAC at 28 Gbaud, followed by a polarization- and mode-multiplexing stage. One conclusion from this experiment was that the 4d formats have less implementation penalty than their cubic-lattice counterparts (e.g., when comparing 128-SP-QAM with PM-16-QAM).

³This format is erroneously referred to as the optimum (in the γ sense) 8-ary 2d format in [40, p. 277], i.e., as the cluster $C_{2,8}$. However, as shown by Foschini in 1974 [33], $C_{2,8}$ is a hexagonal structure comprising the 7-point kissing constellation plus an extra point.

The record SP-QAM experiments in terms of constellation size are impressive 512- and 2048-SP-QAM, which are obtained from the related PAM constellations are PM-32-QAM (cross constellation) and PM-64-QAM, as recently demonstrated by Fischer et al. [112]. The formats were generated by carefully co-optimizing multilevel analog to digital converters with I/Q modulators, but the 2048 case suffered, not surprisingly, from 3-6 dB implementation penalty. The 512-SP-QAM had a more moderate 2 dB of implementation penalty.

More complex 4d formats The $\mathcal{C}_{4,16}$ cluster was experimentally demonstrated independently at OFC 2013 by Karout et al. [113] and by Bülow et al. [114]. The experiment [113] was based on an optical OFDM link, with 81 subcarriers generated by 2 synchronized DACs, comparing $\mathcal{C}_{4,16}$ with PM-QPSK. The signal bandwidth was 6.5 GHz, and the resulting data rate of 25.6 Gbit/s was obtained. A small overhead was allocated for training sequences and guard-band. A small performance gain for $\mathcal{C}_{4,16}$ could be seen, in good agreement with theoretical BER curves. Only back-to-back measurements, i.e., no transmission, were carried out in this demonstration. The experiment [114] used a baseband signal realized by a 28-Gbaud DAC in the transmitter. The data was transmitted 480 km, but the theoretical performance gain of $\mathcal{C}_{4,16}$ was shadowed by a larger implementation penalty than PM-QPSK. The experiments indicated an increased nonlinear tolerance of $\mathcal{C}_{4,16}$, but this was not conclusive.

An extension of 6PolSK-QPSK to 8PolSK-QPSK was proposed and demonstrated by Chagnon et al. [115]. This format uses 8 different polarizations, put on the cube corners in Stokes space, and each with QPSK modulation. The format was generated with a 4-channel DAC driving a dual-polarization I/Q modulator, yielding a data rate of 129 Gbit/s. The reach was 3800 km (including WDM channel loading), well above the 2800 km of the PM-8QAM it was benchmarked against.

Bülow et al. [116] also demonstrated experiments of formats obtained from spherical cuts from the D_4 -lattice, with $M = 64$ and $M = 256$. Power efficiencies over PM-8-QAM and PM 16-QAM of 1.5–1.7 dB were reported. The experiments also included bit-interleaved coded modulation using a 17 % overhead LDPC code with these formats, and PM-64QAM showed a 35 % reach improvement over PM-8-QAM.

Eriksson et al. demonstrated the 256-ary D_4 format, in [66], and compared it to PM-16-QAM at 56 Gbaud. As could be expected the D_4 format were better back-to-back, for low ($< 10^{-3}$) bit error rates, but for higher BERs (corresponding to transmission distances above 1500 km) PM-16-QAM had the edge.

Higher-dimensional formats As the only way of improving performance (increasing both power efficiency and spectral efficiency) beyond the limits of 4d formats is to increase the dimensionality of the constellations, there has been work in that direction as well. A natural extension is to move to eight dimensions (8d), and biorthogonal 8d modulation was implemented and evaluated by Eriksson et al. in [117, 118]. In [117], the 8 dimensions were formed by using two phase-locked neighboring frequency (or wavelength) channels, and then performing coherent detection of both 4d

channels in parallel. The transmitter in [117] was a generalization of the corresponding transmitter for PS-QPSK, i.e., a PS-QPSK transmitter in cascade with a pair of push-pull modulators to select frequency. The format was referred to as 4-ary frequency and polarization switched QPSK, 4FPS-QPSK. The transmission properties (at 10 Gbaud) showed a reach of 14000 km for 4FPS-QPSK, compared to 7500 km for PM-QPSK.

In [118], PPM (see section 2.3.1.5 and below) was used instead of frequency to realize the 8 dimensions. A pair of two PS-QPSK symbols formed a PPM frame, giving a 2PPM-PS-QPSK format, equivalent to the 8d biorthogonal format. The transmitter becomes notably simpler, requiring a single modulator to select time slot, followed by a PS-QPSK transmitter. The implementation penalty at a BER of 10^{-3} and a symbol rate of 21.4 Gbaud was 0.6 dB, slightly more than the 0.3 dB for PM-QPSK. The data rate was 85.6 Gbit/s for both formats, meaning that the 2PPM-PS-QPSK format used 42 GHz of bandwidth, twice that of the PM-QPSK transmission. The transmission reach was almost doubled; 2PPM-PS-QPSK reached 12300 km and PM-QPSK 6700 km, which is in reasonable agreement with a simple GN-model-based theory, predicting a doubling of the reach with a 3 dB improved sensitivity.

Shiner et al. [119] also demonstrated the 8d biorthogonal format by using two subsequent temporal symbols to for the 8 dimensions. Their experiment demonstrated transmission over 5000 km at 35 Gbaud, including WDM channels. They also rotated the constellation aiming to reduce the nonlinear effects, noticing a 1 dB improvement in the nonlinear system margin.

In a spatial-division multiplexed context, and in particular multi-core fibers, Putnam et al. [120] investigated modulation over several cores in parallel. Especially formats based on a single-parity-check scheme (as outlined above) showed good results. The idea was to coherently transmit 4 bits per symbol in K parallel cores, and then use a single bit as a parity check bit, giving a total rate $4K - 1$ parallel BPSK streams. It may be shown that this scheme has both the spectral efficiency and the power efficiency equal to $2 - 1/(2K)$. The special case $K = 1$ is equivalent to PS-QPSK. An experiment at 10 Gbaud over 28 km of multi-core fiber demonstrating the concept for $K = 7$ cores were performed, and the results were in good agreement with theoretical expectations.

PPM implementations Liu et al. [71, 121, 122] suggested combining pulse position modulation (PPM) with PM-QPSK to show a record sensitivity for data transmission of 16-PPM at 2.5 Gbit/s [71] and 4-PPM at 6.23 Gbit/s [121]. The main benefit of these formats are the increased power efficiency, requiring only 2 photons per bit at 2.5 Gbit/s, making them suitable in, e.g., single-span long-distance links that demands high sensitivity. The transmitter was a standard PM-QPSK transmitter, but driven with a more complex data signal, including PPM framing and also some synchronization overhead signals. The transmission was done over a single span of 370 km ultra-large-effective area fiber, with a total loss of 69 dB.

Slightly outside the main topic of this chapter, we could mention the free-space link experiment by Ludwig et al. [123], which used 64-PPM overlaid with PS-QPSK to require only 2.2 photons per bit for data transmission at 0.56 Gbit/s.

In this context, we shall also reiterate the experiments by Sjödin [124], Eriksson [118], and van Uden [111], who implemented PS-QPSK as 2-PPM-QPSK, i.e., by using two adjacent symbol time frames instead of two polarizations.

2.5.2 Receiver realizations and digital signal processing

The conventional coherent receiver digital signal processing for PM-QPSK operates according to the following flow. (See [15, 125] for more details and examples.)

1. Compensation for static errors in the receiver front end (timing skew, power imbalance, I/Q phase error, usually using Gram–Schmidt orthogonalization).
2. Static channel equalization (mainly chromatic dispersion compensation, usually with an finite impulse response (FIR) filter).
3. Dynamic channel equalization (mainly polarization tracking and polarization mode dispersion compensation, usually using the constant modulus algorithm, CMA).
4. Interpolation and timing (clock) recovery.
5. Frequency estimation.
6. Carrier phase estimation (usually using the Viterbi and Viterbi (VV) algorithm)
7. Symbol estimation and decoding.

Traditionally, *blind algorithms* have been mostly used to estimate and compensate for channel effects, meaning that the modulation format is known, but the exact symbol transmitted at each time is unknown. Much is based on the knowledge of the used format, so when changing modulation format, a number of these steps have to be altered or modified. Most “sensitive” in this respect are the final stages: symbol estimation and decoding, and the carrier phase estimation. For example, when moving from PM-QPSK to PM-16QAM, these stages obviously need to be modified. However, also the dynamic channel estimation stage (the CMA) may need to be modified when changing format. For PS-QPSK, this is the case; the CMA needs to be modified, whereas the other stages can be kept the same as for PM-QPSK. More recently, however, there is a research trend towards the use of *nonblind* schemes, where the channel estimation is based on known signals, *training sequences*, which are transmitted regularly. Especially in optical OFDM systems, this is popular, but also in conventional baseband transmission it is becoming used [126, 127]. In the following, we will describe how these stages need to be modified for some of the 4d formats.

3d simplex The experiment of Dochhan et al. [41] used more or less standard coherent receiver algorithms, although slightly modified CMA and phase tracking algorithms were used. A blind algorithm similar to the one presented by Yan et al. [128] was used together with the standard CMA for polarization tracking.

PS-QPSK Naively, one might think that since PS-QPSK is a subset of PM-QPSK, the conventional DSP should work also for PS-QPSK. This is true for all DSP stages except the polarization equalization, the CMA, which needs to be modified. This dynamic equalizer has a cost function J that is minimized in an iterative process, aiming to optimize FIR-filter coefficients in a Jones-matrix-like filter. Figuratively speaking, the cost function is minimized when the detected samples lie on a circle of unit radius in both polarizations. This scheme works surprisingly well also to compensate for transmission impairments such as polarization mode dispersion (PMD) and polarization dependent losses (PDL), and in fact it also works (albeit with reduced performance) “out-of-the-box” for formats without a constant, but with a nonzero, modulus, such as PM-16-QAM.

However, for PS-QPSK it fails, due to an ambiguity making the cost function minimum nonunique. In addition, it requires the polarizations to be independently phase-tracked by two separate VV algorithms [92]. A number of ways to resolve this issue have been reported. Johannisson et al. [129] suggested a modified cost function J according to

$$J = E[(|E_x|^2 + |E_y|^2 - P)^2 + 2q|E_x|^2|E_y|^2] \quad (2.19)$$

where $E[\cdot]$ denotes the expectation operator (i.e., averaging over a number of symbols), $E_{x,y}$ are the complex amplitudes of the symbols in the x and y polarizations, and P is the total signal power. The parameter q is set to -1 for PM-QPSK and $+1$ for PS-QPSK. This enables the same CMA to be used with both formats, by just changing the parameter q in the cost function, which should facilitate implementations.

Alternative CMA approaches were independently suggested. Millar and Savory [130] proposed a modified CMA where the magnitude of the polarization components is compared, and the weaker allowed to reach zero. This scheme was also used in [90, 91, 94, 97].

An experiment using OFDM with PS-QPSK [99] and an outer LDPC code showed that a 4d demapper in the iterative decoder loop gave increased decoder performance.

Renaudier et al. [92] proposed instead to use a combined transmitter/receiver-based solution for the CMA ambiguity. By introducing a time offset equal to an integer number of symbol times between the two polarization components in an XOR-based PM-QPSK transmitter, the polarization ambiguity outlined above is suppressed and the received signal “looks” like a PM-QPSK signal to the receiver CMA. Note that this offset must be larger than maximum differential delay of the multi-tap CMA filter. Then the standard PM-QPSK CMA can be used in the receiver, but a modified VV algorithm must be used instead [92]. Also, Alreesh et al. [131] presented an alternative VV algorithm for PS-QPSK, enabling joint phase tracking in both polarizations.

6PolSK-QPSK The coherent receivers for 6PolSK-QPSK followed the standard coherent receiver, with modifications for the polarization and phase tracking algorithms. The implementation of Buchali [100] to detect 6PolSK-QPSK used a modi-

fied CMA with a new cost function for the polarization tracking and a more complicated VV scheme, raising the signals to the 8th power for the phase estimation.

Fischer et al. [101] used a pilot sequence with 1.1 % overhead, which also helped with the local oscillator frequency offset estimation, to do the polarization tracking, thus avoiding a CMA. For phase estimation, they used a standard VV algorithm, but on a subset of the detected symbols. Tanimura et al. [103] also considered nonlinear compensation via digital back-propagation as well as with an inner Reed–Solomon FEC.

In addition to the experiments [100, 101, 103] discussed above, a couple of later extensions of the 6PolSK-QPSK work should be mentioned. In [132], Bülow calculated mutual information and estimated its performance with an outer FEC. In [133], Bülow and Masalkina investigated coded modulation based on, among other formats, 6PolSK-QPSK. Chen et al. [134] extended 6PolSK-QPSK to a 32-point constellation, enabling a 5-bits-to-symbol mapping by extending the constellation with 8 additional points outside the 24-cell. Even if this format was shown to have a sensitivity improvement over, e.g., star-shaped 8-QAM, it has gives a penalty relative to better 4d constellations with 32 points such as, e.g., 32-SP-QAM.

M-SP-QAM The DSP required for 32- and 128-SP-QAM is very similar to the requirements for PM-16-QAM, which is not obvious, but it works according to simulations [61] and experiments [105, 117]. In implementations with the same receiver DSP (intended for PM-16-QAM), one generally finds that 32- and 128-SP-QAM have less implementation penalty, which is likely due to the better separation of levels [105, 117].

Sun et al. [109] compared the linewidth tolerance of 32-SP-QAM and the hybrid PM-QPSK/8QAM format and found it to be comparable, provided that the VV algorithm was slightly modified for 32-SP-QAM.

The work by Zhang et al. [106, 108] also added coded modulation (of the BICM flavor) to the use of 128-SP-QAM with an LDPC(18360,15300) code and an interleaver over 30 independent bitstreams demultiplexed from 2 separate wavelengths. The decoding process then used 10 inner and 5 outer iterations to achieve the required performance. In [108], the performance was further improved by adding nonlinear back-propagation to the DSP, thus enabling longer transmission distances.

In van Uden's work on few-mode transmission [111], the 6 spatial channels (2 polarization modes in 3 fiber modes) were first optically polarization- and mode-demultiplexed, and then in a 6×6 MIMO (multiple input multiple output) structure optimized by an iterative least mean squares scheme, before entering the coherent phase estimator, demapper, and BER counter. This MIMO structure can be seen as a generalization of the CMA scheme in a conventional coherent receiver. The receiver was also simplified by the fact that the 4d modulation was performed in the time domain, rather than the polarization domain (which will not affect the theoretical spectral efficiency or sensitivity for linear transmission).

More complex formats The $\mathcal{C}_{4,16}$ cluster was transmitted using OFDM in [113], and the synchronization issues, like polarization and phase tracking are then ad-

dressed in the OFDM receiver, using the channel estimators in the OFDM DSP. For example, one of the OFDM subcarriers were left unmodulated to act as a pilot channel, aiding the synchronization DSP. In addition, three training symbols were used in every 512 OFDM-symbol block. The overall spectral efficiency was thus 3.82 bits/symbol/pol, rather than 4 for the raw format. On the other hand, PM-QPSK was transmitted over the same OFDM channel, so the two formats were compared in a fair fashion. The implementation of Bülow et al. [114] to detect $\mathcal{C}_{4,16}$ (referred to as “OPT16”) was a more conventional coherent baseband receiver. The CMA was similar to the one used in their 6PolSK-QPSK experiments. The phase tracking was carried out by using a decision-directed least mean squares scheme, aided by a training sequence. After the training sequence set up the initial starting point, the decision-directed scheme could take over.

Higher-dimensional formats In the 8d implementation of frequency and polarization-switched QPSK [117], the receiver used only one local oscillator (centered between the two channels) and one optical front-end to detect both channels. The two channels were filtered in DSP, down-converted to baseband and then processed in parallel using the standard DSP flow. The CMA had to be modified with a power threshold to estimate in which frequency a given symbol was sent. The PPM implementation of the 8d format [118] was simpler in that it used a standard coherent DSP throughout, with the exception of the CMA that used the power threshold to judge which PPM frame was used.

PPM implementations The experiments by Liu et al. [71, 121, 122] on PPM made extensive use of pilot sequences both for frame synchronization and channel estimation. The frame structure consisted of 3 frames (48 symbol slots) of training sequences followed by 16100 PPM symbol slots. The pilot sequences helped with both polarization tracking and phase estimation. The rate overhead due to this was small, less than 1 %.

2.5.3 Formats overview

In table 2.3, we summarize the first proposals and implementations of some relevant optical 4d modulation formats.

2.5.4 Symbol detection

For nonregular and high-order constellations, the *detection* process, i.e., determining which of the constellation points that was transmitted, can be quite cumbersome and computationally intensive. The best detector, in the sense of minimizing the SER, is the ML detector. In the special case of an AWGN channel, the ML detector computes the Euclidean distance between a received vector \mathbf{r} and all constellation points in \mathcal{X} and picks the symbol with the smallest distance, i.e.,

$$\hat{x} = \arg \min_{x \in \mathcal{X}} (\|\mathbf{r} - \mathbf{x}\|). \quad (2.20)$$

Table 2.3 First experimental demonstrations and theoretical proposals of 4d modulation formats in coherent optical transmission links.

Format	Dim. N	Size M	Theoretical proposal	Experimental demo
3d simplex	3	4	Dochhan [41] 2013.	Dochhan [41] 2013.
PS-QPSK	4	8	Welti [12] 1974, Betti [5] 1991, Karlsson [42] 2009.	Sjödín [88], Millar [89], Nelson [90], Fischer [91], all 2011.
$\mathcal{C}_{4,16}$	4	16	Agrell [27] 2009.	Karout [113], Bülow [114], both 2013.
6PolSK-QPSK	4	24	Bülow [55], Agrell [27], both 2009.	Buchali [100], Fischer [101], both 2012.
32-SP-QAM	4	32	Coelho [59] 2011.	Renaudier [105], Sun [109], Rios-Muller [110], all 2013.
128-SP-QAM	4	128	Coelho [59] 2011.	Eriksson [104, 107], Zhang [106], Renaudier [105], all 2013.
4FPS-QPSK 2PPM-PS-QPSK	8	16	Eriksson [117] 2013, Eriksson [73] 2014.	Eriksson [117] 2013, Eriksson [73] 2014.

This will always work, but requires M distance calculations and a number of comparisons, which is computationally costly, and one therefore seeks easier detectors in practice. In this section, we will describe how this detection is done efficiently for some of the common 4d constellations. We will start by describing lattices, and then give examples for the specific formats.

For the n -dimensional integer (or cubic) lattice \mathbb{Z}^n , the problem is particularly simple; the ML detection is equivalent to just rounding each component of the received vector to the nearest integer. As shown by Conway and Sloane [56], this procedure can be modified to work for other lattices of interest, such as D_4 and E_8 .

For the D_4 lattice (2.14), one can use the following algorithm [56]: (i) decode \mathbf{r} to the nearest point in \mathbb{Z}^4 , and check the parity of the lattice point (i.e., the modulo-2 sum of its coordinates). (ii) If the parity is odd, round the component of \mathbf{r} that is farthest away from its closest integer to its second closest integer. The algorithm is described for 128-SP-QAM in [61] and generalizes straightforwardly to all other SP-QAM formats obtained by reduction. An equally simple algorithm applies to SP-

QAM formats obtained by expansion [56]. These algorithms can be generalized to the E_8 lattice as well [56, 135].

For the simpler formats such as PS-QPSK, the symbol detection is straightforward, but depends on the form after synchronization and phase tracking. For the set-partitioned-form, i.e., $\pm\{(1, 1, 1, 1), (1, 1, -1, -1), (1, -1, 1, -1), (1, -1, -1, 1)\}$, one can use the above D_4 scheme. For the polarization-switched form, where the constellation points are $\{(\pm 1, \pm 1, 0, 0), (0, 0, \pm 1, \pm 1)\}$ one can use the following steps: (i) determine the polarization by comparing $|\Re(E_x)| + |\Im(E_x)|$ with $|\Re(E_y)| + |\Im(E_y)|$, (ii) detect QPSK as usual in the chosen polarization. It is noteworthy and somewhat surprising that comparing the magnitude of the complex numbers $|E_x|$ and $|E_y|$ for determining the polarization is *suboptimal* in contrast with this scheme. As a simple example of this, consider the received vector $\mathbf{r} = (0.9, 0.1, 0.6, 0.5)$, or $E_x = 0.9 + i0.1$, $E_y = 0.5 + i0.6$. The closest PS-QPSK point in the Euclidean distance metric is $(0, 0, 1, 1)$, i.e., an ML symbol detector selects the y polarization, despite the received power in x being $|E_x|^2 = 0.82$, which is higher than the power in y , being $|E_y|^2 = 0.61$.

For the 24-cell (6PolSK-QPSK) in the form $\{(\pm 1, \pm 1, \pm 1, \pm 1), (\pm 2, 0, 0, 0)\}$, the following detector is optimal: (i) Find the maximum of $\{(|r_1| + |r_2| + |r_3| + |r_4|)/2, |r_1|, |r_2|, |r_3|, |r_4|\}$, where $(r_1, r_2, r_3, r_4) = \mathbf{r}$. (ii) If the first is maximum, proceed with standard PM-QPSK detection to return $(\text{sgn}(r_1), \text{sgn}(r_2), \text{sgn}(r_3), \text{sgn}(r_4))$. If one of the last four are largest, take the sign of that component and multiply with 2.

2.5.5 Realizing Dimensions

The dimensionality can be taken as one metric of complexity, and often one chooses to compare modulation formats of the same dimensionality, i.e., keeping the complexity similar. Shannon proved that an arbitrarily small SER is achievable for any channel, assuming that the spectral efficiency is below a certain threshold, the *channel capacity*, and that the dimensionality goes to infinity [85, 136]. The work was later extended to quantify how the SER decreases with dimension [137].

The introduction of a channel (FEC) code is the obvious, and prevailing, way of increasing the dimensionality in communications. If used with BPSK mapping (see section 2.4.1), the dimensionality is simply the number n of bits in the FEC frame. There may, however, be practical reasons for keeping the value n down; the latency, complexity, and power consumption incurred by the FEC will increase with n . Moreover, even if a good FEC code is used that can tolerate a low SNR, the pre-decoder signal may be so distorted by noise that DSP and synchronization algorithms will limit the performance rather than the ideal FEC ability.

Thus, in practical systems, it will make sense also to use the degrees of freedom (DOF) of the transmission channel to increase the dimensionality and the constellation distances. However, this often leads to issues with crosstalk and synchronization that need to be resolved in the receiver. Below we will briefly discuss the practical implementation challenges by increasing the dimensionality via the physical DOFs.

We should emphasize that the DOFs used as signaling dimensions usually have independent noise sources, which simplifies their usage for the AWGN channel model, as the noise can be modeled as a hyperspherical cloud around the transmitted symbol.

Quadratures Every carrier wave has two DOFs that can be modulated independently, i.e., the two quadratures usually described by the real and imaginary parts of a complex phasor, or as the “sine” and “cosine”-components of the wave. An alternative decomposition is the amplitude and phase of the wave. Up to around 2000, the amplitude was the only DOF used in commercial optical links, and it still is in short-haul links, due to the cost and complexity associated with modulating the optical phase. To reliably detect the optical phase, a coherent or differential-phase receiver is required. The differential receiver is simpler in optical hardware, but has extra an penalty relative to the coherent counterpart and is limited to mainly PSK modulation.

Since the intradyne coherent receiver was demonstrated [3], the differential optical receivers have faded away. Coherent receivers are challenging due to the required rapid phase tracking on a microsecond time scale, but with the development of fast electronics and DSPs, they are becoming increasingly common and cost effective, and will likely prevail in future optical links. Ways of reducing the coherent receiver complexity by, e.g., co-propagating a local oscillator carrier (so-called self-homodyning [138]) have been proposed.

The quadrature dimensions are, just as polarizations, and contrary to the time- and space-related dimensions discussed below, not scalable to more than 2.

Polarization Electromagnetic waves has a vector property not seen in longitudinal waves (e.g., acoustic waves) or transverse matter waves (e.g., water waves, oscillating strings). The easiest description is that of two independent carrier waves, with orthogonally directed vector field components. They are often referred to as the x and y polarization states, but there are other orthogonal decompositions possible as well. Usually a “polarization state” refers to the relative amplitude and phase between these two waves. Thus, one has two ways of describing the 4 DOFs of the classical electromagnetic wave, either with “amplitude,” “phase,” and “polarization state” or with the I and Q quadratures in the x and y polarizations. The former is the traditional description used in optics, and the latter is the most attractive one used in communications, as those DOFs form a Cartesian system.

The polarization state in a fiber link is slowly (second to millisecond time scale) drifting due to imperfections, micro- and macrobendings, thermal changes, fiber movements etc., and the use of polarization dimensions will thus require polarization tracking in the receiver. A commercially attractive and low-cost polarization tracker was not available before the intradyne coherent receiver, and as a result, polarization was not actively used for modulation in commercial systems (although studied, e.g., in [139]). Also, the intersymbol interference problems related to polarization mode dispersion were long regarded as a significant obstacle (also in conventional links that were not polarization modulated), but elegantly resolved by the adaptive CMA filter in the coherent receiver.

The combined use of quadratures and polarization leads to a real 4d constellation space that is the basis for coherent signalling. It is noteworthy that some key problems and issues with 4d modulation are unresolved or only recently being explored, e.g., channel modeling via 4d rotations [140] and modulation format optimization [27,42].

Time With an AWGN channel, a T -orthogonal pulse (2.3), and a matched-filter receiver, adjacent symbols in time will have independent noise, and can be thus be framed to a “supersymbol” with dimensionality equal to the number of symbols used (possibly times the dimensionality of each symbol). Such a supersymbol can now be modulated with a higher-dimension format. As detailed in section 2.4.1, when using a simple format such as 1d OOK or BPSK, this is equivalent to applying an FEC frame with a binary code. This clearly highlights the close relationship between modulation and coding; indeed, there is no clearcut distinction between coding and modulation in communication theory.

Nonetheless, from a practical and implementation perspective, it makes sense to distinguish between modulation and coding, as they are implemented with very different hardware, meeting different challenges. The time-multiplexing described above (of which PPM discussed in section 2.3.1.5 is one special case) is particularly simple and attractive, since the phase and symbol time synchronization essentially comes for free. The frame synchronization needs to be resolved, however, and usually requires a test sequence and/or use of a specific transmission protocol.

The temporal dimension and its simplicity thus forms a simple playground for testing new formats without challenging synchronization issues. For example, this was used in the experiment by van Uden et al. [111], where two time slots were used to form a 4d symbol rather than the two polarization states, simplifying receiver DSP significantly.

Frequency/Wavelength Different channels transmitting at adjacent wavelengths (frequencies) can be used to form multidimensional supersymbols. The concept of “superchannels” were introduced to denote such multi-wavelength channels, which are routed and detected as one entity [141]. To make use of correlated modulation to increase signal space dimensionality with such superchannels has not yet been realized, but should be possible.

In a similar vein, one could perform joint detection of multi-wavelength channels in a WDM link, thus enabling multidimensional modulation and coding. This meets practical problems with temporal (walk-off related) synchronization and phase synchronization, as well as signal ambiguities of the independent wavelengths, so it has not yet been widely used.

A few limited cases have been reported though, e.g., the 8d format by Eriksson [117] detecting two wavelengths with the same local oscillator.

Space With the recent interest in spatial division multiplexing, i.e., the use of waveguide modes and/or parallel waveguides to increase the data rates of optical links, it seems natural to try to further increase the capacity by moving to *joint* trans-

mission over parallel modes and/or fibers, i.e., increasing signaling dimensionality by making use of also the spatial DOFs. These can be (i) the different modes in a multimode fiber, (ii) different cores in a multi-core fiber, (iii) entirely separate fibers, or a combination of these.

However, spatial multiplexing is associated with severe practical challenges. For example, the different modes in a multimode fiber have different group velocities and hence a significant differential mode delay will arise that needs to be compensated for. Even worse, bends and fiber imperfections gives rise to modal crosstalk that mixes the delayed modes, further complicating the reception of individual modes. As a result, MIMO signal processing is needed to compensate for the modal crosstalk (see, e.g., [111]), which is very DSP-heavy.

The single-mode (linear-polarized, LP_{01} -mode) fiber has, as was outlined above, four dimensions due the polarization and quadratures. The next mode in the weakly-guiding, step-index, circular fiber mode hierarchy, the (linearly polarized) LP_{11} -mode, is doubly degenerate. This means that it can be of two orientations that are mathematically orthogonal, usually referred to as LP_{11a} and LP_{11b} . A linear combination of these modes is popularly referred to as the lowest orbital angular momentum (OAM) mode, with azimuthal index ± 1 . Thus OAM modes fall well within the conventional modal description, and offer limited novelty and no principal extension of the existing DOFs, as recently shown for radio frequency (RF) transmission [142].

This means that inclusion of the next higher order mode in a circular fiber totals to three orthogonal modes, of 4d each, leading to a 12d space. Format optimization in this 12d space has been discussed by Bülow et al. [143].

A more straightforward approach seems to be to use different (single-mode) cores in a multicore fiber to increase the dimensionality. This relaxes the crosstalk and walk-off penalties. Novel formats exploiting these DOFs were recently discussed by Eriksson et al. [73] and some initial experiments were reported by Puttnam et al. [120, 144].

Spatial frequencies Due to the space–time duality [145], which states that free-space propagation of optical beams is similar to dispersive propagation of optical pulses, one could just as well use spatial frequencies as temporal frequencies. Spatial frequencies translate into physically different propagation directions, so their use would be of most interest in free-space rather than in guided-wave propagation. A deeper discussion of this topic is therefore not included in this chapter. It suffices to say that the challenges and solutions connected with spatial frequency (physical beam direction in the paraxial propagation limit) are similar to the MIMO technologies used in wireless communications, cf. [142].

2.6 Summary and Conclusions

In this chapter, we have overviewed the relatively large body of work (experimental and theoretical) on modulation formats for optical coherent links that have emerged over the last 5 years. We have also shown the performance limits of formats in 2, 4,

and 8d by reviewing sphere packing simulations, lattice-cuts, and code-based format design. To reach higher dimensions, formats based on codes are probably the most straightforward approach rather than numerically optimized sphere packings, and we showed a few examples of this as well.

The results summarized in this chapter are somewhat idealized, in that they to a large extent (i) neglect the impact of FEC, (ii) focus on the asymptotic high-SNR behavior (low SER and BER), and (iii) emphasize the AWGN channel. This is the classical starting point for modulation theory research and should be viewed as a first step towards a fuller understanding of optical link design. Broadening the scope in all three directions are presently active areas of research.

First, to separate modulation format and coding is (while unnecessary from the information theorist's point of view) necessary from the system engineer's perspective. The modulation format dictates the transmitter and receiver hardware optoelectronics, the complexity, and thus to some extent the cost. It dictates the DSP and the complexity of the receiver electronics, especially if blind equalization is used. It also dictates the attainable spectral efficiency. Thus, when selecting formats, one makes critical system choices and it is important for all the trade-offs made in system design to know how well the formats behave—even if it is an ideal or asymptotic behavior.

Moreover, there are communication applications that are latency-critical, where the use of FEC is prohibited or at least limited, and there good formats are very important. Examples may be control systems, video conferencing or telephony, and transfer of stock market trading data.

Another issue, seldom studied or emphasized, is the performance of synchronization algorithms, channel estimation algorithms, and adaptive equalizers for various formats. Those tend to operate worse at low SNR (which might arise if strong FEC codes are used), but better if constellations with well-separated points are used. In such situations, the clusters or lattice-based constellations described in this chapter might be a better choice than standard QAM.

An important application for the use of many different formats is the emergence of *elastic networking*. In future optical networks, where an increased *flexibility* is desired (often called elastic optical networks), one strives to adapt the data rate provided to customers after the available bandwidth, SNR, and demand. In such systems, it is of great value to be able to switch between different modulation formats to provide the sought flexibility, and a good overview of the performance and trade-offs of formats is needed. A recent overview of 4d modulation formats from this perspective was recently provided by Fischer et al. [62].

Second, the study of asymptotic performance metrics should be complemented by studies of the performance of modulation formats at more practically relevant SNR values. Unfortunately, no analytic instruments are available for this purpose. To numerically optimize multidimensional modulation formats for specific SNRs is computationally complex, and it may be difficult to get the full picture. Asymptotic metrics such as γ , CFM , and G are attractive in that they give one number to compare formats by, which quickly can be used to compare and select formats. They also provide an intuitive interpretation in terms of sphere packing, and they give an upper limit of the performance gains, which is valuable. However, once a set of for-

mats are selected, they should be compared via SER or BER simulations as well as complexity estimates. Even if the asymptotic gains are never achieved, the current design paradigm in optics is to compare formats with the same FEC, which has a waterfall region around 10^{-3} for the uncoded SER, and then, quite often, the optimized multidimensional formats will outperform (e.g., in term of system reach) the standard QAM formats.

Third, more advanced channel models to go beyond the AWGN model are currently being studied (as discussed in section 2.3.2) and depending on their nature, the results from AWGN modeling may or may not be useful. For example, the GN model, in links offering high SNR, can benefit from AWGN-optimized constellations, whereas nonlinear phase-noise channels benefit from radically different constellations.

Finally we note that, as we touched upon previously in this chapter, the price we pay when going to higher-dimensional formats (and codes) is *complexity*, which is a notoriously difficult quantity to quantify. Even if it can be quantified in terms of component cost, number of floating point operations, chip area for DSP implementation, or dimensionality of codes and formats, it is hard to provide generic results, since the most reliable complexity metrics are implementation-specific. Nevertheless, the complexity factor must enter the system design and hopefully the contents of this chapter may somewhat aid system designers when dealing with the trilemma of limited spectral efficiency, low noise tolerance, and complexity.

REFERENCES

1. T. Pfau, S. Hoffmann, R. Peveling, S. Ibrahim, O. Adamczyk, M. Porrmann, S. Bhandare, R. Noé, and Y. Achiam. Synchronous QPSK transmission at 1.6 Gbit/s with standard DFB lasers and real-time digital receiver. *Electronics Letters* 42(20):1175–1176, 2006.
2. G. Charlet, M. Salsi, J. Renaudier, O. Pardo, H. Mardoyan, and S. Bigo. Performance comparison of singly-polarised and polarisation-multiplexed coherent transmission at 10 Gbauds under linear impairments. *Electronics Letters* 43(20):1109–1111, 2007.
3. H. Sun, K. Wu, and K. Roberts. Real-time measurements of a 40 Gb/s coherent system. *Optics Express* 16(2):873–879, 2008.
4. S. Betti, F. Curti, G. De Marchis, and E. Iannone. Exploiting fibre optics transmission capacity: 4-quadrature multilevel signalling. *Electronics Letters* 26(14):992–993, 1990.
5. S. Betti, F. Curti, G. De Marchis, and E. Iannone. A novel multilevel coherent optical system: 4-quadrature signaling. *Journal of Lightwave Technology* 9(4):514–523, 1991.
6. S. Betti, G. De Marchis, E. Iannone, and P. Lazzaro. Homodyne optical coherent systems based on polarization modulation. *Journal of Lightwave Technology* 9(10):1314–1320, 1991.
7. R. Cusani, E. Iannone, A. Salonic, and M. Todaro. An efficient multilevel coherent optical system: M-4Q-QAM. *Journal of Lightwave Technology* 10(6):777–786, 1992.
8. F. Derr. Optical QPSK homodyne transmission of 280 Mbit/s. *Electronics Letters* 26:401, 1990.

9. A. H. Gnauck and P. J. Winzer. Optical phase-shift-keyed transmission. *Journal of Lightwave Technology* 23(1):115, 2005.
10. L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson. Rate-adaptive coded modulation for fiber-optic communications. *Journal of Lightwave Technology* 32(2):333–343, 2013.
11. L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson. Coded modulation for fiber-optic networks: Toward better tradeoff between signal processing complexity and optical transparent reach. *IEEE Signal Processing Magazine* 31(2):93–103, 2014.
12. G. Welsch and J. Lee. Digital transmission with coherent four-dimensional modulation. *IEEE Transactions on Information Theory* 20(4):497–502, 1974.
13. L. Zetterberg and H. Brändström. Codes for combined phase and amplitude modulated signals in a four-dimensional space. *IEEE Transactions on Communications* 25(9):943–950, 1977.
14. E. Biglieri. Advanced modulation formats for satellite communications. In J. Hagenauer, editor, *Advanced Methods for Satellite and Deep Space Communications*, pages 61–80. Springer, Berlin, Germany, 1992.
15. S. J. Savory. Digital coherent optical receivers: algorithms and subsystems. *IEEE Journal of Selected Topics in Quantum Electronics* 16(5):1164–1179, 2010.
16. X. Zhou. Efficient clock and carrier recovery algorithms for single-carrier coherent optical systems. *IEEE Signal Processing Magazine* 31(2):35–45, 2014.
17. A. P. T. Lau, Y. Gao, Q. Sui, D. Wang, Q. Zhuge, M. H. Morsy-Osman, M. Chagnon, X. Xu, C. Lu, and D. V. Plant. Advanced DSP techniques enabling high spectral efficiency and flexible transmissions. *IEEE Signal Processing Magazine* 31(2):82–92, 2014.
18. E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn. Coherent detection in optical fiber systems. *Optics Express* 16(2):753–791, 2008. (erratum, vol. 16, no. 26, p. 21943, Dec. 2008).
19. K. Kikuchi and S. Tsukamoto. Evaluation of sensitivity of the digital coherent receiver. *Journal of Lightwave Technology* 26(13):1817–1822, 2008.
20. P. Poggiolini. The gn model of non-linear propagation in uncompensated coherent optical systems. *Journal of Lightwave Technology* 30(24):3857–3879, 2012.
21. L. Beygi, E. Agrell, and M. Karlsson. Optimization of 16-point ring constellations in the presence of nonlinear phase noise. In *Optical Fiber Communication Conference*, page OThO.4. 2011.
22. P. Johannisson and M. Karlsson. Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system. *Journal of Lightwave Technology* 31(8):1273–1282, 2013.
23. A. Splett, C. Kurtzke, and K. Petermann. Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities. In *European Conference and Exhibition on Optical Communication*. Montreux, Switzerland, 1993.
24. E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson. Capacity of a nonlinear optical channel with finite memory. *Journal of Lightwave Technology* 32(16):2862–2876, 2014.
25. K.-P. Ho. *Phase-Modulated Optical Communication Systems*. Springer, 2005.

26. L. Beygi, E. Agrell, M. Karlsson, and P. Johannisson. Signal statistics in fiber-optical channels with polarization multiplexing and self-phase modulation. *Journal of Lightwave Technology* 29(16):2379–2386, 2011.
27. E. Agrell and M. Karlsson. Power-efficient modulation formats in coherent transmission systems. *Journal of Lightwave Technology* 27(22):5115–5126, 2009.
28. E. Agrell. Database of sphere packings. Online: <http://codes.se/packings>, 2014.
29. G. D. Forney, Jr. Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations. *IEEE J. Sel. Areas Commun.* 7(6):877–892, 1989.
30. F. R. Kschischang. Optimal nonuniform signaling for Gaussian channels. *IEEE Transactions on Information Theory* 39(3):913–929, 1993.
31. M. Karlsson and E. Agrell. Power-efficient modulation schemes. In S. Kumar, editor, *Impact of Nonlinearities on Fiber Optic Communications*, chapter 5, pages 219–252. Springer, 2011.
32. S. Benedetto and E. Biglieri. *Principles of Digital Transmission: With Wireless Applications*. Kluwer Academic Publishers, 1999.
33. G. Foschini, R. Gitlin, and S. Weinstein. Optimization of two-dimensional signal constellations in the presence of gaussian noise. *IEEE Transactions on Communications* 22(1):28–38, 1974.
34. R. L. Graham and N. J. A. Sloane. Penny-packing and two-dimensional codes. *Discrete and Computational Geometry* 5(1):1–11, 1990.
35. N. J. A. Sloane, R. H. Hardin, T. S. Duff, and J. H. Conway. Minimal-energy clusters. Online: <http://neilsloane.com/cluster>.
36. J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*, 3rd ed. Springer-Verlag, New York, 1998.
37. J. R. Pierce. Comparison of three-phase modulation with two-phase and four-phase modulation. *IEEE Transactions on Communications* COM-28(7):1098–1099, 1980.
38. N. Ekanayake and T. Tjhung. On ternary phase-shift keyed signaling. *IEEE Transactions on Information Theory* 28(4):658–660, 1982.
39. J.-E. Porath and T. Aulin. Design of multidimensional signal constellations. *IEE Proceedings-Communications* 150(5):317–323, 2003.
40. J. G. Proakis. *Digital Communication*. McGraw-Hill, 2001.
41. A. Dochhan, H. Griesser, and M. Eiselt. First experimental demonstration of a 3-dimensional simplex modulation format showing improved OSNR performance compared to DP-BPSK. In *Optical Fiber Communication Conference*, page JTh2A.40. 2013.
42. M. Karlsson and E. Agrell. Which is the most power-efficient modulation format in optical links? *Optics Express* 17(13):10814–10819, 2009.
43. M. K. Simon, S. M. Hinedi, and W. C. Lindsey. *Digital Communication Techniques: Signal Design and Detection*. PTR Prentice Hall, 1995.
44. E. Agrell and M. Karlsson. On the symbol error probability of regular polytopes. *IEEE Transactions on Information Theory* 57(6):3411–3415, 2011.

45. D. Saha and T. G. Birdsall. Quadrature-quadrature phase-shift keying. *IEEE Transactions on Communications* 37(5):437–448, 1989.
46. D. Slepian. Permutation modulation. *Proceedings of the IEEE* 53(3):228–236, 1965.
47. M. Karlsson and E. Agrell. Multilevel pulse-position modulation for optical power-efficient communication. *Optics Express* 19(26, SI):B799–B804, 2011.
48. B. Krongold, T. Pfau, N. Kaneda, and S. C. J. Lee. Comparison between PS-QPSK and PDM-QPSK with equal rate and bandwidth. *IEEE Photonics Technology Letters* 24(3):203–205, 2012.
49. P. Poggiolini, G. Bosco, A. Carena, V. Curri, and F. Forghieri. Performance evaluation of coherent WDM PS-QPSK (HEXA) accounting for non-linear fiber propagation effects. *Optics Express* 18(11):11360–11371, 2010.
50. P. Serena, A. Vannucci, and A. Bononi. The performance of polarization switched-QPSK (PS-QPSK) in dispersion managed WDM transmissions. In *European Conference on Optical Communication*, page Th.10.E.2. 2010.
51. P. Serena, N. Rossi, and A. Bononi. PDM-iRZ-QPSK vs. PS-QPSK at 100 Gbit/s over dispersion-managed links. *Optics Express* 20(7):7895–7900, 2012.
52. H. S. M. Coxeter. *Regular Polytopes*. Dover, New York, NY, 1973.
53. O. Musin. The kissing number in four dimensions. *Annals of Mathematics* 168:1–32, 2008.
54. G. Taricco, E. Biglieri, and V. Castellani. Applicability of four-dimensional modulations to digital satellites: A simulation study. In *Proc. IEEE Global Telecommunications Conference*, volume 4, pages 28–34. 1993.
55. H. Bülow. Polarization QAM modulation (POL-QAM) for coherent detection schemes. In *Optical Fiber Communication Conference*, page OWG.2. 2009.
56. J. H. Conway and N. J. A. Sloane. Fast quantizing and decoding and algorithms for lattice quantizers and codes. *IEEE Transactions on Information Theory* 28(2):227–232, 1982.
57. M. Karlsson and E. Agrell. Spectrally efficient four-dimensional modulation. In *Optical Fiber Communication Conference*, page OTu2C.1. Optical Society of America, 2012.
58. G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory* 28(1):55–67, 1982.
59. L. Coelho and N. Hanik. Global optimization of fiber-optic communication systems using four-dimensional modulation formats. In *European Conference on Optical Communication*, page Mo.2.B.4. 2011.
60. J. Renaudier, A. Voicila, O. Bertran-Pardo, O. Rival, M. Karlsson, G. Charlet, and S. Bigo. Comparison of set-partitioned two-polarization 16QAM formats with PDM-QPSK and PDM-8QAM for optical transmission systems with error-correction coding. In *European Conference on Optical Communication*, page We.1.C.5. Optical Society of America, 2012.
61. M. Sjödin, P. Johannisson, J. Li, E. Agrell, P. A. Andrekson, and M. Karlsson. Comparison of 128-SP-QAM with PM-16-QAM. *Optics Express* 20(8):8356–8366, 2012.
62. J. K. Fischer, S. Alreesh, R. Elschner, F. Frey, M. Nölle, C. Schmidt-Langhorst, and C. Schubert. Bandwidth-variable transceivers based on four-dimensional modulation formats. *Journal of Lightwave Technology* 32(16):2886–2895, 2014.

63. M. Karlsson and E. Agrell. Four-dimensional optimized constellations for coherent optical transmission systems. In *European Conference on Optical Communication*, page We.8.C.3. 2010.
64. M. Sjödin, E. Agrell, and M. Karlsson. Subset-optimized polarization-multiplexed PSK for fiber-optic communications. *IEEE Communication Letters* 17(5):838–840, 2013.
65. J. H. Conway and N. J. A. Sloane. A fast encoding method for lattice codes and quantizers. *IEEE Transactions on Information Theory* 29(6):820–824, 1983.
66. T. A. Eriksson, S. Alreesh, C. Schmidt-Langhorst, F. Frey, P. W. Berenguer, C. Schubert, J. K. Fischer, P. A. Andrekson, M. Karlsson, and E. Agrell. Experimental investigation of a four-dimensional 256-ary lattice-based modulation format. In *Optical Fiber Communication Conference*, page W4K.3. 2015.
67. T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons. Eight-dimensional modulation for coherent optical communications. In *European Conference and Exhibition on Optical Communication*, page Tu.3.C.3. 2013.
68. D. S. Millar, T. Koike-Akino, S. Ö. Arik, K. Kojima, K. Parsons, T. Yoshida, and T. Sugihara. High-dimensional modulation for coherent optical communications systems. *Optics Express* 22(7):8798–8812, 2014.
69. D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons. A 24-dimensional modulation format achieving 6 dB asymptotic power efficiency. In *Signal Processing in Photonic Communications*, page SPM3D.6. Optical Society of America, 2013.
70. D. S. Millar, T. Koike-Akino, R. Maher, D. Lavery, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel. Experimental demonstration of 24-dimensional extended Golay coded modulation with LDPC. In *Optical Fiber Communication Conference*, page M3A.5. Optical Society of America, 2014.
71. X. Liu, T. H. Wood, R. Tkach, and S. Chandrasekhar. Demonstration of record sensitivity in an optically pre-amplified receiver by combining PDM-QPSK and 16-PPM with pilot-assisted digital coherent detection. In *Optical Fiber Communication Conference*, page PDPB.1. 2011.
72. H. Selmy, H. M. H. Shalaby, and Z.-I. Kawasaki. Proposal and performance evaluation of a hybrid BPSK-modified MPPM technique for optical fiber communications systems. *Journal of Lightwave Technology* 31(22):3535–3545, 2013.
73. T. A. Eriksson, P. Johannisson, B. J. Puttnam, E. Agrell, P. A. Andrekson, and M. Karlsson. K -over- L Multidimensional position modulation. *Journal of Lightwave Technology* 32(12):2254–2262, 2014.
74. A. P. T. Lau and J. M. Kahn. Signal design and detection in presence of nonlinear phase noise. *Journal of Lightwave Technology* 25(10):3008–3016, 2007.
75. E. Agrell and M. Karlsson. Satellite constellations: towards the nonlinear channel capacity. In *25th IEEE Photonics Conference, IPC 2012*, pages 316–317. 2012.
76. C. Häger, A. Graell i Amat, A. Alvarado, and E. Agrell. Constellation optimization for coherent optical channels distorted by nonlinear phase noise. In *IEEE Global Communications Conference (GLOBECOM)*, pages 2870–2875. 2012.
77. F. Kayhan and G. Montorsi. Constellation design for memoryless phase noise channels. *IEEE Transactions on Wireless Communications* 13(5):2874–2883, 2014.

78. M. Steiner. The strong simplex conjecture is false. *IEEE Transactions on Information Theory* 40(3):721–731, 1994.
79. F. Chang, K. Onohara, and T. K.-A. Mizuochi. Forward error correction for 100 G transport networks. *IEEE Communications Magazine* 48(3):S48–S55, 2010.
80. I. B. Djordjevic, M. Arabaci, and L. L. Minkov. Next generation FEC for high-capacity communication in optical transport networks. *Journal of Lightwave Technology* 27(16):3518–3530, 2009.
81. B. P. Smith and F. R. Kschischang. Future prospects for fec in fiber-optic communications. *IEEE Journal of Selected Topics in Quantum Electronics* 16(5):1245–1257, 2010.
82. A. J. Viterbi. On coded phase-coherent communications. *IRE Transactions on Space Electronics and Telemetry* SET-7(1):3–14, 1961.
83. S. Lin and D. J. Costello, Jr. *Error Control Coding*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 2004.
84. ITU-T. Forward error correction for submarine systems. Recommendation G.975, International Telecommunication Union, 2000.
85. C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal* 27:379–423, 623–656, 1948.
86. A. Alvarado and E. Agrell. Achievable rates for four-dimensional coded modulation with a bit-wise receiver. In *Optical Fiber Communication Conference*, page M2C.1. Optical Society of America, 2014.
87. H. Yamazaki, Y. Hashizume, and T. Saida. Simple three-dimensional simplex modulator. In *Optical Fiber Communication Conference*, page W11.3. 2014.
88. M. Sjödin, P. Johannisson, H. Wymeersch, P. A. Andrekson, and M. Karlsson. Comparison of polarization-switched QPSK and polarization-multiplexed QPSK at 30 Gbit/s. *Optics Express* 19(8):7839–7846, 2011.
89. D. S. Millar, D. Lavery, S. Makovejs, C. Behrens, B. C. Thomsen, P. Bayvel, and S. J. Savory. Generation and long-haul transmission of polarization-switched QPSK at 42.9 Gb/s. *Optics Express* 19(10):9296–9302, 2011.
90. L. E. Nelson, X. Zhou, N. Mac Suibhne, A. D. Ellis, and P. Magill. Experimental comparison of coherent polarization-switched QPSK to polarization-multiplexed QPSK for 10 x 100 km WDM transmission. *Optics Express* 19(11):10849–10856, 2011.
91. J. K. Fischer, L. Molle, M. Nölle, D.-D. Gross, and C. Schubert. Experimental investigation of 28-GBd polarization-switched quadrature phase-shift keying signals. In *European Conference on Optical Communications*, page Mo.2.B.1. Optical Society of America, 2011.
92. J. Renaudier, P. Serena, A. Bononi, M. Salsi, O. Bertran-Pardo, H. Mardoyan, P. Tran, E. Dutisseuil, G. Charlet, and S. Bigo. Generation and detection of 28 Gbaud polarization switched-QPSK in WDM long-haul transmission systems. *Journal of Lightwave Technology* 30(9):1312–1318, 2012.
93. H. Yamazaki, T. Goh, T. Saida, Y. Hashizume, and S. Mino. IQ-coupling-loss-free polarization-switched QPSK modulator. In *Optical Fiber Communication Conference*, page PDP5A.8. 2012.

94. C. Behrens, D. Lavery, D. S. Millar, S. Makovejs, B. C. Thomsen, R. I. Killey, S. J. Savory, and P. Bayvel. Ultra-long-haul transmission of 7x42.9Gbit/s PS-QPSK and PM-BPSK. In *European Conference on Optical Communication*, page Mo.2.B.2. 2011.
95. D. Lavery, C. Behrens, S. Makovejs, D. S. Millar, R. I. Killey, S. J. Savory, and P. Bayvel. Long-haul transmission of PS-QPSK at 100 Gb/s using digital backpropagation. *IEEE Photonics Technology Letters* 24(3):176–178, 2012.
96. M. Sjödin, B. J. Puttnam, P. Johannisson, S. Shinada, N. Wada, P. A. Andrekson, and M. Karlsson. Transmission of PM-QPSK and PS-QPSK with different fiber span lengths. *Optics Express* 20(7):7544–7554, 2012.
97. M. Nölle, J. K. K. Fischer, L. Molle, C. Schmidt-Langhorst, and C. Schubert. Influence of channel spacing on ultra long-haul WDM transmission of 8×112 Gb/s PS-QPSK signals. In *Optical Fiber Communication Conference*, page OTh3A.4. 2012.
98. M. Sjödin, B. J. Puttnam, P. Johannisson, S. Shinada, N. Wada, P. A. Andrekson, and M. Karlsson. Comparison of PS-QPSK and PM-QPSK at different data rates in a 25 GHz-spaced WDM system. In *Optical Fiber Communication Conference*, page OTu2A.3. Optical Society of America, 2012.
99. E. Masalkina, R. Dischler, and H. Bülow. Experimental study of polarization-switched-QPSK subcarrier modulation and iterative demapping on optical OFDM systems. In *Optical Fiber Communication Conference*, page OThO.6. 2011.
100. F. Buchali and H. Bülow. Experimental transmission with POLQAM and PS-QPSK modulation format using a 28-Gbaud 4-D transmitter. In *European Conference on Optical Communication*, page We.3.A.1. Optical Society of America, 2012.
101. J. K. Fischer, S. Alreesh, R. Elschner, F. Frey, C. Meuer, L. M., C. Schmidt-Langhorst, T. Tanimura, and C. Schubert. Experimental Investigation of 126-Gb/s 6PolSK-QPSK signals. *Optics Express* 20(26):B232–B237, 2012.
102. D. Ding, Y. Zhang, X. Yuan, J. Zhang, Y. Li, and Y. Huang. Generation of 6polsk-qpsk using dual-drive mach-zehnder modulators. In *Asia Communications and Photonics Conference*, page ATH3A.130. 2014.
103. T. Tanimura, S. Alreesh, J. K. Fischer, C. Schmidt-Langhorst, F. Frey, C. Meuer, R. Elschner, L. Molle, and C. Schubert. Nonlinear transmission of 6PolSK-QPSK signals using coded modulation and digital back propagation. In *Optical Fiber Communication Conference*, page OTu3B.3. 2013.
104. T. A. Eriksson, M. Sjödin, P. A. Andrekson, and M. Karlsson. Experimental demonstration of 128-SP-QAM in uncompensated long-haul transmission. In *Optical Fiber Communication Conference*, page OTu3B.2. 2013.
105. J. Renaudier, O. Bertran-Pardo, A. Ghazisaeidi, P. Tran, H. Mardoyan, P. Brindel, A. Voicila, G. Charlet, and S. Bigo. Experimental transmission of Nyquist pulse shaped 4-D coded modulation using dual polarization 16QAM set-partitioning schemes at 28 Gbaud. In *Optical Fiber Communication Conference*, page OTu3B. 2013.
106. H. Zhang, H. G. Batshon, D. G. Foursa, M. Mazurczyk, J. X. Cai, C. R. Davidson, A. Pilipetskii, G. Mohs, and N. S. Bergano. 30.58 Tb/s transmission over 7,230 km using PDM half 4D-16QAM coded modulation with 6.1 b/s/Hz spectral efficiency. In *Optical Fiber Communication Conference*, page OTu2B.3. 2013.
107. T. A. Eriksson, M. Sjödin, P. Johannisson, P. A. Andrekson, and M. Karlsson. Comparison of 128-SP-QAM and PM-16QAM in long-haul WDM transmission. *Optics Express* 21(16):19269–19279, 2013.

108. H. Zhang, J.-X. Cai, H. G. Batshon, M. Mazurczyk, O. Sinkin, D. Foursa, A. Pilipetskii, G. Mohs, and N. Bergano. 200 Gb/s and dual wavelength 400 Gb/s transmission over transpacific distance at 6 b/s/Hz spectral efficiency. In *Optical Fiber Communication Conference*, page PDP5A.6. 2013.
109. H. Sun, R. Egorov, B. E. Basch, J. McNicol, and K.-T. Wu. Comparison of two modulation formats at spectral efficiency of 5 bits/dual-pol symbol. In *European Conference on Optical Communication*, page Th.2.D.3. 2013.
110. R. Rios-Muller, J. Renaudier, O. Bertran-Pardo, A. Ghazisaeidi, P. Tran, G. Charlet, and S. Bigo. Experimental comparison between hybrid-QPSK/8QAM and 4D-32SP-16QAM formats at 31.2 GBaud using Nyquist pulse shaping. In *European Conference on Optical Communication*, page Th.2.D.2. 2013.
111. R. van Uden, C. Okonkwo, H. Chen, H. de Waardt, and A. Koonen. 6×28 GBaud 128-SP-QAM transmission over 41.7 km few-mode fiber with a 6×6 MIMO FDE. In *Optical Fiber Communication Conference*, page W4J.4. 2014.
112. J. K. Fischer, C. Schmidt-Langhorst, S. Alreesh, R. Elschner, F. Frey, P. W. Berenguer, L. Molle, M. Nölle, and C. Schubert. Generation, transmission and detection of 4d set-partitioning qam signals. *to appear in Journal of Lightwave Technology* 2015.
113. J. Karout, X. Liu, S. Chandrasekhar, E. Agrell, M. Karlsson, and R.-J. Essiambre. Experimental demonstration of an optimized 16-ary four-dimensional modulation format using optical OFDM. In *Optical Fiber Communication Conference*, page OW3B.4. 2013.
114. H. Bülow, T. Rahman, F. Buchali, W. Idler, and W. Kuebart. Transmission of 4-D modulation formats at 28-Gbaud. In *Optical Fiber Communication Conference*, page JW2A.39. 2013.
115. M. Chagnon, M. Osman, Q. Zhuge, X. Xu, and D. V. Plant. Analysis and experimental demonstration of novel 8PolSK-QPSK modulation at 5 bits/symbol for passive mitigation of nonlinear impairments. *Optics Express* 21(25):30204–30220, 2013.
116. H. Bülow, X. Lu, L. Schmalen, A. Klekamp, and F. Buchali. Experimental performance of 4d optimized constellation alternatives for PM-8QAM and PM-16QAM. In *Optical Fiber Communication Conference*, page M2A.6. 2014.
117. T. A. Eriksson, P. Johannisson, M. Sjödin, E. Agrell, P. A. Andrekson, and M. Karlsson. Frequency and polarization switched QPSK. In *European Conference on Optical Communication*, page Th.2.D.4. 2013.
118. T. A. Eriksson, P. Johannisson, E. Agrell, P. A. Andrekson, and M. Karlsson. Biorthogonal modulation in 8 dimensions experimentally implemented as 2PPM-PS-QPSK. In *Optical Fiber Communication Conference*, page W1A.5. 2014.
119. A. D. Shiner, M. Reimer, A. Borowiec, S. Oveis Gharan, J. Gaudette, P. Mehta, D. Charlton, K. Roberts, and M. O’Sullivan. Demonstration of an 8-dimensional modulation format with reduced inter-channel nonlinearities in a polarization multiplexed coherent system. *Optics Express* 22(17):20366–20374, 2014.
120. B. J. Puttnam, T. A. Eriksson, J.-M. Delgado Mendieta, R. S. Luís, Y. Awaji, N. Wada, M. Karlsson, and E. Agrell. Modulation formats for multi-core fiber transmission. *Optics Express* 22(26):32457–32469, 2014.
121. X. Liu, S. Chandrasekhar, T. H. Wood, R. W. Tkach, P. J. Winzer, E. C. Burrows, and A. R. Chraplyvy. M-ary pulse-position modulation and frequency-shift keying

- with additional polarization/phase modulation for high-sensitivity optical transmission. *Optics Express* 19(26):B868–B881, 2011.
122. X. Liu, T. H. Wood, R. W. Tkach, and S. Chandrasekhar. Demonstration of record sensitivities in optically preamplified receivers by combining PDM-QPSK and M-ary pulse-position modulation. *Journal of Lightwave Technology* 30(4):406–413, 2012.
 123. A. Ludwig, M.-L. Schulz, P. Schindler, R. Schmogrow, A. Mustafa, B. Moos, S. Brunsch, T. Dippon, D. Malsam, D. Hillerkuss, F. Roos, W. Freude, C. Koos, and J. Leuthold. Stacking PS-QPSK and 64PPM for long-range free-space transmission. In *Advanced Photonics Congress*, page NW2C.2. 2013.
 124. M. Sjödin, T. A. Eriksson, P. A. Andrekson, and M. Karlsson. Long-Haul Transmission of PM-2PPM-QPSK at 42.8 Gbit/s. In *Optical Fiber Communication Conference*. 2013.
 125. S. J. Savory. Digital filters for coherent optical receivers. *Optics Express* 16(2):804–817, 2008.
 126. M. Kuschnerov, M. Chouayakh, K. Piyawanno, B. Spinnler, E. De Man, P. Kainzmaier, M. S. Alfiad, A. Napoli, and B. Lankl. Data-aided versus blind single-carrier coherent receivers. *IEEE Photonics Journal* 2(3):387–403, 2010.
 127. R. Elschner, F. Frey, C. Meuer, J. K. Fischer, S. Alreesh, C. Schmidt-Langhorst, L. Molle, T. Takahito, and C. Schubert. Experimental demonstration of a format-flexible single-carrier coherent receiver using data-aided digital signal processing. *Optics Express* 20(27):28786–28791, 2012.
 128. M. Yan, Z. Tao, H. Zhang, W. Yan, T. Hoshida, and J. Rasmussen. Adaptive blind equalization for coherent optical BPSK system. In *European Conference on Optical Communication*, page Th.9.A.4. 2010.
 129. P. Johannisson, M. Sjödin, M. Karlsson, H. Wymeersch, E. Agrell, and P. A. Andrekson. Modified constant modulus algorithm for polarization-switched QPSK. *Optics Express* 19(8):7734–7741, 2011.
 130. D. S. Millar and S. J. Savory. Blind adaptive equalization of polarization-switched QPSK modulation. *Optics Express* 19(9):8533–8538, 2011.
 131. S. Alreesh, J. K. Fischer, M. Nolle, and C. Schubert. Joint-polarization carrier phase estimation for PS-QPSK signals. *IEEE Photonics Technology Letters* 24(15):1282–1284, 2012.
 132. H. Bülow. Ideal POL-QAM modulation for coherent detection schemes. In *Signal Processing in Photonic Communications*, page SPTuB.1. 2011.
 133. H. Bülow and E. S. Masalkina. Coded modulation in optical communications. In *Optical Fiber Communication Conference*, page OThO1. 2011.
 134. C. Chen, C. Li, M. Zamani, and Z. Zhang. Coherent detection of a 32-point 6PolSK-QPSK modulation format. In *Optical Fiber Communication Conference*, page OTh3C.4. 2013.
 135. J. H. Conway and N. J. A. Sloane. Soft decoding techniques for codes and lattices, including the Golay code and the Leech lattice. *IEEE Transactions on Information Theory* 32(1):41–50, 1986.
 136. C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE* 37(1):10–21, 1949.

137. C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell Syst. Tech. J* 38(3):611–656, 1959.
138. M. Sjödin, E. Agrell, P. Johannisson, G.-W. Lu, P. A. Andrekson, and M. Karlsson. Filter optimization for self-homodyne coherent WDM systems using interleaved polarization division multiplexing. *Journal of Lightwave Technology* 29(9):1219–1226, 2011.
139. S. Benedetto and P. Poggiolini. Theory of polarization shift keying modulation. *IEEE Transactions on Communications* 40(4):708–721, 1992.
140. M. Karlsson. Four-dimensional rotations in coherent optical communications. *Journal of Lightwave Technology* 32(6):1246–1257, 2014.
141. S. Chandrasekhar, X. Liu, B. Zhu, and D. W. Peckham. Transmission of a 1.2-Tb/s 24-carrier no-guard-interval coherent OFDM superchannel over 7200-km of ultra-large-area fiber. In *European Conference on Optical Communications*, page Th.13.B.1. 2009.
142. O. Edfors and A. J. Johansson. Is orbital angular momentum (OAM) based radio communication an unexploited area? *IEEE Transactions on Antennas and Propagation* 60(2):1126–1131, 2012.
143. H. Bülow, H. Al-Hashimi, and B. Schmauss. Coherent multimode-fiber MIMO transmission with spatial constellation modulation. In *European Conference on Optical Communications*, page Tu.5.B.3. 2011.
144. B. J. Puttnam, J.-M. Delgado Mendinueta, R. S. Luis, W. Klaus, J. Sakaguchi, Y. Awaji, N. Wada, T. A. Eriksson, E. Agrell, P. A. Andrekson, and M. Karlsson. Energy-efficient modulation formats for multi-core fibers. In *Optoelectronics and Communications Conference, OECC/ACOFT*, page We9B.1. 2014.
145. B. Kolner. Space-time duality and the theory of temporal imaging. *IEEE Journal of Quantum Electronics* 30(8):1951–1963, 1994.