Link-Layer Retransmission-Based Error-Control Protocols in FSO Communications: A Survey
Hoang D. Le, Member, IEEE, and Anh T. Pham, Senior Member, IEEE

Abstract—Free space optical (FSO) communication has established a reputation for itself capable of delivering high-speed data services over long distances without exhausting radio frequency (RF) resources. FSO communication can be considered in different network scenarios, including inter-satellite/deep-space links, ground-station/vehicles, satellite/aerial links, and terrestrial links. It is expected to be one of the key enabling technologies for the next generation of 6G wireless networks. Nevertheless, despite the great potential of FSO communications, its performance suffers from various limitations and challenges: atmospheric turbulence, weather conditions, pointing misalignment. The error-control solutions, including physical layer (PHY) and link-layer methods, aim to mitigate the transmission errors caused by such adverse issues. While the existing surveys on error-control solutions in FSO systems primarily focussed on the PHY methods, we instead provide a review of link-layer solutions. In particular, we conduct an extensive literature survey of state-of-the-art retransmission protocols, both automatic repeat request (ARQ) and hybrid ARQ (HARQ), for various FSO communication scenarios, including point-to-point terrestrial, cooperative, multi-hop relaying, hybrid FSO/RF, satellite/aerial, and deep-space systems. Furthermore, we provide a survey of recent literature and insightful discussion on the cross-layer design frameworks related to link-layer retransmission protocols in FSO communication networks. Finally, the lessons learned, design guidelines, related open issues, and future research directions are exposed.

Index Terms—Free-space optical (FSO) communications, link layer error-control protocols, automatic repeat request (ARQ), hybrid ARQ (HARQ), cross-layer design.

NOMENCLATURE

5G Fifth Generation Wireless Mobile Network
6G Sixth Generation Wireless Mobile Network
ACK acknowledgment
AI Artificial Intelligence
AMC Adaptive Modulation and Coding
AoA Angle of Arrival
AP Adaptive Power
AR Adaptive Rate
ARQ Automatic Repeat Request
BER Bit Error Rate
C-ARQ Cooperative Automatic Repeat Request
CC Chase Combining
C-HARQ Cooperative Hybrid Automatic Repeat Request
CRC Standard Cyclic Redundancy Check
CSI Channel State Information
DF Decode-and-Forward
DLR German Aerospace Centre
ECC Error Correction Code
EE Energy Efficiency
FC Frame Combining
FEC Forward Error Correction
FER Frame Error Rate
FLR Frame Loss Rate
FSO Free-Space Optical
GBN Go-Back-N
HAPs High Altitude Platforms
HARQ Hybrid Automatic Repeat Request
IM/DDIntensity Modulation/Direct Detection
IoV Internet of Vehicles
IR Incremental Redundancy
IRS Intelligent reflecting surface
LEO Low Earth Orbit
LT Luby Transform
M-C-ARQ Modified Cooperative ARQ
mmWave Millimeter-Wave
MRC Maximum Ratio Combining
mURLLC Massive ultra-reliable and low latency communications
NAK Negative acknowledgment
NOMA Non-Orthogonal Multiple Access
OSI Open Systems Interconnection
PAA Point-Ahead-Angle
PHY Physical Layer
PPM Pulse Position Modulation
QKD Quantum Key Distribution
QoS Quality of Service
Qubit Quantum Bit
RCPC Rate-Compatible Punctured Convolutional
RF Radio Frequency
RS Reed-Solomon
RTT Round Trip Time
SACK TCP-Selective acknowledgment
SE Spectral Efficiency
SI Scintillation Index
SLP Segment Loss Probability
SNR Signal-to-Noise Ratio
SR Selective Repeat
I. INTRODUCTION

The rapid development of various emerging applications, such as virtual reality (VR), augmented reality, Internet of Vehicles (IoV), or smart applications with the aid of artificial intelligence (AI) and machine learning, has produced a massive volume of data traffic that requires extremely high-speed wireless connectivity [1]. According to the International Telecommunication Union (ITU), the global mobile traffic volume was 7.462 EB/month in 2010 and is predicted to be 5016 EB/month in 2030 [2]. Indeed, the support of this extremely high volume data poses a significant challenge for the forthcoming fifth-generation (5G) and beyond wireless communication networks.

It is well established that the current communications based on radio frequency (RF) are becoming more restricted due to the limited spectrum resources and may not satisfy this growing demand [3]–[5]. As a result, recent years have witnessed an increasing interest in free-space optical (FSO) communications research and development. FSO has been proposed as an alternative or complementary solution to the current RF, thanks to the enormous available unlicensed bandwidth and the capability of transmission at very high data rates over long distances [6]. Furthermore, as compared to existing RF-based wireless systems, the narrow and directional characteristics of a laser beam employed in FSO communications enable a high level of security, a low power consumption, and an immunity to electromagnetic interference [7].

FSO systems can be classified into two broad categories, i.e., terrestrial and space links [8]. As an example illustrated in Fig. 1, the terrestrial links can be a connection between building-to-building. In addition, the space links include inter-orbital (e.g., satellite-to-vehicles, between unmanned aerial vehicles (UAVs) or high-altitude platforms (HAPs)), inter-satellite, and deep space links. The FSO transmissions through the atmosphere are, nevertheless, not without challenges. The primary concerns of FSO links is briefly summarized in Table I, in which different systems may experience different adverse issues [9], [10]. Indeed, those adverse issues pose various challenges to the performance of FSO systems, which requires a lot of research efforts for tackling.

A. Error-Control Methods in FSO Communications

Extensive studies have been devoted to error-control solutions, which can be mainly categorized into two groups: physical (PHY) layer and link-layer methods. Notably, the PHY methods widely used in FSO communications include adaptive modulation and coding (AMC) [11], adaptive rate (AR)/power (AP) transmissions [12], forward error correction (FEC) code [13], multi-hop transmissions [14], cooperative diversity technique [15], aperture averaging [16], hybrid RF/FSO [17], and signal processing techniques [18]. Such methods can considerably mitigate the transmission errors caused by the aforementioned adverse issues on FSO links, thus improving the system’s reliability and availability. For instance, as reported in [13], by using FEC-based Turbo codes with a coding rate of 1/3, the terrestrial FSO systems can retain a level of bit error rate (BER) of $10^{-6}$ in the moderate-to-strong turbulence conditions. Indeed, a lower coding rate is required to maintain a lower BER level, leading to the inefficiency in system throughput performance under time-varying channel conditions, as many redundancy bits are required. As a result, it is difficult and not cost-effective for the PHY layer to guarantee error-free reception.

To further enhance the reliability and efficiency in communication systems, link-layer error-control methods, including redundancy and retransmission mechanisms, have been widely investigated in the context of FSO communications. A typical redundancy mechanism is the error correction code (ECC), which guarantees transmission reliability by adding some redundancies to the original message so that receivers can use them to recover the erroneous data [19]. In addition, ARQ is one of the well-known retransmission mechanisms, which facilitates the retransmission of erroneously received frames via feedback from the receiver to the transmitter [20], [21]. Under the impact of severe channel impairments, a more
robust retransmission-based error-control method, i.e., hybrid ARQ (HARQ), which achieves better reliability by combining ARQ and ECC, is preferable to standard ARQ and ECC in some scenarios, e.g., long-distance satellite communication systems [22]. However, the main drawbacks of HARQ protocols are system complexity, additional signaling, and large overhead.

B. Relevant Survey/Tutorial Articles

A couple of surveys and tutorials related to error-control methods appeared in the literature of FSO communications, which mainly focused on the PHY layer solutions [6], [23]–[33]. Khalighi and Uysal reviewed several PHY layer error-control methods used in FSO-based terrestrial systems, including FEC code, spatial diversity technique, adaptive transmissions, relay-assisted cooperative transmissions, and hybrid FSO/RF schemes in [23]. A survey paper by Kaushal and Kadoum [24] summarized and reviewed recent works on the PHY error-control solutions in FSO-based space communications, including satellite-to-ground, ground-to-satellite, and inter-satellite systems. Various PHY methods, e.g., aperture averaging, diversity, relay transmissions, adaptive optics, signal processing techniques, hybrid FSO/RF, were presented for such systems. Another survey by Son and Mao [25] focused on both terrestrial and space communications, in which a short review of PHY error-control methods was provided, i.e., cooperative diversity technique, relay-assisted communications, hybrid FSO/RF, and signal processing techniques. In [26], Kaur and Singh provided a short survey on hybrid FSO/RF studies published from 2015 to 2019. Vavoulas et al. [27] presented a survey on ultraviolet C-band for FSO communications and PHY methods, i.e., signal processing techniques, diversity techniques, and FEC code for such systems. The surveys and tutorials on signal processing technique of PHY error-control methods were reported in [28] and [30]. In addition, Raj and Majumder [29] presented the state-of-art developments of FSO communications, wherein several PHY solutions, i.e., aperture averaging, waveform correction, and hybrid FSO/RF, were reviewed. In [6], Chowdhury et al. provided a comprehensive overview of existing literature on optical wireless hybrid networks, such as RF/optical and optical/optical systems. A relay-assisted technology in FSO communications was surveyed by Liu et al. [31] for various systems, from the terrestrial to space communications. They also presented a literature review regarding the signal processing techniques, channel coding, adaptive transmission, cooperative diversity for FSO communications. Trichili et al. [32] provided an up-to-date review of PHY error-control solutions, containing channel coding, diversity, adaptive optics, relay transmission, signal processing techniques, and hybrid FSO/RF, in terrestrial FSO communications. Most recently, Jahid et al. [33] conducted a comprehensive survey on several PHY error-control solutions, including aperture averaging, adaptive optics, relay-aided transmissions, channel coding, cooperative diversity, and hybrid FSO/RF schemes, for both terrestrial and space networks. The aforementioned surveys with regards to error-control methods are summarized at a glance in Table II, which allows readers to capture the major contributions of each of the existing surveys.

At the time of writing this paper, we realized that a detailed survey of link-layer error-control solutions and their cross-layer design is not available in the literature of FSO communications. Although these issues were briefly discussed in [24], [26], [33], these studies lack a comprehensive/in-depth survey/tutorial for the issues because their main focus was the PHY error-control methods. In addition, among link-layer error-control solutions, retransmission-based methods, including ARQ and HARQ, do not apply to FSO communications, making them a fertile research area in the domain of FSO communications. The FSO links operate at extremely high data rates (several Gb/s) over the slow fading channels (the channel coherence time in the order of milliseconds). As a result, it is not inherently allowed the efficient use of ECC schemes, especially when the probability of the channel in deep fade is non-zero [43]. As a result, the main focus of this paper is the link-layer retransmission protocols and their cross-layer design for various FSO communication scenarios.

C. Motivations and Contributions

While the link-layer retransmission protocols have been widely surveyed in the literature of RF communications [34]–[42] as summarized in Table III, the adoption of these protocols for RF systems can not be straightforwardly applied to the FSO ones. This is because there exist many fundamental specificities in the transmission protocols, fading channels, and modulation schemes of FSO systems that make them different from RF ones [33]. More explicitly, we have to point out the challenging issues when applying link-layer retransmission protocols in FSO systems compared to the RF ones. The design of such protocols needs to be examined and investigated carefully in the context of FSO, as follows.

- **Huge bandwidth of FSO communications:** As the bandwidth of FSO systems is much larger than that of RF ones, it may not be suitable for conventional error-control designs to exploit such massive bandwidth fully. For example, the stop-and-wait (SW) ARQ protocol performs well in RF communications [44]–[46]. However, it becomes unsuitable for high-speed FSO systems due to its inefficient bandwidth utilization. The reason is that the idle time waiting for the acknowledgment of a data frame (twice the propagation delay) in SW-ARQ becomes significant compared with the transmission delay, which becomes very small thanks to the high data rate of FSO communications [21]. This leads to the poor throughput performance of SW-ARQ over FSO communications, for which a proper design needs to be considered.

- **Turbulence-induced FSO fading channels:** In the design and analysis of FSO’s link-layer retransmission protocols, one of the critical issues comes from the modeling of the time-varying behavior of turbulence-induced fading channels [47]. The temporal coherence time of the atmospheric turbulence process is typical of the order of milliseconds (i.e., slow fading). For FSO links operating at high data rates, the error probabilities of frames
are highly correlated. That means the frame errors tend to occur in burst patterns during the transmission [48], [49]. Therefore, it is necessary to understand the correlated error transmission structure to provide an accurate performance analysis of the FSO systems. It is important to note that the conventional uniform error model for RF systems uses the error structure of different frame transmissions that are assumed to be independent [42]. Moreover, as the FSO fading channel models are entirely different from those of RF [33], it would be essential to have a proper protocol design under the atmospheric turbulence conditions in terms of system performance optimizations.

Besides, unlike RF systems, the feedback issue is relatively easy to implement in FSO, making the cross-layer design of link-layer retransmission protocols popular, as follows.

- **Feedback implementation:** The issue of imperfect/outdated channel state information (CSI) feedback is critical in the implementation of link-layer retransmission protocols with PHY adaptation schemes (e.g., rate/power...
adaptation) in RF systems [50]. Nevertheless, the CSI feedback is relatively easy to implement in FSO systems. Practically, transmitters can estimate the forward link’s CSI estimation by using the reciprocal channel [51]. In addition, the feedback channel reliability can be guaranteed by a robust error correction code thanks to the abundantly available bandwidth in FSO systems.

Looking at the popularity and above critical differences, it is of importance and necessity to provide an overview of current studies and tutorials associated with the design and performance of link-layer retransmission protocols and their cross-layer frameworks in various FSO communication scenarios.

In this paper, our focus is a comprehensive review of the design and performance evaluation of link-layer retransmission-based error-control protocols in the context of FSO communications. The main contributions of the paper can be summarized as follows.

1) It is an up-to-date review of link-layer retransmission-based error-control protocols, both ARQ and HARQ, in various FSO communication scenarios, including point-to-point terrestrial, cooperative, multi-hop relaying, hybrid FSO/RF, satellite/aerial, and deep-space systems.

2) A survey and insightful discussion on the cross-layer design frameworks of link-layer retransmission protocols with PHY methods and/or transport layer protocols in the context of FSO communication networks are also provided.

3) From the holistic survey, we dedicate elaborated lessons learned section and outline future research directions to develop link-layer retransmission protocols and their cross-layer design frameworks in FSO communication networks.

4) Based on the lessons learned from this survey paper, we derive generic design guidelines recommended for effectively designing such protocols in each FSO network scenario.

5) Finally, we identify the important research challenges and discuss the open issues for such protocols in the vision of future sixth-generation (6G) wireless communication networks.

It is noteworthy that a highlight of our contributions compared to existing surveys relating to error-control methods in the context of FSO communications is illustrated in Table II. Additionally, the difference of our work compared to ones in RF communications is briefly summarized in Table III. The details of basic components of a generic FSO link, such as light sources, photodetectors, and modulation schemes, together with channel modeling and standardization activities, are, however, beyond the scope of this paper. Interested readers can refer to the detailed studies in [7], [23], [24], [32], [33] for the background and projects of FSO communication systems.

D. Paper Organization

The remainder of this paper is organized as follows. Section II presents state-of-the-art literature on ARQ protocols for different FSO communication scenarios, including point-to-point terrestrial, cooperative, hybrid FSO/RF, satellite/aerial, and deep-space systems, and then outlines future research directions. In Section III, we provide an extensive review of the HARQ protocols in various FSO systems, i.e., point-to-point terrestrial, cooperative, multi-hop relaying, hybrid FSO/RF, satellite/aerial, and then point out the potential research topics for such protocols. We dedicate Section IV to introduce and discuss cross-layer design frameworks related to the link-layer retransmission protocols, including the joint design of physical-layer/link-layer and link-layer/transport-layer in FSO communication networks. As a result of our extensive survey, Section V provides lessons learned, design guidelines, and discusses the challenges as well as open issues for link-layer retransmission protocols in FSO communications. Finally, Section VI concludes the paper. For the sake of explicit clarity, the organization of this paper is depicted in Fig. 2.

II. ARQ AIDED FSO COMMUNICATIONS

The ARQ protocol, one of the most popular link-layer error-control schemes, is an efficient solution to increase link
reliability. The advantages of ARQ protocols have been proven in RF-based wireless systems for a long time. Nevertheless, the high data-rate transmission in FSO systems over time-varying atmospheric turbulence channels poses new challenges to the design and performance of such protocols. A proper design of ARQ protocols for new/future immense-bandwidth FSO systems becomes an essential aspect of being considered. To this end, in this section, we first revisit the background of ARQ protocols. Next, we provide an overview of the state-of-art design and performance evaluation of ARQ protocols in various FSO communication scenarios covering from point-to-point terrestrial, cooperative, hybrid FSO/RF, satellite/aerial, to deep-space systems. Finally, we summarize the primary characteristics of reviewed ARQ protocols investigated in various FSO systems, followed by possible future directions for such protocols.

A. Overview of ARQ Protocols

Background: The ARQ protocol provides a reliable transmission service for the link layer by detecting and retransmitting erroneous frames. More specifically, the transmitter sends a frame consisting of data and an error detection code, such as the standard cyclic redundancy check (CRC). The receiver can verify the integrity of the received frame using the error detection code. Then, based on the verification result, a feedback message, i.e., either a positive acknowledgment (ACK) or negative acknowledgment (NAK), is returned to the transmitter. The transmitter can then decide to retransmit the frame or proceed with the next frame.

Classification: Based on the retransmission strategies, there are two basic types of ARQ schemes: stop-and-wait (SW) and sliding window protocols as depicted in Fig. 3(a). The sliding window ARQ is further categorized into go-back-N (GBN) and selective repeat (SR), and the operation of ARQ schemes can be briefly described as follows [52]:

- **Stop-and-wait ARQ:** The transmitter sends a single frame at a time and waits for an ACK from the receiver side. If the ACK is received, a new frame will be transmitted. Otherwise, frame retransmission is activated.
- **Go-Back-N ARQ:** The transmitter sends multiple frames specified by a window size (denoted as \( W \)) and keeps track of the frame index within the window size. When a frame’s index is not as expected (i.e., out of order or duplicate), it will return a NAK message containing the index ACK for the last correct in-order frame. The transmitter starts retransmission of its entire window, commencing from the most recent positively acknowledged frame, and continues the process over again.
- **Selective Repeat ARQ:** The data frames are continuously transmitted without waiting for acknowledgment from the receiver, as in the GBN-ARQ. However, unlike the GBN-ARQ, the receiver accepts out-of-order frames and buffers them, and only missing/negatively acknowledged frames are retransmitted.

Comparison between ARQ protocols: Of the three ARQ protocols, as illustrated in Fig. 3(b), SR-ARQ is able to achieve the highest transmission efficiency but its implementation is the most complicated. Further discussions on the comparison of ARQ protocols regarding the complexity and transmission efficiency are as follows.

- **Complexity:** The complexity comes from the protocol implementation, which is mainly related to (i) frame buffering and (ii) timer management. First, a buffer is required at the transmitter side for GBN-ARQ and SR-ARQ due to the transmission of multiple frames. In SW-ARQ, no transmitter’s buffer is needed. Additionally, as the receiver accepts and keeps out-of-order frames, a receiver’s buffer is required for the SR-ARQ. On the other hand, only one timer is necessary for each transmission round for SW-ARQ and GBN-ARQ. In contrast, multiple timers are required for each transmitted frame in the SR-ARQ. Therefore, it is clear that SR-ARQ is the most complicated protocol, while SW-ARQ is the simplest one in terms of implementation. The comparison of ARQ protocols concerning the complexity is summarized in Table IV.

- **Transmission Efficiency:** The performance of a system using ARQ protocols can be evaluated by its reliability and transmission efficiency [34]. All three basic ARQ protocols achieve the same reliability with robust yet less overhead error-detection code; they, nonetheless, provide

![Figure 3](image-url)
TABLE V
Transmission Efficiency Comparison of ARQ Protocols in High-Speed Satellite Communications [53, Ch. 5], [34]

| ARQ Type | Expression | Bit error rate |
|----------|------------|----------------|
| SW-ARQ   | \[\eta_{SW} = \frac{1 - \frac{n_o}{n_f}}{1 + \frac{n_o}{n_f} + \frac{t_{prop} + t_{proc}}{n_f}} (1 - P_f) \] | \begin{tabular}{c|c|c|c|c|c}
|          | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ | $10^{-8}$ |
|----------|-----------|-----------|-----------|-----------|-----------|
|          | 0.0899%   | 0.2211%   | 0.2419%   | 0.2441%   | 0.2444%   |
| GBN-ARQ  | \[\eta_{GBN} = \frac{1 - \frac{n_o}{n_f}}{1 + (W-1)} (1 - P_f) \] | \begin{tabular}{c|c|c|c|c|c}
|          | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ | $10^{-8}$ |
|----------|-----------|-----------|-----------|-----------|-----------|
|          | 4.9240%   | 45.4359%  | 88.2444%  | 96.9332%  | 97.8923%  |
| SR-ARQ   | \[\eta_{SR} = \left(1 - \frac{n_o}{n_f}\right)(1 - P_f) \] | \begin{tabular}{c|c|c|c|c|c}
|          | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ | $10^{-8}$ |
|----------|-----------|-----------|-----------|-----------|-----------|
|          | 36.0504%  | 88.6740%  | 97.0249%  | 97.9020%  | 97.9902%  |

different transmission efficiencies [34]. The transmission efficiency is defined as the ratio of useful data rate over the channel bit rate. In Table V, the transmission efficiency of SW-ARQ, GBN-ARQ, and SR-ARQ protocols are expressed in (1), (2), and (3), respectively. Here, \( R_b \) is the data bit rate, \( W \) is the window size, while \( t_{prop} \) and \( t_{proc} \) are respectively the propagation and processing delays. Additionally, \( n_o \) is the size of acknowledgment frame (either ACK or NAK), \( n_f \) is the frame size containing the overhead of \( n_o \) and \( n_f - n_o \) of data, and \( P_f \) is the frame error rate (FER) computed as \( P_f = 1 - (1 - BER)^{n_f} \). The details of these expressions can be found in [53, Ch. 5]. In this example, we compare the transmission efficiency of ARQ protocols in high-speed and long-distance satellite communications. Also, \( R_b = 1 \) Gbps, satellite altitude \( H_s = 600 \) km, \( n_f = 1250 \) bytes, \( n_o = 25 \) bytes, \( n_a = 25 \) bytes, and \( W = 11 \) frames. As seen, SW-ARQ is not practically useful in long fat networks as it retains a very low level of transmission efficiency. Among sliding window protocols, SR-ARQ can achieve higher transmission efficiency than GBN-ARQ because only missing or negatively acknowledged frames are retransmitted in SR-ARQ.

B. ARQ Protocols in FSO Systems

The ARQ protocols have been widely studied for FSO communications, and several variants and modifications were adopted for reliable and very high-throughput systems. A summary of recent studies on ARQ protocols for various FSO system scenarios, including point-to-point terrestrial, cooperative, hybrid FSO/RF, satellite/aerial, and deep-space systems, is given in Table VI. The following provides further details on the design and performance estimation of ARQ protocols studied in such scenarios.

1) Point-to-Point Terrestrial Systems: Atmospheric turbulence is one of the most challenging issues for the widespread deployment of the FSO terrestrial systems, e.g., building-to-building [66]. Early work on ARQ protocols for reliable transmission in FSO terrestrial systems has been reported in [54], [55]. In various turbulence conditions, these studies quantitatively compared two well-known link-layer error-control solutions, including error correction code (ECC) schemes and ARQ protocols. It was shown that ARQ is more efficient than ECC in weak turbulence regimes. At the same time, its performance is degraded in strong turbulence conditions, as many retransmissions are required.

**SW-ARQ Design:** The design of ARQ protocols over weak-to-strong turbulence channels would be beneficial and more effective if combined with PHY’s error-control solutions. In FSO systems, the adaptive rate (AR) scheme is one of the most popular PHY solutions [11], and the joint design between ARQ and AR scheme is attractive to counteract the transmission errors without using costly FEC schemes at the PHY layer. Integration in the design of standard SW-ARQ protocol and AR scheme for FSO system was studied in [20], [59]. For such joint design, the proposal of frame combining (FC) ARQ scheme realized by implementing frame combining at the receiver side was also provided to increase the system reliability [59]. In FSO systems using SW-ARQ with FC, the receiver can store all copies of previously erroneous frames to jointly decode with retransmission ones. This improves the likelihood of successfully retransmitting frames. In practical systems, as the receiver always has a buffer to store frames before delivering them to end-users, this solution does not introduce additional system costs. As a result, the SW-ARQ with FC performs more efficiently than standard SW-ARQ in different turbulence conditions [59].

**Sliding Window ARQ Design:** It is worth noting that the SW-ARQ protocol is practically not useful, especially in point-to-point high-speed systems, due to its inefficient bandwidth utilization. Practical systems employ the sliding window protocols, i.e., GBN and SR, which allow higher efficiency. Novel design of SR-ARQ with pulse position modulation (PPM) hard decision for FSO systems was reported in [63]. The advantages of this novel ARQ design in terms of latency and efficiency over the traditional one come from two remarkable issues. First, it uses the decision result of each PPM symbol to detect errors without the need for parity-check code (reducing the latency caused by the parity-check code’s generation and verification). Second, it relies on the special NAK signals to
retransmit only erroneous PPM symbols instead of the whole data frame. As the number of erroneous PPM symbols is often less than half of the received frame size, the system efficiency can be considerably improved.

In the design and analysis of the sliding window ARQ in FSO systems, one of the key issues comes from modeling the time-varying behavior of turbulence fading channels. For FSO links operating at high data rates, the temporal coherence time of the atmospheric turbulence process is of the order of milliseconds, leading to the highly correlated frame errors, i.e., frame errors tend to happen in bursts [48], [49]. This issue, nonetheless, was not mentioned for the design of FSO systems using SW-ARQ as a single frame is transmitted per round trip delay. Studies in [21], [47], [67] report the design and analysis of sliding window ARQ protocols considering that critical issue. These studies also investigated the joint design between ARQ protocols (both GBN and SR mechanisms) with AR burst transmissions, where the window size of ARQ is designed based on the burst duration. The obtained results demonstrated the outperformance of sliding window ARQ protocols compared to the SW-ARQ one in high-speed FSO communications [21].

2) Cooperative Systems: Cooperative diversity has been widely studied for FSO communications to realize spatial diversity advantages [68]–[70]. The key idea behind cooperative diversity is based on observing the FSO turbulence channels. The signal transmitted by a source node is overheard by others nodes, i.e., called relay nodes. This introduces additional degrees of freedom in the spatial domain, thus significantly reducing the impact of atmospheric turbulence. For cooperative FSO systems, parallel relaying can be implemented through the use of multiple transmitter apertures directed to relay nodes, as shown in Fig. 4.

As reported in [71], the cooperation through relay nodes is beneficial only if the received signal-to-noise ratio (SNR) is high enough; otherwise, relay nodes likely forward the corrupted copies of data, leading to performance deterioration. To tackle the issue, an effective design of ARQ protocols, together with the benefit of cooperative diversity, could offer significant performance enhancement for FSO systems. The design and performance evaluation of cooperative ARQ (C-ARQ) protocol was reported in [56], where the SW mechanism was used for C-ARQ. Figure 4 illustrates an FSO system model for C-ARQ protocol, in which the communication between a source (S) node and a destination (D) node is

| Ref. | Year | Objective | ARQ Type | System | Remarks |
|------|------|-----------|----------|--------|---------|
| [54] | 2008 | FLR       |          |        |         |
| [55] | 2013 | Throughput, PER |        |        | A comparison between SR-ARQ and ECC using LT codes in different turbulence conditions |
| [56] | 2014 | Goodput, Delay, PER | ✓ | ✓ | The proposed M-C-ARQ outperforms the conventional C-ARQ in cooperative FSO systems |
| [57] | 2014 | Transmission Efficiency, PER | ✓ | ✓ | An investigation of ARQ protocols in long-distance HAP-to-HAP FSO links |
| [58] | 2015 | Goodput, Outage Probability, PER | ✓ | ✓ | Analysis of SW-ARQ in adaptive multi-rate hybrid FSO/RF systems |
| [59] | 2016 | SE, Outage Probability, PER | ✓ | ✓ | Two cross-layer designs are considered: standard AR/ARQ and AR/ARQ with PC |
| [60] | 2016 | Throughput, Delay, FLR |        | ✓ | An investigation of disruption-tolerant SR-ARQ protocols in the context of deep-space systems |
| [61] | 2017 | Throughput |        | ✓ | Study the tradeoff between ARQ feedback rate and ARQ protocol efficiency in satellite systems |
| [62] | 2017 | Throughput, Outage Probability | ✓ | ✓ | Evaluate the effect of adaptive power allocation between the ARQ retransmissions |
| [63] | 2017 | Throughput | ✓ | ✓ | A proposal of novel ARQ with PPM hard decision for FSO systems |
| [21] | 2019 | Throughput, Delay, PER | ✓ | ✓ | Highlight the cross-layer design of sliding window ARQ and AR in FSO systems |
| [64] | 2019 | Throughput |        | ✓ | Experimentally demonstrate the use of SR-ARQ protocol in FSO-based satellite systems |
| [65] | 2020 | Goodput, Outage Probability, PER | ✓ | ✓ | Analysis of SW-ARQ in hybrid FSO/RF systems with FSO links described by Malaga model |

Fig. 4. An FSO system model for cooperative ARQ protocols over atmospheric turbulence channels [56].
achieved through $N$ relay nodes ($R_i, 1 \leq i \leq N$) placed between the S node and the D node. This study introduced two design approaches: conventional C-ARQ and modified C-ARQ (M-C-ARQ) schemes.

**Conventional C-ARQ Scheme:** In the conventional C-ARQ, the S node broadcasts a data frame to all R nodes for the first transmission. These R nodes, which can successfully decode the data from S, forward the frame to D. At the destination side, if the frame is successfully received by the D node, it will send an ACK message to the S node through R nodes; otherwise, a NAK message is sent. The failure occurs when either no R nodes can decode the frame successfully or the presence of errors is detected by CRC at the D node. In that case, the S node will retransmit that frame until reaching a predefined maximum number of retransmission attempts for a frame, i.e., $N_t$ times. If a frame does not get through the FSO links after the ARQ’s persistent level of $N_t$ transmission attempts, the S node gives up, and that frame is clarified to be lost.

**M-C-ARQ Scheme:** The main drawback of the conventional C-ARQ protocol is the additional delay and energy consumption, especially in strong turbulence conditions. This is because the S node has to re-broadcast frames to all R nodes, including those successfully received by R nodes before, if frame errors are detected at the D node. In that case, the S node will retransmit that frame until reaching a predefined maximum number of retransmission attempts for a frame, i.e., $N_t$ times. If a frame does not get through the FSO links after the ARQ’s persistent level of $N_t$ transmission attempts, the S node gives up, and that frame is clarified to be lost.

**Performance Comparison:** Figure 5 shows the performance comparison of conventional C-ARQ and M-C-ARQ over FSO turbulence channels. Transmitted powers vary when the number of relay nodes $N = 2$, ARQ’s persistent level $N_t = 4$, the turbulence strength $C_n^2 = 10^{-6}m^{-2/3}$, and the distance from the S node to the D node as $l_{SD} = 3000$ m. Interested readers can refer to the study in [56] for the details of derived equations for obtained results. In addition, several performance metrics, including FER, Goodput, and energy efficiency (EE), are investigated in Figs. 5(a), 5(b), and 5(c), respectively. As is expected, the M-C-ARQ protocol offers better performance than the C-ARQ one over FSO turbulence channels.

3) **Hybrid FSO/RF Systems:** The presence of atmospheric turbulence, fogs, clouds, and pointing errors on the FSO links leads to frequent link outages and dramatic degradation of system performance. An efficient solution is the hybrid system, where the RF link serves as a backup link in case of FSO link failure [72]. This is because the RF link is less subject to the atmospheric turbulence, pointing errors [73], and is also much less affected by fogs and clouds [74]. For example, the impact of fog and rain on the corresponding FSO and RF links is drastically, but these factors rarely happen simultaneously [23]. As a result, two links can function in a complementary manner.

Driven by the potential of hybrid FSO/RF systems, several studies addressed the performance of ARQ protocols in such systems [58], [62], [65], [75]. The challenge on the design of ARQ protocols for hybrid FSO/RF systems comes from the fact that the corresponding data rate in the RF channel is slower than that in the main FSO link. Moreover, the relative fading channel coherence time for such two links is different, e.g., typically on hundreds of microseconds for millimeter-Wave (mmWave) links [76] and a few milliseconds for FSO links [77]. Therefore, there is a critical need to design ARQ protocols considering those issues properly.

In many existing studies of ARQ protocol, the SW mechanism is often used thanks to its analysis simplicity. Notably, the performance of the SW-ARQ protocol was analyzed in adaptive multi-rate hybrid FSO/RF systems using the link-switching scheme [58], [65], [75]. In the design of ARQ protocol handling one frame at a time, its frame size is adjusted accordingly to transmission modes with different data rates to satisfy a predefined link-layer FER. The design of rate adaptation allows switching between two links gradually to reduce the frequent link switching in error-prone environments of conventional fixed-rate design, thus significantly improving the
performance of SW-ARQ protocol [58], [75]. Besides the popular Gamma-Gamma model for FSO turbulence channel [58], [75], the Malaga distribution model is also used to offer a more accurate SW-ARQ protocol performance estimation [65].

Unlike [58], [65], [75], Makki et al. analyzed the performance of SW-ARQ in fixed-rate hybrid FSO/RF systems, where ARQ frames in FSO and RF modes are simultaneously transmitted to the receiver. These frames are then combined and jointly decoded at the receiver side. In addition, the adaptive power allocation scheme for ARQ retransmissions was also investigated. Instead of using the same power for each retransmission, the basic idea for this scheme is to weigh the energy in each ARQ’s retransmission round by its consumption probability. Additional energy is assigned to the last retransmissions, which are rarely used. This results in a significant improvement in system performance.

4) Satellite/Aerial Systems: The advancement in space technology and the development of sophisticated space-based instruments have opened a new chapter for FSO-based space communications [78], [79], e.g., FSO connections from satellites to ground stations, for inter-satellite, or between high platform altitudes (HAPs) [24]. It is worth noting that the challenging issues involved in FSO-based space links are different in different scenarios, as summarized in Table I.

For reliable satellite/aerial communications, ARQ protocols have been investigated for inter-HAPs [57], [80] and satellite [64] systems. The link distances of such communication systems are typically a few hundreds of kilometers, resulting in high latency [81]. Therefore, the biggest challenge on the design of ARQ protocols is the satisfactory delay and throughput performance, as they require retransmissions for erroneous frames under the time-varying turbulence fading channels. The frequent retransmissions also degrade the system’s energy efficiency, which is especially important for the limited power budget of satellite/aerial-based FSO systems. As mentioned earlier, an ARQ scheme cannot be designed generically for all FSO applications. They have to be tailored for each specific network scenario experienced with different challenging issues.

**ARQ Over HAP-to-HAP FSO Links:** The use of ARQ protocols for FSO-based inter-HAP communications was addressed by the German Aerospace Centre (DLR) [57], [80]. The performance of different ARQ variants, including SW, GBN, and SR, was estimated under the combined impact of pointing error and atmospheric turbulence. Table VII shows the minimum FER required to retain maximum TE performance, for which a suitable ARQ scheme could be chosen for different link distances and turbulence conditions indicated by the scintillation index (SI) [57]. As seen, the SW-ARQ protocol fails in these scenarios, while the minimum required FERs for GBN-ARQ and SR-ARQ are respectively $10^{-5}$ and $10^{-2}$ in different turbulence conditions. This is because the effect of turbulence is generally small at the HAP’s altitude, i.e., 20 km, resulting in the same FER level requirement. Moreover, the SR-ARQ protocol can work well with higher FER levels, but it is more complex in terms of implementations than GBN-ARQ.

**ARQ Over Satellite-Ground FSO Links:** A challenging issue on the design of ARQ protocols is the capability to support extremely high data rate (i.e., the order of tens of Gbps up to Tbps) over the long-distance of satellite communications, e.g., the NASA’s Terabyte Infrared Delivery program [82]. In [64], Schieler et al. successfully demonstrated the error-free FSO communications using SR-ARQ protocol at the data rate close to 100 Gbps in various atmospheric turbulence conditions for low Earth-orbit (LEO) satellite systems. To achieve that extremely high data rate, the ARQ’s frame size is designed to be as large as 10 MB so that the window size is shorter and close to the fading channel coherence time, which varies from 1 ms to 100 ms (a condition generally observed at night). In addition, due to the limitation of feedback channel capacities in FSO-based satellite communications, the tradeoff between ARQ’s feedback rate and its throughput performance was investigated in [61].

**5) Deep-Space Systems:** The demand for high data rate transmissions is expected in future deep-space missions (on Mars, Mercury, etc.), which enables a wide range of services to support remote human operations [35]. With the rapid development of deep space exploration, a new era of study on FSO communications in deep space is of widespread concern, thanks to its high-speed connectivity and low power consumption compared to the RF counterpart [83]. Several missions on the optical communication test and space exploration have been conducted with an FSO deep space communication system (see references therein [84]). Nevertheless, various adverse issues, such as path loss, pointing errors, and coronal solar wind turbulence, pose various challenges to the FSO system design for deep space missions [85]. Indeed, the selection of proper error-control methods plays an essential role in such systems with long propagation delay, limited onboard storage, and time-constrained visibility window between spacecraft and ground stations [86].

One of the reliability options currently available for FSO deep space systems is the use of ARQ protocols reported by NASA in [60]. In this study, the reliability of the FSO system for envisioned deep space mission scenarios is assured by the SR-ARQ protocol. Here, reliable RF-based communication is used for the feedback channel. The disruption-tolerant ARQ, such as Licklider Transmission Protocol [87] of the Disruption-Tolerant Networking suite [88], is assumed, in which ARQ process is aware of the link schedule with the delay prediction and correspondingly adjusts timeout timers. Clearly, for such FSO-based deep space systems, the deployment of ARQ protocols strongly affects the latency.
performance, as erroneous data frames require a random number of retransmission attempts until they are successfully received. Such delay may also reorder the arrival of data units, which may impose additional delay for the high-layer information context. For example, the receiver must wait until all data frames forming a single TCP segment are received.

C. Summary and Future Work

This section reviewed state of the art on design and performance evaluation of ARQ protocols in different FSO system scenarios. The main findings of this section can be summarized as follows.

- Most of the existing works in FSO-based terrestrial systems, i.e., point-to-point, cooperative, and hybrid FSO/RF communications, focussed on the design and performance evaluation of SW-ARQ protocol for the sake of simplicity in the analysis. In addition, the SW-ARQ protocol with the concept of frame combining [59] was shown to be more efficient than the standard SW-ARQ one in terrestrial FSO communications.
- Sliding window protocols, including GBN-ARQ and SR-ARQ, were shown the potential for point-to-point terrestrial FSO systems by promising obtained results [21], [63]. The outperformance of novel SR-ARQ-based PPM compared to the standard SR-ARQ was confirmed in [63].
- Many studies showed exciting results to confirm the effectiveness of FSO systems integrating the design of adaptive multi-rate schemes and ARQ protocols, e.g., point-to-point terrestrial [21], [59] and hybrid FSO/RF [58], [65] systems.
- The SR-ARQ protocol, thanks to offering higher achievable throughput performance than other variants, was widely applied for long-distance communications of satellite/aerial as well as deep space systems. While only throughput performance was investigated, other crucial performance metrics for such systems, including delay and energy efficiency, are still lacking in the literature.
- Table VI summarizes the existing works on ARQ protocols in different FSO systems. From the viewpoint of current issues addressed on the design of ARQ protocols in FSO communications, some potential future research directions can be further investigated, for example,
  - Taking advantages of frames combining, adaptive multi-rate, and sliding window mechanisms, a novel design of ARQ protocols, a challenge but interesting, would offer better performance than SW-ARQ with FC [59] or sliding window protocols without FC [21] in FSO communications.
  - While the M-C-ARQ protocol can achieve better performance than conventional C-ARQ one in cooperative diversity FSO systems [56], its operation is still based on the SW mechanism, which is not an optimal design for ARQ protocols in high-speed systems as pointed out in [21]. Sliding window mechanisms handling multiple frames at a time should be considered for designing such an M-C-ARQ protocol.
  - The existing design of ARQ in hybrid FSO/RF systems was based on SW mechanism and simultaneous/parallel FSO and RF transmissions. Addressing the design of sliding window ARQ in more efficient solutions of hybrid FSO/RF systems, e.g., switching schemes [89], would be challenging but increase the system performance.
  - An investigation of the design and performance of ARQ protocols over all-optical multi-hop relaying systems is not available. For future work, sliding window ARQ with adaptive multi-rate schemes could be potentially considered for this network scenario.

III. HARQ AIDED FSO COMMUNICATIONS

The literature on the ARQ protocol revealed that it may fail to provide acceptable latency and throughput performance in some network scenarios, e.g., in long-distance FSO systems, due to many frame retransmissions required over time-varying turbulence channels. A more robust error-control method, i.e., hybrid ARQ (HARQ), which offers better reliability by combining ARQ and error correction code (ECC), is preferable to standard ARQ protocols in such scenarios. In this section, we first briefly present the background of HARQ protocols. Then, we review existing studies on the design and performance evaluation of HARQ for FSO communications in various scenarios, including point-to-point terrestrial, cooperative, multi-hop relaying, hybrid FSO/RF, and satellite/aerial systems. Finally, we summarize the main issues of reviewed HARQ protocols studied in different FSO system scenarios, followed by future research directions.

A. Overview of HARQ Protocols

Background: The ARQ protocol can be combined with the ECC to enhance the achievable ARQ efficiency over time-varying wireless fading channels. Such a combination of the two basic error-control schemes is referred to as hybrid ARQ (HARQ), which was introduced in the 1960s by Wozencraft and Horstein [90], [91]. Unlike the ARQ, the HARQ protocol can do both frame correction and retransmission of uncorrectable frames. In particular, the receiver handles error detection, error correction, and retransmission requests simultaneously. Retransmission is requested only when the receiver detects an uncorrectable error. Furthermore, the receiver can also store previously received frames for the joint decoding with the last received frame to improve decoding reliability. By combining error correction and retransmission and appropriately selecting an ECC scheme, the HARQ protocol can offer higher reliability than an ECC system alone and better throughput performance than a system with pure ARQ over time-varying turbulence fading channels.

Classification: HARQ protocols can be mainly classified into two categories [34], [38], namely type-I (TI) and type-II schemes, as shown in Fig. 6(a). Compared to the TI scheme, type-II is an advanced form of HARQ, which uses the concept of frame combining. The critical idea of frame combining is that even if a received signal detection fails, it still contains some useful information about the transmitted frame. Therefore, it is stored in the receiver buffer to be combined
with other retransmissions, resulting in enhanced detection performance. Depending on the retransmission scheme and the combining method, type-II HARQ can be further partitioned into chase combining (CC) based on the retransmission of the same coded frame; and incremental redundancy (IR), which is based on the retransmission of additional redundancy bits [92]. In the following, we give a brief description of HARQ schemes.

- **Type I HARQ**: The straightforward combination of ARQ and ECC is referred to as TI-HARQ. More explicitly, an ECC-encoded frame is the original one complemented by parity bits. This frame is (re)transmitted for all (re)transmissions. All corrupted frames are discarded at the receiver side, and each ECC decoding action is performed solely for a single encoded frame. As reported in [93], the TI-HARQ is capable of improving the achievable throughput performance of conventional pure ARQ in a relatively high error-rate environment, as its error-correction capability reduces the frequent retransmissions. Nevertheless, if the channel conditions are sufficiently good, the resultant reduced number of errors may not require any ECC parity bits. In this case, the throughput performance of the system using TI-HARQ is reduced by the unnecessary parity bits and becomes lower than that of conventional ARQ.

- **Chase Combining HARQ**: The CC-HARQ was proposed by D. Chase [94] for combining an arbitrary number of erroneous frames in a single frame based on a maximum likelihood criterion. In CC-HARQ, the modulated bits of the frame \( x \), which is created from the coded bits \( c \), is initially transmitted to the receiver. If the receipt of a NAK is confirmed, the same bits of frame \( x \) is retransmitted. This process for the retransmission of the same frame is continued until it is decoded successfully or reaches a predefined maximum number of transmission attempts of a frame, denoted by \( M \). Then, the previously received copies of the same frame are combined at the receiver in a single signal, which contains the accumulated information of the transmitted frame from all received signals. The operation of CC-HARQ with maximum-ratio combining (MRC) method is illustrated in Fig. 7, in which HARQ persistent is given by \( K \). The received signal at the \( i \)-th transmission, denoted by \( y^{(i)} \), is expressed by

\[
y^{(i)} = h^{(i)} x + n^{(i)},
\]

where \( x \) is the transmitted signal, \( h^{(i)} \) is the channel fading coefficient, and \( n^{(i)} \) is signal-independent additive white Gaussian noise with variance \( \sigma_n^2 \). After the \( K \)-th HARQ round (\( K \leq M \)), the receiver combines the \( K \) received frames at the symbol level in a single frame denoted by \( y_K \). The combined frame is obtained by weighting each frame by an estimation of its reliability before being summed with the other frames, i.e.,

\[
y_K = \sum_{i=1}^{K} \alpha^{(i)} y^{(i)},
\]

where \( \alpha^{(i)} \) is the reliability of each received frame, which is given under maximum-likelihood decoding [95]. The CC-HARQ takes advantage from the accumulated SNRs of the individual transmissions to enhance the decoding performance of the transmitted frame.

- **Incremental Redundancy HARQ**: The IR-HARQ scheme, as depicted in Fig. 8, generalizes the CC-HARQ by considering that each transmission is a punctured version of a low rate mother FEC code, denoted by \( C_0 \). In IR-HARQ, the information frame, which includes frame header, message, and CRC, is first encoded by the original rate \( C_0 \) encoder (called mother code). For the first transmission,
only some coded bits of $C_0$ are chosen according to a predetermined puncturing pattern $P_1$. The selected bits are grouped in a single block $c^{(1)}$ which is modulated and then transmitted to the receiver side. Here, we denote the transmitted codeword by $C_1 = c^{(1)}$. At the receiver side, after decoding for error correction with code $C_1$, error detection using CRC is performed. If the presence of errors is detected, IR-HARQ only retransmits the additional redundancies for uncorrectable frames based on a NAK message from the receiver. Accordingly, the transmitter sends the second block $c^{(2)}$ containing the additional coded bits taken from the remaining bits of $C_0$, which have not been transmitted yet. The newly received redundancy is combined with the previously received frames to construct coded bits of code $C_2 = [c^{(1)}, c^{(2)}]$ for decoding. This process for retransmission of a frame is continued until it is decoded successfully or reaches a predefined maximum number of transmission attempts of a frame $M$. Thus, we have $C_1 \subseteq C_2 \subseteq \cdots \subseteq C_M \subseteq C_0$ which expresses the rate-compatibility restriction of the punctured codes in IR-HARQ protocols.

**Comparison Between HARQ Protocols:** Of the three HARQ protocols, TI-HARQ is the most straightforward implementation but provides a deficient throughput performance over time-varying fading channels. The CC and IR-HARQ protocols outperform the TI-HARQ thanks to the concept of frame combining. The IR-HARQ protocol can achieve better performance than the CC-HARQ because only incremental redundancies are retransmitted for each event of the erroneous frame. However, the CC-HARQ offers a lower complexity than IR-HARQ. This is because the use of IR-HARQ protocol requires additional signaling and a large-size buffer [96]. Additionally, since each retransmission is identical, the CC-HARQ can easily combine with other techniques, e.g., spacetime coding and coded modulations. Moreover, as reported in [97], the gain of the IR-HARQ scheme compared to the CC-HARQ one is more significant with higher code rates, while with the lower code rate, the performance of the CC-HARQ protocol is almost as good as that of IR-HARQ.

**B. HARQ Protocols in FSO Systems**

The design of HARQ protocols has been widely investigated in FSO systems, in which a summary of existing studies on HARQ protocols addressed in various FSO scenarios is given in Table VIII. In the following, we aim to review works on HARQ protocols in each FSO communication system in detail.

1) **Point-to-Point Terrestrial Systems:** The design and performance of HARQ protocols in point-to-point terrestrial FSO systems have been addressed in [98], [100], [101], [103], [107], [110]. In these studies, the SW mechanism was used for HARQ, and the performance analysis was based on the information-theoretic approach.

Early works on HARQ, with IR and CC schemes, in FSO systems, were reported in [98], [100]. Kose et al. addressed the design of IR-HARQ with TrellisWare’s flexible low-density parity-check (LDPC) code family over turbulence fading channels [98]. Initially, the IR-HARQ operates by transmitting a high-rate codeword obtained from the family code. In the event of a NAK, the additional (i.e., punctured) parity bits required to form lower-rate codewords are subsequently transmitted. Besides the IR scheme, the design of CC-HARQ with PPM was mentioned in [100]. In this design, retransmissions of HARQ are attempted upon the reception of NAK using a random delay with a minimum value exceeding the coherence time of turbulence channels. The CC-HARQ could achieve a significant performance improvement over the pure ARQ in FSO systems.

Compared to the simple and cost-effective intensity modulation/direct detection (IM/DD) mentioned in [98], [100], the FSO coherent receivers are relatively more complex in the implementation but they provide more flexibility and performance improvement [115]–[117]. By mixing the received signal with the strong local oscillator field, coherent receivers have better spatial and frequency selectivity than their non-coherent counterparts. Aiming to take advantage of the inherent features of coherent FSO systems, for the first time, Aghajanzadeh et al. analyzed the performance of HARQ protocols, including TI, CC, and IR schemes, in terms of outage probability and throughput over turbulence fading channels [101]. While only turbulence effect was considered in [101], the analysis carried out therein for the performance of HARQ is mathematically intractable with pointing errors. Consequently, and additionally inspired by the promising results obtained in [101], several studies aimed to provide a more accurate performance estimation of HARQ in FSO coherent systems, as follows.

- Zedini et al. analyzed the ergodic capacity and outage performance of HARQ under the combined effect of atmospheric turbulence (modeled by Gamma distribution for tractable analysis) and pointing errors [103].
- Using the same approach in [103], Touati et al. investigated the delay performance of HARQ protocols in [107].
- Verma et al. provided a more comprehensive analysis of HARQ throughput under the combined Gamma-Gamma distributed turbulence and pointing errors [110].

The above-mentioned studies showed that HARQ using IR is the most efficient scheme in various turbulence strengths and pointing jitter conditions. This is because only incremental redundancies are retransmitted to combine with the previously received ones for the correction of the corrupted frame. Additionally, the CC scheme, which outperforms the TI one, can be used in weak turbulence and pointing error conditions, as its performance is close to IR one yet simpler in the implementation.

2) **Cooperative Systems:** Several studies on the design of HARQ protocols for cooperative FSO communications, either for terrestrial broadcasting systems using SW mechanism [109] or long-distance satellite systems using sliding window mechanism [111], have been reported.

C-HARQ-Based SW: Figure 9 illustrates a typical broadcasting FSO system, where the center node broadcasts the same information frames to $N$ fixed users, i.e., $D_1, D_2, \ldots, D_N$, with $N \geq 2$, via separate FSO links. Hosseini et al. presented a novel cooperative HARQ (C-HARQ) using SW mechanism for FSO broadcasting systems [109]. Two well-known HARQ
TABLE VIII
LITERATURE REVIEW OF HARQ PROTOCOLS IN DIFFERENT FSO SYSTEM SCENARIOS, INCLUDING POINT-TO-POINT TERRESTRIAL (PP), COOPERATIVE (CP), MULTI-HOP RELAYING (MH), HYBRID FSO/RF (HB), AND SATELLITE/AERIAL (SA) SYSTEMS

| Ref | Year | Objective | HARQ Type | System | Remarks |
|-----|------|-----------|-----------|--------|---------|
| [98] | 2009 | Throughput | TI | CC | IR | PP | CP | MH | HB | SA |
| [99] | 2010 | Achievable Rate | | | | | | | | |
| [100] | 2012 | Outage Probability | | | | | | | | |
| [101] | 2013 | Throughput, PER | | | | | | | | |
| [102] | 2014 | Ergodic Capacity, Outage Probability | | | | | | | | |
| [103] | 2016 | Throughput, Outage Probability | | | | | | | | |
| [105] | 2016 | Transmission Efficiency | | | | | | | | |
| [106] | 2017 | Ergodic Achievable Rate, Outage Probability | | | | | | | | |
| [107] | 2018 | Throughput, Delay, Outage Probability | | | | | | | | |
| [108] | 2019 | Throughput, BER, Outage Probability | | | | | | | | |
| [109] | 2020 | Average Sum Rate | | | | | | | | |
| [110] | 2021 | Outage Probability, Throughput | | | | | | | | |
| [111] | 2021 | Throughput, Delay | | | | | | | | |
| [112] | 2022 | Throughput, Delay, Energy Efficiency | | | | | | | | |
| [113] | 2022 | Outage Probability, PER | | | | | | | | |
| [114] | 2022 | Energy Efficiency | | | | | | | | |

variants, both CC and IR, were considered. Besides, error-free feedback using RF links was assumed. As shown in Fig. 9, two design approaches were considered in [109]: (i) conventional approach and (ii) novel cooperative approach.

- **Conventional Approach**: The center node handles frame retransmissions, and users only confirm the correct or erroneous nature of received frames via feedback channels. The users do not participate in the retransmission process, even though they could effectively cooperate in this process.

- **Novel Cooperative Approach**: Instead of relying solely on the center node, the frame retransmission process could be done by a neighboring user with a shorter distance than that from the center node and successfully decoded the original data frame. To this end, a novel C-HARQ was introduced for broadcasting FSO systems, where selected neighbors would assist users in the retransmission process.

**Example**: An example of novel C-HARQ protocol is depicted in Fig. 9, where D1 and D2 are the neighboring users. Let dSD1 and dD1D2 denote the respective transmission distances from S to D1 and D1 to D2, then dSD1 > dD1D2. At the initial round, the S node broadcasts the frame $F^{(1)}_1$ corresponding the sub-codeword $C_1$ to all users, and then users return feedback signals, i.e., either ACK or NAK, to the S node. For the neighboring users, if the D2 node is unable to decode the received frame $F^{(1)}_1$, while the D1 node successfully decoded that frame after requesting a redundancy $F^{(2)}_1$ from the S node. Then, the S node informs D1 to encode data and retransmit the frame $F^{(2)}_1$ to D2 for the next retransmission. At the D2, the newly received redundancy, i.e., $F^{(2)}_1$, is combined with the previously received frame ($F^{(1)}_1$) to construct coded bits of code $C_2$ for decoding. The D1 node keeps retransmitting the redundancy to the D2 node until it is decoded successfully or reaches a predefined maximum number of transmission attempts of a frame, denoted by $M$. Here, the D2 node requests retransmissions by sending NAK signals to the S node, and then the S node informs the D1 node to retransmit frames to the D2 node. It is noted that each user is able to receive data frame from either the S node or only one neighboring user in each (re)transmission round.

**C-HARQ-Based Sliding Window**: The design of the C-HARQ based sliding window mechanism was addressed in HAP relay-assisted satellite FSO systems for the Internet of Vehicle (IoV) applications [111]. The satellite-HAP-vehicle
A typical broadcasting FSO system using HARQ protocols with (1) conventional approach and (2) novel cooperative approach in [109].

The system consists of two links: satellite-to-HAP and HAP-to-vehicle links. As the altitude of HAP is in the stratosphere with less susceptibility to the weather effects, the pure sliding window ARQ mechanism was used for this link. Otherwise, the IR-HARQ schemes were employed for the HAP-to-vehicle links, which experience weather-related issues like atmospheric turbulence and clouds. In such proposed HARQ protocols, the relay node, i.e., HAP, plays a role in the retransmission of erroneous frames received by vehicles instead of requesting retransmission from the source, i.e., satellite. Moreover, using a pure sliding window ARQ mechanism for satellite-to-HAP links can avoid the complexity of decoding at HAP-based relaying nodes.

3) Multi-Hop Relaying Systems: Multi-hop transmission is an alternative relay-assisted scheme, which employs relays in a serial configuration [118]–[120]. The scheme could mitigate the distance-dependent turbulence strength and atmospheric loss by utilizing shorter distances in resulting hops. The adoption of multi-hop relaying in FSO systems has been investigated for amplify-and-forward [120] and decode-and-forward (DF) [119] relaying techniques. The multi-hop schemes often fail to provide reliability in turbulence conditions, as they will be forwarding the noisy replicas of the received data [24]. To maintain reliable transmissions, several studies on HARQ protocols, mainly for mixed dual-hop FSO-RF systems, have been reported in [106], [113]. In these studies, the SW mechanism, which processes one frame at a time, was used.

IR-HARQ Aided Dual-Hop FSO-mmWave: The combination of FSO and mmWave links is considered as a powerful candidate for high data-rate multi-hop communications [121]. Makki et al. addressed the design of IR-HARQ for multi-hop FSO-mmWave system operating in the DF mode [106]. For the operation of HARQ, the successfully received data frame at each hop is decoded and re-encoded for the transmission to a next-hop. Otherwise, only sub-codewords divided from a parent codeword (low code rate) are retransmitted and combined with all previously received sub-codewords to decode that frame.

At each hop, the retransmission continues until the frame is correctly decoded or the maximum permitted transmission round is reached. In the latter case, the frame is clarified to be lost. Therefore, a frame is successfully received by the destination if it is correctly decoded in all hops. With typical parameter settings of the RF-FSO dual-hop systems, i.e., the outage probability of $10^{-4}$ and code rate of 3 nats-per-channel-use, the IR-HARQ with the persistent level of 2 and 3 reduce the required powers, compared to cases without HARQ, by 13 and 17 dB, respectively.

CC-HARQ Aided Dual-Hop FSO-RF With IRS: Intelligent reflecting surface (IRS) has recently emerged as a promising solution to expand wireless communication coverage. In [113], Verma et al. analyzed the performance of CC-HARQ over dual-hop FSO-RF system operating in the DF mode. Here, the IRS with phase errors was considered for the relay-to-destination RF link. It was shown that the RF fading, pointing errors, and atmospheric turbulence effects could be compensated by using the CC-HARQ on both links.

4) Hybrid FSO/RF Systems: The state-of-art hybrid FSO/RF systems using HARQ protocols were based on the parallel/simultaneous FSO/RF links [99], [102], [104], [122]. In these studies, realizing the potential of incremental redundancy schemes, most of the works focussed on the IR-HARQ, which offers the highest efficiency compared to other HARQ types. In addition, the simple SW mechanism was used for the operation of HARQ.

The puncturing technique plays an essential role in the design of IR-HARQ, which allows an encoder/decoder pair to adjust code rates without changing their structure [123].
Several studies presented the novel punctured structures for the operation of IR-HARQ in hybrid FSO/RF systems [99], [102]. In [99], AbdulHussein et al. addressed the design of IR-HARQ using rateless Fountain codes for hybrid FSO/RF systems. This design involves the use of Raptor codes as described in [99], in which the data is partitioned into two sets of bits, separately encoded, and sent through FSO/RF channels. Also, the authors established the pertinent information-theoretic limits showing that coding schemes with code-rate selection may suffer from rate loss or outages depending on channel conditions. It was demonstrated that HARQ-based rateless coding scheme using an off-the-shelf Raptor code [124], well approaches the information-theoretic limits regardless of channel conditions. The advantages of such a HARQ design are twofold.

- Firstly, it is an elegant approach for data transmissions over the time-varying behavior of FSO/RF channels, as it requires neither a bank of codes with various rates nor explicit code selection. As a result, other considerations aside, one may choose this design simply due to its ease of operation.
- Secondly, no rate mismatch due to outdated or inaccurate channel estimation can happen. This is a distinct performance benefit over schemes with code-rate selection, e.g., [105], [125], which may fail if the channel quality varies notably from one codeword to another.

Another design approach on the puncturing technique for IR-HARQ was mentioned in [102]. A novel IR scheme with a joint puncturing pattern and FSO/RF bit selection optimized for a hybrid FSO/RF system was introduced. In contrast to the rate-compatible punctured codes for conventional IR-HARQ, this design also involves bit splitting between FSO and RF in each retransmission, affecting the puncturing pattern selection. The algorithm uses separate puncturing and RF bit selection patterns in each retransmission based on minimizing the BER, which can apply to several ECC codes, including non-systematic/systematic convolutional codes and turbo codes.

Differently from [99] and [102], the performance of IR-HARQ using the conventional punctured structure was analyzed in [104], [122]. The data sequence is encoded into parallel FSO and RF bit streams, which are then simultaneously sent to the receiver side for joint decoding. Moreover, the adaptive power allocation was exploited to improve the performance of hybrid FSO/RF systems using the HARQ protocol. For high received SNR values, it was recommended to use uniform power allocation for HARQ-based RF-FSO links due to the complexity of adaptive power allocation.

5) Satellite/Aerial Systems: For long-distance FSO systems, HARQ is an efficient error-control solution, which circumvents the limitations of other link-layer solutions, including pure ARQ and ECC. This is because ECC schemes often require excessive redundancies to guarantee the transmission reliability leading to inefficient throughput performance [126]. Besides, pure ARQ usually fails to provide satisfactory delay and throughput performance for high-latency systems, as they require many retransmissions for erroneous frames [61]. Several studies on the design of HARQ protocols have been reported for HAP-to-HAP [105], [125] and LEO satellite [108], [112], [114], [127] systems.

TI-HARQ Aided HAP-to-HAP: The HAPs are located at altitudes of 17-25 km, where the impact weather is negligible [128]. As a result, using the simplest TI variant for the HARQ protocol is sufficient for FSO-based inter-HAP systems. Parthasarathy et al. introduced a design of adaptive TI-HARQ, which is the combination of Reed-Solomon (RS) codes and SR-ARQ, over weak turbulence of inter-HAP links [105], [125]. In the adaptive TI-HARQ, the proper code rates are adaptively selected based on the channel condition via feedback signals, i.e., channel state information (CSI), which is illustrated in Fig. 10.

It is noted that the estimated CSIs in practical long-distance FSO links are delayed due to the channel propagation time. This kind of delay is in the order of milliseconds, which is similar to the channel coherence time. The authors also investigated this issue by analyzing the transmission efficiency (TE) in the presence of perfect and delayed CSIs. Table IX [105] illustrates the coding gain for the adaptive TI-HARQ protocols with different CSI qualities compared to the pure SR-ARQ in the maintenance of TE of 30% and 90%. From Table IX [105], we observe a considerable gain improvement for the adaptive TI-HARQ in comparison with the uncoded non-adaptively pure ARQ for the scenarios of FSO-based inter-HAP communications.

Type II-HARQ aided LEO satellites: Unlike inter-HAP systems, where the simple TI-HARQ can be used, a more robust HARQ variant should be considered for FSO-based LEO-to-ground links experiencing various adverse issues, such as atmospheric turbulence, cloud coverage, atmospheric attenuation, and pointing misalignment. The design of type II-HARQ, including CC and IR schemes, was considered for such links [108], [112], [114], [127], in which a summary of these works is as follows.

- CC-HARQ Design: Xiang et al. proposed the multi-frame interleaving scheme for CC-HARQ protocol [108]. The conventional CC-HARQ generally uses interleaving schemes to disperse the serial burst errors and then
computing solutions to enhance the performance of IR-HARQ protocols.

### Table IX

| Distance (km) | Coherence Time (ms) | Gain for TE = 30% | Gain for TE = 90% |
|--------------|---------------------|-------------------|-------------------|
|              |                     | Delayed CSI       | Perfect CSI       | Delayed CSI | Perfect CSI |
| 300          | 10.95               | 3.6               | 4.8               | 2.6         | 3.4         |
|              | 2.7                 | 2.9               | 4.5               | 0.7         | 3.3         |
| 600          | 13.75               | 1.2               | 4.4               | 2.4         | 5.4         |
|              | 2.7                 | 1.9               | 4.1               | 0.3         | 3.6         |

A joint design of adaptive transmissions, either adaptive coding schemes [105], [125] or adaptive rate transmissions [112], with HARQ protocols can offer a significant performance improvement in FSO-based satellite/aerial communications. In addition, the issues of imperfect estimated CSIs on the joint design between adaptive schemes and HARQ protocols using the simplest TI variant were addressed in [105], [125].

• The summary of existing works on HARQ protocols in different FSO communication systems is given in Table VIII.

In addition, some potential research topics on HARQ protocols can be further investigated, for example:

• The HARQ protocols can be potentially considered for the emerging scenarios of FSO-based satellite/aerial-assisted IoV to maintain reliable connectivity. The strict delay requirement is essential for such network scenarios, while the current HARQ design’s primary focus is on throughput performance. One of the possible solutions is using HARQ without waiting for feedback based on channel conditions. The promising results obtained in the most recent study in RF systems [129] confirmed the effectiveness of using fast HARQ protocols to minimize the end-to-end delay for RF systems. Considering such HARQ protocol design in the context of FSO-based satellite/aerial-assisted IoV networks would be exciting and a possible direction for future study.

• Machine learning techniques have effectively advanced the state-of-the-art for many research problems in wireless networks [130]. It has been recently applied to enhance the performance of RF systems using HARQ [131]. This allows providing the HARQ feedback earlier by predicting the decoding outcome, enabling the original transmitter to react faster to the current channel conditions and send additional redundancy at an earlier point. In this way, more HARQ iterations are encouraged to improve system reliability under strict delay constraints. For future work, this approach could be potentially considered to enhance the FSO systems using HARQ protocols.

C. Summary and Future Work

This section reviewed the existing studies on the design and performance estimation of HARQ protocols in different FSO system scenarios. The remarkable points of this section can be summarized as follows.

• The state-of-art on HARQ protocols revealed that IR-HARQ is the most efficient variant for different FSO system scenarios, thanks to retransmissions of only redundancy codewords for jointly decoding with the previously received ones.

• Aside from [111], [112], [127], most of the studies addressed the use of SW mechanism for the operation of HARQ protocols for the sake of simplicity in the analysis.

• A lot of studies on the HARQ protocol focussed on analyzing its performance over turbulence fading channels from an information-theoretic approach, e.g., [100], [101], [103], [106], [109], [110].

• Using rateless codes [99], modified punctured structure [102], and adaptive power allocations [104] are existing solutions to enhance the performance of IR-HARQ protocols in the context of hybrid FSO/RF systems.

IV. CROSS-LAYER DESIGN APPROACH

Cross-layer design, which allows information sharing among layers, has attracted research efforts in the domain of FSO networks. Indeed, realizing cross-layer design is essential for improving the performance of FSO systems over atmospheric turbulence channels. This section addresses the cross-layer design approach for considered link-layer retransmission protocols in FSO networks. First, we review the background of design methodologies, including the classification, motivation, and challenging issues. Next, we introduce several cross-layer design frameworks of link-layer retransmission protocols incorporating protocols at the PHY and transport layers in FSO networks. Finally, we summarize key issues and present potential research topics related to this design approach.
A. Overview of Design Methodologies

Design Methodologies: For designing the link-layer retransmission protocols in FSO networks, two methodologies have been investigated so far, i.e., (1) layered design and (2) cross-layer design. In the following, we briefly present the concept of these methodologies.

- **Layered Design:** The protocols on one layer are designed independently from the protocols on other layers. In this design, the communication among non-adjacent layers is not permitted [132]. Although the strict boundary between the layers makes the networks easy to deploy, the layers’ encapsulation prevents some necessary information sharing between the layers. As a result, this design approach does not provide a mechanism for the performance optimization between different protocol layers, which can significantly improve the system performance.

- **Cross-Layer Design:** Unlike the layered design, the cross-layer approach allows the interaction between layers by permitting one layer to access the data of another layer for the joint optimization of protocols across the communication stack [133]. In fact, the cross-layer design provides inter-layer communication between non-adjacent layers without destroying the existing network reference models. In other words, the cross-layer design implies that each layer can share parameters, status, and additional information with the different layers without breaking the layers structure of computer networks. Indeed, networks employing the cross-layer design can provide a better performance, and many studies have confirmed the poor performance of layered design in wireless networks, e.g., [134]–[136].

**Cross-Layer Design for FSO Networks:** The motivation of the cross-layer design originates from the unique and novel characteristics of FSO networks. As a matter of fact, the traditional layered architecture reference models, including the open systems interconnection (OSI) and Transmission Control Protocol/Internet Protocol (TCP/IP), were initially designed for wired networks [137]. Nevertheless, as the fading channel models in FSO networks are entirely different from that of wired networks, the problems created by the FSO networks but solved with the layered architecture may result in unsatisfactory results and poorly achievable performance.

In addition, an individual TCP/IP protocol usually aims to solve one specific set of problems without considering the end-to-end network performance leading to the deployment of these protocols does not always satisfy the increased performance requirements [132]. Moreover, the time-varying turbulence fading and new issues in FSO networks motivate the consideration of cross-layer design, which can address these characterizations by exploiting the interactions and dependencies among layers.

For example, the performance of TCPs over FSO networks mainly depends on the congestion loss caused by the network’s buffer overflows and transmission errors due to the uncertainty of atmospheric FSO last-mile channels [138], [139]. Instead of executing the TCP’s congestion control in isolation at the transport layer, by using the cross-layer design approach, the congestion control can be jointly managed with buffer management at the network layer, the error-control methods at the PHY and link layers [140]. As a result, the performance of TCP could be improved significantly.

**Challenging Issues of Cross-Layer Design:** Cross-layer designs offer a significant number of benefits, e.g., joint optimization of protocols from multiple layers to increase the system throughput, reduce latency, and minimize the transmission error rate [132]. Nevertheless, there are several drawbacks/challenges of cross-layer designs that are inevitable due to the characteristics of these designs, as follows [134]:

- **Cross-layer overhead/signaling:** It is inevitable to result in an extra overhead and control signaling when exchanging the cross-layer information in networks [141]. A large amount of such information may occupy a great deal of bandwidth, leading to a burden on the network performance.

- **Coexistence of multiple cross-layer designs:** It is not straightforward to integrate different cross-layer designs into a uniform design due to the specific communication standard of each cross-layer design [137]. As a result, the coexistence problem is a challenge that cross-layer designs have to deal with.

- **Universal cross-layer design:** The cross-layer model for a single application may not be suitable for different applications. For instance, video streaming requires a small ARQ persistent level, while the data transmission can endure a high ARQ persistent level in the error-prone environment of wireless networks [142]. Finding a universal cross-layer standard for different applications poses various challenges and is an open research issue.

- **Destruction of layered architecture:** Cross-layer frameworks may break the encapsulation of layers in the conventional layered architecture. As a matter of fact, a slight modification in one layer results in a series of changes in the other layers [143]. Therefore, the destruction of layered architecture is one of the most challenging issues and the fundamental drawbacks of the cross-layer design.

**Cross-Layer Frameworks in FSO Networks:** The cross-layer design frameworks of link-layer retransmission protocols in FSO networks are mainly partitioned into two groups: (i) physical layer and link layer (PHY/Link design) and (ii) link layer and transport layer (Link/Transport design), as illustrated in Fig. 11. As for the cross-layer PHY/Link, the integration of AR and ARQ/HARQ is the most popular framework in FSO networks [21], [59], [65], [112]. Regarding the cross-layer Link/Transport, most of the studies focused on the joint design of ARQ/HARQ with loss-based TCPs [144]–[148]. A summary of these cross-layer frameworks in FSO networks is provided in Table X. In the following, we present the details of each cross-layer design group.

B. PHY/Link Cross-Layer Frameworks

Atmospheric turbulence, which results in a very slowly-varying fading, is a major degrading factor in FSO systems.
As reported in [149], [150], the channel coherence time is typically 1 to 10 ms or longer, and a Gbps-transmission-rate period may cover up to thousands of consecutive frames. This quasi-static channel property makes providing reliable feedback possible, and the available CSI at the transmitter can be used to design the adaptive transmission schemes for considerable performance enhancement in FSO systems. Moreover, the feedback information required in the adaptive transmission is relatively feasible to implement in FSO systems. The reason is that commercially available FSO units have full-duplex (bi-directional) capabilities, and a small portion of the large available bandwidth could be allocated for feedback purposes without much effect on the data rate [12].

As a result, adaptive transmission, one of the most popular error-control PHY schemes, emerges as a promising solution for FSO systems. Adaptive transmission has been extensively studied in the context of FSO communications [151]–[153] and involves the change of system parameter settings, e.g., transmission rate, coding rate, modulation order, transmitted power, or the combination of those according to the channel conditions. In this regard, the cross-layer design between link-layer retransmission protocols and adaptive transmissions, an exciting but challenging research problem, would effectively improve the overall system performance over turbulence-induced fading channels.

1) Joint Design of AR and ARQ: As shown in Fig. 12, the cross-layer frameworks incorporating AR scheme at PHY and different link-layer ARQ protocols have been addressed in [21], [59], [65] for the context of FSO terrestrial communications.

Design of AR scheme: The objective of AR scheme is to maximize the data rate over the turbulence fading channels while satisfying a required QoS at the link layer, i.e., targeted FLR (denoted as $FLR_{\text{target}}$) [59], [65] and targeted FER (denoted by $FER_{\text{target}}$) [21]. Using the assumption of perfect estimated CSIs, multiple transmission modes with different data rates can be adjusted dynamically at the transmitter. For the design of AR scheme, the entirely received SNR range is partitioned into $K + 1$ non-overlapping consecutive intervals with boundary points denoted as $\gamma_k = \{0, 1, 2, \ldots, K\}$. Each channel-state is, then, assigned by a specific transmission mode, which can bring the highest possible data rate while maintaining the targeted $FER_{\text{target}}$.

AR with SW-ARQ: The cross-layer design between AR scheme and truncated SW-ARQ protocol, which processes one frame at a time, was introduced in [59], [65]. While the Gamma-Gamma turbulent model was used in [59], a more comprehensive Malaga-$M$ turbulent model was considered in [65]. Two cross-layer frameworks, including (i) AR and standard-ARQ (AR/ARQ) and (ii) AR and ARQ with FC (AR/ARQ-FC), were investigated [59]. In this study, the frame duration of ARQ operation was adjusted based on the transmission modes (with different data rates) of the AR scheme as depicted in Fig. 12. In addition, the number of available AR scheme’s modes, denoted by $N_t$, is supposed to be equal to the number of channel states, i.e., $N = K$. The selection of SNR thresholds satisfies the condition that the FLR for each channel-state is exactly the predefined $FLR_{\text{target}}$. Using the truncated SW-ARQ with $N_t$ persistent level (e.g., $N_t = 2$ in
Fig. 12. Cross-layer frameworks incorporating AR scheme and different ARQ mechanisms, i.e., SW-ARQ [59] with (a) standard SW and (b) SW with FC, and Sliding Window ARQ [21] with (c) GBN and (d) SR for FSO terrestrial systems.

As is evident, the AR/ARQ-based sliding window mechanisms, AR/ARQ-GBN, and AR/ARQ-SR, are shown in Fig. 13. The channel-state \( R_{k} \) at the state \( n \), where \( M = 2^{n} \) with \( n \in \{1, 2, \ldots, N\} \), and FER\(_{n,k}(M, \gamma)\) is the instantaneous FER when using mode \( n \) on the \( l \)-th transmission attempt. Given the equation of FER\(_{n,k}(M, \gamma)\) in [59, eq. (10)], the SNR thresholds can be obtained as [59]

\[
\gamma_n = \begin{cases} 
-\frac{2^{(2^{n}-1)} \ln \left( \frac{5 \text{FER}_{\text{target}}}{N_t} \right)}{3}, & \text{for AR/ARQ,} \\
-\frac{4(2^{n}-1) \ln \left( \frac{\text{FER}_{\text{target}}}{0.2 N_t} \right)}{3(N_t+1)}, & \text{for AR/ARQ - FC,} 
\end{cases}
\]

(7)

where \( N_t \) is the frame size.

**AR with Sliding Window ARQ:** Differently from [59], the sliding window ARQ protocols, both GBN and SR, was jointly designed with the AR scheme in [21]. Two cross-layer frameworks, including (i) AR/ARQ-GBN and (ii) AR/ARQ-SR, are illustrated in Fig. 12. 

**Design:** To effectively facilitate sliding window protocols with the AR scheme under the impact of atmospheric turbulence, the ARQ’s window size is designed based on the PHY’s burst transmission. Also, the burst size is designed considering the time-varying behavior of the turbulence-induced fading channels. In particular, the data are transmitted in fixed-time bursts, in which each channel-state covers a burst transmission. The selection of SNR thresholds satisfies the condition that intervals of all channel states, which are shorter than the fading channel coherence time \( T_c \), are equal to the burst duration \( T_{\text{burst}} \). It is important to note that channel thresholds \( \{\gamma_k\}_{k=1}^{K} \) are not the same as \( \gamma_n^*, n \in \{1, 2, \ldots, N\} \), which are used for the selection of transmission modes obtained by a least-square curve fitting of FER in [154].

The channel-state \( k \)-th with the SNR interval of \( [\gamma_{k-1}, \gamma_k) \) is said to be in the transmission mode \( n \) if it brings the highest possible data rate, and the corresponding average FER satisfies FER\(_{\text{target}}\). If we denote by \( \phi_k = \{n|k - \text{th state is assigned by mode } n\} \), the design problem can be formally formulated as

\[
\text{maximize } R_{\phi_k}(n) \\
\text{subject to } \text{FER}_k(M_{\phi_k}, \gamma) \leq \text{FER}_{\text{target}}, \\
M_{\phi_k} = 2^n, \quad n = 1, 2, \ldots, N,
\]

(8)

where \( R_{\phi_k}(n) \) and \( \text{FER}_k(M_{\phi_k}, \gamma) \) are the data rate and average FER at the state \( k \)-th using mode \( n \), respectively. Here, \( R_{\phi_k}(n) = R_s \log_2 M_{\phi_k} \) with \( M_{\phi_k} = 2^n \) and \( R_s \) the symbol rate.

**Performance:** The throughput comparison among cross-layer designs for FSO terrestrial systems, including AR/ARQ-SW, AR/ARQ-GBN, and AR/ARQ-SR, are shown in Fig. 13. As is evident, the AR/ARQ-based sliding window mechanisms outperform AR/ARQ-SW thanks to the transmission of many frames without waiting for an acknowledgment. The essential point is that the throughput improvement becomes much more considerable when the symbol rates are high. The reason is that the time waiting for the acknowledgment of a data frame (twice the propagation delay) in AR/ARQ-SW becomes significant compared with the transmission delay, which becomes very small thanks to the high data rate of FSO systems. This leads to inefficiency in the throughput performance of AR/ARQ-SW in high-speed FSO systems.
2) Joint Design of AR and IR-HARQ: A cross-layer design framework of AR scheme and IR-HARQ based sliding window mechanism over FSO-based satellite channels was introduced in [112].

Design: The IR-HARQ with AR scheme is applied to each burst, containing multiple frames depending on AR's modes. From a low mother code rate of $R_{C_1}$ (e.g., 1/2 or 1/3), a family of punctured codes, including $N_1$ different code rates, is obtained by using the punctured technique, i.e., $(1 \geq) R_{C_1} > R_{C_2} > \cdots > R_{C_{N_1}}$. For the initial round, frames containing only the coded bits of code $R_{C_1}$ are selected to be transmitted in a burst. At the receiver side, after decoding for error correction with code $R_{C_1}$, the error detection using CRC is performed. If the presence of errors is detected, the additional redundancies for uncorrected frames are retransmitted in the next round based on a NAK message indicating the sequence numbers of those frames. The newly received redundancies are then combined with the previously received frames to construct coded bits of code $R_{C_2}$ for decoding. The process for retransmissions of a frame is terminated when decoding the frame successfully or reaching the persistent level of IR-HARQ, i.e., $N_1$ transmission attempts. If a frame does not get through the FSO link after $N_1$ transmission attempts, it is clarified to be lost.

Channel-state Modeling: The design of the channel-state model is similar in [21], where the channel is divided into equal intervals defined by a range of SNR thresholds. Nonetheless, each channel-state covers not only a burst transmission but also the feedback signals, as the propagation delay is significant in long-distance satellite systems, which can not be simply ignored as in [21] for FSO terrestrial systems. Using the same approach in [21], where the average channel-state duration $\tau_k$ is set to $T_{\text{burst}} + 2t_{\text{prop}}$ with $t_{\text{prop}}$ the propagation delay, the SNR thresholds for channel states can be obtained. For the AR scheme, each channel-state is then assigned a specific transmission mode, which can bring the highest possible data rate while maintaining a targeted $\text{BER}_{\text{target}}$ at PHY. The channel-state $k$-th with the SNR interval of $[\gamma_{k-1}, \gamma_k), k \in \{0, 1, \ldots, K\}$, is said to be in the transmission mode $n$, where $n \in \{1, 2, \ldots, N\}$, if the corresponding average BER at PHY satisfies $\text{BER}_{\text{target}}$. If we denote by $\phi_k = \{n|k-\text{th state is assigned by mode } n\}$, the design problem can be formally formulated as

$$\begin{align*}
\text{maximize} & \quad n | M_{\phi_k} = M_n \\
\text{subject to} & \quad \text{BER}_k(M_{\phi_k}, \gamma) \leq \text{BER}_{\text{target}},
\end{align*}$$

where $\text{BER}_k(M_{\phi_k}, \gamma)$ is the average BER at state $k$-th using mode $n$, which can be found in [112].

Example: An example of the joint design between AR scheme and IR-HARQ based sliding window mechanism is depicted in Fig. 14. Here, the window size is set to $W = 1$ burst time, and the persistent level of IR-HARQ is $N_1 = 3$. Also, the number of frames per burst in the transmission mode $n$ is determined by $n \times n_2$, where $n_2 = 100$ blocks/burst. In this example, frame 2 is supposed to be uncorrectable in burst 1 at the initial transmission round. The additional redundancies adopted from puncturing FEC family codes for this frame are then transmitted in the next bursts (i.e., bursts 2 and 3) together with the other new frames for the joint decoding. These redundancies are combined with the previously received frames to construct the more robust codes (i.e., lower code rates) for error correction. After reaching the HARQ’s persistent level, if frame 2 is still uncorrected, it is clarified to be lost.

Performance: Figure 15 quantitatively highlights the effectiveness of the IR-HARQ protocol using a sliding window mechanism by comparing their throughputs with those of traditional IR-HARQ using SW mechanism and pure sliding window ARQ over a range of received powers. A considerable throughput enhancement of link-layer solutions using sliding window mechanism compared to ones using SW mechanism is confirmed. This phenomenon is due to the time waiting for the acknowledgment of a data frame (twice the propagation delay) in satellite communications is significant compared to the transmission delay, which becomes very small thanks to the high data rate of FSO communications. Moreover, IR-HARQ outperforms the pure sliding window ARQ because only incremental redundancies are required for retransmission to create a robust FEC code for frame error correction in IR-HARQ. In contrast, repetitive frames without error correction are retransmitted in the pure sliding window ARQ.

C. Link/Transport Cross-Layer Frameworks

Transmission control protocol (TCP) is, by far, the most essential transport protocol for Internet applications that require reliable delivery, e.g., HTTP, email, file transfer, and some streaming media applications [155]. It is widely known that TCP tends to perform poorly in the error-prone environment of wireless channels, primarily because it misinterprets the losses due to the poor channel conditions as an indication of network congestions [156], [157]. The issues regarding the behavior of TCP over wireless links have been extensively addressed in the domain of RF-based communication networks [158] and recently investigated in the context of FSO networks [159].
A common approach to enhance the performance of TCPs over wireless fading channels is the use of link-layer retransmission protocols, including ARQ [160] and HARQ [161]. The benefit of this solution is that a transparent transmission path over wireless links is seen by TCP variants, provided most wireless channel errors can be mitigated by link-layer error control schemes. Driven by that fact, many research efforts have addressed the cross-layer design between link-layer retransmission protocols and TCP variants to improve the TCP throughput performance over the FSO last-mile access [144]–[148]. The typical FSO-based last-mile access networks are illustrated in Fig. 16, e.g., building-to-building or satellite-assisted IoV of autonomous cars and UAVs.

1) FSO-Based Terrestrial Networks: Link/Transport frameworks for FSO terrestrial networks have been reported in [144]–[146].

Framework based Uniform Error Model: The joint design of link-layer SR-ARQ with standard TCP variants, including TCP Tahoe, TCP Reno, and TCP-Selective acknowledgment (SACK) was studied in [144]. Each TCP segment is divided into $L_f$ smaller link-layer data frames. If a link-layer frame does not get through the FSO link after the ARQ’s persistent level, i.e., $N_t$, SR-ARQ gives up, and the corresponding TCP segment is clarified to be lost. While a higher value of ARQ persistent level can reduce the effects of frame or segment losses and increase the TCP throughput, it also increases the overall end-to-end latency or round-trip time (RTT) of TCP. The TCP segment loss probability (SLP) and average RTT can be written as [144]

$$\text{SLP} = 1 - \left(1 - P_{\text{others}}\right) \times \left(1 - FER^{N_t}\right)^{L_f}, \quad (10)$$

$$E[RTT] = \frac{2t_{\text{others}}}{\text{RTT of other parts}} + \left(\frac{N_t}{R_b} \cdot L_f\right) \frac{1}{N_t} \cdot \frac{L_f}{N_t} \cdot \text{RTT of FSO last-mile}, \quad (11)$$

where $E(\cdot)$ denote the expected value, $P_{\text{others}}$ is the SLP over other parts of network, $N_f$ is the frame size, $R_b$ is the bit rate, and $N_t$ is the average number of frame retransmissions. From the tradeoff between TCP throughput and end-to-end latency, the ARQ’s persistent level was suggested to be $N_t = 6$ for the considered FSO networks in [144].
Framework based Markov Error Model: A key weakness of the framework in [144] is the use of a uniform error model, which may not accurately estimate the TCP performance. For such models, the error structure of different frame transmissions is assumed to be independent. In FSO systems, during the transmission within a coherence time of turbulence fading channel (i.e., order of milliseconds [77]), the frame errors tend to occur in a burst pattern, and the error structure is correlated. It is, therefore, essential to investigate the correlated error structure to provide a more accurate performance analysis of TCP. The cross-layer design framework incorporating the PHY’s AMC scheme, link-layer SR-ARQ, and standard TCP variant (TCP Reno), was investigated in [145]. Also, instead of using the uniform error model, a more accurate performance analysis of TCP was provided, taking into account the nature of burst error by using a Markov error model. A considerable TCP throughput improvement was observed in [145] by jointly design of lower-layer schemes, both AMC and SR-ARQ.

Framework based Network Simulation: The advantage of the analytical approach in [145] is the provision of a good approximation of network performance with various channel conditions in a relatively quick fashion, which supports the optimization purposes. However, simplifying assumptions are often used at the cost of accuracy for mathematical modeling, e.g., the slow-start phase is usually ignored in the TCP performance analysis [162]. Another approach for the TCP performance evaluation is the use of network simulator, e.g., NS-2 [163], which allows an accurate evaluation of a wide range of network protocols. A framework using network simulator was introduced in [146], where TCP performance was evaluated over FSO turbulence channels in the presence of PHY error models and link-layer SR-ARQ. Using ARQ, the optimal TCP’s window size could be maintained over time as ARQ retransmissions considerably reduce the TCP’s SLP.

2) FSO-Based Satellite Networks: Link/Transport frameworks for satellite networks have been reported in [147], [148]. A cross-layer design framework between link-layer IR-HARQ based sliding window protocol and TCP variants, including NewReno, Cubic and Hybla, was mentioned for the scenario of satellite-assisted Internet of vehicles in [148]. Similar to [145], a burst loss model was developed for the cross-layer performance analysis, which can accurately estimate the TCP throughput performance. Figure 17 quantitatively compares the use of different link-layer solutions, including IR-HARQ, pure sliding window ARQ, and no ARQ/FEC, for the throughput performance of TCP Cubic [148]. As expected, the TCP throughput could be significantly enhanced when using IR-HARQ based sliding window protocol, which requires retransmitting only the incremental redundancies for erroneous frames. It can mitigate the effect of transmission losses compared to the case of no ARQ/FEC and reduce the number of retransmissions compared to using pure sliding window ARQ, thus increasing the achievable TCP performance.

Another cross-layer framework incorporating the FEC code at PHY, link-layer SR-ARQ, and TCP variants, for satellite-based hybrid FSO/RF vehicular networks was introduced in [147]. Unlike [148], where only the congestion avoidance phase was considered, a complete model taking into account other phases of TCP operation was presented in [147]. In addition, a novel channel-state model was developed considering relatively different coherence times of FSO and RF to facilitate the operation of SR-ARQ, which processes multiple frames simultaneously. Using the same approach as in [21] for both FSO and RF links, a proper design of PHY and link-layer error-control methods in [147] could improve the TCP throughput performance significantly.

D. Summary and Future Work

This section provided insightful discussions on the cross-layer frameworks of link-layer retransmission protocols with PHY methods and transport layer protocols in FSO networks. Remarkable points and future directions are summarized as follows.

- For the joint design of AR and ARQ/HARQ-based sliding window mechanism, the ARQ/HARQ’s window size is designed by the AR’s burst size, in which the burst duration is chosen to be shorter than the turbulence fading coherence time [21], [112].
- The channel-state model considering the time-varying behavior of fading channels plays an essential role in the cross-layer design of the link-layer ARQ/HARQ-based sliding window mechanism. A proper selection of channel-state duration could significantly improve the overall system performance, e.g., refer to in [47, Fig. 4] for ARQ and [112, Fig. 8] for HARQ.
- The ARQ/HARQ’s frame errors tend to happen in a burst pattern over the slowly time-varying FSO turbulence channels. Correlated error structure by Markov models can offer a more accurate analysis for PHY/Link [21], [112] and Link/Transport [145], [148] frameworks.
- Most of Link/Transport frameworks are for loss-based TCP variants, where TCP Cubic using link-layer HARQ is an efficient solution for FSO-based last-mile networks [148]. Frameworks considering the link-layer
retransmission protocols with emerging congestion control mechanisms [155], such as delay-based and recent proposals from Google, e.g., TCP BBR, would be a future direction.

V. FINDINGS, CHALLENGES, AND OPEN ISSUES

This section presents our findings as a result of the comprehensive survey of link-layer retransmission protocols, both ARQ and HARQ, and their cross-layer frameworks in the context of high-speed FSO communications. First, we elaborate on the lessons learned from the aforementioned survey of the literature. Then, we suggest the design guidelines for developing link-layer retransmission protocols in various FSO network scenarios. Finally, we identify and discuss the non-exhaustive list of research challenges, then shed light on future research directions and applications.

A. Lessons Learned

1) ARQ Aided FSO Communications: From the comprehensive literature for design and performance evaluation of ARQ protocols in the context of FSO communications in Section II, we now identify several valuable lessons learned regarding such protocols.
- **The ARQ protocol, which is sufficiently reliable yet simple, and has less overhead, is an efficient error-control solution for short-range FSO systems.** There are two main reasons [59]. Firstly, the impact of channel impairments in short-range FSO systems is less severe than that of the RF; hence using another link-layer robust method, e.g., HARQ, results in more system complexity, additional signaling, and overhead. Secondly, with the aid of PHY error-control methods, e.g., FEC and/or AR schemes, the latency caused by the ARQ’s retransmissions for residual errors in short-range FSO transmissions does not significantly degrade the system performance. In contrast, ARQ is an inefficient solution for long-range FSO systems because its performance is severely deteriorated due to the increased latency.
- **The SW-ARQ is not practically valuable for high-speed FSO systems, and sliding window protocols are preferable for such systems.** This is because the waiting time for the acknowledgment of a data frame (twice the propagation delay) in SW-ARQ becomes significant compared with the transmission delay, which is minimal in high-speed FSO systems [21]. This results in inefficiency in terms of the throughput performance of SW-ARQ in FSO systems. Sliding window protocols, which allow higher efficiency by letting data frames be continuously transmitted without waiting for acknowledgment, are the effective solution for FSO systems.
- **SR-ARQ-based PPM hard decision offers a significant latency and efficiency improvement for FSO systems.** The advantages of this novel ARQ design come from two issues [63]. First, it uses the decision result of each PPM symbol to detect errors without the need for CRC, thus reducing the latency by CRC’s generation and verification. Second, the system efficiency is significantly enhanced because only erroneous PPM symbols, which are often less than half of the whole received frame size, are retransmitted.
- **ARQ protocols using the concept of frame combining can offer considerable performance improvement without introducing a significant system cost increase.** The ARQ-based FC realized by implementing frame combining at the receiver could be used to improve the reliability of standard-ARQ protocols [59]. The main idea behind this scheme is to buffer the copies of previously received data frames for joint decoding, which enhances the likelihood of successful frame decoding.
- **The adaptive power allocation between the ARQ retransmissions significantly improves the performance of hybrid FSO/RF systems.** This fact was confirmed in [62] for the design of ARQ protocols in hybrid FSO/RF systems using parallel/simultaneous schemes. The intuition behind the considered power allocation is to weigh the energy in each ARQ’s round by its consumption probability. Additional energy is assigned to the last retransmissions, which are rarely used.

2) HARQ Aided FSO Communications: Next, we highlight the following lessons learned from the design and performance perspective of HARQ protocols over FSO turbulence-induced fading channels reviewed in Section III.
- **HARQ protocols are preferable to the pure ARQ in long-range FSO systems.** The pure ARQ protocols, which facilitate retransmissions of repetitive data frames, often fail to provide satisfactory delay and throughput performance over time-varying channel conditions, as many retransmissions are required. Benefiting from their inherent advantages, HARQ is an indispensable and effective solution to maintain reliable transmissions for long-range FSO systems.
- **The IR scheme is the most efficient HARQ variant for various FSO communication scenarios.** Although the CC scheme could enhance the link performance by coherently combining multiple copies of the received frames, the IR strategy can benefit from these as well as certain coding gains. The reason is that the different coded frames can jointly form a lower code rate with more powerful error protection capabilities.
- **It is not possible to jointly maximize the energy efficiency and throughput performance in FSO systems using HARQ protocols.** In fact, when transmitted powers are high enough, the FER saturates, and any further increase of transmitted power only leads to additional energy consumption. As a result, an optimal value of transmitted power exists, at which the energy efficiency is maximized. These values are, nonetheless, not always the optimal ones for throughput performance, and the tradeoff between them can be found in [112].
- **Information-theoretic approach is the most popular for analyzing the performance of HARQ in FSO systems.** In general, the performance analysis of HARQ protocols, which depends on the types of codes and decoding/error detection technique, is based on the Markov chain.
Modeling the system by a Markov model might be complicated since each state must convey all the information about the memory of systems [164]. As a result, the information-theoretic approach [165] could be used to gain insight into the implication on the performance of HARQ, resorting to the relatively simple models which lend themselves to the analytical investigation.

- **The HARQ-based protocol using rateless codes is an efficient solution for hybrid FSO/RF systems.** The HARQ protocols using the rateless codes adapt seamlessly to the changes in rate supported by the channels without the rate mismatch due to the imprecise/outdated channel estimation [99]. This is a distinct performance benefit over the conventional HARQ-based code-rate selection, which may fail if the channel conditions vary notably from one codeword to another.

- **The inter-user cooperative HARQ protocol is an effective solution for broadcasting FSO systems.** In FSO broadcasting systems, instead of relying on the center node, a user, who successfully decoded the original data frame and whose distance from the NAK issuing user is shorter than that of the center node, is invited to retransmit this frame to the latter user. The outperformance of the inter-user cooperative HARQ scheme compared to the conventional schemes is confirmed in [109].

3) **Cross-Layer Approach:** Finally, the lessons learned from the cross-layer design frameworks of link-layer retransmission protocols with PHY methods and/or transport layer protocols presented in Section IV can be summarized as follows.

- **The joint design of the AR scheme and ARQ/HARQ protocols is the most popular PHY/Link framework in FSO networks.** There are two main reasons. First, the CSI feedback for the AR scheme is relatively easy to implement in FSO systems due to the reciprocity in bidirectional slowly varying FSO channels, making the AR scheme more popular among PHY solutions [59]. Second, this joint design is attractive to counteract transmission errors with no additional physical infrastructure required.

- **Channel-state model considering the time-varying behavior of FSO turbulence channels plays an essential role in the cross-layer design of the AR scheme with sliding window ARQ/HARQ.** In these frameworks, the channel-state model is designed to cover the fixed-time burst transmission containing multiple frames. The channel-state and burst durations should be carefully chosen based on the FSO channel coherence time [21], [112].

- **Link-layer retransmission protocols could effectively improve the performance of TCPs, in which the selection of their persistent levels should be considered carefully.** In fact, higher values of the persistent level of the ARQ/HARQ can mitigate the impact of frame losses caused by FSO turbulence channels, thus improving the TCP throughput performance. However, such a high value of persistent level also increases the overall end-to-end latency of TCP in networks. A proper selection of persistent levels for ARQ/HARQ plays an essential role in effectively optimizing the overall performance of TCPs [144].

- **Correlated error structure by the Markov model can offer a more accurate performance analysis of TCPs in Link/Transport frameworks.** For FSO links operating at high data rates, frame errors tend to occur in a burst pattern (correlated error structure) over the slowly time-varying FSO turbulence channels. Conventional uniform error model does not reflect this critical issue, leading to inaccurate TCP throughput evaluation over turbulence-induced fading channels [145].

- **The IR-HARQ based sliding window mechanism is the most efficient link-layer solution in improving the TCP throughput over FSO-based satellite networks.** By transmitting only incremental redundancies to create a more robust code to correct erroneous frames, IR-HARQ can mitigate the effect of transmission losses compared to the case of no ARQ/ECC. Furthermore, compared with pure sliding window ARQ, it reduces the number of retransmissions required, thus increasing the achievable TCP throughput performance [148].

**B. Design Guidelines**

We now attempt to provide a design guideline of link-layer retransmission protocols recommended for several FSO scenarios, as explicitly shown in Fig. 18. It is noteworthy that this guideline comprises plausible and reasonable observations gleaned from the intensive survey of ARQ, HARQ, and the cross-layer designs, in the context of FSO communications. Based on the trade-off between complexity and performance, our recommendations for such protocol designs in different FSO network scenarios are listed below.

- **Short-range FSO systems:** According to the investigations in [21], [47], [67], the joint design of AR transmissions and sliding window protocols, i.e., either GBN-ARQ or SR-ARQ, could offer a significant throughput improvement over turbulence-induced fading channels. While SR-ARQ can provide a higher throughput performance than GBN-ARQ, it is more complex in terms of implementation. In addition, as shown in [47, Fig. 6], their throughputs were close in weak turbulence conditions. Hence, we recommend the use of AR/GBN-ARQ and AR/SR-ARQ in respectively weak and strong turbulence conditions for short-range FSO communications, e.g., building-to-building.

- **Long-range FSO systems:** For HAP-to-HAP and satellite-to-satellite FSO systems, which are less susceptible to weather, such as atmospheric turbulence and clouds, the adaptive HARQ protocols are suitable, as reported by DLR [105], [125]. In addition, it is suggested to employ a joint design of AR schemes and HARQ-based sliding window mechanism for long-range FSO systems under the impact of atmospheric turbulence, e.g., those with satellite-to-vehicles and ground-to-satellite links. As highlighted in [112], the joint design of the AR/CC-HARQ and AR/IR-HARQ could offer a considerable performance enhancement, including throughput, energy efficiency, and delay, for satellite-based FSO systems.
Hybrid FSO/RF systems: Both ARQ-based adaptive power allocation [62] and HARQ-based rateless codes [99] are promising candidates for hybrid FSO/RF systems. Notably, the former is based on the power allocation for retransmission rounds of ARQ, while the latter uses the concept of rateless code for HARQ, which does not require rate adjustment before transmissions.

Cooperative FSO systems: As introduced in [109], the inter-user cooperative HARQ protocol, using the idea of frame retransmissions by neighboring nodes instead of the center one, is an efficient solution for FSO broadcasting systems. In addition, the M-C-ARQ protocols investigated in [56] can achieve a better performance than the conventional C-ARQ ones in the context of cooperative FSO-based terrestrial systems. Regarding the high-latency satellite-based FSO communications, the use of cooperative IR-HARQ is preferable to maintain high performance, as confirmed in [111] for HAP-based relaying satellite-to-vehicles FSO systems.

C. Challenges and Open Issues

We now identify the research challenges and discuss the open issues for link-layer retransmission protocols in the vision of future sixth-generation (6G) wireless networks. Interested readers can refer to recent papers discussing the vision, applications, requirements, and technologies for future 6G wireless networks [166]–[169].

1) Vertical FSO Networks Using ARQ/HARQ: Regarding the demand for global coverage (near 100% geographical coverage) in 6G wireless networks, non-terrestrial networks, including satellites, HAPs, and UAVs, are needed to complement the terrestrial networks for the cost-effective, seamless, and ubiquitous service availability [170]. While the narrow beam is often used for FSO communications, it is possible to have a wide coverage beam footprint by FSO-based satellite networks [16]. This, in turn, leads to an extended reach of FSO networks for multiple users. An example of FSO-based satellite networks that can support multiple users within a single laser beam is illustrated in Fig. 19.

For such network scenarios, the health safety issue with eye safety restrictions must be considered carefully [24, Sec. I-C]. Regarding the design of link-layer retransmission protocols, energy efficiency and security are the critical concerns in such networks. Particularly, energy-efficient communications reduce energy consumption and extend the battery life of wireless terminals, especially important in FSO-based satellite/aerial systems (with limited powers) [171]. A low transmission error rate using retransmission protocols may require high energy consumption due to frequent retransmissions. Therefore, green communications for 6G-based non-terrestrial networks become a challenging issue for the design of link-layer retransmission protocols.
In addition, security and privacy are essential to the success of envisioned 6G wireless networks [172]. As reported in [173], retransmission is an effective way to improve reliability, but it may also compromise security due to the provision of additional diversity for eavesdroppers. For instance, using HARQ protocols, the transmitter needs to provide sufficient redundancies for the legitimate receiver to decode its message successfully. Too many redundancies, nevertheless, may help the adversarial eavesdropping [174]. Unlike conventional FSO networks using the narrow beam, security is a challenging issue for the design of link-layer retransmission protocols in wide-coverage FSO networks.

2) ARQ/HARQ in IRS-Aided FSO Systems: IRS has recently emerged as a new approach to expand the FSO communication coverage [175]. IRS is constructed by a planar meta-surface composed of a large number of reflecting elements. It is adapted by integrated electronics to control the phase shifts, amplitude, and polarization of incoming light in a programmable manner. The IRS module, which is installed on the building walls or carried by UAV/HAP/Satellite, serves as a reflector to the incident light and ensures that the transmitted light points to the receiver when the direct FSO link is not available. In fact, IRS requires less complex additional hardware than conventional relay nodes, thus realizing it as a key technology for 6G wireless networks [176].

For reliable transmissions, link-layer ARQ/HARQ protocols have been recently considered in IRS-assisted RF systems [113]. It is worth noting that the design and performance of such protocols have remained relatively unexplored in high-speed IRS-assisted FSO systems, where the key differences of IRS in FSO and RF were reported in [175]. In IRS-assisted FSO systems, the narrow FSO beam-induced pointing error and atmospheric turbulence are the most challenging issues for the widespread deployment [177]. In addition, reliable feedback of ARQ/HARQ protocols is challenging to maintain. Secondly, although link-layer retransmission protocols can fully exploit time diversity to increase reliability, they usually suffer from high latency. This latency is the result of multiple retransmissions, multiple frames decoding at the receiver, and the ACK/NAK signals transmission/processing delay [178]. It is difficult to support, for example, more than one retransmission within 1 ms end-to-end latency constraint of mURLLC at least for the initial mURLLC specification in Rel. 15 [131]. Reducing the end-to-end latency to a minimum for enabling ARQ/HARQ protocols in mURLLC becomes a critical issue.

Strategies for reducing the feedback delay using prediction mechanisms powered by machine learning techniques have been recently studied in the context of RF-based wireless communications [131], [179], [180]. From the machine learning perspective, the main task is to predict the decoding result of a given transmission using data, which is available after the first few decoder iterations. In the context of FSO communications, it is worth noting that machine learning in ARQ/HARQ protocols has not been well investigated yet. Indeed, extensive research is needed to exploit the potential of machine learning to enhance the performance of link-layer retransmission protocols. Machine learning for retransmission protocols in high-speed FSO systems would undoubtedly be a hot topic and raise interest for future studies.
4) Quantum ARQ/HARQ: In the fast technological advances over the last couple of decades, quantum technology has emerged as a promising candidate that has the potential of radically revolutionizing the way we compute and communicate [181]–[183]. Quantum communications support secure data dissemination since any measurement or observation by eavesdroppers perturbs the quantum superposition. Real-world quantum channels, as well as quantum systems, are, nevertheless, not perfect, in which quantum bits (qubits) may experience both channel-induced and quantum processing impairments [182]. For instance, the secret key transmission rate of a quantum key distribution (QKD) system deteriorates significantly over the turbulence fading channels [184].

One of the essential prerequisites to build reliable quantum communications is to employ error-control solutions. However, the law of quantum mechanics prevents us from transplanting classical error-control protocols directly into the quantum domain. The initial studies on quantum ARQ/HARQ protocols for quantum communication systems have been reported in [185], [186]. Extensive investigation is needed to exploit the potential quantum ARQ/HARQ protocols in quantum communication systems.

VI. CONCLUSION

This paper presents a state-of-the-art survey on the design and performance evaluation of link-layer retransmission-based error-control protocols, both ARQ and HARQ, in the emerging high-speed FSO communication networks. The survey was conducted extensively in various FSO communication scenarios, including point-to-point terrestrial, cooperative, multihop relaying, hybrid FSO/RF, satellite/aerial, and deep-space systems. Also, we provided a survey of recent works and insightful discussion on the cross-layer design frameworks related to link-layer retransmission protocols in FSO communication networks. The critical lessons from the survey, followed by potential research directions for each domain, have been derived. Moreover, we derived a design guideline comprising plausible and reasonable observations from the intensive survey of ARQ, HARQ, and cross-layer design frameworks. In addition, based on the holistic survey, we have pointed out the fundamental research challenges to be considered carefully for further investigation of such protocols in the context of future 6G wireless communication networks. Finally, we have discussed and outlined potential open issues toward future 6G wireless networks.

REFERENCES

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjoland, and F. Tufvesson, “6G wireless systems: Vision, requirements, challenges, insights, and opportunities,” Proc. IEEE, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.
[2] ITU. “The data traffic estimates for the years 2020 to 2030.” [Online]. Available: https://www.itu.int/pub/R-REP-M.2370-2015s
[3] Z. Ghassemlooy, S. Arnon, M. Uysal, Z. Xu, and J. Cheng, “Emerging optical wireless communications-advances and challenges,” IEEE J. Sel. Areas Commun., vol. 33, no. 9, pp. 1738–1749, Sep. 2015.
[4] M. Li, “Orbital-angular-momentum multiplexing optical wireless communications with adaptive modes adjustment in Internet-of-Things networks,” IEEE Internet Things J., vol. 6, no. 4, pp. 6134–6139, Aug. 2019.
[5] L. E. M. Matheus, A. B. Vieira, L. F. M. Vieira, M. A. M. Vieira, and O. Gnawali, “Visible light communication: Concepts, applications and challenges,” IEEE Commun. Surveys Tuts., vol. 21, no. 4, pp. 3204–3237, 4th Quart., 2019.
[6] M. Z. Chowdhury, M. K. Hasan, M. Shahjalal, M. T. Hossan, and Y. M. Jang, “Optical wireless hybrid networks: Trends, opportunities, challenges, and research directions,” IEEE Commun. Surveys Tuts., vol. 22, no. 2, pp. 930–966, 2nd Quart., 2020.
[7] A. S. Hamza, J. S. Deogun, and D. R. Alexander, “Classification framework for free space optical communication links and systems,” IEEE Commun. Surveys Tuts., vol. 21, no. 2, pp. 1346–1382, 2nd Quart., 2019.
[8] E. Yaacoub and M. Alouini, “A key 6G challenge and opportunity—Connecting the base of the pyramid: A survey on cultural connectivity,” Proc. IEEE, vol. 108, no. 4, pp. 533–582, Apr. 2020.
[9] D. R. Kolev and M. Toyoshima, “Received-power fluctuation analysis for LEO satellite-to-ground laser links,” J. Lightw. Technol., vol. 35, no. 1, pp. 103–112, Jan. 1, 2017.
[10] S. A. Al-Gailani et al., “A survey of free space optics (FSO) communication systems, links, and networks,” IEEE Access, vol. 9, pp. 7345–7373, 2021.
[11] J. B. Djordjevic, “Adaptive modulation and coding for free-space optical channels,” IEEE/OSA J. Opt. Commun. Netw., vol. 2, no. 5, pp. 221–229, May 2010.
[12] M. Karimi and M. Uysal, “Novel adaptive transmission algorithms for free-space optical links,” IEEE Trans. Commun., vol. 60, no. 12, pp. 3808–3815, Dec. 2012.
[13] A. T. Pham, T. A. Lau, and N. T. Dang, “Performance bound for turbo-coded 2-D FSO/HARQ systems over atmospheric turbulence channels,” IETE Trans. Fundam., vol. E93-A, no. 12, pp. 2696–2699, Dec. 2010.
[14] T. V. Pham and A. T. Pham, “Performance analysis of amplify-decode-and-forward multihop binary phase-shift keying/free-space optical systems using avalanche photodiode receivers over atmospheric turbulence channels,” IET Commun., vol. 8, no. 9, pp. 1518–1526, Jun. 2014.
[15] C. Abou-Rjeily and S. Haddad, “Cooperative FSO systems: Performance analysis and optimal power allocation,” J. Lightw. Technol., vol. 29, no. 7, pp. 1058–1065, Apr. 1, 2011.
[16] H. D. Le and A. T. Pham, “Level crossing rate and average fade duration of satellite-to-UAV FSO channels,” IEEE Photon. J., vol. 13, no. 1, pp. 1–14, Feb. 2021.
[17] M. Z. Hassan, M. J. Hossain, J. Cheng, and V. C. M. Leung, “Hybrid RF/FSO Backhaul networks with statistical-QoS-aware buffer-aided relaying,” IEEE Trans. Wireless Commun., vol. 19, no. 3, pp. 1464–1483, Mar. 2020.
[18] N. D. Chatzidiamantis, G. K. Karagiannidis, and M. Uysal, “Generalized maximum-likelihood sequence detection for photon-counting free space optical systems,” IEEE Trans. Commun., vol. 58, no. 12, pp. 3381–3385, Dec. 2010.
[19] I. Liu, X. Zhang, K. Blow, and S. Fowler, “Performance analysis of packet layer FEC codes and interleaving in FSO channels,” IET Commun., vol. 11, no. 13, pp. 2042–2048, Oct. 2017.
[20] V. V. Mai, T. C. Thang, and A. T. Pham, “Cross-layer design and analysis for FSO links using automatic repeat request and adaptive modulation/coding schemes,” in Proc. Int. Symp. Commun. Syst. Netw. Digit. Signal Process., 2014, pp. 1176–1180.
[21] H. D. Le, V. V. Mai, C. T. Nguyen, and A. T. Pham, “Design and analysis of sliding window ARQ protocols with rate adaptation for burst transmission over FSO turbulence channels,” IEEE/OSA J. Opt. Commun. Netw., vol. 11, no. 5, pp. 151–163, May 2019.
[22] K. Kiasaleh, “Hybrid ARQ communications for severely degraded hermite-Gaussian FSO link,” in Proc. IEEE Int. Conf. Space Opt. Syst. Appl., 2016, pp. 1–7.
[23] M. A. Khalighi and M. Uysal, “Survey on free space optical communication: A communication theory perspective,” IEEE Commun. Surveys Tuts., vol. 16, no. 4, pp. 2231–2258, 4th Quart., 2014.
[24] H. Kaushal and G. Kaddoum, “Optical communication in space: Challenges and mitigation techniques,” IEEE Commun. Surveys Tuts., vol. 19, no. 1, pp. 57–96, 1st Quart., 2017.
[25] I. K. Son and S. Mao, “A survey of free space optical networks,” Digit. Commun. Netw., vol. 3, no. 2, pp. 67–77, May 2017.
[26] N. Kaur and H. Singh, “A survey on recent work in hybrid FSO/RF communication link,” Int. J. Electron. Commun. Instrum. Eng. Res. Develop., vol. 9, no. 1, pp. 25–32, May 2019.
et al. [168] L. Bariah et al., “A prospective look: Key enabling technologies, applications and open research topics in 6G networks,” *IEEE Access*, vol. 8, pp. 174792–174820, 2020.

[169] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, “The road towards 6G: A comprehensive survey,” *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[170] X. You et al., “Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts,” *China Inf. Sci.*, vol. 64, no. 1, pp. 1869–1919, Jan. 2021.

[171] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, “A survey of energy-efficient wireless communications,” *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, 1st Quart., 2013.

[172] P. Porambage, G. Gur, D. P. M. Osorio, M. Liyanage, A. Gurtov, and M. Ylianttila, “The Roadmap to 6G security and privacy,” *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1094–1122, 2021.

[173] X. Guan, Y. Cai, and W. Yang, “On the reliability-security tradeoff and secrecy throughput in cooperative ARQ,” *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 479–482, Mar. 2014.

[174] X. Tang, R. Liu, P. Spasojevic, and H. V. Poor, “On the throughput of secure hybrid-ARQ protocols for Gaussian block-fading channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1575–1591, Apr. 2009.

[175] V. Jamali, H. Ajam, M. Najafi, B. Schmauss, R. Schober, and H. V. Poor, “Intelligent reflecting surface assisted free-space optical communications,” *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 57–63, Oct. 2021.

[176] S. Basharat, S. A. Hassan, H. Pervaiz, A. Mahmood, Z. Ding, and M. Gidlund, “Reconfigurable intelligent surfaces: Potentials, applications, and challenges for 6G wireless networks,” *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 184–191, Dec. 2021.

[177] M. Najafi, B. Schmauss, and R. Schober, “Intelligent reflecting surfaces for free space optical communication systems,” *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6134–6151, Sep. 2021.

[178] G. J. Sutton et al., “Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2488–2524, 3rd Quart., 2019.

[179] G. Qiu, M.-M. Zhao, M. Lei, and M.-J. Zhao, “Throughput Maximization for polar coded IR-HARQ using deep reinforcement learning,” in *Proc. Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2020, pp. 1–6.

[180] D. Liu, G. Zhu, Q. Zeng, J. Zhang, and K. Huang, “Wireless data acquisition for edge learning: Data-importance based retransmission,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 406–420, Jan. 2021.

[181] M. Grassl, L. Kong, Z. Wei, Z.-Q. Yin, and B. Zeng, “Quantum error-correcting codes for Qudit amplitude damping,” *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4674–4685, Jun. 2018.

[182] Z. Babar et al., “Duality of quantum and classical error correction codes: Design principles and examples,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 970–1010, 1st Quart., 2019.

[183] Z. Babar et al., “Polar codes and their quantum-domain counterparts,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 123–155, 1st Quart., 2020.

[184] N. Hosseinidehaj, Z. Babar, R. Malaney, S. X. Ng, and L. Hanzo, “Satellite-based continuous-variable quantum communications: State-of-the-art and a predictive outlook,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 881–919, 1st Quart., 2019.

[185] A. Ashikhmin, “Fidelity of a quantum ARQ protocol,” in *Proc. IEEE Inf. Theory Workshop*, 2006, pp. 42–46.

[186] J. Li, Z. Guo, and H. Ma, “Development of hybrid ARQ protocol for the quantum communication system on stabilizer codes,” *China Commun.*, vol. 18, no. 2, pp. 40–48, Feb. 2021.

**Hoang D. Le** (Member, IEEE) received the B.E. degree in electronics and communications from the Hanoi University of Science and Technology, Vietnam, in 2016, and the M.E and Ph.D. degrees in computer science and engineering from the University of Aizu (UoA), Japan, in 2018 and 2021, respectively. He is currently a Postdoctoral Researcher with the Computer Communications Laboratory, UoA. His current research interests include optical wireless communications, hybrid FSO/RF networks, quantum cryptography, cross-layer design, and error-control protocols. He was a recipient of a number of awards, including the 2018 KICS/IEEE ICTC Best Paper Award, the 2019 IEEE VTS Tokyo Chapter Young Researcher Encouragement Award, and the 2019 IEEE ComSoc Sendai Chapter Student Excellent Research Award.

**Anh T. Pham** (Senior Member, IEEE) received the B.E. and M.E. degrees in electronics engineering from the Hanoi University of Technology, Vietnam, in 1997 and 2000, respectively, and the Ph.D. degree in information and mathematical sciences from Saitama University, Japan, in 2005. From 1998 to 2002, he was with NTT Corporation, Vietnam. Since 2005, he has been a Faculty Member with the University of Aizu, where he is currently a Professor and the Head of the Computer Communications Laboratory, Division of Computer Engineering. He has authored/coauthored over 200 peer-reviewed articles on these topics. His research interests include the broad areas of communication theory and networking with a particular emphasis on modeling, design, and performance evaluation of wired/wireless communication systems and networks. He is a member of IEICE and Optica.