Power-Efficient Modulation Formats in Coherent Transmission Systems

Erik Agrell and Magnus Karlsson

***Abstract*—Coherent optical transmission systems have a four-dimensional signal space (two quadratures in two polarizations). These four dimensions can be used to create modulation formats that have a better power efficiency (higher sensitivity) than the conventional binary and quaternary phase-shift keying (BPSK/QPSK) signals. Several examples are given, with some emphasis on a 24-level format and an 8-level format, including descriptions of how they can be realized and expressions for their symbol and bit error probabilities. These formats are, respectively, an extension and a subset of the commonly used 16-level dual-polarization QPSK (DP-QPSK) format. Sphere-packing simulations in 2, 3, and 4 dimensions, up to 32 levels, are used to verify their optimality. The numerical results as the number of levels increases are shown to agree with lattice-theoretical results. Finally we point out that the use of these constellations will lead to improved fundamental sensitivity limits for optical communication systems, and they may also be relevant as a way of reducing power demands and/or nonlinear influence.**

I. Introduction

THE last few years have seen a remarkably increased research activity on high-speed ($\geq$Gb/s) coherent fiber-optic links. There are many reasons for this interest, mainly the many benefits offered by coherent detection with respect to sensitivity, spectral efficiency, and equalization potential. The most important reason is perhaps the feasibility of electronic signal processing technologies to perform phase tracking algorithms, thus enabling coherent receivers. Recently coherent systems based on online synchronization at symbol rates of 1.4 Gbaud [1], 2.2-2.5 Gbaud [2], [3], and 10 Gbaud [4] have been reported. In addition, a plethora of experiments with off-line processing (post processing of data samples on a computer) were carried out, early on, as proof of concepts [5], [6].

Beside offering multiple advantages with respect to, e.g., post-processing of signal impairments, coherent systems also enable independent use of both quadratures and both polarization components of the electromagnetic field for data transmission. This is because the polarization tracking algorithms may also (just as the quadrature synchronization) be performed electronically, with reasonable complexity. This means that one may use all four degrees of freedom of the electromagnetic field to transmit data. Even if polarization multiplexed transmission has been demonstrated in microwave links at lower data rates, it is accompanied with too many free-space (as well as antenna-related) transmission obstacles to make it a commodity. On the other hand, transmission in a fiber modifies the polarization of the carrier wave in a controlled way that can be tracked in a coherent receiver, so the optical coherent systems may well be the first communication systems that are naturally suited for using four-dimensional (4D) signal constellations. Indeed, in recent experiments the dual-polarization quaternary phase shift keying (DP-QPSK) modulation format, which is independent QPSK modulation in each polarization component, has been demonstrated [4], [7]–[9]. This may be interpreted as one bit in each degree of freedom of the electromagnetic carrier, i.e., in total 4 bits per symbol.

When fundamental sensitivity limits for different modulation formats are discussed in textbooks and reviews [10]–[14], an additive white Gaussian noise (AWGN) model is assumed, for which the sensitivity for the chosen modulation format can be calculated. The sensitivity is defined as the signal-to-noise ratio (SNR) required to reach a bit error rate (BER) of 10^{-9} or (which is increasingly common) 10^{-3}. As most fundamental noise sources in optical channels are quantum mechanical in origin, the required SNR is often expressed in terms of photons per bit.

In most cases, binary phase shift keying (BPSK) is chosen as a reference, and as a result, BPSK is often believed to have the best sensitivity among all possible modulation formats. The DP-QPSK format (whose constellation diagram form the 16 vertices of a 4D cube) can be seen as four parallel and independent BPSK channels. Thus, the sensitivity for DP-QPSK (and in fact all cubic constellations, regardless of dimension [15]) is the same as that for BPSK. What is probably less well known, at least in the optical communication community, is that there are other modulation formats, taking full advantage of the 4D signal space, that have improved sensitivities over BPSK. The improvement sometimes comes at the expense of bandwidth, but there exist also constellations that improve both sensitivity and bandwidth over BPSK.

As the AWGN model is the foundation for the quantum sensitivity limits, these modulation formats will also provide improved quantum limits. Besides being of fundamental interest by providing new ultimate sensitivity limits, such power-efficient modulation formats may be of practical relevance as they provide means to reduce nonlinear fiber transmission impairments.

Modulation in a 4D constellation space has been investigated previously in the communication theory literature [15]–[18]. In [16], constellations with more than 12 levels were analyzed in terms of symbol error rates (SER). Some simpler systems, including 5-, 8- and 16-level systems, were analyzed

E. Agrell is with the Communication Systems Group, Dept. of Signals and Systems, and M. Karlsson is with the Photonics Laboratory, Dept. of Microtechnology and Nanoscience, and both at Chalmers University of Technology, SE-412 96 Göteborg, Sweden.

in [17]. For reasons that will be apparent later on in this article, the 5-, 8-, 16-, and 24-level schemes are of most interest to us. However, these studies (with the notable exception of Taricco *et al.* [18]) did not compute asymptotic bit energy sensitivities, which would reveal what modulation schemes are most power efficient. In those studies, the 4D constellation space was realized by using two different carrier waves or two time slots, which is a bit different from optical transmission, where the vector nature of the electromagnetic field creates a natural 4D signal space. However, also in the optical communication context, 4D modulation was investigated in the early 1990s [19]–[22], when coherent optical communications had a brief period of popularity. These papers demonstrated theoretically how optical transmission systems could benefit from 4D modulation techniques, by showing how transmitters and receivers could be realized. Some fundamental sensitivity limits were given in [19], [20]. However, it is not entirely clear from these works under what circumstances the constellations were optimized (for example under an average or maximum symbol energy constraint). Nor do they point out that sensitivity improvements over BPSK could be achieved, which in our opinion is a most important observation.

We showed recently [23] that an 8-level format, which we referred to as polarization-switched QPSK (PS-QPSK), is the format with overall best sensitivity for uncoded transmission[1], being asymptotically 3/2 (1.76 dB) better than BPSK. In this paper we will put the results of [23] in a broader context by elaborating on the general problem of power-efficient modulation and its relation to polarization states, symbol and bit error rates, sphere packings, lattice theory, and quantum limits. Several other formats, in addition to PS-QPSK, are given that have improved sensitivities over BPSK. In particular, a 24-level modulation scheme is described in detail, including a discussion of its nontrivial bit-to-symbol mapping. It should be noted that we are not the first to point out that multilevel formats with sensitivities better than BPSK exist. Rather, their asymptotic sensitivity gains were originally given in [15], [18] in the context of increasing the dimensionality of the signal by using two carrier waves instead of the two polarization components that can be used in fiber communications. In addition, our discussion is more general as we show the performance of the optimal constellations for all numbers of levels up to 32.

As stated above, we will give a number of examples of modulation formats (e.g., based on 5, 8, and 24 levels) that have *improved* receiver sensitivities over BPSK and DP-QPSK. Two of these (the 8- and 24-level formats) have a reasonable complexity and, contrary to the 5-level system, the transmitter and the bit-to-symbol mapping problem can be solved without too much loss of performance. Furthermore, the systems of today utilize slightly different setups and modulators than in the early 1990s, so here we will also anchor our findings in the present research.

The overall structure of the paper is the following. In Sec. II, we lay out the basic definitions and notation, discuss the relation between polarization states and signals in 4D space, and explain the relation between dense sphere packings and power-efficient constellations. In Sec. III we present several DP-QPSK improvements, including bit-to-symbol mappings. Symbol- and bit-error probability performances are given in Sec. IV. In Sec. V, we present optimum (in the sense of best power efficiency) constellations in 2–4 dimensions for all number of levels up to 32, while in Sec. VI we discuss the implications of these findings for the quantum-limited sensitivities of optical systems. Finally, in Sec. VII, we summarize the findings of the paper.

[1]"Uncoded" means that all 4D symbols transmit independent information.

II. Definitions and preliminaries

This section describes the basic properties of the electromagnetic field and how we interpret it as a 4D signal. Some of it is standard textbook material, but we wish to include it for completeness. The four subsections are devoted to first how the two quadratures and two polarization states can be interpreted as a 4D optical signal; secondly, the properties of the 4D equivalent of QPSK, its transmitter structures and polarization states; thirdly, the transmission in additive Gaussian noise, including definitions of signal-to-noise ratio and spectral efficiency; and finally, the importance of the constellation distance properties, the asymptotic power efficiency, and the interpretation as sphere packings in 4D space.

A. The four-dimensional optical signal

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

$$\boldsymbol{E} = \begin{pmatrix} E_{x,r} + iE_{x,i} \\ E_{y,r} + iE_{y,i} \end{pmatrix} = \begin{pmatrix} |E_x| \exp(i\varphi_x) \\ |E_y| \exp(i\varphi_y) \end{pmatrix}, \tag{1}$$

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

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

$$\boldsymbol{E} = \|\boldsymbol{E}\| \exp(i\varphi_a) J = \|\boldsymbol{E}\| \exp(i\varphi_a) \begin{pmatrix} \cos\theta \exp(i\varphi_r) \\ \sin\theta \exp(-i\varphi_r) \end{pmatrix}, \tag{2}$$

where $\|\boldsymbol{E}\|^2 = |E_x|^2 + |E_y|^2$ and $\theta = \sin^{-1}(|E_y|/\|\boldsymbol{E}\|)$. J denotes the Jones vector, which is usually normalized to unity, i.e., $J^+J = |J|^2 = 1$. Note the distinction between the absolute phase $\varphi_a = (\varphi_x + \varphi_y)/2$ of the field and the relative phase $\varphi_r = (\varphi_x - \varphi_y)/2$ between the field vector components. The relative phase $\varphi_r \in (-\pi, \pi]$ describes the *ellipticity* of the polarization state, with the special cases $\varphi_r = 0, \pm\pi/2, \pi$ for linear polarization and $\varphi_r = \pm\pi/4, \pm 3\pi/4$ for circular polarization, and all other cases are called elliptical states of polarization. The angle $\theta \in [0, \pi/2]$ is usually called the *azimuth* as it describes the orientation in the xy plane of the linear polarization states, or, more generally, the major axis of the polarization ellipse.

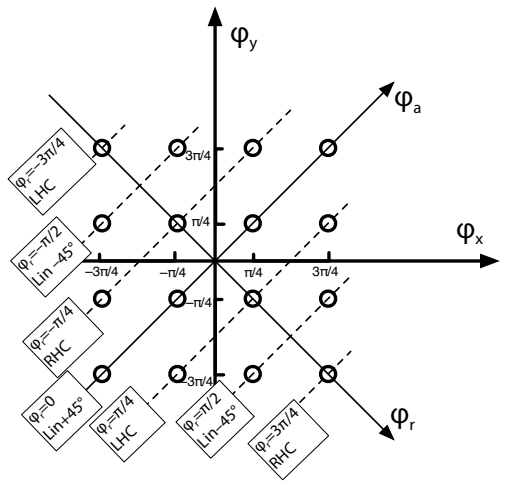

Fig. 1. The phase values used for DP-QPSK modulation. The diagonal axes show the φ_r and φ_a phases. For the φ_r levels, the corresponding states of polarization are denoted as linear ±45°, LHC, or RHC.

A final way of expressing the signal is as a 4D vector $\boldsymbol{c}$ with real components

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

The transmitted optical power is $P = \|\boldsymbol{s}\|^2 = \|\boldsymbol{E}\|^2 = E_{x,r}^2 + E_{x,i}^2 + E_{y,r}^2 + E_{y,i}^2$. Note that this 4D vector should not be confused with the Stokes vector description of polarization states, which is defined in a completely different way and proportional to the intensity rather than being linear in the field. The three-dimensional (3D) Stokes space was used as a signal space for polarization shift keying modulation in the 1990s [24]. However, the lack of an absolute phase description makes constellation points with different absolute phase but the same polarization coincide in Stokes space, and it is therefore less useful as a signal space in a coherent communication system with additive noise (see Sec. II-C). Yet the Stokes space description of the optical field is useful when discussing the polarization properties of the different modulation formats.

B. DP-QPSK modulation

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

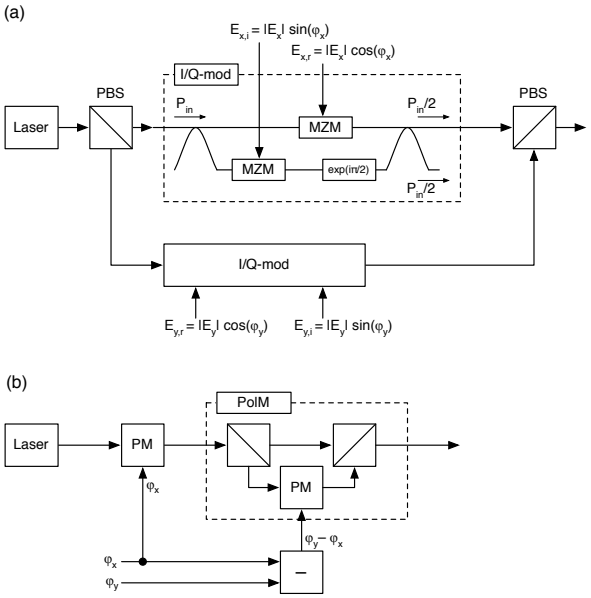

Fig. 2. (a) The conventional DP-QPSK transmitter, based on one I/Q modulator in each polarization state. The small arrows indicate the 3 dB loss of the I/Q modulators. (b) An alternative transmitter structure, based on a single PM followed by a PolM that retards the y polarization component the indicated phase value. This transmitter has no intrinsic 3 dB loss.

The standard DP-QPSK transmitter (see, e.g., [9]) is based on the simultaneous modulation of the real and imaginary parts of both polarization components, as shown in Fig. 2 (a). This transmitter is entirely general, in the sense that it can transmit any I/Q modulation format in each polarization state, including quadrature amplitude modulation (QAM), by letting each Mach–Zehnder modulator (MZM) generate one element of the 4D vector (3). Another transmitter structure, exploiting the fact QPSK is a constant-intensity format, is shown in Fig. 2 (b). This transmitter has a lower optoelectronic hardware complexity, as it requires only two phase modulators (PM) and no external polarization beam splitter (PBS). However, one PM is part of a polarization modulator (PolM) which is schematically shown in Fig. 2 (b) as consisting of a PM sandwiched between two PBSs, but in reality it can be realized as a single-waveguide structure with a birefringent electrooptic PM that modulates one polarization state only. In some cases a standard PM with misaligned input polarization may be used, although not at very high symbol rates, due to the walk-off between the optical and electrical signals.

Other PolM structures and transmitter structures are described in [20]–[22]. A more general PolM was realized in a $LiNbO_3$ structure [25]. Such a modulator has the possibility of producing any output polarization state for a given specific input polarization state, and its transfer (Jones) matrix can be written as

$$\begin{pmatrix} \cos\theta \exp(i\varphi) & -\sin\theta \exp(i\varphi) \\ \sin\theta \exp(-i\varphi) & \cos\theta \exp(-i\varphi) \end{pmatrix}. \tag{4}$$

The alternative transmitter described in Fig. 2 (b) has potentially lower cost as it is a simpler structure, and it also avoids the intrinsic 3 dB loss of the I/Q modulators indicated in Fig. 2 (a), which arises when the in-phase and quadrature compo-

nents are mixed in the output coupler. This will imply that the PBS-I/Q transmitter in Fig. 2 (a) suffers a power loss of $\sqrt{2}$ when transmitting purely x- or y-polarized light, compared with when both polarization components are present, and this will make other modulation formats (such as some suggested in this paper) more difficult to realize. Still, the I/Q modulator configuration has a couple of important advantages. The first is that the nonlinear characteristic of the transfer function makes the transmitter less susceptible to noisy or inaccurate drive signals. The second is that the inter-symbol interference (ISI) induced by the transmitter/receiver pair, which plagues PM-based QPSK and DQPSK transmitters [26]–[28], is absent. However, for a return-to-zero (RZ) transmitter, the transitions between the transmitted symbols are suppressed, making this second issue insignificant.

C. Digital transmission over a noisy channel

In general, all entities in (3) vary continuously with time. For the purpose of digital communications, $\boldsymbol{s}(t)$ is designed to transmit a sequence of information symbol $(\boldsymbol{s}_0, \boldsymbol{s}_1, \boldsymbol{s}_2, \ldots)$, one symbol every T seconds. The symbol $\boldsymbol{s}_n$ is taken from a finite set, or *constellation,* $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$ of N-dimensional vectors.[2] We assume all constellation vectors to be equally likely. Thus, $\log_2 M$ information bits are transmitted every T seconds, yielding an information bit rate of $R = \log_2 M/T$.

With linear modulation as in Fig. 2 (a), $\boldsymbol{s}(t)$ is generated as

$$\boldsymbol{s}(t) = \sum_n \boldsymbol{s}_n p(t - nT), \tag{5}$$

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

The signal $\boldsymbol{s}(t)$ is transmitted over a noisy channel. In a coherent system, the dominating noise source is usually either amplified spontaneous emission (ASE) noise from in-line optical amplifiers or shot noise from the local oscillator in the receiver [13], [14], [29]. Both these noise sources are accurately modeled by the additive white Gaussian noise (AWGN) channel, for which the received N-dimensional signal is $\boldsymbol{r}(t) = \boldsymbol{s}(t) + \boldsymbol{z}(t)$, where $\boldsymbol{z}(t)$ is a vector of N independent, white, and Gaussian noise processes, each with a double-sided spectral density of $N_0/2$ (which is the standard notation in communications literature).

The purpose of the receiver is to recover the sequence $(\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots)$ as reliably as possible, given an observation of the signal $\boldsymbol{r}(t)$. It is well known (see [30, Sec. 2.6] or [31, Sec. 5.1]) that in the absence of ISI, the optimal receiver operates by filtering $\boldsymbol{r}(t)$ and sampling, creating a sequence of so-called *received vectors* $(\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots)$ where

$$\boldsymbol{r}_n = \int_{-\infty}^{\infty} \boldsymbol{r}(t + nT)p(t)dt. \tag{6}$$

[2]The theory in this subsection holds for any dimension N; e.g., single-quadrature BPSK has $N = 1$, QPSK has $N = 2$, and DP-QPSK has $N = 4$.

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

$$E_s = \frac{1}{M}\sum_{k=1}^{M} \|\boldsymbol{c}_k\|^2 = PT. \tag{7}$$

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

The *spectral efficiency*, SE, is generally defined either as the information bitrate per bandwidth (in bits/s/Hz) or as information bits per channel use, where a "channel use" refers to the transmission of two (or sometimes one) real vectors over the discrete-time channel, i.e., to two (or one) dimensions in signal space [30, p. 219]. We follow the latter approach, defining the spectral efficiency as the number of transmitted bits per polarization, where each polarization represents a dimension pair. Formally,

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

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

D. Symbol error rates and sphere packing

If the pulse $p(t)$ is suitably chosen, there is no ISI and $\boldsymbol{s}_n$ can be optimally estimated from the single received vector $\boldsymbol{r}_n$. The N-dimensional additive noise means that the received vector $\boldsymbol{r}_n$ has an isotropic distribution around $\boldsymbol{s}_n$ in a 4D space, and for a maximum likelihood receiver, the symbol decision is based on which signal in the constellation set is closest (in the Euclidian sense) to the received vector. To put this on more solid mathematical grounds, consider the constellation $\mathcal{C}$ of M signaling points, or symbols, $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$. Each symbol $\boldsymbol{c}_k$ is surrounded by a decision region, also known as *Voronoi region,* defined as all points closer to $\boldsymbol{c}_k$ than to any $\boldsymbol{c}_j \neq \boldsymbol{c}_k$. The probability of receiving symbol $\boldsymbol{c}_k$ in error is then the probability for a Gaussian variable centered at $\boldsymbol{c}_k$ to be outside the Voronoi region. For constellations in many dimensions, this calculation is intractable, since the Voronoi regions may be very complex in shape and an exact computation of error rates usually requires numerical simulations.

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

Finally, averaging over all equiprobable symbols $\boldsymbol{c}_k$, the union bound on the SER can be expressed as [30]

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

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

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

We wish to compare the performance of constellations with different numbers of levels at a fixed bit rate R. We therefore rewrite the dominant term in (9) as $\operatorname{erfc}(\sqrt{P\gamma/(RN_0)})$, where

$$\gamma = \frac{d_{\min}^2}{4E_b} \tag{10}$$

and $E_b = E_s/\log_2 M$ is the average *energy per bit*. The parameter γ, which captures the constellation's influence on the SER and is usually given in dB, is called the *asymptotic power efficiency* [30, p. 220], because the power needed for a certain required SER is proportional to $1/\gamma$. Another interpretation of γ is as the *sensitivity gain* over BPSK to transmit the same data rate, since $\gamma = 0$ dB for BPSK, QPSK, and DP-QPSK.

In fact, most common modulation formats have a penalty with respect to BPSK; for example, M-PSK and M-QAM have [30, pp. 226, 234]

$$\gamma_{M-PSK} = \sin^2(\pi/M)\log_2 M, \tag{11}$$

$$\gamma_{M-QAM} = \frac{3\log_2 M}{2(M-1)}, \tag{12}$$

where (12) is valid for M being a power of 4. We can show from these expression that both M-PSK and M-QAM have efficiencies $\gamma \leq 0$ dB for all values of M (with the notable exception of 3-PSK, which will be discussed in Sec. IV). A natural question is then: At a given dimension N, which modulation format has the highest asymptotic power efficiency γ? Or, equivalently, which modulation format has the lowest sensitivity? Remarkably, we have not found the answer to this question anywhere in the communications (or photonics) literature, and in Section V of this paper we return to it. Before that, however, we will present a couple of of promising modulation formats in four dimensions that have $\gamma > 0$ dB, i.e., lower sensitivities than the DP-QPSK format.

III. Selected four-dimensional constellations

Considering the DP-QPSK format, we may ask whether it is possible to increase the number of levels M, without increasing the average power or reducing the minimum distance of the constellation. The answer is yes [32], and it can be understood in the following way.

We noted in Fig. 1 that only four polarization states (±45°, LHC, RHC) and four phase levels per polarization state are involved in the DP-QPSK format. A simple retarder wave plate can change those formats to the linear states (x, y, ±45°) without affecting the performance. Or, by using a polarization rotator, the states could be transformed to (x, y, LHC, RHC). This shows that purely x- or y-polarized light may be used just as well as the 45° or circular states, and obviously this opens up for using two additional polarization states with four more phase levels each, so that eight more levels can be used without changing the minimum distance or the average power. These additional modulation levels were recently suggested by Bülow [32] to be utilized for forward error correction (FEC) overhead. Here we will instead take a step back and discuss how they can improve the fundamental power efficiencies of uncoded systems.

We denote the DP-QPSK constellation with $\mathcal{C}_1$, i.e., the set of 16 4D vectors given by any combination of signs in the set $(\pm1, \pm1, \pm1, \pm1)$. Then the extra levels we can use is formed by the set $\mathcal{C}_2$, which is defined as the eight vectors $(\pm2, 0, 0, 0)$ with any sign and any permutation of coordinates. By now forming a set of 24 levels $\mathcal{C}_3 = \mathcal{C}_1 \cup \mathcal{C}_2$, we have a modulation format that uses four phase levels for each of the six polarization states (x, y, ±45°, LHC, RHC). The 24 levels in $\mathcal{C}_3$ form the vertices of 4D polytope sometimes referred to as the *24-cell*. This constellation was also investigated in [16]–[18].

In sphere-packing theory, the maximum number of nonoverlapping spheres in N-dimensional space that can touch a given sphere with the same size is known as the *kissing number* K_N. The relevance of the kissing number in communications is that it directly shows how many symbols can be stacked at the same distance from the origin with maintained minimum distance, i.e., without modulating the amplitude. For two and three dimensions, one has $K_2 = 6$ and $K_3 = 12$, respectively [33], and in four dimensions one has $K_4 = 24$. Like many sphere-packing problems, rigorous proof are difficult, and although $K_4 = 24$ was long conjectured [33], it was only recently proven formally [34]. The centers of these 24 spheres are given by the set $\mathcal{C}_3$.

Theoretically, by using the set $\mathcal{C}_3$, the sensitivity of the DP-QPSK format is improved by $\log_2(24)/\log_2(16) = 0.59$ dB, as already noted in [18]. It is noteworthy that we can do even better by including also the all-zero vector, so by forming the set $\mathcal{C}_4 = \mathcal{C}_3 \cup \{(0,0,0,0)\}$ we have 25 levels with $d_{\min} = 2$, $E_s = 4 \cdot 24/25$, and $\gamma = (25/96)\log_2 25 = 0.83$ dB. This

is the densest way to pack 25 spheres in four dimensions. However, to practically implement the constellation $\mathcal{C}_4$ meets a number of problems, so we will disregard this constellation and focus on $\mathcal{C}_3$ instead. There are still a few practical problems that need to be solved in order to build a modulation format from the constellation $\mathcal{C}_3$; they are (i) the bit-to-symbol mapping and (ii) the modulator implementation. We will address these problems shortly.

An alternative and more compact description of the 24 levels of $\mathcal{C}_3$ is $\boldsymbol{c}' \in (\pm\sqrt{2}, \pm\sqrt{2}, 0, 0)$, again allowing for arbitrary sign choices and coordinate permutations. This representation, denoted $\mathcal{C}_3'$, can be obtained from the previous set by applying the coordinate transformation [33]

$$\boldsymbol{c}' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \boldsymbol{c} \tag{13}$$

for any vector $\boldsymbol{c} \in \mathcal{C}_3$. This is an equally common representation of the 24-cell. In fiber-optics language, we may interpret this transformation as $\boldsymbol{E}' = \boldsymbol{E} \exp(i\pi/4)$, i.e., a 45° rotation of the carrier phase of the electric field.

To make use of the eight extra modulation levels in $\mathcal{C}_3$, we propose to use two symbols in sequence, which gives us $24^2 = 576$ levels which is fairly close to $2^9 = 512$. Therefore we propose to transmit nine bits over two symbols, which corresponds to an asymptotic efficiency of $\gamma = 9/8 = 0.51$ dB. We refer to this modulation format as 6P-QPSK, to indicate that it has six polarization states with four phase levels each.

The symbol mapping for these nine bits is explained in Fig. 3. Denoting the two consecutive symbols as $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$, we distinguish between four cases, depending on whether $(\boldsymbol{c}_a, \boldsymbol{c}_b)$ belongs to $(\mathcal{C}_1, \mathcal{C}_1)$, $(\mathcal{C}_1, \mathcal{C}_2)$, $(\mathcal{C}_2, \mathcal{C}_1)$, or $(\mathcal{C}_2, \mathcal{C}_2)$. A symbol in set $\mathcal{C}_1$ can encode four bits and a symbol in set $\mathcal{C}_2$ can encode three bits. If the first bit in the 9-bit block to be transmitted, b_1, equals 0, then the remaining 8 bits are used to encode two symbols $(\boldsymbol{c}_a, \boldsymbol{c}_b)$ in $(\mathcal{C}_1, \mathcal{C}_1)$. If $b_1 = 1$, then the next bit, b_2, determines whether the remaining seven bits should encode a pair $(\boldsymbol{c}_a, \boldsymbol{c}_b)$ in $(\mathcal{C}_1, \mathcal{C}_2)$ or $(\mathcal{C}_2, \mathcal{C}_1)$. The case when both symbols belong to $\mathcal{C}_2$ is not used, so that effectively only $2^{9/2} = 22.6$ of the 24 constellation points in $\mathcal{C}_3$ are used. The unused constellation points can provide some error-detection capability. This is how we from now on distinguish between the $\mathcal{C}_3$ constellation and the more "practical" 6P-QPSK format: The number of bits per symbol is 4.5 for the 6P-QPSK and $\log_2 24 = 4.58$ for $\mathcal{C}_3$.

The second problem is how to modulate the extra eight levels in the transmitter. As stated in Sec. II, the standard DP-QPSK transmitter based on I/Q-modulators, shown in Fig. 2 (a), cannot be used directly, since the output amplitude with only a single polarization component present is (in normalized terms) $\sqrt{2}$ less than the when both are present. One solution of this problem is to precede the transmitter in Fig. 2 (a) with a variable PolM, which rotates the input state between x or y polarization for the set $\mathcal{C}_2$ or $\pm 45°$ polarization for $\mathcal{C}_1$. Alternatively, the transmitter of Fig. 2 (b) can be used, but then a full PolM with transfer characteristics according to (4) instead of the indicated partial PolM would be a viable solution. A third alternative, suggested in [32], is to use an amplitude modulator in addition to the transmitter of Fig. 2 (a) that will force all levels to have the same amplitude.

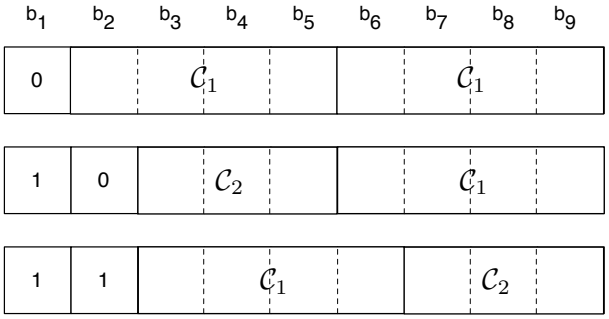

Fig. 3. The bit-to-symbol mapping proposed for 6P-QPSK. 9 bits $b_1, \ldots, b_9$ are mapped onto 2 symbols, each taken from the set $\mathcal{C}_1$ (16 levels) or $\mathcal{C}_2$ (8 levels), but not both from $\mathcal{C}_2$.

TABLE I
THE STRUCTURE OF SOME OPTIMAL M-LEVEL 4D CONSTELLATIONS.

M	Structure (see Section V)
2–5	$(M-1)$-dimensional simplex
6	Two 4D simplexes with a 3D facet in common
7	$\mathcal{C}_2$ with one point removed
8	$\mathcal{C}_2$
10	The *rectified simplex*, which consists of the midpoints between the 10 pairs of the 5 simplex points. All those points lie on the 4D unit sphere. An alternative description is as one tetrahedron and one octahedron, placed in parallel 3D hyperplanes.
16	A remarkable structure comprising 9 points from the D_4 lattice and 7 points from a rotated version of the same lattice
24	$\mathcal{C}_4$ with one corner point removed
25	$\mathcal{C}_4$

A very efficient modulation format, PS-QPSK, is obtained by using the set $\mathcal{C}_2$ only. Compared with DP-QPSK, it has only eight levels, so the spectral efficiency is reduced to 3 bits per symbol (1.5 bits per dimension pair), but this is more than compensated for by the minimum distance increasing by a factor of $\sqrt{2}$. Thus the asymptotic power efficiency becomes $\gamma = 3/2 = 1.76$ dB better than the DP-QPSK format [15], [18]. This is a significant improvement, and it is noteworthy that it comes from reducing the complexity from 16 to 8 levels. The format was originally proposed for optical systems in [20]. Its transmitter realizations were discussed in [23], where it was also identified as the format with the overall highest power efficiency, for uncoded transmission and any number of levels.

Table I briefly describes the structures of some other 4D constellations that are more power efficient than DP-QPSK, although some of them have a lower spectral efficiency. As explained in Sec. V, they all attain the optimum power efficiency for their respective sizes M. However, these constellations might be of limited interest in practical systems, as their generation or bit-to-symbol mapping may be hard to realize.

IV. BIT AND SYMBOL ERROR RATE PERFORMANCES

We will now quantify the two proposed modulation formats in terms of bit and symbol error rates. For $\mathcal{C}_1$ (DP-QPSK), $\mathcal{C}_2$

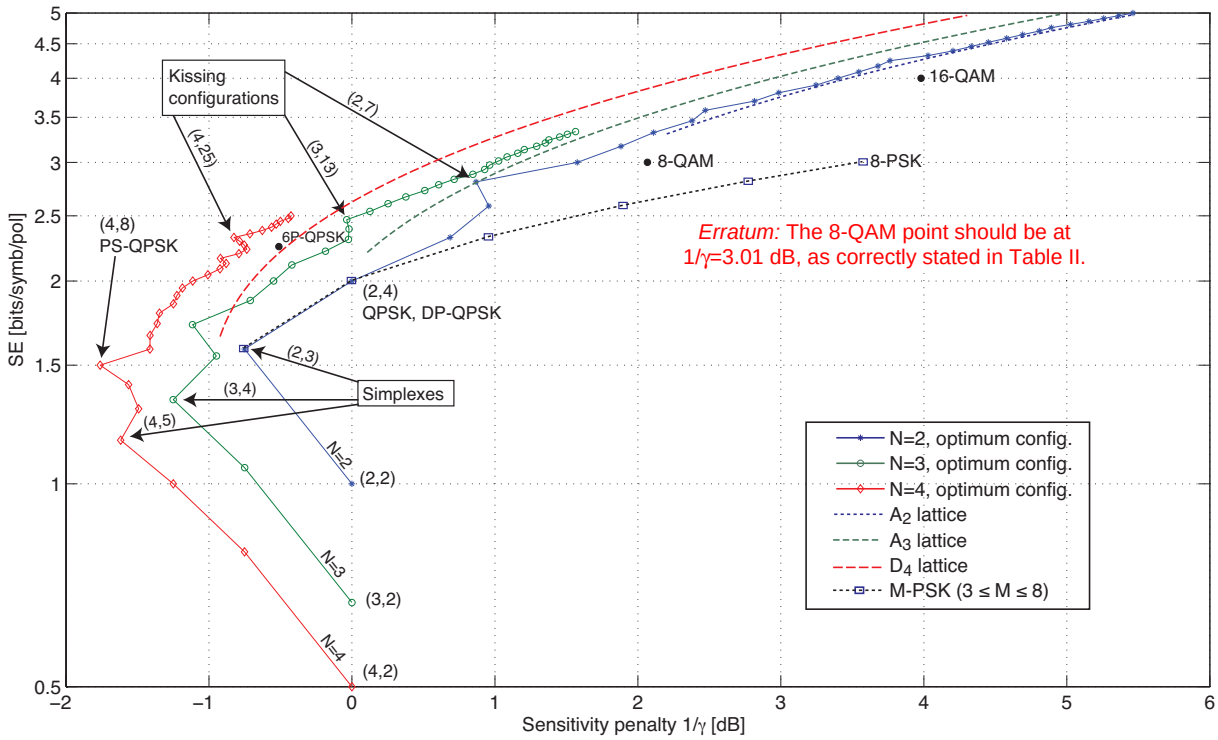

Fig. 6. Spectral efficiency vs. sensitivity penalty $1/\gamma$ for several modulation formats. Points (N, M) refer to the optimum M-ary constellations for $M = 2, 3, \ldots, 32$ and $N = 2, 3, 4$, based on data from [35], [36]. Simplexes and kissing configurations are marked (see text), as are the asymptotes (lattices) for high M. Also included for comparison are the M-PSK, 8-QAM, 16-QAM, and 6P-QPSK formats.

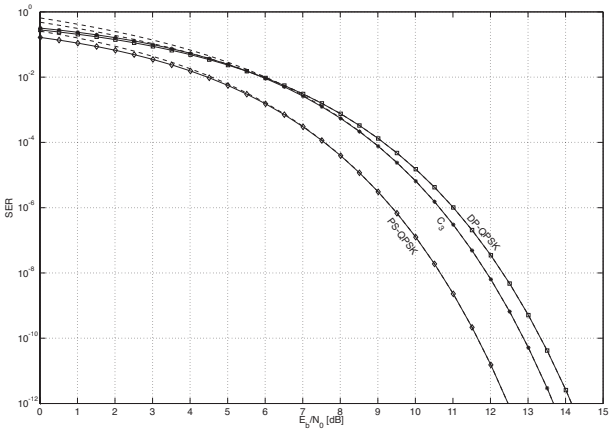

Fig. 4. SER versus E_b/N_0 for PS-QPSK, $\mathcal{C}_3$, and DP-QPSK. The dashed lines are union bound calculations, whereas the solid lines are exact calculations from (14)–(16). The expected asymptotic improvements are 1.76 dB for PS-QPSK and 0.59 dB for $\mathcal{C}_3$.

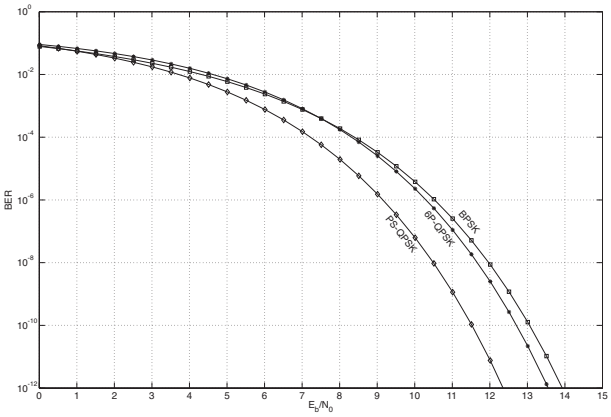

Fig. 5. BER versus E_b/N_0 for PS-QPSK, 6P-QPSK, and BPSK. QPSK and DP-QPSK have the same BER performance as BPSK. The improvement of PS-QPSK over BPSK is 0.97 dB at a BER of 10^{-3} and 1.51 dB at 10^{-9}. The asymptotic gains are again 1.76 dB for PS-QPSK but only 0.51 dB for 6P-QPSK.

(PS-QPSK), and $\mathcal{C}_3$, the exact SER expressions are, resp.,

$$SER_1 = 1 - \left[1 - \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_s}{4N_0}}\right)\right]^4 \tag{14}$$

$$SER_2 = 1 - \frac{1}{\sqrt{\pi}}\int_0^\infty (1 - \operatorname{erfc} x)^3 e^{-\left(x - \sqrt{\frac{E_s}{N_0}}\right)^2} dx \tag{15}$$

$$SER_3 = 1 - \frac{1}{\sqrt{\pi}}\int_0^\infty (1 - \operatorname{erfc} x)^2 \operatorname{erfc}\left(x - \sqrt{\frac{E_s}{2N_0}}\right) \cdot e^{-\left(x - \sqrt{\frac{E_s}{2N_0}}\right)^2} dx. \tag{16}$$

Eq. (14) is straightforward to derive due to the simple geometry of the cubic constellations. The SER_2 expression (15) can be found in standard textbooks [30, p. 210], [37, p. 201] by recognizing $\mathcal{C}_2$ as an 8-ary biorthogonal constellation. The derivation of the SER_3-expression (16) is more cumbersome and we omit the details.

We do not recommend (14)–(16) for numerical evaluation at high E_s/N_0. Cancellation occurs when subtracting two almost equal numbers. As observed in [38] for the case of $\mathcal{C}_2$, expanding the polynomials in $\operatorname{erfc} x$ and integrating out

the constant term yields

$$SER_1 = \frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_s}{4N_0}}\right)\left[4-\operatorname{erfc}\left(\sqrt{\frac{E_s}{4N_0}}\right)\right] \cdot\left[8-4\operatorname{erfc}\left(\sqrt{\frac{E_s}{4N_0}}\right)+\operatorname{erfc}^2\left(\sqrt{\frac{E_s}{4N_0}}\right)\right] \quad (17)$$

$$SER_2 = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_s}{N_0}}\right)+\frac{1}{\sqrt{\pi}}\int_0^\infty \operatorname{erfc} x \cdot(3-3\operatorname{erfc} x+\operatorname{erfc}^2 x)e^{-\left(x-\sqrt{\frac{E_s}{N_0}}\right)^2}dx \quad (18)$$

$$SER_3 = \operatorname{erfc}\left(\sqrt{\frac{E_s}{2N_0}}\right)\left[1-\frac{1}{4}\operatorname{erfc}\left(\sqrt{\frac{E_s}{2N_0}}\right)\right] + \frac{1}{\sqrt{\pi}}\int_0^\infty \operatorname{erfc} x(2-\operatorname{erfc} x)\operatorname{erfc}\left(x-\sqrt{\frac{E_s}{2N_0}}\right) \cdot e^{-\left(x-\sqrt{\frac{E_s}{2N_0}}\right)^2}dx. \quad (19)$$

In Fig. 4, we show the SER as a function of E_b/N_0 for some of the constellations discussed in this paper. Union bounds from (9) are also shown. It is noteworthy that the union bound becomes indistinguishable from the exact values when the SER is less than 10^{-3}.

To get the BER performance, the bit-to-symbol mapping must be considered. Throughout this paper, we assume symbolwise detection, which is what is implemented in all optical communication systems. That the BER can be somewhat reduced by detecting each bit individually [39], which is mainly of theoretical interest, does not appear to be well known.

For the DP-QPSK format, one can map each quadrature individually. This will lead to a Gray mapping and the corresponding BER performance will be equivalent to that of the BPSK channel, which is $(1/2)\operatorname{erfc}(\sqrt{E_b/N_0})$. This property holds for any N-dimensional cubic modulation format, such as BPSK, QPSK, or DP-QPSK, as was pointed out in [15].

The eight levels of the PS-QPSK format are not possible to Gray code, since each point has six nearest neighbors. The best one can do is to encode the levels so that the pairs of points farthest away from each other have inverted binary code words. In such a situation, the six most likely symbol errors will have one or two bits wrong, of the transmitted three bits. Ignoring the seventh possible symbol error, which is much less probable, $BER_2 \approx SER_2/2$. Or, if the exact expression is desired, the seventh symbol error can be determined by a separate integral [37, pp. 198–203], which after some simplification yields

$$BER_2 = \frac{1}{2}-\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}(1-\operatorname{erfc} x)^3 e^{-\left(x-\sqrt{\frac{E_s}{N_0}}\right)^2}dx \quad (20)$$

$$= \frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}\operatorname{erfc} x \cdot(3-3\operatorname{erfc} x+\operatorname{erfc}^2 x)e^{-\left(x-\sqrt{\frac{E_s}{N_0}}\right)^2}dx \quad (21)$$

where (21) is numerically stable. A comparison between (20) and (15) verifies that indeed $BER_2 \approx SER_2/2$.

For the 6P-QPSK format, we use the bit-to-symbol mapping of Fig. 3. However, depending on how blocks of four and three bits are mapped to points in $\mathcal{C}_1$ and $\mathcal{C}_2$, resp., many inequivalent 6P-QPSK mappings can be generated, whose BER performance differ slightly. For the purpose of this study, we map $0 \to 1$ and $1 \to -1$ to generate a point in $\mathcal{C}_1$ from four bits. For $\mathcal{C}_2$, we assign the vectors $(0,0,0,2)$, $(0,0,2,0)$, $(0,2,0,0)$, and $(2,0,0,0)$ to, resp., 000, 001, 010, and 011, while the negative of these vectors are assigned to inverted bit patterns. By enumerating all nine-bit blocks and their corresponding symbol pairs with this mapping, it can be shown that the minimum-distance symbol error events generate on average 5/2 bit errors[3], disregarding the case when both received symbols are in $\mathcal{C}_2$, and a transmission error is detected. Thus, $BER_{\text{6P-QPSK}} \approx (5/18)SER_3$.

The BER performance of DP-QPSK (or, equivalently, BPSK), PS-QPSK (exact), and 6P-QPSK (approximation) are compared in Fig. 5. One may note the value of E_b/N_0 at a BER of 10^{-3}, which is 5.82 dB for PS-QPSK and 6.79 dB for BPSK. At 10^{-9}, we have 11.04 dB for PS-QPSK and 12.55 dB for BPSK.

V. N-DIMENSIONAL SPHERE PACKING

We have in this paper exemplified two modulation formats that can improve the sensitivity over the well-known BPSK, QPSK, or DP-QPSK optical coherent systems. One may then ask if there are even better modulation formats (in terms of asymptotic power efficiency γ) and obviously which modulation format that has the highest sensitivity.

Before embarking on the general question we will make some historical notes. As we noted in Sec. II, the problem of finding the constellation with maximum asymptotic efficiency is equivalent to finding the densest (in the sense of the second moment) packing of M N-dimensional spheres. It is actually challenging enough to find the best constellations for a fixed number of levels M in a given dimension N. In general, no formal mathematical proof that a certain constellation is the densest is known, but rather, empirical evidence in the sense that "no better constellations have been found."

For modulation in the plane ($N = 2$) and for selected values of M up to 16, some optimal constellations were presented by Foschini *et al.* [40], and they are typically hexagonal packings of M levels centered around the origin. This was further demonstrated by Graham *et al.* [36], who numerically computed conjectured optimum packings up to $M \leq 500$. In $N = 3$ dimensions, the best sphere packings, including images of the cases $M \leq 20$, were originally reported in [41]. Moreover, tables of the second moment of the best known packings for $N = 3$, $M \leq 99$ and $N = 4$, $M \leq 32$ are available online [35] as a result of extensive simulation work by Sloane *et al.* around 1994. Also, other tables based on numerical simulations have been reported, e.g., in [42], but it is noteworthy that some of the constellations reported

[3]We believe that this is the fewest possible bit errors at the given mapping complexity. A lower value can be obtained by optimizing the four $\mathcal{C}_1$ mappings and the two $\mathcal{C}_2$ mappings in Fig. 3 separately, but the gain is marginal and asymptotically 0 dB.

there are inferior to those of [35] (one such example is the case $M = 8$, $N = 4$ which is of particular interest to us). We performed sphere-packing simulations of our own up to $M \leq 16$ that verified the reported values from [35]. Moreover, since simulations provide numerical coordinates only, no exact values or geometrical understanding, we identified the underlying structure of the optimal 4D constellations of certain sizes M. Cases where M is a power of 2 or the power efficiency is especially high are summarized in Table I.

Based on all these reported minimum second moments $U(N,M) = \sum_{k=1}^{M} \boldsymbol{c}_k^2$, we computed the asymptotic efficiency $\gamma = (M/(4U)) \log_2 M$, which is the gain relative to BPSK of the best modulation format for a given (N,M). In Fig. 6, we show the results, in terms of spectral efficiency vs. power efficiency (to be exact, sensitivity penalty), which is a common way to compare modulation formats [30], [31]. The optimum constellations (those with the largest γ) are shown for dimensions N from 2 to 4 and for M from 2 to 32. The common formats 8-QAM[4] and 16-QAM have been included for completeness, as well as M-PSK for $M = 5, \ldots, 8$.

A particularly important structure in this context is the *simplex*. The simplex in dimension N is the unique (up to scaling and rotation) set of $M = N + 1$ vectors $\boldsymbol{c}_k$ with the same length and the same pairwise angular separation. For $N = 2$, the simplex is the equilateral triangle, and for $N = 3$ it is the tetrahedron. The power efficiency of the N-dimensional simplex is

$$\gamma_{\text{simp}} = \frac{N+1}{2N} \log_2(N+1), \tag{22}$$

which is always larger than 1 for $N > 1$. Thus, the simplex constellations are always better than the cubic (or BPSK). Indeed, the overall best modulation format in the plane ($N = 2$) occurs for the simplex, i.e., the 3-PSK format, which was suggested for modulation in [43], [44]. It has a $\gamma = (3/4) \log_2 3 = 0.75$ dB sensitivity gain over BPSK. Due to the moderate gain as well as the difficulty of mapping bits to three levels, this format has gained little attention, however. In three dimensions, the most efficient constellation arises also for the simplex ($M = 4$), by placing the signal points on the vertices of a tetrahedron. The 4D simplex is called the *pentachoron* or *pentatope,* and it was discussed in several papers analyzing 4D modulation [16]–[18], [20]. Its power efficiency is $\gamma = (5/8) \log_2 5 = 1.61$ dB, which is nevertheless less than that of the cross-polytope (PS-QPSK), which is the best constellation for $(N,M) = (4,8)$. Moreover, the cross-polytope is the overall most power-efficient constellation of all 4D constellations. This means, perhaps a bit surprising, that $N = 4$ is the lowest dimension where the simplex is *not* the most power-efficient modulation format. Since $(N-1)$-dimensional hyperplanes are embedded in the N-dimensional Euclidean space, the power efficiency cannot decrease with dimension, and we conclude that the PS-QPSK format has the lowest sensitivity of all formats in dimension $N \leq 4$. By the same argument, the simplex is not optimal in any dimension $N \geq 4$.

[4]8-QAM, which in the literature refers to several different constellations, denotes in this paper a regular 3×3 grid minus the central point.

It is also interesting to note that the *kissing configurations*, i.e., the configurations involving the K_N spheres touching the central sphere, are local minima for the power efficiency at $M = K_N + 1$. The reason is probably that these constellations are particularly efficient ways of packing spheres, and they are marked in the figure. The 4D kissing configuration (4,25) is the same as the set $\mathcal{C}_4$ discussed earlier. Although the 6P-QPSK format is based on the kissing configuration (with the central sphere removed), it is not the same as the (4,24) point; in fact, it is slightly below and to the right of this point in Fig. 6. Its bit-to-symbol mapping leads to penalties in both power and spectral efficiency.

As M increases for a given (low) dimension N, the best (densest) packings are known to approach a regular structure called a *lattice*. In two dimensions, the best lattice is generated by placing three circles in a regular triangle (simplex) and extending the pattern indefinitely in all directions. This generates the well-known honeycomb, or hexagonal lattice, usually denoted A_2. Its density is $\Delta(2) = \pi/(2\sqrt{3}) = 0.91$, which means that the circles cover 91% of the plane. The 3D analogy is the face-centered cubic lattice A_3, obtained by extending a regular tetrahedron (3D simplex), with the density $\Delta(3) = \pi/(3\sqrt{2}) = 0.74$. In four dimensions, however, something unexpected happens. Even though a 4D lattice A_4 can be generated from a 4D simplex in perfect analogy with A_2 and A_3, it is not the densest lattice possible. The densest lattice is instead obtained by extending $\mathcal{C}_4$. The lattice, which can be seen as a 4D analogy of the checkerboard pattern, is denoted D_4 and has the density $\Delta(4) = \pi^2/16 = 0.62$.

The power efficiency of a lattice is [33, eq. (32)]

$$\gamma_{\text{lat}} = \log_2(M) \left(1 + \frac{2}{N}\right) \left(\frac{\Delta(N)}{M}\right)^{2/N}, \tag{23}$$

where the densities $\Delta(N)$ are tabulated in [33, Tab. 1.2]. The performance of the densest lattices, A_2, A_3, and D_4, is included as dashed-line asymptotes in Fig. 6.

VI. Implication for Fundamental Sensitivity Limits

We will now discuss how these power-efficient modulation formats will improve the fundamental quantum-limited sensitivities of optical systems.

Consider a coherent optical link limited by ASE noise from N_a in-line amplifiers with gain G and spontaneous emission factor n_{sp}. It has been shown [45] that ASE noise is additive and Gaussian in nature, i.e., that the AWGN model applies to such a system. The optical noise at the receiver has a power spectral density of $N_0 = N_a n_{sp} h\nu (G-1)/G \approx N_a n_{sp} h\nu$ per polarization [11], [14], where $h\nu$ is the photon energy. In a polarization diversity homodyne coherent receiver, the optical amplitude is directly mapped to the electrical signal, so our AWGN results can be interpreted by using $E_b/N_0 = n_b/N_a n_{sp}$, where n_b is the average number of photons per bit. In the limit of a single amplifier with 3 dB noise figure ($N_a = n_{sp} = 1$), this implies that E_b/N_0 has a physically appealing interpretation as the number of photons per bit of the received signal. This can now be used to translate the

TABLE II
PERFORMANCE OF ALL CONSTELLATIONS MENTIONED IN THIS PAPER. FOR COHERENT HOMODYNE RECEPTION IN THE ASE LIMIT, THE SENSITIVITY IN UNITS OF PHOTONS PER BIT EQUALS E_b/N_0. THE SENSITIVITIES FOR 8-PSK, 8-QAM AND 16-QAM ARE OBTAINED FROM REFS [12] AND [14], UNDER ASSUMPTIONS OUTLINED IN THOSE PAPERS. NO BER RESULTS ARE GIVEN FOR $\mathcal{C}_3$ AND $\mathcal{C}_4$, SINCE THESE CONSTELLATIONS DO NOT HAVE A WELL-DEFINED BIT-MAPPING.

Name	Nbr. of pts. M	Nbr. of dims. N	Pow. Eff. γ [dB]	Spectral Eff. [bits/symb/pol]	Sens. at $BER = 10^{-3}$ E_b/N_0 [dB]	Sens. at $BER = 10^{-9}$ E_b/N_0 [dB]
BPSK	2	1	0	2	6.8	12.5
QPSK	4	2	0	2	6.8	12.5
8-PSK	8	2	–3.57	3	10.0	16.2
8-QAM	8	2	–3.01	3	9.0	14.6
16-QAM	16	2	–3.98	4	10.5	16.6
DP-QPSK = $\mathcal{C}_1$	16	4	0	2	6.8	12.5
PS-QPSK = $\mathcal{C}_2$	8	4	1.76	1.5	5.8	11.0
6P-QPSK	$2^{9/2} = 22.6$	4	0.51	2.25	6.9	12.2
$\mathcal{C}_3$	24	4	0.59	2.29	N/A	N/A
$\mathcal{C}_4$	25	4	0.83	2.32	N/A	N/A

results from Fig. 5 to sensitivities (i.e., the number of photons per bit required to get $BER = 10^{-9}$). For BPSK we get the well-known result $E_b/N_0 = 12.5$ dB $= 18$ photons per bit [11], [12]. The most sensitive format, PS-QPSK, improves this with 1.5 dB to 13 photons per bit [23]. 6P-QPSK is with 17 photons per bit slightly better than BPSK. All sensitivities (including other formats discussed in this paper and those at $BER = 10^{-3}$) are found in Table II.

We believe that these relative improvements of PS-QPSK and 6P-QPSK over BPSK will translate also to other coherent optical channels where the AWGN model applies, such as, e.g., the shot-noise limit [13], [14]. Neglecting pulse position modulation (which have been shown to provide unbounded capacity but is impractical in high-speed links [46]), we have shown that the PS-QPSK modulation format gives the best sensitivity in uncoded optical links [23].

To get some real numbers into these sensitivities, we may note that at a bit rate of $1/T = 10$ Gbit/s, one photon per bit equals a received optical power of –59 dBm, and the sensitivity for BPSK in the ASE limit is then 12.5 dB above this, at –46.5 dBm. Recent experiments, based on offline synchronization algorithms, have succeeded in reaching remarkably close, within 4 dB, of this limit [29]. At higher rates, e.g., 100 Gbit/s, the sensitivity power levels become 10 dB higher in absolute power terms. Eventually, at this and higher rates, the nonlinear distortions of optical fibers will limit the BER, and power-efficient modulation formats such as those outlined in this paper may play an important role in improving the performance.

VII. SUMMARY AND CONCLUSIONS

In this paper, we have discussed the asymptotic power efficiency and receiver sensitivities of coherent optical systems. We have found that the conventional quantum-limited sensitivity of 18 photons per bit for BPSK can be surpassed by using more power-efficient modulation formats. This is possible only by jointly optimizing quadrature modulation formats for both polarization components. The geometric interpretation is that a 4D constellation space enables denser sphere packings than two independent planar constellations.

In particular, we presented two modulation formats, PS-QPSK and 6P-QPSK, which have (asymptotically) 1.76 and 0.51 dB higher sensitivity, resp., than conventional BPSK or QPSK systems. The 6P-QPSK constellation utilizes 24 points, which form the vertices of the 24-cell, and this is largest number of constellation points that can be put on a 4D hypersphere without sacrificing power or minimum distance.[5] The 24 levels can be viewed as six different polarization states with four phase levels each. However, the best trade-off between number of levels and transmitted power is given by the 8-level PS-QPSK format, obtained by extending conventional QPSK with the addition of a polarization-switching bit. We also exemplified how transmitters and bit mapping for these modulations formats could be implemented. Table II summarizes the properties of some of the most common modulation formats as well as the ones discussed in this paper.

By using numerically optimized sphere constellations, we computed the best sensitivities of 4D modulation formats up to 32 levels, which resulted in the conclusion that PS-QPSK is the format with the overall best sensitivity, 1.76 dB better than BPSK. We have shown, based on these simulations, that no more power-efficient modulation format is available for fiber communications, unless the dimension is somehow increased. This can be done for example by using error-correcting codes, wavelength/space/time division multiplexing, or different modes in multimode fibers. Other possible generalizations include optimizing constellations when coding or coded modulation is added to the system, or when the SNR is finite rather than asymptotically high. However, these topics are all beyond the scope of this paper.

ACKNOWLEDGEMENTS

This work was financially supported by The Swedish Governmental Agency for Innovation Systems (VINNOVA) within the IKT program, and the Swedish Foundation for Strategic Research (SSF). MK wishes to acknowledge the gracious hospitality of Stojan Radic's group at UCSD, where part of the writing of this paper took place. He also wishes to

[5]Accounting for the bit-to-symbol mapping, only $2^{9/2} = 22.4$ levels are effectively used for transmission.

acknowledge discussions with Martin Sjödin, Mats Sköld and Peter Andrekson at Chalmers, Nicola Alic, Stojan Radic, George Papen and John Proakis at UCSD, as well as some helpful comments from Peter Winzer and Shiva Kumar during OFC 2009.

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*, vol. 42, no. 20, pp. 1175–1176, 2006.

[2] A. Leven, N. Kaneda, U. Koc, and Y. Chen, "Coherent receivers for practical optical communication systems," in *Proceedings of Optical Fiber Communication and National Fiber Optic Engineers Conference, OFC/NFOEC*, 2007, p. OThK4.

[3] A. Leven, N. Kaneda, and Y.-K. Chen, "A real-time CMA-based 10 Gb/s polarization demultiplexing coherent receiver implemented in an FPGA," in *Optical Fiber Communication/National Fiber Optic Engineers Conference, 2008.*, Feb. 2008, p. OTuO2.

[4] H. Sun, K. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Optics Express*, vol. 16, no. 2, pp. 873–879, 2008.

[5] D. Ly-Gagnon, K. Katoh, and K. Kikuchi, "Unrepeatered optical transmission of 20 Gbit/s quadrature phase-shift keying signals over 210 km using homodyne phase-diversity receiver and digital signal processing," *Electronics Letters*, vol. 41, no. 4, pp. 206–207, 2005.

[6] G. Charlet, N. Maaref, J. Renaudier, H. Mardoyan, P. Tran, and S. Bigo, "Transmission of 40 Gb/s QPSK with coherent detection over ultra-long distance improved by nonlinearity mitigation," in *Proceedings of European Conference on Optical Communications, ECOC*, 2006, p. PDP Th.4.3.6.

[7] S. Tsukamoto, D. Ly-Gagnon, K. Katoh, and K. Kikuchi, "Coherent demodulation of 40-Gbit/s polarization-multiplexed QPSK signals with 16-GHz spacing after 200-km transmission," in *Proceedings of Optical Fiber Communication and National Fiber Optic Engineers Conference, OFC/NFOEC*, vol. 6, 2005, p. PDP29.

[8] 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*, vol. 43, no. 20, pp. 1109–1111, 2007.

[9] J. Renaudier, G. Charlet, M. Salsi, O. Pardo, H. Mardoyan, P. Tran, and S. Bigo, "Linear fiber impairments mitigation of 40-Gbit/s polarization-multiplexed QPSK by digital processing in a coherent receiver," *Journal of Lightwave Technology*, vol. 26, no. 1, pp. 36–42, 2008.

[10] G. Jacobsen, *Noise in Digital Optical Transmission Systems*. Artech House Publishers, 1994.

[11] L. Kazovsky, S. Benedetto, and A. Willner, *Optical Fiber Communication Systems*. Artech House Publishers, 1996.

[12] J. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 10, no. 2, pp. 259–272, Mar.–Apr. 2004.

[13] K.-P. Ho, *Phase-Modulated Optical Communication Systems*. Springer, 2005.

[14] E. Ip, A. Lau, D. Barros, and J. Kahn, "Coherent detection in optical fiber systems," *Optics Express*, vol. 16, no. 2, pp. 753–791, 2008, erratum vol. 16, no. 26, p. 21943, 2008.

[15] D. Saha and T. Birdsall, "Quadrature-quadrature phase-shift keying," *IEEE Transactions on Communications*, vol. 37, no. 5, pp. 437–448, May 1989.

[16] G. Welti and J. Lee, "Digital transmission with coherent four-dimensional modulation," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 497–502, 1974.

[17] L. Zetterberg and H. Brändström, "Codes for combined phase and amplitude modulated signals in a four-dimensional space," *IEEE Transactions on Communications*, vol. 25, no. 9, pp. 943–950, 1977.

[18] G. Taricco, E. Biglieri, and V. Castellani, "Applicability of four-dimensional modulations to digital satellites: a simulation study," in *Proceedings of IEEE Global Telecommunications Conference, GLOBECOM*, vol. 4, Nov.–Dec. 1993, pp. 28–34.

[19] S. Betti, F. Curti, G. De Marchis, and E. Iannone, "Exploiting fibre optics transmission capacity: 4-quadrature multilevel signalling," *Electronics Letters*, vol. 26, no. 14, pp. 992–993, July 1990.

[20] ——, "A novel multilevel coherent optical system: 4-quadrature signaling," *Journal of Lightwave Technology*, vol. 9, no. 4, pp. 514–523, Apr. 1991.

[21] S. Betti, G. De Marchis, E. Iannone, and P. Lazzaro, "Homodyne optical coherent systems based on polarization modulation," *Journal of Lightwave Technology*, vol. 9, no. 10, pp. 1314–1320, Oct. 1991.

[22] R. Cusani, E. Iannone, A. Salonico, and M. Todaro, "An efficient multilevel coherent optical system: M-4Q-QAM," *Journal of Lightwave Technology*, vol. 10, no. 6, pp. 777–786, 1992.

[23] M. Karlsson and E. Agrell, "Which is the most power-efficient modulation format in optical links?" *Optics Express*, vol. 17, no. 13, pp. 10 814–10 819, 2009. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-17-13-10814

[24] S. Benedetto and P. Poggiolini, "Theory of polarization shift keying modulation," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 708–721, Apr. 1992.

[25] S. Benedetto, A. Djupsjöbacka, B. Lagerström, R. Paoletti, P. Poggiolini, and G. Mijic, "Multilevel polarization modulation using a specifically designed $LiNbO_3$ device," *IEEE Photonics Technology Letters*, vol. 6, no. 8, pp. 949–951, Aug. 1994.

[26] H. Zhao, E. Agrell, and M. Karlsson, "Unequal bit error probability in coherent QPSK fiber-optic systems using phase modulator based transmitters," *European Transactions on Telecommunications*, vol. 19, pp. 895–906, 2007.

[27] H. Zhao, M. Karlsson, and E. Agrell, "Transmitter comparison and unequal bit error probabilities in coherent QPSK systems," in *Proceedings of Optical Fiber Communication and National Fiber Optic Engineers Conference, OFC/NFOEC*, 2007, p. OTuH2.

[28] H. Zhao, E. Agrell, and M. Karlsson, "Intersymbol Interference in DQPSK Fiber-Optic Systems," *to appear in European Transactions on Telecommunications*, 2009.

[29] K. Kikuchi and S. Tsukamoto, "Evaluation of sensitivity of the digital coherent receiver," *Journal of Lightwave Technology*, vol. 26, no. 13, pp. 1817–1822, 2008.

[30] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Kluwer Academic Publishers, 1999.

[31] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2001.

[32] H. Bülow, "Polarization QAM Modulation (POL-QAM) for Coherent Detection Schemes," in *Proceedings of Optical Fiber Communication and National Fiber Optic Engineers Conference, OFC/NFOEC*, 2009, p. OWG2.

[33] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*. Springer, 1999.

[34] O. Musin, "The kissing number in four dimensions," *Annals of Mathematics*, vol. 168, pp. 1–32, 2008.

[35] N. J. A. Sloane, R. H. Hardin, T. S. Duff, and J. H. Conway. (1997) Minimal-energy clusters, library of 3-d clusters, library of 4-d clusters. [Online]. Available: http://www.research.att.com/∼njas/cluster/

[36] R. L. Graham and N. J. A. Sloane, "Penny-packing and two-dimensional codes," *Discrete and Computational Geometry*, vol. 5, no. 1, pp. 1–11, 1990.

[37] M. Simon, S. Hinedi, and W. Lindsey, *Digital Communication Techniques: Signal Design and Detection*. PTR Prentice Hall, 1995.

[38] L. Xiao and X. Dong, "New analytical expressions for orthogonal, biorthogonal, and transorthogonal signaling in Nakagami fading channels with diversity reception," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1418–1424, July 2005.

[39] M. K. Simon and R. Annavajjala, "On the optimality of bit detection of certain digital modulations," *IEEE Transactions on Communications*, vol. 53, no. 2, pp. 299–307, Feb. 2005.

[40] G. Foschini, R. Gitlin, and S. Weinstein, "Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise," *IEEE Transactions on Communications*, vol. 22, no. 1, pp. 28–38, 1974.

[41] N. J. A. Sloane, R. H. Hardin, T. S. Duff, and J. H. Conway, "Minimal-energy clusters of hard spheres," *Discrete and Computational Geometry*, vol. 14, no. 3, pp. 237–259, 1995.

[42] J.-E. Porath and T. Aulin, "Design of multidimensional signal constellations," *IEE Proceedings-Communications*, vol. 150, no. 5, pp. 317–323, Oct. 2003.

[43] J. R. Pierce, "Comparison of three-phase modulation with two-phase and four-phase modulation," *IEEE Transactions on Communications*, vol. COM-28, no. 7, pp. 1098–1099, July 1980.

[44] N. Ekanayake and T. Tjhung, "On ternary phase-shift keyed signaling," *IEEE Transactions on Information Theory*, vol. IT-28, no. 4, pp. 658–660, July 1982.

[45] J. P. Gordon, L. R. Walker, and W. H. Louisell, "Quantum statistics of masers and attenuators," *Phys. Rev.*, vol. 130, no. 2, pp. 806–812, Apr. 1963.

[46] J. R. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Transactions on Communications*, vol. 26, no. 12, pp. 1819–1821, Dec. 1978.

PLACE PHOTO HERE

Erik Agrell received the M.S. degree in electrical engineering in 1989 and the Ph.D.degree in information theory in 1997, both from Chalmers University of Technology, Sweden.

From 1988 to 1990, he was with Volvo Technical Development as a Systems Analyst, and from 1990 to 1997, with the Department of Information Theory, Chalmers University of Technology, as a Research Assistant. In 1997–1999, he was a Postdoctoral Researcher with the University of Illinois at Urbana-Champaign and the University of California, San Diego. In 1999, he joined the faculty of Chalmers University of Technology, first as an Associate Professor and since 2009 as a Professor in Communication Systems. His current research interests include coding, modulation, and equalization for fiber-optic channels, bit-interleaved coded modulation and multilevel coding, bit-to-symbol mappings in coded and uncoded systems, and multidimensional geometry.

Prof. Agrell served as Publications Editor for IEEE Transactions on Information Theory from 1999 to 2002.

PLACE PHOTO HERE

Magnus Karlsson received his Ph.D. in Electromagnetic Field Theory in 1994 from Chalmers University of Technology, Gothenburg, Sweden. The title of his Ph.D thesis was "Nonlinear propagation of optical pulses and beams". Since 1995, he has been with the Photonics Laboratory at Chalmers, first as assistant professor and since 2003 as professor in photonics. He has authored or co-authored over 170 scientific journal and conference contributions, served as guest editor for the Journal of Lightwave Technology, and holds two patents. He has served in the technical committees for the Optical Fiber Communication Conference (OFC) (2009 as subcommittee chair), the Asia-Pacific Optical Communications Conference (APOC), the Optical Amplifiers and Applications (OAA), and the Nonlinear Photonics (NP) meetings.

His research has been devoted to a variety of aspects of fiber optic communication systems, in particular transmission effects such as fiber nonlinearities and polarization effects, but also applied issues such as high-capacity data transmission and all-optical switching. Currently he is devoted to parametric amplification, multilevel modulation formats, and coherent transmission in optical fibers.