Survey on Free Space Optical Communication: A Communication Theory Perspective

Mohammad Ali Khalighi*, Murat Uysal**

* Institut Fresnel, UMR CNRS 7249, École Centrale Marseille, Marseille, France
E-mail: Ali.Khalighi@fresnel.fr
** Department of Electrical and Electronics Engineering, Özyeğin University, Istanbul, Turkey, 34794
E-mail: Murat.Uysal@ozyegin.edu.tr

Abstract—Optical wireless communication (OWC) refers to transmission in unguided propagation media through the use of optical carriers, i.e., visible, infrared (IR), and ultraviolet (UV) bands. In this survey, we focus on outdoor terrestrial OWC links which operate in near IR band. These are widely referred to as free space optical (FSO) communication in the literature. FSO systems are used for high rate communication between two fixed points over distances up to several kilometers. In comparison to radio-frequency (RF) counterparts, FSO links have a very high optical bandwidth available, allowing much higher data rates. They are appealing for a wide range of applications such as metropolitan area network (MAN) extension, local area network (LAN)-to-LAN connectivity, fiber back-up, backhaul for wireless cellular networks, disaster recovery, high definition TV and medical image/video transmission, wireless video surveillance/monitoring, and quantum key distribution among others.

Despite the major advantages of FSO technology and variety of its application areas, its widespread use has been hampered by its rather disappointing link reliability particularly in long ranges due to atmospheric turbulence-induced fading and sensitivity to weather conditions. In the last five years or so, there has been a surge of interest in FSO research to address these major technical challenges. Several innovative physical layer concepts, originally introduced in the context of RF systems, such as multiple-input multiple-output communication, cooperative diversity, and adaptive transmission have been recently explored for the design of next generation FSO systems. In this paper, we present an up-to-date survey on FSO communication systems. The first part describes FSO channel models and transmitter/receiver structures. In the second part, we provide details on information theoretical limits of FSO channels and algorithmic-level system design research activities to approach these limits. Specific topics include advances in modulation, channel coding, spatial/cooperative diversity techniques, adaptive transmission, and hybrid RF/FSO systems.

I. INTRODUCTION

A. Overview of Optical Wireless Communication

The proliferation of wireless communications stands out as one of the most significant phenomena in the history of technology. Wireless devices and technologies have become pervasive much more rapidly than anyone could have imagined thirty years ago and they will continue to be a key element of modern society for the foreseeable future. Today, the term “wireless” is used almost synonymously with radio-frequency (RF) technologies as a result of the wide-scale deployment and utilization of wireless RF devices and systems. The RF band of the electromagnetic spectrum is however fundamentally limited in capacity and costly since most sub-bands are exclusively licensed. With the ever-growing popularity of data-heavy wireless communications, the demand for RF spectrum is outstripping supply and the time has come to seriously consider other viable options for wireless communication using the upper parts of the electromagnetic spectrum.

Optical wireless communication (OWC) refers to transmission in unguided propagation media through the use of optical carriers, i.e., visible, infrared (IR) and ultraviolet (UV) band. Signalling through beacon fires, smoke, ship flags and semaphore telegraph [1] can be considered the historical forms of OWC. Sunlight has been also used for long distance signaling since very early times. The earliest use of sunlight for communication purposes is attributed to ancient Greeks and Romans who used their polished shields to send signals by reflecting sunlight during battles [2]. In 1810, Carl Friedrich Gauss invented the heliograph which involves a pair of mirrors to direct a controlled beam of sunlight to a distant station. Although the original heliograph was designed for geodetic survey, it was used extensively for military purposes during the late 19th and early 20th century. In 1880, Alexander Graham Bell invented the photophone, known as the world’s first wireless telephone system [1]. It was based on the voice-caused vibrations on a mirror at the transmitter. The vibrations were reflected and projected by sunlight and transformed back into voice at the receiver. Bell referred to the photophone as “the greatest invention [he had] ever made, greater than the telephone” [3], but it never came out as a commercial product. The military interest on photophone however continued. For example, in 1935, the German Army developed a photophone where a tungsten filament lamp with an IR transmitting filter was used as a light source. Also, American and German military laboratories continued the development of high pressure arc lamps for optical communication until the 1950s [4].

In modern sense, OWC uses either lasers or light emitting diodes (LEDs) as transmitters. In 1962, MIT Lincoln Labs built an experimental OWC link using a light emitting GaAs diode and was able to transmit TV signals over a distance of 30 miles. After the invention of laser, OWC was envisioned to be the main deployment area for lasers and many trials.
were conducted. In fact, just months after the first public announcement of the working laser on July 1960, Bell Labs scientists were able to transmit signals 25 miles away using a ruby laser [5]. A comprehensive list of OWC demonstrations performed during 1960-1970 using different types of lasers and modulation schemes can be found in [6]. However, the results were in general disappointing due to large divergence of laser beams and the inability to cope with atmospheric effects. With the development of low-loss fiber optics in the 1970’s, they became the obvious choice for long distance optical transmission and shifted the focus away from OWC systems.

Over the decades, the interest in OWC remained mainly limited to covert military applications [7], [8] and space applications including inter-satellite and deep-space links3. OWC’s mass market penetration has been so far limited with the exception of IrDA which became a highly successful wireless short-range transmission solution [16]. With the growing number of companies offering terrestrial OWC links in recent years and the emergence of visible light communication (VLC) products [17]–[23], the market has begun to show future promise [24], [25]. Development of novel and efficient wireless technologies for a range of transmission links is essential for building future heterogeneous communication networks to support a wide range of service types with various traffic patterns and to meet the ever-increasing demands for higher data rates. Variations of OWC can be potentially employed in a diverse range of communication applications ranging from optical interconnects within integrated circuits through outdoor inter-building links to satellite communications. Based on the transmission range, OWC can be studied in five categories (see Fig. 1 for some application examples):

1) Ultra-short range OWC, e.g., chip-to-chip communications in stacked and closely-packed multi-chip packages [26]–[29].
2) Short range OWC, e.g., wireless body area network (WBAN) and wireless personal area network (WPAN) applications [30], underwater communications [31], [32].
3) Medium range OWC, e.g., indoor IR and VLC for wireless local area networks (WLANs) [22], [33], [34], inter-vehicular and vehicle-to-infrastructure communications [35], [36].
4) Long range OWC, e.g., inter-building connections.
5) Ultra-long range OWC, e.g., inter-satellite links [37], deep space links [14].

In this survey, we focus only on outdoor terrestrial OWC links (i.e., the fourth category), which are also widely referred to as free space optical (FSO) communication in the literature. This terminology will be adopted hereinafter.

B. Advantages and Applications of FSO

FSO systems are used for high rate communication between two fixed points over distances up to several kilometers. In comparison to RF counterparts, the FSO link has a very high optical bandwidth available, allowing much higher data rates. Terrestrial OWC products with transmission rates of 10 Gbps are already in the market [38] and the speeds of recent experimental OWC systems are competing with fiber optic [39]–[43]. FSO systems use very narrow laser beams. This spatial confinement provides a high reuse factor, an inherent security, and robustness to electromagnetic interference. Furthermore, the frequency in use by the FSO technology is above 300 GHz which is unlicensed worldwide. Therefore, FSO systems do not require license fees [44]. FSO systems are also easily deployable and can be reinstalled without the cost of dedicated fiber optic connections.

FSO systems have initially attracted attention as an efficient solution for the “last mile” problem to bridge the gap between the end user and the fiber optic infrastructure already in place. Telecom carriers have already made substantial investments to augment the capacity of their fiber backbones. To fully utilize the existing capacity, and therefore to generate revenue, this expansion in the backbone of the networks should be accompanied by a comparable growth at the network edge where end users get access to the system. FSO systems are also appealing for a wide range of applications some of which are elaborated in the following [44]–[47] (see Fig. 2).

- **Enterprise/campus connectivity:** Today’s corporations and school/university campuses are experiencing a heterogeneous network traffic (i.e., voice, data, fax, multimedia traffic) that is overwhelming the typical connections. FSO systems can bridge multiple buildings in corporate and campus networks supporting ultra-high speeds without the cost of dedicated fiber optic connections.

- **Video surveillance and monitoring:** Surveillance cameras are widely deployed in commercial, law enforcement, public safety, and military applications. Wireless video is convenient and easy to deploy, but conventional wireless technologies fail to provide high throughput requirements for video streams. FSO technology presents a powerful alternative to support high quality video transmission.

- **Back-haul for cellular systems:** Wireline connections such as T1/E1 leased lines and microwave links are typically deployed between the base stations and the mobile switching center in a cellular system. The growing number of bandwidth-intensive mobile phone services
now requires the deployment of technologies such as FSO which allow much higher throughput.

- **Redundant link and disaster recovery:** Natural disasters, terrorist attacks, and emergency situations require flexible and innovative responses. Temporary FSO links can be readily deployed within hours in such disaster situations in which local infrastructure could be damaged or unreliable. A tragic example of the FSO deployment efficiency as a redundant link was witnessed after 9/11 terrorist attacks in New York City. FSO links were rapidly deployed in this area for financial corporations which were left out with no landlines.

- **Security:** Today’s cryptosystems are able to offer only computational security within the limitations of conventional computing power and the realization of quantum computers would, for example, make electronic money instantly worthless. Based on the firm laws of physics, quantum cryptography provides a radically different solution for encryption and promises unconditional security. Quantum cryptography systems are typically considered in conjunction with fiber optic infrastructure. FSO links provide a versatile alternative in cases where the fiber optic deployment is costly and/or infeasible.

- **Broadcasting:** In broadcasting of live events such as sports and ceremonies or television reporting from remote areas and war zones, signals from the camera (or a number of cameras) need to be sent to the broadcasting vehicle which is connected to a central office via satellite uplink. The required high-quality transmission between the cameras and the vehicle can be provided by a FSO link. FSO links are capable of satisfying even the most demanding throughput requirements of today’s high-definition television (HDTV) broadcasting applications. For example, during 2010 FIFA World Cup, UK TV station BBC deployed FSO links for Ethernet-based transport of high definition video between temporary studio locations set up in Cape Town, South Africa.

Currently, there are several companies which are working on the design and manufacturing of FSO systems as outdoor wireless transmission solutions such as Canon (Japan), Cassidian (Germany), FSONA (Canada), GeoDesy (Hungary), Laser ITC (Russia), LightPointe Communications (USA), MRV (USA), Northern Hi-Tec (UK), Novasol (USA), Omnitek (Turkey), Plaintree Systems (Canada), and Wireless Excellence (UK) among others.

### II. FSO Channel Modeling

The optical power launched from the transmitter is affected by various factors before arriving at the receiver. These include system loss, geometric loss, misalignment loss, atmospheric loss, atmospheric turbulence induced fading, and ambient noise. The system loss highly depends on the design specifications and is usually specified by the manufacturers. Details on the system loss can be found in [48]. In the following, we provide further details on the other factors.
A. Geometric and Misalignment Losses

The geometric loss is due to the divergence of the beam when propagating through the atmosphere. It can be calculated given the divergence angle, the link distance, and the receiver lens aperture size. In calculating the geometric loss, an important factor is the optical wave propagation model. For horizontal FSO transmissions, a good approximation is to consider a Gaussian profile for the beam intensity. When a Gaussian beam has a relatively large divergence, its statistical properties are close to the case of a point source. In such a case, the approximations of plane or spherical wave can effectively be used.

The degree of beam divergence also affects transmitter-receiver alignment and beam tracking at the receiver. Misalignment occurs in practice mostly due to beam wander, building sway, or errors in the tracking system. Beam wander is the result of inhomogeneities of large-scale atmosphere eddies that cause random deflections of the optical beam, and as a result, the beam deviates from its original path. This phenomenon is in particular important for long distance paths. On the other hand, building sway is the result of a variety of factors, including thermal expansion, wind loads, small earthquakes, and vibrations. Because of the narrowness of the transmitted beam and the usually small receiver field of view (FOV), building sway can effectively cause a communication interrupt.

When no tracking mechanism is used at the receiver side, which is typically the case for entry model FSO links with a range of several hundred meters, the misalignment loss can be alleviated by increasing the beam divergence at the transmitter. The use of spatially partially coherent Gaussian beams has been further proposed in [59]–[61] to mitigate the misalignment-induced pointing errors. It was shown in [57] that beam optimization allows significant gains in the channel capacity. Similar studies [54], [62] showed that the transmitter beam radius can be optimized to maximize the average link capacity and to minimize the outage probability (see Section IV). Deployment of variable wavelengths by using quantum cascade lasers is also proposed in [63] to mitigate the effect of building sway. For long distances (i.e., more than one kilometer), as a narrower beam should be used to avoid suffering from a high geometric loss, the use of automatic pointing and tracking at the receiver becomes necessary to remove or reduce the effects of pointing errors.

Statistical modeling of the pointing errors and its impact on the system performance has been studied in several recent works. Under the assumption that the building sway statistics follow an independent Gaussian distribution for elevation and for horizontal directions, the radial pointing error angle is modeled by a Rayleigh distribution in [56], [63], [64]. The combined effect of pointing errors and atmospheric turbulence (see Subsection II-C) has been further studied in several works. In [57], it is proposed to consider the random attenuation of the channel as the product of path loss, geometric spread and pointing errors, and atmospheric turbulence. Also, considering a Gaussian beam profile and Rayleigh distributed radial displacement at the receiver, a statistical model is derived for the misalignment loss that takes the detector size, beam width, and jitter variance into account. The same model was used in [65] to study the effect of pointing errors on the FSO link capacity. Also, an analytical expression for the average bit-error-rate (BER) is derived in [66], [67], and the performance of coded FSO links is studied in [68]. The effect of pointing errors on the performance of space-diversity and relayed FSO systems (see Sections VII and IX) has also been considered in [69]–[71] and [72], respectively.
B. Atmospheric Loss

The physics and the transmission properties of the radiation penetrating the atmosphere are very similar in the visible and the near-IR wavelength ranges. Therefore, visibility can be used to characterize particles that absorb or scatter light for near-IR radiations as well. The particles affecting the visibility include rain, snow, fog, but also pollution, dust, aerosols, smoke, etc. They absorb to some degree the laser light energy, causing an attenuation of the optical power. In near-IR, absorption occurs primarily due to water particles [47], [73]–[76]. They cause light scattering, which is the deflection of incident light from its initial direction, causing spatial, angular, and temporal spread. For rain and snow, the size of the particles is much larger than the wavelength, and consequently, the FSO transmission is relatively unaffected [77]. In the case where FSO systems are deployed in metropolitan areas over distances less than 1 km, typical rain attenuation values are typically on the order of 3 dB/km. Only for very severe rain, the attenuation can become an issue in deployments beyond the distance scale of a typical metropolitan area [74], [78]. For snow, the attenuation can be more severe than rain due to a much larger droplet size. In fact, the impact of light snow to blizzard falls approximately between light rain to moderate fog (see below) [74], [79].

When the particle diameter is on the order of the wavelength, the resulting scattering coefficient is very high. That is why the most detrimental environmental conditions are fog and haze [47], [73], [74], [80] as they are composed of small particles with radii close to the near-IR wavelengths. Even modest fog conditions can highly attenuate IR signals over shorter distances. Experimental tests have reported about 90% loss in the transmit power over a distance of 50 m in moderate fog [74]. Channel modeling for FSO communication through fog is studied in [58], [77], [81]. The experimental measurements in [80] revealed that the atmospheric attenuation is almost independent of the wavelength between 785 and 1550 nm for fog, but it is wavelength dependent in haze conditions [80]. Typically, haze particles have a size between 0.01 and 1 μm, whereas fog droplets have radius between 1 to 20 μm, and hence, the beam light suffers from less attenuation in haze conditions [80]. Also, different scatterer sizes result in wavelength dependence of light extinction in haze and dense fog conditions [82]. A detailed analysis based on the Mie scattering theory is presented in [82] where a wavelength dependent model for the attenuation coefficient is proposed for fog and haze situations.

An interesting point to note is that RF wireless technologies that use frequencies above approximately 10 GHz are adversely impacted by rain and little impacted by fog [74], [78]. This motivates the design of hybrid RF/FSO systems which will be later discussed in Section X.

An important consideration in FSO channel modeling is the channel coherence bandwidth which is defined as the inverse of the channel delay spread [83]. Whereas under clear weather conditions, the FSO channel has a negligible delay spread [84], fog, moderate cloud, and rain can potentially result in temporal broadening of optical pulses. This, in turn, results in inter-symbol interference (ISI) and degrades the system performance [85]. However, given the typical data rates of FSO links, the channel delay spread as a result of beam scattering due to fog or rain is practically negligible. This is shown recently in [86] where numerical Monte Carlo-based simulations are used to quantify the channel root mean square (RMS) delay spread. In particular, the RMS delay spread due to rain under realistic conditions is less than 10 picoseconds for a 1 km link. Also, under moderate and dense fog, the delay spread is typically limited to 50 picoseconds [86]. Consequently, in any case, the channel can effectively be considered as frequency non-selective, introducing no ISI.

C. Atmospheric Turbulence Induced Fading

Under clear atmosphere conditions, the atmospheric loss associated with visibility is negligible, but we are faced to another adverse effect known as scintillation or fading. Inhomogeneities in the temperature and the pressure of the atmosphere, caused by solar heating and wind, lead to the variations of the air refractive index along the transmission path [51], [87]. The resulting atmospheric turbulence causes random fluctuations in both the amplitude and the phase of the received signal, i.e., channel fading. This results in a considerable degradation of the system performance, especially in long-distance transmissions of about several kilometers [51], [88].

A comprehensive study of turbulence modeling for terrestrial FSO links can be found in [51]. Atmospheric turbulence is mainly characterized by three parameters: the inner and the outer scales of turbulence denoted respectively by \( l_0 \) and \( L_0 \), and the index of refraction structure parameter \( C_n^2 \), sometimes called the turbulence strength [88]. According to the Kolmogorov theory, \( L_0 \) is the largest cell size before the energy is injected into a region and \( l_0 \) is associated with the smallest cell size before energy is dissipated into heat [51], [90]. The energy distribution of the turbulence cells can be described by the spatial power spectrum of refractive-index fluctuations. Kolmogorov and Tarasci models are two spectra that are usually considered [87]. For moderate to strong turbulence regimes, a modified spectrum is used by considering two spatial filters which remove the contribution of the turbulent eddies of size between the coherence radius and the scattering disc [51], [91]. Usually, the outer scale is approximated as \( L_0 \to \infty \) as it has a negligible impact on turbulence in practice [88]. On the other hand, the inner scale \( l_0 \) has a significant impact on the turbulence [92]; in particular, larger values of \( l_0 \) result in a higher irradiance variance in the strong turbulence regime [93], [94].

The refraction structure parameter \( C_n^2 \) is altitude dependent and is larger at lower altitudes due to the more significant heat transfer between the air and the surface [88]. In general, it also depends on the link distance [95]. However, usually the conditions of homogeneous turbulence are considered in terrestrial FSO systems and it is assumed that \( C_n^2 \) does not depend on distance. Typical values for \( C_n^2 \) vary from \( 10^{-17} \) to \( 10^{-14} \) for turbulence modeling in over-water and coastal environments can be found in [89].
10^{-13} \text{ m}^{-2/3} \text{ [96]. Its variations can be extremely important during daytime at a given location that can attain four orders of magnitude [97]. On the other hand, it becomes almost constant at night [98] and its dependence on height decreases, compared with daytime [97]. At near ground level, \( C_n^2 \) has its peak value during midday hours whereas its minima occur near sunrise and sunset [98]. An important question is how the meteorological conditions affect the refraction structure parameter. In [99], experimental models were proposed to predict \( C_n^2 \) according to the weather forecast. The performed measurements show that scintillation is affected by aerosols, particularly when their total cross-sectional area is relatively large. Similar studies are presented in [89], [96], and tables reporting \( C_n^2 \) for different weather conditions can be found in [73], [100].

To quantify the fluctuations resulting from atmospheric turbulence, the scintillation index (SI) is frequently used in the literature. It is defined as \( \sigma_I^2 = E(I^2)/E(I)^2 - 1 \) [101], where \( I \) is the intensity of the received optical wave and \( E(\cdot) \) denotes the expected value. While SI provides a characterization of the turbulence strength based on the first and the second moments of the intensity, full statistical characterization has been further investigated in the literature and several statistical channel models have been proposed for the distribution of turbulence-induced fading in FSO systems. The most widely accepted model under weak turbulence conditions is the log-normal model. This model was derived based on the first-order Rytov approximation several decades ago [51], [101], [102]. It applies to the FSO systems deployed over relatively short ranges in urban areas and has been considered in several works such as [103]–[105]. However, experimental data over long propagation paths have shown that the log-normal model is not appropriate for moderate-to-strong turbulence regime [88], [106]–[110]. The negative exponential distribution is a limit distribution for the intensity in the saturation regime [110] and is used in several works considering strong turbulence conditions [105], [111]–[114]. The Rayleigh distribution has been used in [115] to model limiting cases of severe atmospheric turbulence. The K distribution, originally proposed as not appropriate for moderate-to-strong turbulence propagation paths has shown that the log-normal model is still appropriate for moderate turbulence conditions with the log-normal Rice (also known as Beckmann) have been further proposed [110], [119]–[121]. In particular, the log-normal, log-normal exponential and exponential distributions can be considered as special cases of the log-normal Rice model [122]. Another doubly-stochastic scintillation model is the Gamma-Gamma distribution [51], [110] which has gained a wide acceptance in the current literature. In the Gamma-Gamma model, the received intensity \( I \) is considered as the product of two independent Gamma random variables \( X \) and \( Y \), which represent the irradiance fluctuations arising from large- and small-scale turbulence, respectively. The PDF of \( I \) is:

\[
p(I) = \frac{2(ab)^{(a+b)/2}}{\Gamma(a) \Gamma(b)} I^{(a+b)/2-1} K_{a-b}(2\sqrt{ab}I), \quad I > 0,
\]

where the parameters \( a \) and \( b \) represent the effective numbers of large- and small-scale turbulence cells, and \( \Gamma(\cdot) \) is the Gamma function. Also, the SI by this model is given by \( \sigma_I^2 = \frac{1}{b} + \frac{1}{a} + \frac{1}{\alpha} \).

The Double Weibull distribution is another doubly-stochastic model for atmospheric turbulence channels that has been shown to be more accurate than the Gamma-Gamma model, particularly for the cases of moderate and strong turbulence [123]. The M-distribution is another recent model that includes most of the already proposed statistical models, e.g. K and Gamma-Gamma, as special cases [124], [125]. One of the latest attempts in atmospheric turbulence modeling based on the doubly stochastic theory is reported in [126] which proposes the Double Generalized Gamma (Double GG) model, which is slightly more accurate than Double Weibull. The superiority of Double GG over Gamma-Gamma is particularly obvious in the strong turbulence when considering the spherical wave propagation model, as well as in the moderate turbulence regime considering plane wave propagation.

In addition to modeling the intensity fluctuations, an important point is the temporal characterization of turbulence. In most practical cases, the channel fading is very slowly varying and the channel coherence time is typically 0.1 to 10 ms [44]. As in FSO systems we are concerned with very high transmission rates on the order of several tens of Mbps to several Gbps, the channel fading coefficient remains constant over thousands up to millions of consecutive bits. Therefore, the quasi-static channel fading model [83] applies to FSO links.

As implicitly mentioned above, the beam (wave) model can also impact the effect of atmospheric turbulence. General beam types, namely Gaussian, cos-Gaussian, cosh-Gaussian, and annular beams are compared in [127]. The three latter can be considered as general beam shapes, which can reduce to simpler models such as plane and spherical propagation models or the classical Gaussian beam model by setting some specific parameters [128], [129]. It is shown that for small source sizes and when transmitting over long propagation distances, the best performance is obtained for annular beams [127], [130]. On the other hand, for relatively large source sizes and when transmitting over short propagation distances,
the best performance is achieved using cos-Gaussian beams. Furthermore, higher-order beams provide better performances than the zero-order beams at longer propagation distances [127]. In [131], the flat-topped Gaussian beam is studied, which can be represented as a superposition of several Gaussian beams of different scales. It is shown in particular that the turbulence effect reduces by using flat-topped Gaussian beams, compared to single Gaussian beams, for source sizes much larger than the first Fresnel zone [131]. However, except for very small and very large source sizes, the effect of turbulence increases by increasing the number of Gaussian beams used for flattening out the overall beam profile [132].

D. Background Radiation

Last but not least, background radiation, also called background noise or ambient noise, can degrade the performance of FSO links. In fact, in addition to the useful signal, the receiver lens also collects some undesirable background radiations that may consist of direct sunlight, reflected sunlight, or scattered sunlight from hydrometeor or other objects [48], [133]–[136]. Their effect can be reduced by means of narrow spectral bandpass and spatial filtering, prior to photo-detection. Nevertheless, a non-negligible background noise may fall within the spatial and frequency ranges of the detector that can limit the system performance by causing a variable offset in the converted electrical signal [134]. This, in turn, results in a reduced signal-to-noise ratio (SNR) [137] and effective receiver sensitivity [135]. In some circumstances, background radiation can even cause link outages because of the saturation of the receiver [134].

In the (theoretical) case of a diffraction-limited receiver, the received background noise level is independent of the receiver aperture size [133], [138]. In practice, an FSO receiver uses a lens and a photo-detector of a given size and, hence, has a FOV much larger than the diffraction limit. In fixed FOV receivers, the background noise power is proportional to the receiver pupil area [133]. Experimental measurements indicate that while the received optical signal power is typically about 10 mW for direct sunlight [138], up to about 10 mW for direct sunlight [136]. This latter case can statistically occur less than 1 hour per year, however.

Background noise can be statistically modeled by a Poisson random process [134], [139]. When the background radiation level is relatively high, the average number of the corresponding received photons is large enough to allow the approximation of the Poisson distribution by a Gaussian distribution [139]. Since the mean value of the background noise is rejected by the ac-coupled receiver circuitry, the noise has zero mean. Furthermore, the contributions from the interaction of the signal with background radiations due to the non-linear characteristic of the photo-detector [133] can practically be neglected [140], and a signal-independent Gaussian model can be used.

III. FSO TRANSCEIVER

In an FSO communication system, a source produces information waveforms which are then modulated onto an optical carrier. The generated optical field is radiated through the atmosphere towards a remote destination. At the receiver, the field is optically collected and a photo-detector transforms the optical field to an electrical current. The receiver processes the detected electrical current to recover the original transmitted information. Current FSO systems typically operate in the near-IR wavelengths, i.e., from 750 to 1600 nm. Although the (clear) atmosphere is considered as highly transparent in the near-IR wavelength range, certain wavelengths can experience severe absorption due to the presence of different molecules in the atmosphere [48]. For some special wavelength windows, located around four specific wavelengths of 850, 1060, 1250, and 1550 nm, an attenuation of less than 0.2 dB/km is experienced [141]. Interestingly, the 850 and 1550 nm windows coincide with the standard transmission windows of fiber communication systems. That is why most of commercially available FSO systems operate at these two windows so as to use the corresponding available off-the-shelf components. Other wavelengths such as 10 μm [48], [142] and UV wavelengths [143] have been recently considered for FSO systems. The 10 μm wavelength is known to have better fog transmission characteristics [48]. UV transmissions, on the other hand, are more robust against pointing errors and beam blockage and have a lower sensitivity to solar and other background interferences [143].

A. Transmitter

As illustrated in Fig. 3, the transmitter consists of an optical source, a modulator, an optical amplifier (if required), and beam forming optics. Channel coding can be optionally used before modulation (see Section VI). Data bits from the information source are first encoded, then modulated. The modulated laser beam is then passed through the optical amplifier to boost the optical intensity. The light beam is collected and refocused by means of beam forming optics before being transmitted.

The typical optical source in FSO systems is a semiconductor laser diode (LD) [34], although some manufacturers use high power LEDs with beam collimators [144]. The optical source should deliver a relatively high optical power over a wide temperature range. Moreover, it should have a long mean time between failures (MTBF) and the corresponding components should be small in footprint and have low power consumption [48], [141]. Consequently, vertical-cavity surface-emitting lasers (VCSEL) are mostly used for operation

![Fig. 3. The general block diagram of the transmitter.](image-url)
around 850 nm, and Fabry-Perot (FP) and distributed feedback (DFB) lasers are mostly used for operation at 1550 nm. An important factor for laser transmitters is the safety issues. The primary safety concern is the potential exposure of the eye to the laser beam. Several standards have been developed to limit the transmitted optical power, which rely on parameters such as the laser wavelength and the average and peak transmission power [145]. In fact, only certain wavelengths in the near-IR wavelength range can penetrate the eye with enough intensity to damage the retina. Other wavelengths tend to be absorbed by the front part of the eye before the energy is focused on the retina. In fact, the absorption coefficient at the front part of the eye is much higher for longer wavelengths (>1400 nm) [48], [145]. For this reason, the allowable transmission power for lasers operating at 1550 nm is higher [146], and hence, they are considered for longer distance transmissions.

B. Receiver

FSO systems can be broadly categorized into two classes based on the detection type: non-coherent and coherent. In coherent systems (Fig. 4), amplitude, frequency, or phase modulation can be used. At the receiver side, the received field is optically mixed before photo-detection with a locally generated optical field.

In non-coherent systems (Fig. 5), the intensity of the emitted light is employed to convey the information. At the receiver side, the photo-detector directly detects changes in the light intensity without the need for a local oscillator. These systems are also known as intensity-modulation direct-detection (IM/DD) systems. Although coherent systems offer superior performance in terms of background noise rejection, mitigating turbulence-induced fading, and higher receiver sensitivity [44], [149]–[151], IM/DD systems are commonly used in the terrestrial FSO links due to their simplicity and low cost. In the following, we will focus on IM/DD systems while a discussion on advances in coherent FSO systems is provided in Section XI.

The receiver front-end in an IM/DD FSO systems consists of optical filters and a lens which has the role of collecting and focusing the received beam onto the photodiode (PD). The PD output current is next converted to a voltage by means of a trans-impedance circuit, usually a low-noise Op-Amp with a load resistor. This latter is determined based on the transmission rate, the dynamic range of the converted electrical signal, the generated receiver thermal noise, and impedance matching with the other receiver parts. It is typically about several hundreds of kΩ in deep-space applications [152] down to about 50-100Ω in high-rate terrestrial FSO links [153]. The output of the trans-impedance circuit is then low-pass filtered in order to limit the thermal and background noise levels.

Concerning the PD, solid-state devices are mostly used in commercial FSO systems since they have a good quantum efficiency for the commonly used wavelengths [133], [154]. The junction material can be of Si, InGaAs, or Ge, which are primarily sensitive to the commonly used wavelengths and have an extremely short transit time, which leads to high bandwidth and fast-response detectors [133]. Si PDs have a maximum sensitivity around 850 nm, whereas InGaAs PDs are suitable for operation at longer wavelengths around 1550 nm. Ge PDs are rarely used, however, because of their relatively high level of dark current [48].

The solid state PD can be a P-i-N (PIN) diode or an avalanche photodiode (APD). PIN diodes are usually used for FSO systems working at ranges up to a few kilometers [155]. The main drawback of PIN PDs is that the receiver performance becomes very limited by the thermal noise. For long distance links, APDs are mostly used which provide a current gain thanks to the process of impact ionization. The drawback of APDs, in turn, is the excess noise at their output, which models the random phenomenon behind the generation of secondary photo-electrons. Due to this reason, the APD gain is usually optimized with respect to the received signal power in order to maximize the received SNR [156]. The advantage of APD comes at the expense of increased implementation complexity. In particular, we need a relatively high voltage for APD reverse biasing that necessitates the use of special electronic circuits. This also results in an increase in the receiver power consumption [157].

The use of optical pre-amplifiers has also been proposed in long range FSO links to improve their performance [158], [159]. In the 1550 nm wavelength, an Erbium-doped fiber amplifier (EDFA) is a good choice. Semiconductor optical amplifiers (SOAs) can also be used in a variety of wavelengths (including 1550 nm). However, apart from the problems associated with coupling to the receiver optics, especially when using a multimodal fiber, the optical amplifier introduces an amplified spontaneous emission (ASE) noise, usually modeled as additive white Gaussian noise (AWGN), which can degrade the receiver performance [139]. More specifically, in direct detection receivers, an optical pre-amplifier can degrade the SNR by at least 3 dB [154]. Nevertheless, when the receiver performance is limited by the electronic noise (see the following subsection), optical pre-amplification can be highly beneficial [154]. The use of an EDFA or an SOA when gain-saturated by the input signal has also been proposed to reduce the scintillation effect in the weak turbulence regime [160].

C. Receiver Noise and Modeling

The noise sources at the receiver [133], [139], [161] consist of the PD dark current, the transmitter noise, thermal noise, and the photo-current shot-noise (which arises from input signal and/or background radiations). The PD dark current can be neglected for most practical purposes. The transmitter
Fig. 4. Coherent FSO receiver block diagram.

Fig. 5. IM/DD FSO receiver block diagram.

noise arises from the instability of the laser intensity and the resulting fluctuations of the photo-current at the receiver, which are modeled by considering the so-called laser relative intensity noise (RIN) [139]. However, RIN has usually a negligible effect on the receiver performance [140].

If the background illumination level is negligible, the two main noise sources affecting the receiver are thermal and shot noises. A PIN-based receiver is usually thermal-noise limited. On the other hand, APD-based receivers are usually shot-noise-limited except for relatively small values of the load resistor where the thermal noise also affects the performance [140]. Thermal noise originates from the receiver electronic circuitry, mainly the load resistor, and is modeled as a zero-mean Gaussian random process. On the other hand, shot noise, also called the quantum noise, arises from random fluctuations of the current flowing through the PD and is modeled by a Poisson process. In the case of using a PIN PD, if the mean number of absorbed photons is relatively large, the shot noise can be approximately modeled by a Gaussian process [139]. In most FSO applications, the received photon flux is high enough to allow this approximation. In the case of using an APD, on the other hand, the distribution of the number of generated electrons is given by McIntyre in [162], (16a)] and experimentally verified by Conradi in [163]. However, it has been shown in [140], [164] that this distribution can be approximated by a Gaussian. So, whatever the PD type, the receiver shot noise can be modeled as Gaussian distributed. Notice that this is also true when background radiations cannot be neglected [103], [105], [113], [122], [133], [165]–[167].

IV. INFORMATION THEORETICAL LIMITS

The Shannon-Hartley theorem determines the (theoretical) maximum data rate that can be transmitted with an arbitrarily small BER over a channel for a given average signal power [168]. This maximum achievable rate is known as channel capacity. Numerous works have considered the capacity of a “classical” optical channel, i.e., in the absence of turbulence. The earliest works have considered a Poisson channel model for the quantum-noise limited receivers, assuming negligible thermal and background noise. It was shown in [169], [170] that the capacity of these photon counting receivers (in nats/photon) under an average optical power constraint is unbounded. In such channels, the Q-ary pulse position modulation (PPM) (see Section V) can achieve arbitrarily small probability of error for any rate [171], [172]. Under an additional constraint of fixed peak optical power, it was shown in [173], [174] that binary level modulation schemes are capacity-achieving. The capacity of a PPM channel was also studied in [175] for the case of deep-space communication using a photon counting receiver. Also, [152] studied the capacity of the PPM channel assuming a receiver with an APD. Nevertheless, PPM-based photon counting schemes require an exponential increase in bandwidth as a function of the rate [171]. To avoid the need to increased bandwidth, one solution is to use pulse amplitude modulation (PAM) and to increase the corresponding number of levels [176]. These general conclusions are also valid for the case of FSO links affected by background and thermal noises, where the noise is modeled as Gaussian distributed [177].

In practice, as FSO channels are subject to atmospheric turbulence, the channel capacity should be considered as a random variable due to the randomness of the channel fading coefficient [178]. In general, for channels subject to fading, the definitions of ergodic or outage capacities are used [179]. Ergodic (also called average) capacity is the expectation of the instantaneous channel capacity and is useful when the channel varies very fast with respect to the symbol duration [180]. The ergodic capacity can be calculated through the expectation of the mutual information expression with respect to random fading coefficients. For FSO channels where the channel coherence time is relatively large, the outage capacity becomes more meaningful [92]. In this case, communication is declared successful if the mutual information exceeds the information rate. Otherwise, an outage event is declared. The probability of an outage event is commonly referred to as outage probability or the probability of fade. Based on this outage definition, \( \theta \)-outage capacity is the largest rate of transmission such that the outage probability is less than \( \theta \), where the value of \( \theta \) depends on the intended application. Note that another definition of channel capacity that has been proposed for fading channels

1553-877X (c) 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
is the delay-limited capacity, which corresponds to the zero-outage capacity, i.e., the capacity conditioned to a zero outage probability [179]. For a turbulent channel, when no diversity technique is employed, the delay-limited capacity equals zero, and at the limit of infinite diversity order, it tends to the ergodic capacity [113], [178]. Similar to ergodic capacity, this definition is not useful in the case of FSO channels and the outage capacity is quite more appropriate for these channels.

Several works have investigated the capacity of turbulent FSO channels. The ergodic capacity of an FSO link was studied in [181]–[183] for the cases of log-normal, Gamma-Gamma, negative exponential, and I-K fading models and considering the AWGN model for the receiver noise. Outage capacity of I-K fading channels with AWGN was also studied in [184] while the outage probability is investigated under the assumption of a log-normal fading channel in [88] and for a Gamma-Gamma channel in [110]. Other works have considered FSO systems with transmit and/or receive diversity (see Section VII). For instance, outage capacity for aperture averaging and multiple aperture receivers was studied in [92] considering Gamma-Gamma fading and AWGN at the receiver. For instance, considering Gamma-Gamma modeled strong turbulence with Rytov variance 19.18, background-noise-limited receiver, uncoded OOK modulation, an outage probability of $10^{-9}$, and a moderate average received SNR of 15 dB, the outage capacity of an FSO system increases from 0.05 to 0.86 bit/symbol by increasing the receiver aperture diameter from 20 to 100 mm, respectively [92]. Under the same conditions, for a four-aperture FSO system of aperture diameter 10 and 50 mm (with the same total receiver aperture size as for the SISO case), the outage capacity equals 0.61 and $\sim 1$ bit/symbol, respectively.

The outage probability of MIMO FSO systems was also derived in [113], [122] for AWGN model and different channel models including exponential, log-normal, Gamma-Gamma, log-normal Rice, and I-K fading. Ergodic and outage capacities of a MIMO Poisson channel subject to log-normal fading were also studied in [185], and the ergodic MIMO capacity was studied in [186] for the case of a PIN-based receiver assuming AWGN and Gamma-Gamma fading. Lastly, the outage and ergodic capacities of FSO systems with pointing errors were studied in [57] and [187]–[189], respectively, for the case of Gamma-Gamma fading and AWGN model.

Table I summarizes the contribution of the most relevant works in the literature by specifying the considered capacity definitions and channel models.

V. MODULATION

The most commonly used IM technique due to its implementation simplicity is on-off keying (OOK), which is a binary level modulation scheme. In OOK signaling, modulated data is represented by the presence (“on”) or absence (“off”) of a light pulse in each symbol interval. At the receiver, for optimal signal detection, we need to know the instantaneous channel fading coefficient to perform dynamic thresholding [190]. The channel state information (CSI) can be estimated with good accuracy by using a few pilot symbols in practice [191]. Alternative solutions include symbol-by-symbol maximum likelihood (ML) detection based on the availability of distribution of the channel fading (not the full knowledge of the instantaneous channel fading coefficient) [103] and ML sequence detection based on the knowledge of the joint temporal statistics of the fading [165]. In addition to the need to dynamic thresholding at the receiver, OOK has relatively poor energy and spectral efficiency. Indeed, these are two important factors relative to the choice of a modulation scheme. Energy efficiency refers to the maximum achievable data rate at a target BER (or the minimum BER at a target data rate) for a given transmit energy irrespectively of the occupied bandwidth. As its definition indicates, in particular, it does not take into account the increase in the switching speed of the electronics that can be an important point regarding the implementation complexity. Spectral or bandwidth efficiency, on the other hand, refers to the information transmission rate for a given bandwidth without taking the required transmit energy into account.

Several other IM schemes have been proposed to overcome some disadvantages. To address energy efficiency, PPM becomes a powerful solution [104]. It is shown in [174] that, for a classic optical channel under peak and average power constraints, a slotted binary modulation can nearly achieve the channel capacity. Furthermore, it is proved in [133] that, under such constraints, PPM can attain the near-optimum channel capacity. When performing hard signal detection at the receiver, PPM has the advantage that, in contrast to OOK, it does not require dynamic thresholding for optimal detection [111], [192], [193]. PPM is in particular proposed for deep space communication (together with photon-counting receivers), where energy efficiency is a critical factor [14], [194].

In comparison to PPM, multipulse PPM (MPPM) brings the further advantages of having a reduced peak-to-average power ratio (PAPR) and a higher spectral efficiency [195], [196] while it has an increased demodulation complexity [192]. Note that, although there is a large bandwidth available in the optical band, spectral efficiency is still an important design consideration since it is directly related to the required speed of the electronic circuitry in an FSO system from a practical point of view. Under a constraint on the peak transmit power, MPPM outperforms PPM. Conversely, when a constraint is imposed on the average transmit power, PPM outperforms MPPM [195], [197].

Two other well-known modulation schemes are pulse width modulation (PWM), and digital pulse interval modulation (DPIM). Compared with PPM, PWM requires a lower peak transmit power, has a better spectral efficiency, and is more resistant to ISI, especially for a large number of slots per symbol ($Q$) [198]. Nevertheless, these advantages are counterbalanced by higher average power requirements of PWM that increases with $Q$. By DPIM, for each symbol, a pulse is sent followed by a number of empty slots, depending on the input bits [199], [200]. An additional guard slot is also usually added to avoid sending consecutive “on” pulses.

PPM and PWM are usually called synchronous modulations because they map the input bits on a symbol of fixed duration.
Both schemes require slot and symbol-level synchronization. In contrast, DPIM is an asynchronous modulation scheme with variable symbol length, and does not require symbol synchronization [199]. In addition, it is more spectrally efficient than PPM and PWM, because it does not need to wait the end of a fixed symbol period before sending the next symbol. The main potential problem with DPIM is the possibility of error propagation in signal demodulation at the receiver. In fact, if an “off” slot is detected erroneously as “on,” all the succeeding symbols in the frame will be decoded with error.

Other modulation schemes based on some modifications of either PPM or PWM have also been proposed in the literature. Using the same idea of MPPM, overlapping PPM (OPPPM) constrains the multiple pulses to occupy adjacent slots [201]. In differential PPM (DPPM), the empty slots following a pulse in a PPM symbol are removed, which improves the spectral efficiency of the system [202]. Also, in this way, every DPPM symbol ends with a pulse, which can be exploited for symbol synchronization at the receiver [76]. In digital pulse interval and width modulation (DPIWM), the binary sequence is encoded into the width of the pulses of alternating energy [203]. The PMPW scheme, proposed in [198], is a combination of PPM and PWM with power and spectral efficiencies in mid-way between PPM and PWM. The main drawbacks of all these modulation schemes are the reduced energy efficiency, the relatively high demodulation complexity, and the risk of error propagation in detecting a received frame of symbols.

In the so-called subcarrier intensity modulation (SIM) [204], [205], the data is first modulated onto an RF signal, and then used to change the intensity of an optical source [34], [84], [206]–[208]. When combined with orthogonal frequency division multiplexing (OFDM) [209], [210], it offers the advantages of high capacity and cost effective implementation, as compared with coherent modulation [211]. The main argument for using SIM is to cope with the optical fiber networks employing subcarrier modulation together with wavelength division multiplexing [212], [213]. The main drawback of SIM is its poor optical power efficiency [205] due to the DC bias that should be added to the multiple-subcarrier electrical signal before optical intensity modulation (to avoid negative amplitudes).  

A polarization modulated DD scheme was proposed in [214] based on the extraction of the Stokes parameters of the transmitted light. Such a modulation scheme is not constrained by the nonlinear response of the intensity modulators, as it is the case for IM schemes. Polarization-based modulation has also the advantage of high immunity to the phase noise of lasers [215]. Moreover, it is more resilient to atmospheric turbulence-induced fading because the polarization states are better conserved during propagation than the amplitude and the phase of the optical signal [216]. This can be particularly useful for long range FSO systems [215].

Finally, multi-level modulation schemes could be used in FSO systems to obtain higher spectral efficiencies compared to binary modulations. Once again, the improved spectral efficiency is obtained at the expense of increased system complexity. An example is the PAM, with OOK as its simplest scheme [83], [133], [217], [218]. By Q-ary PAM, the instantaneous intensity of the laser source is modulated on Q levels and, hence, it requires a laser with a variable emission intensity which could be costly. The main advantage of PAM is its higher spectral efficiency with respect to binary-level modulations like PPM [217]. Other multilevel DD schemes include Q-ary differential phase-shift keying (DPSK), differential amplitude-phase-shift keying (DAPS), and differential polarization-phase-shift keying (DPoPSK) [219]. Recently, carrier-less amplitude and phase (CAP) modulation has been considered for OWC that consists in transmitting simultaneously two orthogonal multilevel signals by means of special pulse shaping and without using a carrier [220]. Its main advantages as compared to PAM are its higher energy efficiency and simpler implementation [211].

A summary of the literature related to optical signal modu-

<table>
<thead>
<tr>
<th>Reference</th>
<th>Configuration</th>
<th>Channel Model</th>
<th>Modulation</th>
<th>Ergodic Capacity</th>
<th>Outage Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[169-174]</td>
<td>SISO</td>
<td>Non-fading</td>
<td>PPM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[176]</td>
<td>SISO</td>
<td>Non-fading</td>
<td>PPM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[177]</td>
<td>SISO</td>
<td>Non-fading</td>
<td>PAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[57]</td>
<td>SISO</td>
<td>LN, IT</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[88]</td>
<td>SISO</td>
<td>IT, IT</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[181]</td>
<td>SISO</td>
<td>IT</td>
<td>OOK</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>[182]</td>
<td>SISO</td>
<td>I-K</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[183]</td>
<td>SISO</td>
<td>LN, IT</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[184]</td>
<td>SISO</td>
<td>EXP</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[185]</td>
<td>SISO</td>
<td>I-K, K</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[188]</td>
<td>SISO</td>
<td>IT</td>
<td>OOK</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[92]</td>
<td>SIMO</td>
<td>IT</td>
<td>OOK, PPM</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[113]</td>
<td>MIMO</td>
<td>LN, Gamma-Gamma, EXP</td>
<td>PPM</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[122]</td>
<td>MIMO</td>
<td>LN-Rice, I-K</td>
<td>PPM</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>[180]</td>
<td>MIMO</td>
<td>IT</td>
<td>PPM</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

4Compared with coherent modulation, considered in Section XI, for a given spectral efficiency, SIM offers the advantage of implementation simplicity at the expense of lower energy efficiency as a result of the DC bias added to the signal.
TABLE II
LITERATURE ON FSO SIGNAL MODULATION.

<table>
<thead>
<tr>
<th>Modulation scheme</th>
<th>Related references</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>OOK</td>
<td>[103, 165, 190, 193]</td>
<td>Needs dynamic thresholding at receiver</td>
</tr>
<tr>
<td>PPM</td>
<td>[14, 104, 111, 133, 191, 194]</td>
<td>Optimal in terms of energy efficiency</td>
</tr>
<tr>
<td>MPPM</td>
<td>[192, 195, 196, 197]</td>
<td>Lower PAPR and more bandwidth efficient than PPM</td>
</tr>
<tr>
<td>PWM</td>
<td>[76]</td>
<td>Needs lower peak power, better spectral efficiency, more resistant to ISI than PPM</td>
</tr>
<tr>
<td>PPM/PWM</td>
<td>[198]</td>
<td>Power and bandwidth efficiencies in mid-way between PPM and PWM</td>
</tr>
<tr>
<td>DPIM</td>
<td>[199, 200]</td>
<td>No need to symbol synchronization, more bandwidth efficient than PPM and PAM</td>
</tr>
<tr>
<td>DPPM</td>
<td>[75, 202]</td>
<td>Simpler symbol synchronization and improved bandwidth efficiency than MPPM</td>
</tr>
<tr>
<td>OPPM</td>
<td>[201]</td>
<td>More bandwidth efficient than PPM</td>
</tr>
<tr>
<td>PAM (multilevel)</td>
<td>[83, 177, 216, 217]</td>
<td>Higher bandwidth efficiency than PPM; requires dynamic thresholding at receiver</td>
</tr>
<tr>
<td>SIM</td>
<td>[34, 84, 204, 212]</td>
<td>High capacity, cost effective implementation; low power efficiency</td>
</tr>
<tr>
<td>Pol. mod. &amp; DD</td>
<td>[213-215]</td>
<td>High immunity to laser phase noise and modulator nonlinearity</td>
</tr>
<tr>
<td>CAP</td>
<td>[219]</td>
<td>Higher energy efficiency and simpler implementation than PPM</td>
</tr>
</tbody>
</table>

VI. CHANNEL CODING

The intensity fluctuations on the received signal due to the atmospheric-turbulence-induced channel fading can result in a considerable degradation of the system performance. In fact, the atmospheric optical channel has a very long memory, and a channel fade can cause an abnormally large number of errors that affect thousands of consecutive received channel bits. Mitigating fading in FSO channels has been the subject of intensive research during the last decade. One possible solution is channel coding [83] which is particularly useful under weak turbulence conditions [92]. It is also efficient in moderate and strong turbulence regimes provided that the impact of turbulence can be first significantly reduced, for example, by means of other fading-mitigation techniques such as aperture averaging, diversity techniques, or adaptive optics [92].

Earlier works on coded FSO systems have considered the use of convolutional codes for the atmospheric optical communication channel using OOK or other binary modulation schemes [222]-[224]. Several other works have considered the use of low-density parity check (LDPC) codes for optical communication over atmospheric turbulence channels [180], [225], [226]. These codes, introduced by Gallager in the early 1960’s [227], are constructed by using sparse parity check matrices. The use of LDPC coding together with OFDM modulation is further proposed in [232].

Error performance bounds are derived in [166], [233]–[236], for coded FSO communication systems operating over atmospheric turbulence channels. These works, however, consider an uncorrelated FSO channel requiring the deployment of large interleavers. The channel coherence time is about 0.1-10 ms, therefore fading remains constant over hundreds of thousand up to millions of consecutive bits for typical transmission rates [44]. For atmospheric channels with such long coherence times, this necessitates long delay latencies and the use of large memories for storing long data frames. In addition, since the duration of the fades is random, no single maximum interleaving depth can be used to render the channel completely memoryless. Furthermore, when aperture averaging is employed at the receiver (see the next section), exploiting time diversity through channel coding becomes more difficult and even practically infeasible [92]. Because, under the assumption that the channel time variations are mostly due to the transversal wind (with respect to the optical axis), the use of a relatively large aperture size results in a large channel coherence time [237].

It has recently been proposed that exploiting the FSO channel reciprocity can eliminate the need for interleaving and the amount of the redundancy introduced with channel coding [238]. In fact, given the channel reciprocity, we can estimate the CSI at the transmitter in a full-duplex transmission system. Then, the idea is to use a bank of encoders and decoders and to select the appropriate encoder-decoder pair based on the estimated current CSI [239].

The effect of finite-size interleavers is studied in some works. An LDPC coding scheme combined with interleaving is proposed in [225] for digital video transmission over turbulent temporally-correlated optical channels that satisfies special real-time video delay constraints. There, instead of using a too long interleaver in the physical layer, the data block length is extended in the network layer to benefit from time diversity. The use of interleaved turbo-codes as well as concatenated RS and convolutional codes has been considered in [191], where it is concluded that convolutional codes could be a suitable choice under any turbulence regime as they make a good compromise between complexity and performance.

Rateless codes, also known as fountain codes [242], [243] have been further investigated in the context of FSO [244]. Rateless coding involves the change of the coding rate according to the channel conditions, without using interleaving to exploit channel time diversity. One specific implementation is raptor coding [245], [246] which consists of concatenating rateless codes.

It is worth mentioning that it has been shown for the case of optical fiber communication that LDPC codes outperform block turbo codes [228], with a decoder complexity comparable (or lower) to that of the latter [229]. Their complexity is significantly lower than that of serial/parallel concatenated turbo codes [230] as well [231].

5Early works on time diversity in FSO systems considered the transmission of data streams several times, with large enough delay between them, and performing data detection based on the received delayed copies [138], [240], [241].
an inner code with an outer Luby Transform (LT) code [247]. These codes, although initially designed for erasure channels, have been shown to be quite efficient over binary symmetric and block-fading channels as well [248], [249]. More discussion on these schemes will be provided in Section VIII.

Most of the existing works on coded FSO systems assume binary modulation. There are also some further efforts which consider the deployment of non-binary modulation. For instance, convolutional codes and turbo codes have been applied to the PPM modulation in [250]–[252] and [193], [253]–[256], respectively. In order to perform efficient error correction in the case of non-binary modulations, we should either use non-binary codes, or adapt the binary codes to these modulations. Use of non-binary codes necessitates a considerable decoding computational complexity [133] that can be prohibitively large for a practical implementation in a high rate FSO system. In [257], [258], Reed Solomon (RS) codes are suggested as relatively low-complexity solutions for PPM-based modulations. For example, a \((n, k)\) RS code is matched to \(Q\)-ary PPM for \(n = Q – 1\) [257]. Concatenated convolutional and RS codes were further considered in [191]. However, RS coding cannot provide satisfying performance improvement, in particular, due to hard decoding that is usually performed at the receiver. Note that soft RS decoding is computationally too complex and is rarely implemented.

Some attempts for adapting a binary code to non-binary modulation can be further noted. An example is multilevel coding (MLC) [259] which is a powerful coded modulation scheme [260]. However, the drawback of this technique is the high complexity of the multi-stage decoder that makes its real-time implementation in high-speed applications difficult. Trellis coded modulation is another example investigated in [261]. In [262], an LDPC code is considered in conjunction with DPPM. The use of lattice codes [263] for FSO systems is considered in [202], where higher-dimensional modulation schemes are constructed from a series of one-dimensional constituent OOK constellations. The use of multidimensional lattices is further discussed in [264]. As another solution, it is proposed in [192], [265], [266] to use a classical binary convolutional code and to do iterative soft demodulation and decoding at the receiver. This scheme, which is extended to MPPM in [192], is shown to be quite efficient and suitable for not-too-high transmission rates, so that iterative signal detection can be performed in real time.

VII. SPATIAL DIVERSITY

Spatial diversity can be realized via the use of multiple apertures at the receiver [51], [138], [158], [257], [267], multiple beams at the transmitter [268], [269], or a combination of the two [75], [104], [111], [153], [270]. In contrast to the classical single-beam single-aperture configuration that we will call SISO (for Single-Input Single-Output), these configurations are usually referred to as SIMO (Single-Input Multiple-Outputs), MISO (Multiple-Inputs Single-Output), and MIMO (Multiple-Inputs Multiple-Outputs), respectively. We discuss these techniques in the following.

A. Receive Diversity

A simple solution to reduce the fading effect is to use a relatively large lens at the receiver to average over intensity fluctuations. This technique, usually called aperture averaging, can be considered as “inherent” receive diversity. It is efficient when the receiver lens aperture is larger than the fading correlation length \(\sqrt{\lambda L}\), with \(\lambda\) and \(L\) denoting the wavelength and link distance, respectively [51], [271]. Aperture averaging has widely been studied in the literature and also employed in practical systems [50], [51], [91], [92], [95], [271]–[276], where it is shown that a substantial scintillation reduction can be obtained, especially in the case of moderate-to-strong turbulence. For instance, considering QOK modulation, Gamma-Gamma fading under moderate turbulence conditions with Rytov variance of 2.56, and a target BER of \(10^{-5}\), the SNR gain with respect to a point receiver is about 30, 47, and 60 dB for receiver lens diameters of 20, 50, and 200 mm, respectively [92].

Fading reduction by aperture averaging is usually quantified by considering the so-called aperture averaging factor \(A = \frac{\sigma^2(\lambda)}{\sigma^2(0)}\), where \(\sigma^2(\lambda)\) and \(\sigma^2(0)\) denote the scintillation indexes for a receiver lens of diameter \(\lambda\) and a point receiver (of diameter \(\lambda \approx 0\)), respectively. It is shown in [92], [277] that the performance improvement by aperture averaging is most significant for plane wave and Gaussian-beam propagation models, and also when more complex modulation schemes (e.g., \(Q\)-ary PPM) are used.

It is worth mentioning that the fade statistics change when using aperture averaging. In fact, since averaging is specially performed over small-scale irradiance fluctuations, the PDF of the channel fades shifts toward that of large-scale fluctuations [268]. Experimental results show that the scintillation on the received signal is well described by a log-normal distribution [240], [271]. The Gamma-Gamma and log-normal models become practically equivalent for about \(D > 6\rho_0\), with \(\rho_0\) being the spatial coherence radius [277].

Efficient fading reduction can be also achieved by using multiple apertures at the receiver. In particular, instead of using a large aperture, we can use several smaller apertures at the receiver. This way, each receiver aperture will benefit from some degree of aperture averaging that is smaller than that of the single large aperture case. However, in addition, we also benefit from some degree of spatial diversity after combining the signals of the different apertures. If we assume uncorrelated fading on the different apertures’ signals, the multiple aperture solution provides a better performance than the solution of using a large aperture if we consider the same total effective aperture area in the two cases [92]. For instance, considering background-noise-limited receivers, OOK modulation, Gamma-Gamma fading with Rytov variance of 2.56, and a target BER of \(10^{-5}\), we have an SNR gain of about 1 dB by using four apertures of 50 mm diameter each, compared to using a single aperture of 100 mm diameter [92].

\[7\] The spatial coherence radius is defined as the 1/e point of the wave complex degree of coherence (see [51, Section 6.4]). For the plane wave propagation model, we have \(\rho_0 = \left(\frac{1.46}{2\pi^2 k^2 L}\right)^{\frac{1}{2}}\) with \(k = 2\pi/\lambda\) being the optical wave number. Under weak to moderate turbulence conditions, only eddies of size smaller than \(\rho_0\) contribute to intensity fluctuations [51].
Here, employing a single large aperture would be preferable for the reasons of implementation complexity. The use of multiple apertures is more advantageous in the strong turbulence regime. For instance, for a Rytov variance of 19.18 and the same conditions as above, we have an SNR gain of about 7 dB by using four apertures instead of the single large aperture [92].

It should be noted that, from a practical point of view, the use of a too large lens necessitates a PD with a large active area as well, in order to capture the received photons on the lens focal plane. This will, in turn, impose severe constraints on the system data rate because such a PD will have a relatively large parasitic capacitance.

For SIMO systems, usually equal-gain combining (EGC) is performed at the receiver, which provides performance close to the optimal maximal-ratio combining (MRC) while having the advantage of lower implementation complexity [138], [278].

Lastly, note that apart from diversity techniques, the turbulence effect can also be reduced by adaptive optics [279]. By this technique, the distortion induced in the wave-front by the atmospheric turbulence is reduced through the use of wave-front sensors and deformable mirrors; a technique commonly used in optical astronomy [280], and also envisaged for deep-space optical communication [14]. However, this technique does not seem to be of interest in commercial FSO systems due to its high and unjustified implementation complexity and cost. Also, its effectiveness to compensate turbulence effects is practically limited to relatively short link spans [281].

B. Transmit Diversity

For a MISO FSO system, the simplest signaling scheme is to send the same signal on the different beams; what is usually referred to as repetition coding (RC). This is quite efficient for fading reduction at the receiver. For instance, assuming independent fading conditions, for log-normal fading of standard deviation 0.3, a receiver aperture of 5 cm, a link distance of 2 km, and a target BER of $10^{-5}$, the improvement in the average SNR by using two and three transmit apertures, as compared to a SISO system is about 5 and 7.5 dB, respectively [278]. If CSI is available at the transmitter, it is shown in [282], [283] that selection transmit diversity can exploit full diversity while providing better performance, compared to RC. For the case of imperfect CSI at the transmitter, different transmission strategies are considered in [284]. More complex signaling schemes can be used to increase the coding gain in addition to diversity benefit. For instance, transmit laser selection and space-time trellis coding is proposed in [285].

For a MISO FSO system (or equivalently a SIMO system employing EGC at the receiver), assuming independent fading, fading statistics can be modeled easily [51], [92], [268]. For instance, the received intensity can still be modeled by a Gamma-Gamma distribution, with the variances of large- and small-scales obtained from those of a SISO system divided by the number of sub-channels.

C. MIMO FSO Systems

In RF communication, MIMO systems are very popular as they exploit efficiently the multipath fading to increase the data rate and to reduce the fading effect on the quality of signal transmission [286]. In FSO communication, however, MIMO systems are mostly proposed to reduce the turbulence-induced fading effect by employing RC at the transmitter. Some examples are [104], [111], [153], [167], [287]–[291], where OOK or PPM modulations are considered. Also, multiple-symbol detection is proposed in [105], [112] in the absence of CSI at the receiver, for the case of RC at the transmitter.

A few works have considered the combination of the information bearing symbols at the transmitter in order to optimize the system performance, i.e., employing space-time (ST) coding. This is an extensively-developed subject in RF systems [292]. A fundamental difference between the ST codes for RF and IM/DD-based FSO communication is that the latter employs nonnegative (unipolar) real signals rather than complex signals [293]. In effect, most of the proposed ST schemes for RF applications use phase rotation and amplitude weighting [292], [294], [295], requiring at least bipolar signaling when applied to the FSO context. In general, the ST schemes optimized for RF systems provide full diversity in FSO systems but are not optimized concerning the coding gain [296].

In the following, we first discuss two classical categories of orthogonal and non-orthogonal ST schemes proposed for MIMO FSO systems. The main interest of the orthogonal schemes, which usually provide full diversity, is their low-complexity optimal detection [295]. Most of the orthogonal ST block codes (OSTBCs) proposed for RF systems can be modified in order to adapt to IM/DD FSO systems. For instance, in the case of two transmitter beams, a modified Alamouti scheme [297] for IM/DD optical systems is proposed in [293] by introducing a DC bias to overcome the constraint of unipolar signaling. This idea is then generalized in [298] to OOK modulation with any pulse shape. Due to this DC bias, OSTBC schemes suffer from a degradation in the system performance, compared with the low-complexity RC scheme. Although both RC and OSTBCs provide full diversity, RC is quasi-optimum, as explained in [299]. The difference of the performance of OSTBC and RC schemes increases with increased number of transmitter beams [299].

Non-orthogonal schemes are generally designed to optimize diversity and coding gain but their optimal detection has a relatively high computational complexity. For instance, by the spatial multiplexing (SMux) scheme, the information bearing signals are simply multiplexed at the transmitter. This way, we can attain the maximum transmission rate at the expense of reduced diversity gain. For a number $M$ of beams, the ST coding rate of SMux is equal to $M$. At the receiver, optimal maximum likelihood detection (MLD) can be used for signal detection, which has a relatively high complexity. Otherwise, iterative interference cancelation based on the V-BLAST method [300] can be used, as considered in [270], [301], [302].

Another proposed non-orthogonal ST scheme is the so-called optical spatial modulation (OSM) where only one “on” slot is transmitted from the multiple beams at a given channel-use in order to avoid inter-channel interference [303], [304]. For $M$ transmitting beams, the rate of OSM is $\log_2 M$ symbols
per channel-use. At the receiver, optimal MLD can be used to estimate the corresponding beam [305]. It is shown in [301] that, if we are not limited by practical implementation considerations such as time synchronization and electronic circuitry speed, instead of using non-orthogonal ST schemes, we can alternatively use the simple RC with shorter pulse durations while having a better system performance.

For example, consider a link distance of 5 km, Gamma-Gamma fading with Rytov variance of 24.7, a total receiver aperture diameter of 200 mm, a target BER of $10^{-5}$, uncoded OOK modulation, and MLD detection at the receiver. Fixing the average transmit power as well as the effective transmission data rate for different ST schemes, we modify the pulse duration for each scheme accordingly. Then, for a MIMO structure of two transmit and two receive apertures, the RC duration for each scheme accordingly. Then, for a MIMO structure of two transmit and two receive apertures, the sum of correlated Gamma-Gamma refraction structure parameter $C_2^{\alpha} = 4.58 \times 10^{-13} \text{m}^{-2/3}$, we have $l_c \approx 6.4 \text{cm}$ for $L = 500 \text{m}$ (moderate turbulence regime), and $l_c \approx 37 \text{cm}$ for the case of $L = 1500 \text{m}$ (strong turbulence regime) [92]. In effect, if the required spacing is more or less reasonable under moderate turbulence conditions [289], it becomes too large for the strong turbulence regime.

Evaluation of fading correlation for a space-diversity FSO system can be made by means of experiments or via wave-optics simulations based on the split-step Fourier-transform algorithm [316]. By the latter method, the effect of atmospheric turbulence is taken into account by considering a set of random phase screens. Experimental works for estimating the fading correlation are reported in [289], [317], [318] for MISO and in [318] for SIMO configurations. The study of fading correlation via wave-optics simulations can be found in [268] for the case of a MISO, and in [306], [319] for the case of a SIMO FSO system. It is reported in [268], [319] that fading correlation increases for increased link distance. This is because more atmosphere eddies affect the different receiver apertures at the same time. For the same reason, correlation increases by increased receiver aperture size [268], [289], [319].

Another important question is to see how fading correlation affects the FSO system performance, compared to the “ideal” uncorrelated fading case. For this purpose, it is necessary to develop an appropriate statistical model. A few works have recently considered the effect of fading correlation by considering simplified statistical models. For instance, in [103], [278] and for the case of log-normal distributed fading, the effect of fading correlation on a SIMO system BER is studied by considering the joint distribution of the received signals given the corresponding covariance matrix. More specifically, in [278], the effect of correlation is modeled by an additive correction term to the scintillation index corresponding to the uncorrelated fading case. An exponential correlation model was considered in conjunction with K distributed fading in [234], and with multivariate Gamma-Gamma fading in [320]. However, this correlation model is not appropriate for most practical FSO system configurations. On the other hand, the case of a four-laser single-aperture FSO system is studied in [268] by considering the Gamma-Gamma channel model, where the Gaussian approximation is used to model the correlated fading channels. In [322], for a SIMO system with two receive apertures, the sum of correlated Gamma-Gamma random variables are approximated by an $\alpha-\mu$ distribution [323] in order to evaluate the BER performance of the receiver. This idea is then generalized to the case of multiple diversity in [319], [324]. Also, the Padé approximation method [325], [326] is used in [327] to obtain the PDF of sum of correlated Gamma-Gamma random variables from their moment generating function, which is then used to evaluate the system performance analytically. However, due to the limitation of Padé approximation, this method cannot be used for very low BERs, i.e., lower than $10^{-8}$.

D. Effect of Fading Correlation

Diversity techniques are most efficient under the conditions of uncorrelated fading on the underlying sub-channels. In practice, however, the performance of spatial diversity systems is impaired by fading correlation. As a matter of fact, it is not always practically feasible to satisfy the required spacing between the apertures at the receiver and/or between laser beams at the transmitter to ensure uncorrelated fading. Under weak turbulence conditions, the required aperture side spacing $l_c$ equals the correlation length $\sqrt{\lambda z}$, which is in fact the typical size of scintillation speckles [315]. In the relatively strong turbulence regime, the spatial correlation arises mainly from large-scale fluctuations, where larger aperture spacings are required. Assuming plane wave propagation, we have $l_c = \lambda L/r_0$, where $r_0$ is the Fried parameter. As an example, assuming the wavelength $\lambda = 1550 \text{nm}$ and the index of the refractive index $n = 37$ dB, the received SNR by 2, 31.5, and 37 dB, respectively. For a four-structure of two transmit and two receive apertures, the RC duration for each scheme accordingly.

\[ l_c = \frac{\lambda L}{r_0} \]

where $r_0$ is the Fried parameter. As an example, assuming the wavelength $\lambda = 1550 \text{nm}$ and the index of the refractive index $n = 37$ dB, the received SNR by 2, 31.5, and 37 dB, respectively. For a four-structure of two transmit and two receive apertures, the RC duration for each scheme accordingly.

\[ l_c = \frac{\lambda L}{r_0} \]
Lastly, it should be noted that, when using a doubly-statistic fading model considering separately small- and large-scale fading effects, in most practical cases, we can effectively assume uncorrelated small-scale fading and assign the correlation to the large-scale fading component [328].

VIII. ADAPTIVE TRANSMISSION

A common assumption in the current literature on FSO systems is open-loop implementation in which the transmitter has no knowledge of the channel. The classical approach is then to use at the link layer the automatic repeat request (ARQ) mechanism or hybrid ARQ (HARQ) in the form of incremental redundancy, for example, in order to improve the link reliability [329], [330]. Such open-loop (or low-feedback) designs are favorable in time-varying channels where the feedback of channel estimates becomes problematic. However, particularly for quasi-static channels, providing reliable feedback is possible and the available CSI at the transmitter can be used to design adaptive transmission schemes for significant performance improvements. As a matter of fact, as mentioned previously, atmospheric turbulence results in a very slowly-varying fading in FSO systems. The channel coherence time is about 0.1-10 ms, therefore fading remains constant over up to millions of consecutive bits for typical transmission rates. Therefore, adaptive transmission emerges as a promising solution for FSO systems. Furthermore, the feedback information required in adaptive transmission is relatively easy to implement in FSO systems. This is because commercially available FSO units have full-duplex (bi-directional) capabilities and a small portion of the large available bandwidth can be allocated for feedback purposes without much effect on data rates [331]. In hybrid RF/FSO systems, the RF link can be used as the feedback link to enable CSI knowledge at the transmitter [332], [333].

Adaptive transmission has been extensively investigated in the context of wireless RF networks [334] and involves the change of system parameters such as transmit power, modulation size/type, code rate/type or a combination of those according to the channel conditions. The same ideas have recently been investigated in the FSO context. A simple adaptive power transmission (assuming a fixed modulation) scheme was considered in [335] by taking into account only the path loss that can be time variant on the order of several hours. For the case of Gamma-Gamma turbulent channels and $Q$-ary PAM modulation, power adaptation for maximizing the channel capacity was considered in [333]. Adaptive coding and $Q$-ary PAM modulation was further studied in [332] over Gamma-Gamma channels. Adaptive coding can be performed either by using punctured codes, where the coding rate is varied by puncturing a percentage of parity of information bits [336], [337] or through the use of rate-adaptive codes such as fountain codes. Raptor codes, considered in [244], are a special case, where the coding rate is modified by changing the codeword length. In [338], Section 9.5 a performance comparison is made between Raptor codes and punctured LDPC codes, where it is shown that a punctured LDPC code is useless in the low SNR regime. Also, it is shown that, for the case of imperfect CSI at the transmitter, the performance degradation with Raptor codes is insignificant, compared with the latter approach [338].

The works in [332], [333] build upon the assumption that the modulation size can be changed continuously (i.e., $Q$ can take any real value) and ignore constraints on the peak power. These constraints are particularly important for FSO applications where eye safety standards impose restrictions on the peak of transmit power. In [331], considering $Q$-ary PPM modulation, the design problem of adaptive FSO transmission is revisited under the assumption of practical modulation sizes (i.e., integer values of $Q$) and average/peak power constraints and by considering a joint power and modulation adaptation. Also, it is proposed to quantify the performance improvement in terms of the number of bits carried per chip time (BpC) which is in fact the ratio of bit-rate over the required bandwidth. Considering Gamma-Gamma modeled strong turbulence with Rytov variance of 1.55, for a target outage probability of $10^{-4}$, it is shown for instance that for the average transmit power constrained to $-20$ dB, non-adaptive transmission achieves a BpC of zero, whereas performing adaptive power control (on the instantaneous transmit power) can provide a BpC of 0.15. When the transmit power and modulation are both set adaptively, the BpC can increase to 0.35 [331].

As a practical point, special attention should be paid when optical amplifiers are used at the receiver. Indeed, as explained in [332], adaptive power setting cannot practically be done if an EDFA is used, because the response time of these amplifiers is relatively long (on the order of 10 ms, typically). We do not have such a constraint when an SOA is used, however.

IX. RELAY-ASSISTED (COOPERATIVE) TRANSMISSION

Cooperative diversity has been introduced in the context of RF wireless communication as an alternative way of realizing spatial diversity advantages [339]–[341]. The main idea behind cooperative diversity is based on the observation that in a wireless RF channel, the signal transmitted by the source node is overheard by other nodes, which can be defined as partners or relays. The source and its partners can jointly process and transmit their information, creating a virtual antenna array although each of them is equipped with only one antenna. Multi-hop transmission is an alternative relay-assisted transmission scheme which employs the relays in a serial configuration [105], [342]. Such schemes are typically used to broaden the signal coverage for limited power transmitters and do not offer performance improvement against fading effects in wireless RF environments, i.e., they do not increase the diversity order [167].

Relay-assisted FSO transmission was first proposed by Acampora and Krishnamurthy in [343], where the performance of a mesh FSO network was investigated from a network capacity point of view. In [344] and [345], Tsiftsis et al. considered K and Gamma-Gamma fading models without explicitly taking into account the path-loss and evaluated the outage probability for a multi-hop FSO system. Their results demonstrate the usefulness of relay-assisted transmission as
a method to broaden the coverage area, but do not highlight its use as a fading-mitigation tool. In [346], both path-loss and fading effects are considered and outage probability is derived. It is demonstrated that multi-hop FSO transmission takes advantage of the resulting shorter hops and yields significant performance improvements (in terms of diversity gain) since fading variance is distance-dependent in FSO systems. This is rather different from the RF case where multi-hop transmission is used to extend range, but does not provide diversity advantage. It is further proven in [347] that the outage probability is minimized when the consecutive nodes are placed equidistant along the path from the source to the destination. The diversity gain analysis over log-normal turbulence channels (assuming plane wave propagation) reveals a diversity order of \((K + 1)^{1/6}\), where \(K\) is the number of relays. The performance analysis of multi-hop relaying over Gamma-Gamma channels can be found in [348].

Besides multi-hop (also referred to as “serial”) relaying, parallel relaying is further considered in [346], [349]–[352]. It is obvious that the broadcast nature of wireless RF transmission (i.e., the cost-free possibility of the transmitted signals being received by other than destination nodes) is not present in FSO transmission which is based on line-of-sight transmission through directional beams. Parallel relaying can be therefore implemented through the use of multiple transmitter apertures directed to relay nodes. For parallel relaying, all relays should be located at the same place (along the direct link between the source and the destination) closer to the source, and the exact location of this place turns out to be a function of SNR, the number of relays, and the end-to-end link distance [347]. Parallel relaying with a direct link as a three-way cooperative scheme has been further studied in [349], [350], [353], [354]. It is shown in [351] that cooperation through relay nodes is beneficial only if the SNR is high enough; otherwise, relays are likely to forward too noisy copies of the signal, resulting in a performance degradation.

Inspired by the ideas in the well-known RF counterparts, several signaling strategies have been proposed for relay-assisted FSO links. The classical approaches consider amplify-and-forward (AF) [344], [346], [349], [355], decode-and-forward (DF) [346], [351], [354], and detect-and-forward (Def) [350] protocols. Adaptive Def or adaptive DF have also been proposed in [350], where the relay takes part in the data transmission only if it can receive error-free data frames from the source or when the SNR at the relay is large enough, respectively. When CSI is available at the source and the relays, it is proposed to activate only a single relay in each transmission slot, hence, avoiding the need for relays’ time synchronization. One possible protocol consists in selecting for signal transmission the best relay among multiple parallel relays [352], [354], [356]. Another suboptimal but simple approach when two relays are deployed is to switch the activated relay in the case of too low link SNR [356].

To show more concretely the improvement achieved by relay-assisted transmission, consider for instance the log-normal channel model, an atmospheric attenuation of 0.43 dB/km, \(C_n^2 = 10^{-14} m^{-2/3}\), a total link span of 5 km, and a target outage probability of 10^{-6}. It is shown in [346] that by serial relaying in DF mode, improvements of 18.5 and 25.4 dB are obtained in power margin when one or two (equidistant) relays are inserted between the source and the destination. When AF mode is employed, the improvements are about 12.2 and 17.7 dB, respectively. Also, by parallel relaying, where relays are placed in mid-distance between the source and the destination, the obtained improvements are about 20.3 and 20.7 dB for DF mode, and 18.1 and 20.2 dB for AF mode, for the cases of two and three relays, respectively.

The current literature on AF relaying in FSO systems builds on the assumption that relays employ optical-to-electrical (OE) and electrical-to-optical (EO) converters. The actual advantage of AF relaying over the DF counterpart emerges if its implementation avoids the requirement for high-speed (at the order of GHz) electronics and electro-optics. This becomes possible with all-optical AF relaying where the signals are processed in optical domain and the relay requires only low-speed electronic circuits to control and adjust the gain of amplifiers. Therefore, EO/OE domain conversions are eliminated, allowing efficient implementation. All-optical AF relaying has been considered in recent papers [357]–[359]. In particular, Kazemlou et al. [357] have assumed either fixed-gain optical amplifiers or optical regenerators, and presented BER performance through Monte Carlo simulations. In [358] Bayaki et al. have considered all-optical relays employing EDFAs and presented an outage probability analysis for a dual-hop system taking into account the effect of ASE noise. In [359], the outage performance is re-addressed further taking into account the effect of optical degree-of-freedom (DOF). DOF quantifies the ratio of optical filter bandwidth to the electrical bandwidth and can be on the order of 1000 unless narrow-band optical filtering is employed. It is shown in [359] that even for practical values of DOF in the range of 100 to 1000, a significant performance gain over direct transmission is still maintained.

**X. HYBRID RF/FSO SYSTEMS**

As we discussed in the previous sections, the performance of FSO links can seriously be affected due to several factors such as severe turbulence in long-span links subject to strong winds or hot dry climates, strong attenuation in dense fog and heavy snowfalls, misalignment and pointing errors in mobile links, etc. These factors can result in frequent link failures, and hence, there is an important need to increase the reliability of these links. One efficient solution is to use an RF link in parallel with the FSO link so as to serve as back-up in the case of FSO link outage. Although the corresponding data rate in the RF channel can be less than the main FSO link, it can ensure connectivity when the FSO channel becomes inoperative. In effect, such an RF link is less subject to atmospheric turbulence and pointing errors [360], and furthermore is much less affected by fog. As a matter of fact, fog and rain drastically affect FSO and RF links, respectively, but they rarely occur simultaneously. Therefore, concerning these meteorological phenomena, the two links can function in a complementary manner. The RF link is usually designed in the unlicensed X and Ku bands or millimeter...
waves (MMW) around 60 GHz. The last option is especially interesting because there is a larger bandwidth available in the MMW range [361]. Given the LOS propagation of the signal, the related channel fading is well described by the Rice fading model. For hybrid RF/FSO long-span links, the RF link can also be used for beam acquisition and pointing as well as for the purposes of link control in HARQ scenarios due to its higher reliability [362].

Commercially available hybrid RF/FSO products (like fSONA and MRV products) use the RF link just as a back-up channel. Another simple scheme consists of sending the same data on the two channels and to perform signal detection for each frame at the receiver on the “more reliable” channel [361], [363]. However, these approaches, which can be considered as “hard-switching” between the two channels, are not optimal in the sense of exploiting the available resources. Some research works have hence considered more efficient use of RF and FSO links in parallel. However, optimal signaling and routing on a hybrid RF/FSO link is not an easy task.

In intermediate atmospheric conditions, data can simply be partitioned between the two channels and decoded separately at the receiver side. Monitoring constantly the two channels, transmission can progressively be switched to one link or to another as a result of channel condition deterioration [364]. An experimental set-up has been presented in [365] where hybrid LDPC coding is performed on a wireline low bandwidth link used in conjunction with an FSO link. However, separate data encoding and decoding on RF and FSO links does not fully exploit the available “channel diversity,” and joint data encoding and decoding should be performed over the two channels. The so called hybrid channel coding was considered in [366], where data is encoded over the two channels using non-uniform rate-compatible LDPC coding and jointly decoded at the receiver. This scheme, however, requires the CSI at the transmitter. Also, in [367], joint FSO/RF channel coding using Raptor codes was considered in a HARQ scheme with incremental redundancy coding where the coding rate for each channel is adapted to the data rate that the link can support. The advantage of this method over that of [366], which is based on code-rate selection, is that imperfect CSI does not result in a rate mismatch as it is based only on positive or negative acknowledgments from the receiver (i.e., a single bit feedback). In addition, as it is shown in [367], rateless coding is advantageous over rate adaptation schemes where the code rate is adjusted prior to transmission, especially in the strong turbulence regime where severe fades can occur. A similar work in [100] considered hybrid rateless Raptor encoding with the demonstration of a practical system implementation. Another solution is to partition data over the two channels while performing encoding and interleaving. For instance, a BICM scheme using a convolutional code is proposed in [368] under the assumption of unavailable CSI at the transmitter. Optimized punctured turbo-coding and bit selection patterns for the hybrid channel were further proposed in [369]. Finally, adaptive modulation and coding applied to hybrid RF/FSO channels has been considered in [370].

XI. COHERENT FSO SYSTEMS

In contrast to IM/DD systems, in coherent OWC systems the information is encoded on the optical carrier amplitude and phase. The received beam at the receiver is combined with a local oscillator (LO) beam, as shown in Fig.4. This way, after mixing with the LO, the received signal is amplified and the detection process is rather limited by the shot noise [133]. To this reason, coherent detection is usually considered as a means of increasing the receiver sensitivity in FSO systems [371], [372]. In addition, coherent detection allows the rejection of the background noise and intentional interferences [44]. Another interesting property of coherent systems is that information can be sent on the amplitude, phase, or polarization of the optical field, which permits a considerable increase of the system spectral efficiency [373].

Typical modulation schemes used in coherent systems consist of multilevel phase shift keying (PSK) or quadrature amplitude modulation (QAM), or multilevel Polarization shift keying (PolSK) [374]–[376].

Despite all these advantages, commercial FSO systems rely on IM/DD schemes, as explained previously, due to their lower implementation complexity and cost. However, there is an increasing trend to shift to coherent FSO systems thanks to the recent advances in the fabrication of integrated coherent receivers as well as high-speed digital signal processing integrated circuits, which have greatly increased the practicality of coherent detection [377], [378].

In coherent receivers, there are two approaches of homodyne and heterodyne signal detection. Homodyne detection permits a better detection sensitivity but requires an accurate optical phase-locked loop, which is very expensive to realize. Due to this reason, heterodyne detection has been more widely considered in the literature [371]. Among recent experimental demonstrations of coherent FSO links are the homodyne BPSK transmission at 5.625 Gbps over a distance of 142 km [10], and polarization multiplexed quadrature phase-shift-keying (QPSK) transmission at 112 Gbps over a non-turbulent channel [373]. Coherent detection for amplitude modulation was also considered in [379].

From a practical point of view, an important issue is the performance of coherent FSO system in the presence of atmospheric turbulence. While it is argued in [44] that coherent FSO systems experience improved performance against atmospheric turbulence compared to IM/DD systems, a more detailed study in [380] showed that this is only true when the aperture size is limited or when the equivalent non-coherent receiver suffers from significant thermal noise or interference. Indeed, turbulence distorts the coherency of the received signal field, and the resulting imperfect wavefront match between the incoming signal and the LO reduces the received power [133]. The corresponding phase distortion degrades the system performance, in particular when the diameter of the receive aperture is larger than the coherence length of the received signal wavefront. To compensate this phase distortion, either zonal or modal compensation should be deployed [381]. Phase-compensation aims at adaptive tracking of the beam wavefront in order to correct the turbulence-induced aberrations. In
[151], Belmonte and Kahn proposed a statistical model to characterize the combined effects of turbulence-induced wavefront distortion and amplitude fluctuation in coherent receivers assuming modal phase noise compensation. Modeling phase fluctuations by a Gaussian distribution, they also studied the channel capacity for receive diversity systems in [382] for the case of log-normal turbulence and also investigated the performance of several coherent modulation schemes and PAM in [383]. A more detailed analysis was presented in [384] where the impacts of atmospheric turbulence, configuration and parameters of the transmitted and LO beams, and link misalignment on the heterodyne efficiency of a coherent FSO link were investigated.

Considering K-distributed turbulence, Niu et al. studied the performance of coherent FSO links for binary modulation in [385] and for Q-ary PSK and QAM modulation schemes in [386]. Tsiftsis studied the average BER and outage performance of coherent receivers under Gamma-Gamma turbulence in [387]. The performance of coherent heterodyne systems was further studied in [388] for several phase modulation schemes over Gamma-Gamma turbulent channels. The general M-distributed turbulence model was considered in [389], where the error performance of DPSK coherent systems was evaluated.

On the other hand, receive diversity has been shown of special interest in coherent systems [390], [391]. It is shown in [390] that there is much more diversity benefit against fading, background noise, and interfering signals, compared to non-coherent receivers. The performances of pre-detection and post-detection EGC receivers were compared in [392] under Gamma-Gamma turbulence, where the superiority of the latter scheme was demonstrated. Also, a comparison of the error performance of several coherent and SIM modulation schemes was performed in [393] for Gamma-Gamma distributed turbulence and receive diversity systems. Similar to the non-coherent systems, here, in order to maximize the diversity benefit against the phase noises arising from the transmitter, turbulence, and LO, the PDs at the receiver should be spaced sufficiently apart [138]. This allows for independent electronic phase-noise compensation of the multiple transmitter beams.

A number of works have also considered coherent MIMO systems. ST coding for MIMO coherent systems was considered in [394], where a set of code design criteria assuming a large number of transmitters and receivers was proposed based on the minimization of the pairwise error probability. Also, Bayaki et al. presented in [395] simplified ST code design criteria for coherent and differential FSO links in Gamma-Gamma turbulence. They showed that, in contrast to IM/DD systems, OSTBCs are preferable over RC in coherent and differential systems, since RC does not provide full diversity in coherent systems. Also, the performance gain of ST-coded coherent systems over non-coherent systems was shown to be principally due to the superiority of heterodyne detection. Recently, a special coherent MIMO architecture was proposed in [396] which used wavelength diversity and phase noise estimation. Using laser beams operating at different wavelengths at the transmitter and the receiver, wavelength-selective spatial filters are used at the receiver to separate the different transmitted signals. This allows the combination of multiple received signals with different phase noises.

XII. CONCLUSIONS

The design of pervasive and trustworthy next-generation communication networks is recognized as a major technical challenge that researchers face in the next ten years. Development of novel and efficient wireless technologies for a range of transmission links is essential for building future heterogeneous communication networks to support a wide range of service types with various traffic patterns and to meet the ever-increasing demands for higher data rates. We believe that FSO should be considered as an essential component of such heterogeneous networks. With their large optical bandwidth, FSO systems can be used, in some applications, as a powerful alternative to and, in others, as complementary to the existing RF wireless systems.

Terrestrial FSO links with transmission rates of 10 Gbps (assuming a range of few hundred meters) are already in the market and the speeds of recent experimental FSO systems are promising even more. To further push up the limits of FSO systems and overcome the major technical challenges (particularly atmospheric turbulence fading and adverse weather effects), there have been significant recent research efforts on the physical (PHY) layer design issues of FSO systems. These are mainly inspired by several exciting developments that have been witnessed in the area of PHY layer research for RF wireless communications in the last decade or so. PHY layer methods and techniques such as MIMO communication, cooperative diversity, novel channel codes and adaptive transmission have been explored in recent FSO literature and a detailed account of these research efforts is provided in our survey. We hope that this survey will serve as a valuable resource for understanding the current research contributions in the growing area of FSO communications and hopefully prompt further research efforts for the design of next generation FSO systems as a powerful complementary technology to RF systems in the future heterogeneous wireless networks.

ACKNOWLEDGMENT

The authors would like to acknowledge the support by EU Opticwise COST Action IC1101. They wish also to thank Mr. Hatef Nouri for providing the illustrations of the paper and Mr. Sasan Zhalehpoor for his help in preparing the paper tables.

NOMENCLATURE

AF Amplify-and-Forward
APD Avalanche Photo-Diode
ARQ Automatic Repeat reQuest
ASE Amplified Spontaneous Emission
AWGN Additive White Gaussian Noise
VI


This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COST.2014.2329501, IEEE Communications Surveys & Tutorials


[326x149], vol. 12, no. 8, pp. 545–547, Aug. 2008.


J. Hagenauer, “Rate-compatible punctured convolutional codes (RCPC codes) and their applications,” *IEEE Transactions on Communications*, vol. 36, no. 4, pp. 389–400, Apr. 1988.


This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2014.2329501, IEEE Communications Surveys & Tutorials


