ELSP Adv. Inf. Commun.

Article | Received 9 February 2026; Revised 7 March 2026; Accepted 7 March 2026; Published 31 March 2026
https://doi.org/10.55092/aic20260003

A survey on insertion/deletion error detection and correction: progress and future directions

Peng Zhu[1], **Hui Yang**[1], **Chao Yu**[1,*], **Wenwu Xie**[1] and **Ji Wang**[2]

[1] School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China

[2] Department of Electronics and Information Engineering, College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China

* Correspondence author; E-mail: 12012017@hnist.edu.cn.

Highlights:

- Reviews and categorizes indel error detection and correction methods from traditional to deep learning approaches.
- Focuses on the strengths, limitations, and applicability of these methods in complex communication scenarios.
- Presents current challenges and future directions toward ultra-reliable next-generation communication networks.

Abstract: The detection and correction of insertion/deletion (indel) errors have become increasingly critical in domains such as traditional mobile communication systems, the Internet of Things (IoT), smart homes, smart healthcare, vehicular networks, and large-scale urban infrastructure, establishing it as a prominent research focus. As a typical form of synchronization error, the randomness and asymmetry of indel errors severely disrupt symbol alignment and induce significant synchronization drift, thereby imposing substantial challenges on reliable data transmission. This paper systematically reviews methodologies for detecting and correcting indel errors, tracing their evolution from model-driven to data-driven paradigms. First, we summarize the traditional technical framework, which includes synchronization markers, edit distance (ED) codes, sequence alignment, trellis/convolutional structures, and probabilistic models, with an analysis of their theoretical foundations, representative algorithms, and applicable scenarios. Next, we focus on recent advances in deep learning (DL)-based synchronization recovery methods and semantic communication-driven intelligent error correction frameworks, highlighting their distinct advantages over conventional approaches in handling complex channels and unstructured data. Finally, we outline the current research landscape and key challenges in this field and propose future directions for emerging scenarios such as 6th Generation (6G) ultra-reliable communication, satellite links, and ultra-high-density storage. This review aims to provide comprehensive insights and guidance for the design of synchronization and error correction mechanisms in next-generation communication systems.

Keywords: insertion/deletion error; detection; correction; synchronization; deep learning; semantic communication

1. Introduction

The reliable detection and correction of insertion/deletion (indel) errors is fundamental to ensuring the integrity of information transmission in modern communication systems. These errors manifest as spurious additions or losses of symbols, disrupting transmitter-receiver synchronization and compromising reliable data recovery. A generic indel channel model is illustrated in Figure 1 for reference. With the advent and rapid deployment of fifth-generation (5G) and sixth-generation (6G) mobile networks, the proliferation of the massive Internet of Things (IoT), and the surge in smart devices, the demands on communication systems have increased significantly. Requirements for ultra-precise timing synchronization, stringent data integrity, and robust adaptability to highly dynamic environments have rendered the development of advanced indel error control techniques a critical research frontier [1–3]. Effective solutions in this domain are essential not only for enhancing communication quality and system stability under complex conditions but also for enabling mission-critical applications that depend on high reliability and low latency.

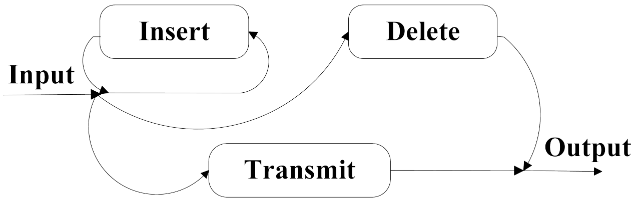

Figure 1. Indel channel model.

Conventional approaches to indel error mitigation have primarily relied on frame header identification, dedicated preamble sequences, or basic coding check mechanisms (e.g., cyclic redundancy checks) to achieve synchronization and error correction [4,5]. Although these methods are advantageous due to their low implementation cost and simplicity, their effectiveness decreases significantly in challenging environments characterized by dynamic channel interference, high error rates, or stringent latency constraints. This performance gap renders them increasingly inadequate for meeting the adaptive and robust demands of next-generation communication systems.

From a methodological perspective, existing research on indel error detection and correction can be broadly categorized into two paradigms: traditional synchronization and recovery-based approaches, and intelligent deep learning (DL)-driven methods. The former aims to explicitly estimate and compensate for synchronization drift caused by indels through structured designs and algorithmic recovery, whereas the latter seeks to implicitly model sequence distortions using data-driven learning frameworks. These two paradigms reflect fundamentally different design philosophies and exhibit distinct advantages and limitations under various system constraints.

Conventional frameworks for mitigating indel-induced synchronization errors primarily rely on explicit synchronization and recovery mechanisms. Historically, a wide range of techniques have been developed to address synchronization loss caused by indel errors, while subsequent research gradually shifted toward more general recovery-oriented frameworks that abstract synchronization restoration as a sequence-level reconstruction problem.

In contrast, the intelligent DL-driven paradigm represents a transformative shift by leveraging the powerful nonlinear modeling and representation learning capabilities of deep neural networks (DNNs). By learning directly from data, these approaches are capable of capturing complex, non-local, and highly nonlinear sequence distortions induced by insertion and deletion errors without relying on explicit analytical channel models. As a result, DL-based methods have shown strong potential in handling complex, highly dynamic, and nonstationary environments, particularly in scenarios where conventional model-based approaches struggle to maintain reliable performance.

Despite these advantages, the adoption of DL-driven indel error correction also introduces a series of practical challenges. Most DL-based approaches rely heavily on large-scale labeled training datasets, which are often costly or difficult to obtain for realistic indel channel conditions. In addition, the training and inference processes typically incur high computational complexity and memory overhead, posing significant challenges for real-time implementation and deployment on resource-constrained edge devices. Furthermore, the generalization capability and interpretability of DL models under varying channel conditions remain open issues, which may limit their reliability in safety-critical and latency-sensitive applications.

The ongoing commercialization of 5G and the proactive development of 6G technologies are further intensifying system requirements, pushing the boundaries of data rates, intelligence, end-to-end latency, and reliability. This evolution inevitably leads to more complex system architectures and, consequently, exacerbates the synchronization challenges induced by indel errors [6]. Against this backdrop, this paper presents a comprehensive survey of indel error detection and correction techniques, with an emphasis on systematic taxonomy, methodological comparison, and open research challenges across traditional and DL-driven paradigms. We examine their technical principles, algorithmic implementations, and respective application scenarios. By elucidating the distinct characteristics, comparative advantages, and inherent limitations of each approach, this survey aims to serve as a valuable reference for researchers and engineers working in this field. Given the unique, non-local, and asymmetric nature of indel errors compared to substitution errors, continued innovation in their detection and correction is poised to play a pivotal role in unlocking the full potential and ensuring the robust foundation of next-generation communication technologies.

1.1. Generation mechanisms and manifestations of indel errors

The intrinsic nature of indel errors stems from structural distortions in sequences caused by synchronization failures. These errors result from interactions between the physical limitations of hardware devices and non-ideal transmission conditions, manifesting through system-specific physical mechanisms.

(1) Storage Systems

In high-density magnetic storage systems, head vibration and misalignment with magnetic islands can result in bit-level insertions or deletions [7]. In Deoxyribonucleic Acid (DNA) storage systems, nucleotide deletions frequently occur during synthesis, while the sequencing stage is prone to base over-calling [8,9]. For racetrack memory, positioning inaccuracies may lead to phenomena such as “sticky insertions” or “unread deletions” [10].

(2) Communication Systems

In high-speed mobile communications, the accumulation of Doppler frequency shift and symbol timing offset may lead to bit-level misalignment. Satellite communications, characterized by long propagation delays and clock discrepancies, are particularly vulnerable to burst synchronization errors. Furthermore, IoT terminals equipped with low-precision clocks and operating under random access protocols demonstrate an increased probability of indel errors.

From a system-level perspective, indel errors may manifest in several typical forms, including burst errors (e.g., consecutive deletions during satellite handover), consecutive errors (e.g., multiple repeated reads in racetrack memory), non-binary errors (e.g., symbol insertion in non-binary channels), and errors with unknown codeword boundaries (e.g., frame start misalignment in asynchronous communication). Each type imposes significantly different requirements on the design of error correction techniques [11,12], as illustrated in Figure 2.

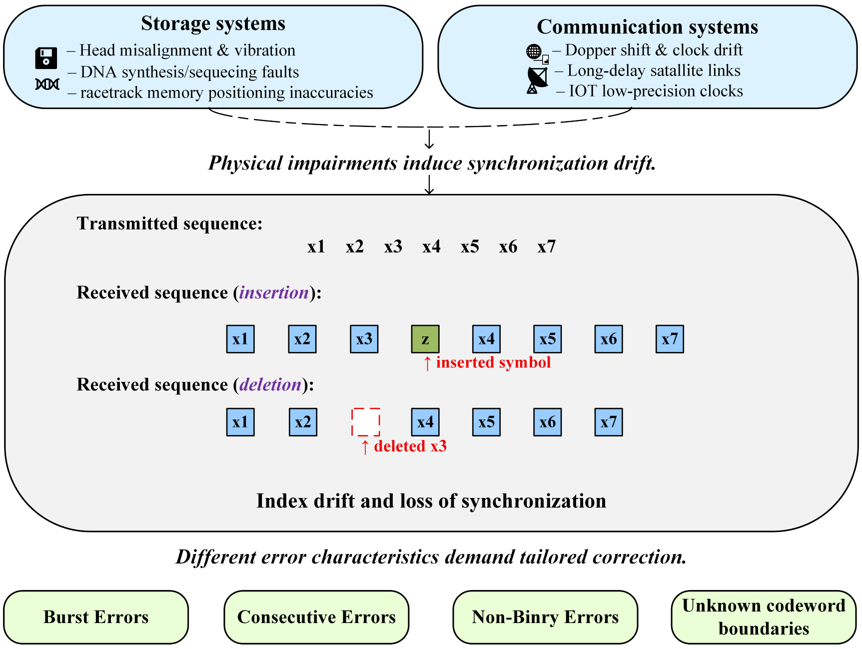

Figure 2. Illustration of physical mechanisms and sequence-level manifestations of insertion and deletion errors.

1.2. Application scenarios

As communication networks evolve toward higher data rates, enhanced reliability, massive connectivity, and greater intelligence, indel errors exhibit new manifestations and impose increasing demands across a wide range of systems. Generally, their typical applications and associated research scenarios can be categorized as follows:

(1) Mobile Communication Scenarios

In mobile communication systems, particularly within 5G and 6G networks, ultra-high-speed mobility scenarios, such as those involving high-speed rail, vehicular networks, and unmanned aerial vehicle (UAV) communications, have emerged as critical target applications. In these environments, the high relative velocity between user terminals and base stations subjects signals to significant Doppler frequency shifts, pronounced fast time-varying fading, and dynamic timing drift. The continuous accumulation of timing and sampling offsets progressively shifts the receiver's sampling points away from their ideal positions, ultimately resulting in structural misalignment at the symbol or even bit level, which manifests as insertion or deletion errors. Furthermore, as communication systems achieve higher data rates, the symbol interval decreases, thereby amplifying sensitivity to clock inaccuracies. Even minor clock drift, front-end circuit jitter, or nonlinear distortion in the radio-frequency chain can induce bit-level misalignment, consequently highlighting the importance of indel errors in high-rate scenarios [13].

(2) Satellite Communication and Space-Air-Ground Integrated Networks

In satellite communication systems, particularly in Low Earth Orbit (LEO) satellite constellations and Space-Air-Ground Integrated Networks (SAGIN), the mechanisms underlying indel errors are more complex. First, due to the high orbital velocity of LEO satellites, receivers experience extreme Doppler frequency shifts and rapidly varying propagation delays. The Doppler effect not only introduces carrier frequency offset but also alters the optimal symbol sampling instants, leading to time-accumulating sampling drift and consequent sequence desynchronization. Without timely clock compensation or frequency tracking, structural indel errors may be induced. Second, LEO satellite links are typically characterized by long propagation delays, significant path loss, and complex shadow fading. These factors collectively result in signal instability and pronounced fluctuations in effective channel bandwidth, rendering the transmitted frame structure susceptible to corruption and thereby inducing synchronization misalignment.

Additionally, in scenarios such as multi-satellite access, satellite handovers, and inter-satellite link coordination, inconsistencies among the local clock systems of different satellites can give rise to synchronization issues resembling indel errors. Consequently, mitigating indel errors has become a critical research focus in satellite communication synchronization, receiver design, and robust coding schemes.

(3) IoT and Large-Scale Low-Power Terminals

In IoT scenarios, particularly within massive machine-type communication (mMTC) and low-power wide-area network (LPWAN) technologies, terminal devices are typically designed with low cost, low power consumption, low complexity, and limited clock precision. Due to their high sensitivity to cost and power constraints, the limited accuracy of local clocks results in cumulative frequency drift. This leads to significant timing deviations between terminals and base stations following uplink wake-up. If the network synchronization mechanism fails to correct these deviations promptly, the receiver may experience bit indels.

Furthermore, factors such as random access mechanisms, shared spectrum usage, multi-user interference, and temporally unaligned bursty access behaviors prevalent in IoT networks at large scale further increase the probability of indel errors, making this domain an important research scenario.

(4) Novel Storage Systems and Specialized Application Scenarios

Beyond the communication domain, a variety of emerging storage and information encoding paradigms inherently operate over channels that are susceptible to insertion and deletion errors. Representative examples include high-density storage technologies such as DNA storage systems, racetrack memory, magnetic/optical hybrid storage, and bit-patterned magnetic recording (BPMR), as well as system-level scenarios such as packet-switched networks and distributed storage architectures.

In DNA storage systems, the processes of sequence synthesis and sequencing, while in racetrack memory systems, positioning and read/write operations, impose stringent requirements on the synchronization recovery capability of error correction techniques.

(5) Smart Healthcare and Industrial Control Scenarios

In healthcare monitoring applications, real-time data transmission from remote sensing and monitoring devices (e.g., heart rate monitors and glucose sensors) to backend processing units requires high timing accuracy and data integrity. Device mobility and in-body signal propagation effects may introduce indel errors, which, if not properly mitigated, can degrade the reliability of diagnostic data.

In industrial control scenarios, real-time command delivery and sensor data synchronization in Industrial Internet of Things (IIoT) systems operate in harsh environments subject to electromagnetic interference and mechanical vibration. Under such conditions, indel errors may disrupt control loops and data consistency. In flexible and automated manufacturing systems, the stringent synchronization requirements for multi-device coordination further highlight the need for robust indel error detection and correction techniques.

Based on the above discussion, insertion and deletion errors occur in a wide range of modern systems with distinct physical origins yet similar synchronization-related manifestations. To provide an intuitive and comprehensive view of these application domains, Figure 3 summarizes integrated application scenarios of indel error detection and correction, encompassing mobile and satellite communications, large-scale IoT networks, emerging storage systems, as well as healthcare and industrial control applications.

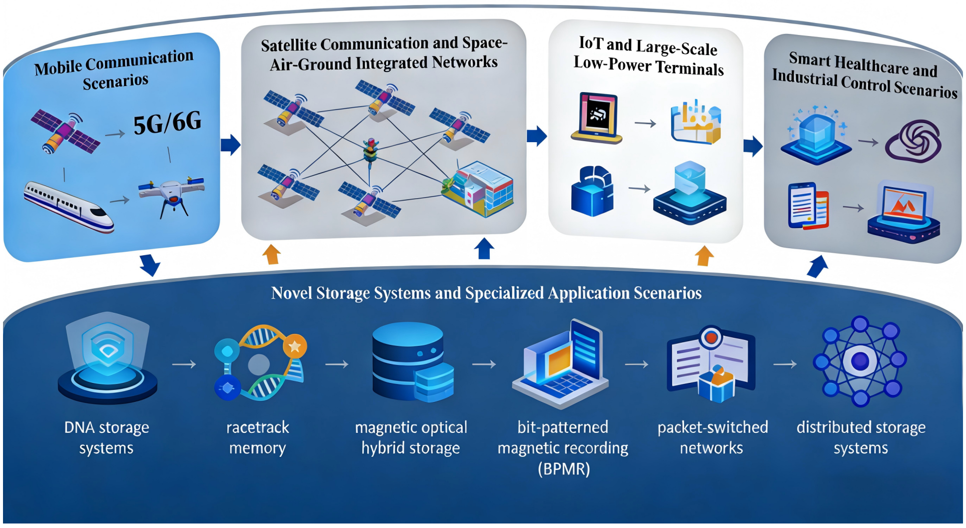

Figure 3. Integrated application scenarios in indel error detection and correction.

2. Traditional synchronization and recovery methods

Building upon the research background and methodological taxonomy introduced in the previous section, this section systematically reviews representative techniques for indel error correction under traditional synchronization and recovery frameworks.

The existing literature in this domain primarily emphasizes explicit synchronization estimation, structured sequence reconstruction, and probabilistic modeling. Accordingly, the related approaches are broadly categorized into five major classes: synchronization marker-based detection and recovery, edit distance (ED)-based error-correcting codes, sequence alignment-based decoding, trellis/convolutional structure-based error correction, and model-based probabilistic error correction, as summarized in Figure 4.

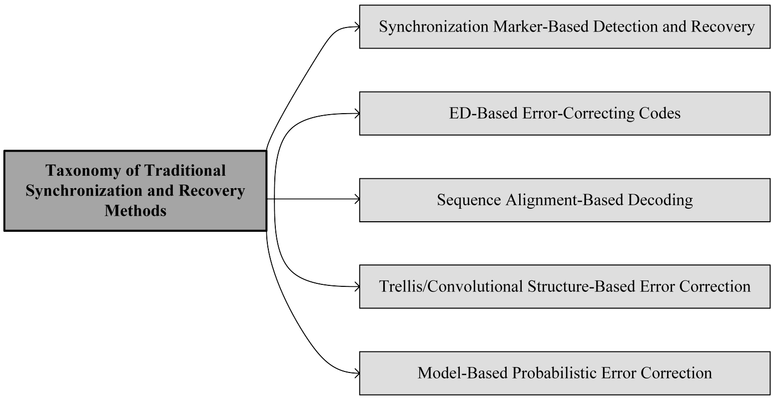

Figure 4. Taxonomy of traditional synchronization and recovery methods for indel errors.

2.1. Synchronization marker-based detection and recovery

Synchronization marker-based detection and recovery is among the earliest and most widely adopted approaches for handling indel errors. The core idea involves embedding unique or periodic synchronization markers into the information sequence. At the receiver, marker positions are identified via sliding-window detection and correlation matching. Missing, repeated, or abnormally shifted markers are typically interpreted as indications of indel errors along the transmission link. Synchronization recovery is then performed by repositioning the detected markers and realigning the received data sequence, often in conjunction with simple compensation or error-checking mechanisms to mitigate the effects of insertions and deletions [14–17].

This approach offers the advantages of straightforward implementation and deterministic procedural design. However, its error correction capability is inherently limited, and the insertion of synchronization markers introduces non-negligible redundancy overhead. As a result, this class of methods is generally suitable for scenarios with low indel error rates and constrained system resources, such as uplink data transmission in low-power IoT terminals or basic synchronization recovery in traditional magnetic storage systems. To address these limitations, subsequent research has increasingly shifted toward exploiting intrinsic sequence structure, thereby motivating the development of ED-based coding frameworks.

2.2. ED-based error-correcting codes

ED-based error-correcting codes achieve indel correction by embedding synchronization robustness intrinsically into the codeword structure, thereby eliminating the need for explicit synchronization markers. The fundamental design objective of such codes is to ensure a sufficiently large minimum ED between any pair of codewords.

Research in this field originated in the 1960s. Levenshtein introduced the concept of ED in 1966 and subsequently proposed Levenshtein codes, which can correct a single insertion or deletion error [18–20]. In the 1990s, Helberg and colleagues developed codes based on overlapping structures, correcting indel errors by ensuring specific overlapping properties among sub-blocks within codewords [21]. Later, Sloane and other researchers investigated Cayley codes based on group theory and combinatorial design. These codes exhibit favorable ED properties and can efficiently correct a limited number of indel errors; however, their construction is complex, and their coding rates are generally low [22,23]. In recent years, efforts have been made to integrate the algebraic structures of established algebraic-geometric codes, such as

Reed-Solomon codes and BCH codes, with ED characteristics, yielding longer codes with moderate correction capability and improved coding efficiency [24–27].

While ED-based codes eliminate the need for external synchronization markers, they are characterized by complex code construction and high decoding complexity. This method is well-suited for static environments where redundancy overhead is not a primary constraint and high correction accuracy is required, such as in low-speed data transmission between fixed devices.

2.3. Sequence alignment-based decoding

From an algorithmic perspective, sequence alignment-based decoding provides a flexible framework for explicitly localizing insertion and deletion events through ED-driven optimization, without imposing strict constraints on the underlying code structure. These methods identify errors and reconstruct sequences by directly comparing the ED between received and candidate sequences, without relying on predefined coding structures or explicit synchronization markers [28].

Technically, sequence alignment-based decoding typically relies on a dynamic programming (DP) framework, encompassing classical algorithms such as the Needleman-Wunsch global alignment algorithm and the Smith-Waterman local alignment algorithm [29–32]. These algorithms are capable of inferring the positions of insertions and deletions, thereby enabling accurate synchronization recovery and sequence reconstruction. Moreover, alignment-based methods exhibit strong extensibility and can be readily integrated with other error-correction modules to improve overall performance in complex channel conditions.

However, conventional DP-based algorithms suffer from high computational complexity when applied to long sequences, which often limits their applicability in real-time systems. In addition, their decoding performance generally falls short of the theoretical upper bound associated with minimum-distance (MD) decoding. To achieve tighter decoding bounds while maintaining practical performance, recent studies have reformulated the ED minimization or MD decoding problem as an integer programming (IP) problem and employed advanced optimization techniques for its solution.

To further address this challenge, semidefinite programming (SDP)-based relaxation methods have been proposed. Compared with linear programming (LP) relaxation, SDP offers the advantage of providing a tighter relaxation bound. By constructing an SDP matrix and incorporating the associated constraints, the relaxed feasible region becomes more compact, enabling the resulting solution to more closely approach optimal MD decoding performance. As a result, SDP-based approaches often outperform LP relaxation and conventional pruning strategies in medium-to-short block-length regimes and have emerged as a representative class of optimization tools in recent alignment-based decoding research.

Representative results reported in the literature, as illustrated in Figure 5, indicate that under a code rate of 1/8, a block length of 16, and a single-deletion scenario, SDP-based relaxation achieves bit error rate (BER) performance close to that of optimal MD decoding. In addition, it significantly outperforms LP relaxation, the list constant depth (LCD) algorithm, and the list constant depth branch-and-cut (LCDC) algorithm. Notably, the performance of the SDP-based method remains relatively insensitive to the subset size M, highlighting its effectiveness in balancing decoding performance and computational complexity. These observations suggest that alignment-based decoding can approach near-optimal performance through optimization-based relaxation, albeit at the cost of increased computational complexity.

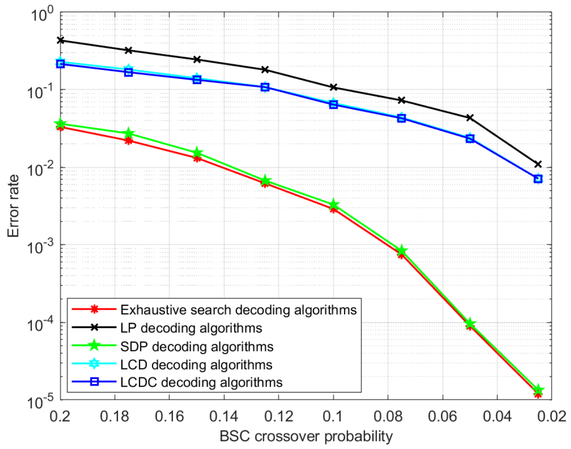

Figure 5. BER performance of rate-1/8 codes over a (16,1) deletion channel.

The theoretical and algorithmic foundations of sequence alignment methods originate from the field of bioinformatics. These studies are reviewed here not for their biological applications per se, but to highlight the strong methodological connections between sequence alignment theory and indel error correction in communication systems.

Early representative work includes the study by Gao *et al.* [33], who systematically evaluated 19 sequence analysis algorithms using a database of human short coding sequences, thereby providing a rigorous benchmark for alignment accuracy and robustness. In 2007, Schwartz *et al.* [34] proposed a multiple sequence alignment framework based on posterior decoding, establishing quantitative evaluation criteria that are closely related to probabilistic inference and optimization-based decoding. More recently, the (genomic pretrained network with multiple-sequence alignment) GPN-MSA model introduced by Benegas *et al.* in Nature Biotechnology [35] integrates multiple sequence alignment with Transformer architectures, demonstrating the effectiveness of learning-based alignment and structured sequence modeling under complex insertion and deletion patterns.

Although these works primarily focus on biological sequences, their contributions to ED modeling, sequence structure analysis, posterior inference, and optimization-based alignment provide valuable theoretical and algorithmic insights for indel error correction in communication systems. In particular, these ideas have influenced the development of alignment-based decoding and optimization frameworks in modern communication receivers, especially in scenarios involving severe synchronization uncertainty.

Overall, sequence alignment-based decoding methods offer several advantages, including strong versatility, independence from explicit synchronization markers, and accurate localization of insertion and deletion events. Furthermore, they can naturally leverage mature bioinformatics analysis tools and algorithmic frameworks.

However, these methods also exhibit inherent limitations. Their computational complexity increases rapidly with sequence length, often necessitating heuristic listing or pruning strategies in long-sequence inference. Additionally, decoding performance is sensitive to the choice of scoring functions and cost parameters, which may result in unstable bit error rate behavior. Consequently, the direct application of sequence alignment-based decoding to long codewords and highly dynamic communication channels remains challenging.

These limitations have motivated subsequent research toward more efficient trellis-based designs, as well as probabilistic and model-driven decoding approaches.

2.4. *Trellis/convolutional structure-based error correction*

These methods represent a shift from explicit alignment estimation toward joint synchronization and decoding within a structured state-space framework. To address the high decoding complexity of indel errors in long sequences, researchers have shifted their focus to trellis/convolutional structure-based error correction methods. These approaches employ convolutional codes as the core encoding mechanism and utilize trellis diagrams to represent state transitions. By leveraging the state memory properties of the encoder and the efficient search capabilities of DP, these methods simultaneously address synchronization recovery and data correction. This methodology is widely implemented in applications requiring stringent real-time performance and high stability, such as satellite and mobile communications.

(a) Indel Error Correction Based on Trellis Structures

To address synchronization disruption caused by indel errors, trellis-based correction schemes have been developed through two primary technical approaches: extended-state trellis methods and iterative synchronization-recovery trellis methods.

The theoretical foundations of the extended-state trellis methodology originate from research on deletion channel capacity and error-correcting coding in the 1960s and 1970s. For instance, Lynch [36] and subsequent researchers [37,38] incorporated trellis concepts into their analyses of deletion channel capacity and coding, establishing a methodological framework for sequence recovery using trellis structures in indel scenarios.

In contrast, the iterative synchronization recovery trellis method represents an optimization strategy for managing dynamic synchronization offsets in complex channels. Its core principle involves tightly coupling synchronization recovery with error-correction decoding, achieving co-optimization through iterative updates of soft information. This approach addresses the limitation of traditional trellis methods, which rely on predetermined synchronization references. Building on Tanner's ideas, Wiberg [39] proposed a general framework for describing codes and iterative decoding, which has since served as the foundation for subsequent iterative trellis and graph-based decoding schemes. Later, Sweeney *et al.* [40] applied this framework to develop iterative soft-decision decoding for linear block codes, demonstrating the feasibility of the iterative trellis framework. Subsequent researchers [41,42] have further employed this framework in various studies, often using convolutional codes as the core component codes, making this a critical direction for the practical implementation of indel error correction.

(b) Coding Applications and Optimization of Convolutional Structure-Based Correction

Given the strong performance of convolutional codes as component codes in iterative decoding and their inherent trellis structure, applying convolutional structures to indel error correction presents a central challenge: shifting the optimization objective from traditional Hamming distance to ED and designing corresponding encoding and decoding schemes to ensure reliable decoding even under symbol synchronization loss. Research efforts have primarily focused on two levels: encoder design and decoding algorithm optimization.

At the encoder design level, the goal is to construct convolutional codes with maximum or near-maximum ED. This builds on the tradition of classical convolutional codes proposed by Elias [43], though the design criteria are more complex. By imposing specific encoding constraints, the impact of consecutive indel errors on the codestream can be mitigated, resulting in more distinguishable paths in the trellis diagram and thereby enhancing synchronization recovery capability.

At the decoding algorithm level, classical algorithms have been significantly modified to accommodate indel channels. The Viterbi algorithm (VA) [44], a fundamental component of sequence detection, cannot be directly applied to indel correction in its standard form. Researchers have redefined state and branch metrics in the trellis to address symbol synchronization offsets, resulting in channel decoding algorithms specifically designed for indel errors. To achieve improved performance within iterative decoding frameworks (e.g., turbo code structures), soft-output decoding algorithms have become essential. Among these, the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [45] and its variants play a critical role due to their ability to provide accurate posterior probability soft information. In indel channels, the classical BCJR algorithm is typically extended by augmenting the trellis with explicit synchronization states and redefining the forward-backward recursions to account for insertion and deletion probabilities, thereby enabling joint synchronization recovery and soft-input soft-output decoding.

Similarly, the soft-output Viterbi algorithm (SOVA) [46] has been extensively studied within iterative decoding frameworks owing to its lower computational complexity. However, its soft-output reliability is generally inferior to that of BCJR-based methods. Consequently, current research emphasizes the trade-off between decoding complexity and soft-information accuracy, facilitating the practical implementation of convolutional structures for indel error correction.

2.5. *Model-based probabilistic error correction*

Although trellis/convolutional structure-based methods effectively reduce complexity in processing long sequences, increasingly complex channel environments and the pursuit of higher correction performance have led researchers to develop model-based probabilistic error correction methods. This approach extends beyond dependence on specific coding structures by modeling the entire communication system as a unified probabilistic generative model.

The core concept involves treating the transmitted sequence, the channel synchronization state, and the received sequence as random variables. Using probabilistic inference techniques, the most probable original data and synchronization positions are jointly estimated, thereby transforming the indel error correction problem into a standard statistical inference problem. A key advantage of this approach is its robust adaptability to unknown error distributions. Consequently, it has emerged as a predominant error correction technique in complex, heterogeneous scenarios such as DNA storage and natural language processing (NLP) [47,48].

(a) Traditional Statistical Model-Driven Correction Schemes

Within the framework of model-based probabilistic error correction, traditional statistical model-driven approaches establish the theoretical and methodological foundation for the field. Here, “model-driven” specifically refers to methods based on explicit probabilistic channel modeling. These approaches fundamentally involve constructing an accurate probabilistic model for the indel channel and subsequently applying rigorous statistical inference algorithms to jointly estimate the transmitted sequence and the channel synchronization state in an optimal manner. Among these methods, the Hidden Markov Model (HMM), due to its inherent capacity to model temporal dependencies and its unified representation of insertion, deletion, and substitution errors, has emerged as the most mature and representative technique, providing a robust probabilistic foundation for sequence labeling problems.

Specifically, the HMM abstracts the indel of the channel as a double random process: the synchronization state of the channel is modeled as a sequence of hidden states that follows the evolution of a Markov chain; the sequence of symbols observed by the receiver, which may be mismatched in length due to synchronization loss, is modeled as the random output of the state sequence. Under this framework, the error correction decoding task is formally defined as: given the observed sequence, solve for the hidden state sequence (*i.e.*, the synchronization trajectory) and the original input sequence that are most likely to generate the observed sequence, thereby transforming a communication problem into a structured statistical inference problem. Based on the HMM, the classic VA can be used to solve for the most likely synchronization path and the transmission sequence, and realize maximum likelihood sequence estimation [49].

To enhance performance and generality, various extensions of the basic HMM have been developed. In particular, the BCJR algorithm and the SOVA can be formulated within the HMM framework to support soft-input soft-output decoding, while parameter-adaptive mechanisms have been introduced to improve robustness against channel variations [50]. Nevertheless, despite their theoretical optimality under accurate prior assumptions, HMM-based methods are highly sensitive to model mismatch. In complex or time-varying channels, inaccurate knowledge of indel statistics and restrictive model assumptions can lead to severe performance degradation, especially in high-error-rate or data-heterogeneous scenarios [51]. These limitations have motivated subsequent research on adaptive optimization and, more recently, deep learning–based approaches that aim to learn probabilistic channel models directly from data, thereby relaxing explicit modeling assumptions [52].

(b) Other Probabilistic Models

Besides HMMs, a variety of probabilistic graphical models (PGMs) have been introduced for indel channels to overcome the limitations of first-order temporal dependency and restrictive generative assumptions inherent in conventional Markov-based frameworks. By offering richer representational capacity and more flexible inference mechanisms, these models enable improved handling of complex synchronization dynamics and indel error patterns.

Factor graphs (FGs), as fundamental representation tools in PGMs, provide significant advantages for joint synchronization and error-correction tasks. An FG explicitly models local probabilistic constraints associated with insertion, deletion, and substitution operations through factor nodes. By employing message-passing algorithms such as the sum-product algorithm (SPA), joint probabilistic inference over synchronization states and symbol sequences can be efficiently performed. In this framework, indel errors are often represented as dedicated “edit factors”, which can be seamlessly integrated with coding constraints, prior distributions, and soft information. Compared with standalone HMM-based formulations, FG-based approaches offer greater modeling flexibility and are particularly suitable for complex channel scenarios involving multiple interacting constraints [53].

Dynamic Bayesian networks (DBNs) extend the HMM framework by introducing multiple interrelated hidden state variables and higher-order temporal dependencies. This extension enables DBNs to capture compound synchronization drift and multifactorial interactions among indel events, making them well suited for channels characterized by complex temporal dynamics and correlated synchronization errors.

Conditional random fields (CRFs) represent a notable departure from generative models such as HMMs by adopting a discriminative modeling paradigm. Rather than modeling the joint distribution of observations and hidden states, CRFs directly estimate the conditional probability of the hidden synchronization state sequence given the observed sequence. This formulation allows CRFs to incorporate global and long-range features related to indel events, thereby substantially improving correction accuracy in scenarios with strong temporal correlations. As a result, CRFs have emerged as an effective technique for sequential indel error correction in communication systems [54,55].

Markov random fields (MRFs), as undirected PGMs, have also been explored for indel error correction, particularly in applications involving spatially correlated data such as images and video. By enforcing local continuity through potential functions defined over neighboring elements, MRFs can effectively mitigate deletion-induced data loss and spurious insertions caused by transmission impairments. Although MRFs are less suitable for modeling temporal synchronization processes, they complement sequential probabilistic models in spatial indel correction tasks.

Overall, the introduction of DBNs, CRFs, and MRFs significantly alleviates the limitations of traditional HMMs in terms of temporal rigidity and restrictive generative assumptions, leading to enhanced modeling flexibility and improved correction performance in complex channels. Nevertheless, exact inference in these advanced probabilistic models often entails high computational complexity and strong dependence on prior model assumptions. To address these challenges, recent research has increasingly incorporated DL-driven techniques, which leverage powerful feature extraction and nonlinear modeling capabilities to enable scalable, data-adaptive, and end-to-end optimization of indel error correction systems [56].

To facilitate a concise overview, Table 1 summarizes the core mechanisms, key advantages, and primary limitations of the aforementioned traditional synchronization and recovery methods.

Table 1. Comparison of traditional methods for indel error correction.

Method Category	Core Mechanism	Key Advantages	Major Limitations
Synchronization Marker-Based Methods	Embedding synchronization markers in the sequence.	Simple and technological maturity.	Error correction capability is limited, and markers introduce redundancy.
ED-Based Codes	The minimum ED is determined by the code set.	Does not rely on external synchronization markers.	Low coding efficiency and high decoding complexity.
Sequence Alignment-Based Decoding	Directly compare the ED between sequences.	No coding structure or synchronization markers are required.	The computational complexity is high, making it difficult to meet real-time requirements for long sequences.
Trellis/Convolutional Structure-Based Methods	By iterating through extended states or soft information.	Decoding robustness and strong adaptability.	Design complexity and sensitivity to channel modeling assumptions.
Model-Based Probabilistic Methods	Modeling the communication system as a probabilistic generative model.	No need to preset the encoding structure.	High inference complexity and strong dependency on accurate prior models.

Note: Table 1 summarizes five categories of traditional techniques for correcting insertion and deletion (Indel) errors, comparing their fundamental principles, strengths, and practical limitations.

3. Intelligent DL-driven methods

Traditional synchronization recovery methods often encounter performance bottlenecks in high-complexity scenarios, such as time-varying channels, high error rates, and long-sequence inference, primarily due to rigid model assumptions and strong dependence on prior parameters. In response to these limitations, researchers have increasingly turned to artificial intelligence (AI) techniques, with DL emerging as a prominent paradigm. By exploiting its nonlinear representation capability, end-to-end optimization framework, and data-driven adaptive learning mechanisms, DL has introduced a new methodological avenue for indel error detection and correction [57].

Compared with conventional model-driven approaches, DL-driven methods alleviate the reliance on precise analytical channel models and predefined coding structures. Instead, they directly learn representative features of channel distortions and indel error patterns from large-scale data, enabling a paradigm shift from model-driven to data-driven solutions. As a result, these methods can adaptively detect and correct sequence desynchronization and structural distortions caused by indel errors, demonstrating superior robustness and performance in complex and dynamic environments.

Driven by this transition, current research in intelligent indel error correction has evolved along three main directions:

(i) DL-based synchronization recovery and iterative correction, where DL modules are integrated into conventional decoding frameworks to enhance soft-information inference and message-passing processes for synchronization states [58];

(ii) semantic communication and DL-driven intelligent correction, which aims to move beyond bit-level fidelity by extracting and recovering semantic information, thereby enabling higher-level reliable communication in unstructured data transmission scenarios [59];

(iii) quantum-coding-based indel correction, which targets the unique requirements of quantum storage and quantum communication systems [60]. Figure 6 Overview of intelligent DL-driven methods for indel error correction, including DL-based synchronization recovery and iterative correction, semantic communication–driven intelligent correction, and quantum-coding-based indel error correction.

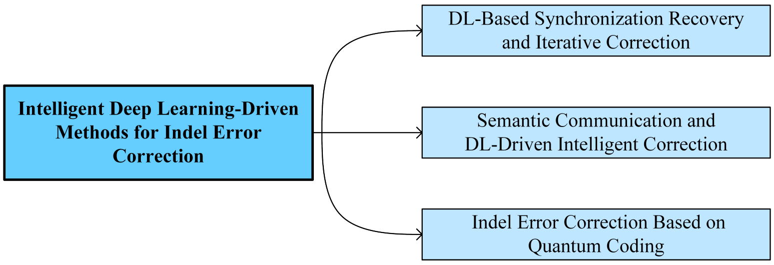

Figure 6. Taxonomy of intelligent DL-driven methods for indel error correction.

3.1. DL-based synchronization recovery and iterative correction

In the context of detecting and correcting indel errors, DL-based synchronization recovery and iterative correction represent a significant advancement from traditional model-driven approaches to intelligent, data-driven methodologies. The core strategy involves leveraging the nonlinear fitting capability of DL by integrating DL modules into the synchronization recovery and iterative correction processes while preserving the overall structure of conventional decoding frameworks. This integration enhances the accuracy of synchronization-state estimation, sequence alignment, and error localization, enabling the system to maintain high robustness and manageable computational complexity even under complex, time-varying channel conditions [61–63]. Consequently, this approach is well-suited for challenging scenarios such as high-speed mobile communications and satellite links.

Firstly, researchers have proposed using neural networks to process information related to synchronization state estimation. For instance, convolutional neural networks (CNNs) capture local sequence-distortion patterns to enhance the detection accuracy of indel trigger points; recurrent neural networks (RNNs) or Transformers model long-range dependencies to predict global alignment paths; and fully connected neural networks adaptively estimate state-transition probabilities to support more robust synchronization searches via dynamic programming. Representative implementations typically adopt supervised or self-supervised training strategies. In such settings, the network is optimized using alignment-aware objective functions, including sequence reconstruction loss, differentiable edit-distance loss, and joint synchronization-decoding likelihood criteria. These carefully designed losses explicitly encode synchronization constraints and structural consistency, thereby enabling stable convergence and effective learning under severe indel distortions. The resulting neurally enhanced synchronizer achieves notable performance gains under unknown or non-ideal channel conditions.

Secondly, owing to the inherent coupling between synchronization detection and error correction in indel channels, recent research has focused on constructing joint feature representations through DL to address both tasks simultaneously. Examples include shared encoder structures for joint alignment and decoding, multi-task learning frameworks that capture shared statistical features, and models based on sequence registration or differentiable edit-distance objectives for end-to-end recovery of sequence structural distortions. For example, CNN-based models are often trained to classify local indel events or predict indel likelihood maps, whereas Transformer-based architectures are commonly optimized for end-to-end sequence reconstruction, leveraging global contextual modeling to achieve robust long-sequence alignment and recovery. By learning statistical patterns of indel errors directly from large-scale data, these approaches significantly improve algorithm adaptability and generalization.

To further enhance performance and robustness, researchers have integrated DL with traditional graph-model decoders, establishing a neural-graph coordinated architecture that implements iterative soft-information feedback. In this architecture, synchronization information generated by the deep model bidirectionally refines the graph-model decoding, while confidence estimates from the graph model improve the deep model's inference. This closed-loop design significantly reduces the probabilities of missed detection and misalignment.

In recent years, to enhance model interpretability, convergence stability, and deployability, researchers have introduced structured deep models, differentiable editing models, and neural network architectures incorporating prior constraints. These advancements facilitate closer alignment between DL-based indel error processing and classical communication theory while improving reliability under extreme channel conditions [64,65].

As summarized in Table 2, CNN-based and RNN-based approaches primarily focus on local distortion detection and sequential alignment modeling, whereas Transformer-based models emphasize global dependency learning and end-to-end sequence reconstruction. In contrast, DNNs are often adopted for soft-information estimation or synchronization-state decision. Such architectural diversity reflects the inherently multi-scale and multi-task nature of indel error correction under synchronization uncertainty.

Table 2. Summary of deep learning models for indel error detection and synchronization

Model Type	Core Functionality	Training Objective	Evaluation Metrics	Representative Works
CNN	Local sequence distortion feature extraction driven by local receptive fields; applied to indel position classification and synchronization anomaly detection.	Indel position classification loss.	Detection accuracy, recall, precision, symbol error rate (SER).	Sliding-window indel position detection and local synchronization refinement.
RNN	Temporal dependency modeling for sequential data; joint estimation of alignment paths and synchronization states.	Sequence alignment likelihood maximization or alignment path prediction loss.	BER, frame error rate (FER).	Sequential alignment modeling for joint synchronization and information recovery under channel distortions.
Transformer	Global dependency modeling based on self-attention mechanism; long-sequence alignment and end-to-end sequence reconstruction.	End-to-end sequence reconstruction loss.	BER, ED.	Sequence-to-sequence reconstruction enabling global symbol resequencing under indel errors.
DNN	High-dimensional feature mapping and nonlinear fusion for synchronization state decision or soft information estimation.	Joint synchronization-decoding loss.	Mean squared error (MSE) of soft information estimation, synchronization state classification accuracy.	Soft-information estimation or synchronization-state decision integrated into decoding algorithms.

Note: Table 2 provides a qualitative summary of four representative DL architectures for indel error detection and synchronization, including their core functionalities, training objectives, and evaluation metrics. The representative works listed in the table are intended to illustrate typical implementation paradigms rather than specific cited studies.

3.2. Semantic-communication and deep-learning-driven intelligent correction

Unlike synchronous recovery and iterative error correction methods based on DL, the intelligent error correction approach grounded in semantic communication and DL seeks to transcend traditional bit-level fidelity. It focuses on extracting, transmitting, and recovering the semantic features of information to achieve co-optimization of semantic validity and reliability in scenarios involving high distortion, high error rates, and unstructured data affected by indel errors. This technical framework integrates the feature extraction capabilities of DNNs with the joint source-channel coding principles of semantic communication, constructing an end-to-end semantic-aware error correction system. This enables the system to maintain semantic consistency even under severely disrupted bit-level alignment.

From a performance evaluation perspective, semantic communication-driven intelligent error correction fundamentally differs from traditional indel error correction schemes. Traditional methods primarily rely on bit-level metrics, such as BER and FER, to quantify symbol-level reconstruction fidelity under synchronization assumptions. However, in indel channels with severe synchronization loss, strict bit-level alignment may be unreliable or even infeasible. In such cases, BER and FER no longer accurately reflect the practical utility of the received information, as high symbol-level error rates do not necessarily correspond to failures in information comprehension or task execution.

Due to this limitation, recent research in semantic communication has increasingly adopted task-oriented evaluation metrics, including semantic similarity and task accuracy. Semantic similarity metrics, typically calculated via embedding distance or cosine similarity, quantify the preservation of semantic content, while task accuracy assesses the effectiveness of recovered information in downstream tasks such as classification, recognition, or decision-making. This shift from symbol-level fidelity to task-level utility highlights the fundamental distinction between synchronization-driven independent correction and semantic-aware communication paradigms [61,62,66,67].

The fundamental differences in evaluation criteria between traditional and semantic-aware methods stem from their divergent technical implementation paths. Semantic communication-driven intelligent error correction primarily revolves around the following core ideas:

First, error recovery and masking mechanisms at the semantic level become the focus of research. Traditional indel error correction relies on bit-by-bit synchronization and alignment, whereas semantic communication leverages DL models to reconstruct and infer information in high-dimensional semantic spaces, thereby reducing dependence on strict bit-level alignment. Researchers have designed semantic encoders, such as convolutional networks, transformers, and cross-modal fusion networks, to extract key semantic embeddings from input data, ensuring robustness in semantic representation even when local structures are distorted by indel [68,69]. On the decoder side, semantic decoders employ attention mechanisms and masking prediction strategies to automatically compensate for missing semantics and filter anomalous segments, achieving semantically consistent recovery [70,71]. This approach is particularly suitable for unstructured information such as text, speech, and images, enabling the system to maintain effective semantic understanding or task usability even in the presence of extensive bit-level misalignment or catastrophic symbol-level loss [68,72].

Second, DL-based joint source-channel coding (JSCC) enhances system robustness by constructing semantic representations inherently tolerant to indel distortions. In this framework, the encoder maps inputs to continuous and redundant semantic vector sequences, where the key feature lies in the intrinsic resilience of continuous vector spaces to perturbations caused by symbol insertions or deletions. Local distortions do not lead to catastrophic failure of semantic representations [72,73]. Through end-to-end training, the network holistically learns semantic compression, channel coding, and implicit synchronization recovery [74,75]. Upon receiving contaminated vector sequences, the decoder adaptively extracts valid semantics and suppresses misaligned interference using mechanisms such as attention, directly reconstructing the original data. This data-driven adaptive coding strategy achieves an optimal balance between semantic fidelity and channel reliability [71–73,76].

Furthermore, in cases where severe indel errors render partial information irrecoverable, generative models offer new technical pathways for semantic restoration [72,77]. Researchers have introduced generative priors such as variational autoencoders (VAEs), generative adversarial networks (GANs), diffusion models, and generative transformers to learn the latent distributions of data [77,78]. During decoding, the system treats corrupted sequences as noisy semantic observations and leverages generative models to produce reconstructions semantically consistent with the original data from the latent space, achieving inferential completion at the semantic level [76,77].

Finally, despite its high potential, this approach still faces several challenges, including high demands for training data and computational resources, performance bottlenecks in transferring semantic models across different data domains, and compatibility issues with existing bitstream-based communication architectures [68,78].

3.3. Quantum-coding-based indel error correction

Beyond classical and semantic communication paradigms, indel error correction has also been explored in emerging quantum communication and storage systems. With advancements in quantum communication and quantum storage technologies, the correction of indel errors in the quantum domain has emerged as a new research direction [79]. Researchers have proposed multiple-insertion error correction algorithms based on quantum Reed-Solomon codes , extending classical error-correction theory to the quantum domain to enhance the reliability of quantum storage and transmission [27,80].

Furthermore, the low-complexity decoding algorithms of polar codes have been applied to the error-correction stage in satellite-based quantum key distribution (QKD) [81]. By aligning with the low-latency requirements of satellite communication scenarios, such integration enhances the practicality of quantum keys [82]. The primary advantage of these methods lies in their adaptation to the distinctive channel characteristics of quantum systems, making them suitable for emerging applications such as quantum storage and quantum communication.

Nevertheless, this field remains in an early stage of development and faces significant challenges in physical implementation. Moreover, coding constructions have yet to achieve fundamental breakthroughs and still largely rely on extensions of classical coding frameworks.

4. Future research directions

This paper systematically reviews technological advancements in indel error detection and correction, tracing the evolution from traditional synchronization recovery methods to intelligent DL-driven approaches. Our analysis encompasses multiple dimensions, including algorithmic design, mathematical optimization, DL techniques, and semantic communication frameworks. Although current research has achieved significant progress, it remains predominantly focused on low- to medium-complexity scenarios. Existing methods continue to face substantial challenges in model interpretability, adaptability to complex channel conditions, and cross-scenario generalization. To address these gaps, future research should prioritize the following directions for further investigation.

Figure 7 illustrates representative application scenarios and physical-layer impairments commonly encountered in complex communication environments, which give rise to severe insertion/deletion errors, including high mobility, multipath propagation, symbol desynchronization, and emerging nontraditional communication paradigms.

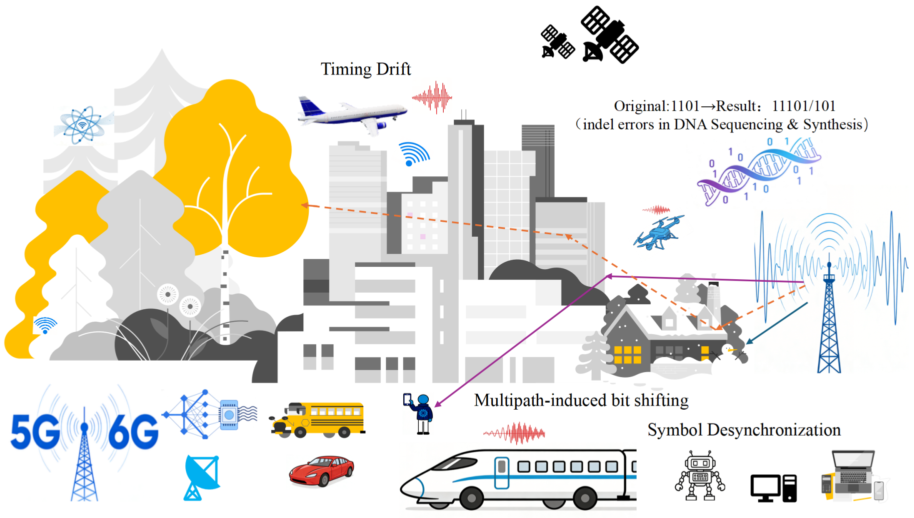

Figure 7. Origins of indel errors in various communication environments.

(1) Synchronization Recovery Mechanisms for Highly Dynamic and Complex Channels

The continuous advancement of highly dynamic channels, wideband multicarrier systems, and integrated space-air-ground networks has resulted in increasingly complex structural characteristics of indel errors. Consequently, it is imperative to develop a unified synchronization recovery framework capable of operating effectively in complex environments, such as high-speed mobility, large-scale multipath propagation, and nonstationary channels. This framework must maintain stable and reliable performance under compound synchronization disturbances, including Doppler shifts, symbol timing offsets, long-delay drifts, and clock inconsistencies.

(2) Lightweight DL Models and Enhanced Cross-Scenario Generalization

Current DL-based detection and correction methods remain constrained by model size, computational complexity, and limited generalization across diverse channel conditions. Future research should prioritize lightweight architectural designs and parameter-sharing strategies to enable efficient deployment on resource-constrained devices. Furthermore, incorporating techniques such as transfer learning and few-shot learning could enhance model adaptability and robustness across varying channel environments, modulation schemes, and system configurations.

(3) Deep Integration of Mathematical Optimization and Learning Algorithms

While methods such as DP, LP, and SDP provide strong theoretical interpretability, they entail high computational costs for long sequences or large-scale problems. Future research should explore the integration of mathematical optimization frameworks with DL, including learned solvers, end-to-end optimization strategies, and adaptive relaxation techniques. By learning to adjust constraint weights, relaxation factors, and solution trajectories, these hybrid approaches can accelerate inference while preserving interpretability, ultimately enhancing performance toward the MD decoding bound.

(4) Semantic Communication-Driven Indel Error Detection and Correction

In semantic communication systems, task-oriented error correction strategies are emerging as an important research direction. Future studies could leverage semantic representations, knowledge structures, and the reasoning capabilities of large-scale models to achieve inference-based, semantic-level recovery of indel errors through semantic redundancy. Additionally, systematic investigations are required to quantify the theoretical limits of semantic reliability and semantic fidelity, thereby establishing a theoretical framework for semantic-layer error correction.

(5) Unified End-to-End Evaluation System and Theoretical Bound Analysis

Currently, no unified end-to-end performance evaluation framework accounts for indel errors across multiple layers, including synchronization, encoding, decoding, and semantics. Future research should focus on developing multidimensional reliability metrics that encompass the bit, sequence, synchronization, and semantic levels. Additionally, in-depth investigations into performance lower bounds, complexity bounds, and Shannon capacity bounds are necessary to establish a robust theoretical foundation for algorithm design and system optimization.

(6) Customized Solution Design for Emerging Application Scenarios

In emerging applications such as DNA storage, quantum communication, random-access interfaces, and near-field high-speed wearable communications, indel errors exhibit statistical properties that differ from those in conventional communication systems. Future research should focus on developing customized coding and fault-tolerance schemes tailored for non-binary errors, unknown codeword boundaries, and cross-modal data characteristics. Additionally, investigating multi-domain fusion correction mechanisms will be essential to meet the heightened demands for robustness and efficiency in these novel scenarios.

In summary, future research should focus on advancing theoretical modeling, improving model generalization, enhancing algorithm interpretability, and refining semantic-layer correction. The objective is to establish a unified, theoretically grounded, and practically deployable framework for indel error detection and correction across heterogeneous communication scenarios.

5. Conclusion

As a quintessential form of synchronization error in communication systems, indel errors are characterized by nonlocality and sequence-level structural disruption. These properties significantly impact synchronization links, making indel errors a critical technical challenge for 5G/6G networks, high-speed mobile communications, satellite links, the IoT, and DNA storage systems.

Current research can be broadly classified into two principal paradigms: traditional synchronization recovery methods and intelligent DL-driven approaches. The former encompasses techniques such as synchronization markers, ED coding, sequence alignment, trellis/convolutional structures, and probabilistic models. These methods benefit from strong interpretability and low implementation complexity, making them suitable for low-error-rate and static channel environments. However, their efficacy is often limited by their reliance on a priori parameters and rigid model assumptions, which restrict their applicability in high-error-rate contexts and long-sequence inference tasks.

In contrast, the intelligent DL-driven paradigm employs data-driven mechanisms and powerful nonlinear modeling to significantly enhance error correction performance. Within this framework, DL-based synchronization recovery integrates neural enhancements into conventional systems, while semantic communication achieves superior robustness in unstructured data and extreme-error scenarios by overcoming bit-level alignment constraints. Although these methods perform effectively in complex, time-varying channels and high-error-rate conditions, they face challenges related to high computational costs, strong data dependency, and limited interpretability.

Due to inherent differences in existing literature regarding channel models, data block lengths, synchronization assumptions, and experimental settings, direct quantitative comparisons of BER or packet error rate (PER) often fail to ensure methodological comparability and may yield misleading conclusions. Consequently, Table 3 establishes a unified qualitative performance comparison framework to facilitate a high-level, structured, and comprehensive analysis of representative insertion and deletion error detection and correction paradigms.

Table 3. Qualitative comparison of representative indel error detection and correction approaches.

Category	Performance Tendency	Computational Complexity	Key References
Synchronization-marker-based methods	Limited-Moderate	Low-Moderate	[17, 83–85]
ED-based coding schemes	Moderate	High	[18–22]
Sequence-alignment-based decoding	Not explicitly quantified	High	[30,31]
Trellis/convolutional structure-based methods	Good	Moderate-High	[36,43,44]
Model-based probabilistic methods	Moderate-Good (strongly model-dependent)	High	[53,85]
DL-based synchronization recovery and iterative correction	Conditional Good	High (training)	[57–61,63,64,75–77,86]
Semantic-communication and DL-driven correction	Task-dependent	High	[65–66,72,73,76–78]
Quantum-coding-based indel correction	Preliminary	High	[79–82,87]

Note: Table 3 provides a qualitative comparison of performance tendency and computational complexity across Indel-correction categories, rather than a unified BER/PER-based quantitative benchmark.

This framework does not rely on heterogeneous numerical benchmarks but instead emphasizes fundamental trade-offs between reliability and efficiency across different design paradigms through cross-category analysis. Specifically, performance trends qualitatively reflect the relative error correction capabilities and robustness reported by representative studies under their respective operating assumptions, while computational complexity summarizes typical algorithm or implementation costs, such as decoding complexity, model size, or training overhead. This approach provides a consistent and interpretable reference for scheme selection under varying communication constraints.

Looking ahead, significant progress is anticipated in several key areas, including the development of unified frameworks for highly dynamic and complex channels, the design of lightweight DL architectures, the enhancement of theoretical interpretability, the exploration of theoretical boundaries for semantic correction, and the formulation of customized solutions for emerging application scenarios. These advancements are expected to provide a robust foundation for the high-quality evolution of next-generation communication networks.

Data availability statement

The data or datasets that support the findings of this study are available from the corresponding author upon reasonable request

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools only to improve language and readability. The authors take full responsibility for the content of the manuscript.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62472169 and Grant 62101205; in part by the Hunan Provincial Department of Public Education under Grant 23C0217 and Grant 22B0676; and in part by the Natural Science Foundation of Hunan Province under Grant 2023JJ50045, Grant 2023JJ50046, Grant 2024JJ7218, and Grant 2024JJ7219.

Authors' contribution

Peng Zhu: conceptualization, methodology, investigation, validation, writing—original and preparation. Hui Yang: methodology, software, writing—original draft preparation and editing. Chao Yu: validation, writing—review, supervision and funding acquisition. Wenwu Xie: formal analysis, data curation, visualization and supervision. Ji Wang: resources and project administration. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Ghogho M, McLernon D, Alameda-Hernandez E, Swami A. Channel estimation and symbol detection for block transmission using data-dependent superimposed training. *IEEE Signal Process Lett.* 2005, 12(3):226–229.

[2] Schmidl TM, Cox DC. Robust frequency and timing synchronization for OFDM. *IEEE Trans. Commun.* 1997, 45(12):1613–1621.

[3] Allam A, Kalnis P, Solovyev V. Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data. *Bioinformatics* 2015, 31(21):3421–3428.

[4] Ratzer EA. Marker codes for channels with insertions and deletions. *Ann. Télécommun.* 2005, 60(1):29–44.

[5] Mitzenmacher M. A survey of results for deletion channels and related synchronization channels. *Probab. Surv.* 2009, 6:1–33.

[6] Lu L, Li G, Swindlehurst AL, Ashikhmin A, Zhang R. An overview of massive MIMO: benefits and challenges. *IEEE J. Sel. Top. Signal Process.* 2014, 8(5):742–758.

[7] Albrecht TR, Arora H, Ayanoor-Vitikkate V, Beaujour JM, Bedau D, *et al*. Bit-patterned magnetic recording: theory, media fabrication, and recording performance. *IEEE Trans. Magn.* 2015, 51(5):1–42.

[8] Yazdi SMHT, Kiah HM, Garcia-Ruiz E, Ma J, Zhao H, *et al*. DNA-based storage: trends and methods. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* 2015, 1(3):230–248.

[9] Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* 2019, 20(8):456–466.

[10] Chee YM, Kiah HM, Vardy A, Vu VK, Yaakobi E. Coding for racetrack memories. *IEEE Trans. Inf. Theory* 2018, 64(11):7094–7112.

[11] Yi C, Zhou J, Li Y, An Z, Li Y, *et al*. Correcting non-binary burst deletions/insertions with De Bruijn symbol-maximum distance separable codes. *IEEE Commun. Lett.* 2023, 27(8):1939–1943.

[12] Yi C, Zou L, Li Y, Qiu Z, Hu Y, *et al*. Error-correcting codes with large field size under non-binary segmented burst deletion/insertion channels and unknown codeword boundaries. *IEEE Trans. Commun.* 2024, 72(10):6001–6013.

[13] Yang BG, Letaief KB, Cheng RS, Cao Z. Timing recovery for OFDM transmission. *IEEE J. Sel. Areas Commun.* 2000, 18(11):2278–2291.

[14] Wang F, Fertonani D, Duman TM. Symbol-level synchronization and LDPC code design for insertion/deletion channels. *IEEE Trans. Commun.* 2011, 59(5):1287–1297.

[15] Smith D, Swart TG, Abdel-Ghaffar KAS, Ferreira HC, Cheng L. Interleaved constrained codes with markers correcting bursts of insertions or deletions. *IEEE Commun. Lett.* 2017, 21(4):702–705.

[16] Munhoz D, de Marca J, Arantes D. On frame synchronization of PCM systems. *IEEE Trans. Commun.* 1980, 28(8):1213–1218.

[17] Levenshtein VI. Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Dokl.* 1966, 10(8):707–710.

[18] Hollmann HDL. A relation between Levenshtein-type distances and insertion-and-deletion correcting capabilities of codes. *IEEE Trans. Inf. Theory* 1993, 39(4):1424–1427.

[19] Levenshtein VI. Bounds for deletion/insertion correcting codes. In *Proceedings IEEE International Symposium on Information Theory(ISIT)*, Lausanne, Switzerland, June 30–July 5, 2002, pp. 370–370.

[20] Helberg ASJ, Clarke WA, Ferreira HC, Vinck AJH. A class of DC free, synchronization error correcting codes. *IEEE Trans. Magn.* 1993, 29(6):4048–4049.

[21] Helberg ASJ, Ferreira HC, Clarke WA. Multiple insertion/deletion correcting codes. In *Proceedings of 1994 IEEE International Symposium on Information Theory(ISIT)*, Trondheim, Norway, June 24–July 1, 1994, pp. 67–67.

[22] Amooshahi M, Taeri B. On Cayley sum graphs of nonn-abelian groups. *Graphs Comb.* 2016, 32:17–29.

[23] Ma X, Feng M, Wang K. Subgroup perfect codes in Cayley sum graphs. *Des. Codes Cryptogr.* 2020, 88:1447–1461.

[24] Dau H, Duursma IM, Kiah HM, Milenkovic O. Repairing Reed-Solomon codes with multiple erasures. *IEEE Trans. Inf. Theory* 2018, 64(10):6567–6582.

[25] Halbawi W, Liu Z, Duursma IM, Dau H, Hassibi B. Sparse and balanced Reed-Solomon and Tamo-Barg codes. *IEEE Trans. Inf. Theory* 2019, 65(1):118–130.

[26] Ding C, Du X, Zhou Z. The Bose and minimum distance of a class of BCH codes. *IEEE Trans. Inf. Theory* 2015, 61(5):2351–2356.

[27] Duc TD, Liu S, Tjuawinata I, Xing C. Explicit constructions of two-dimensional Reed-Solomon codes in high insertion and deletion noise regime. *IEEE Trans. Inf. Theory* 2021, 67(5):2808–2820.

[28] Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 1970, 48(3):443–453.

[29] Rashed AEE-D, Amer HM, El-Seddek M, Moustafa HED. Sequence alignment using machine learning-based Needleman-Wunsch algorithm. *IEEE Access* 2021, 9:109522–109535.

[30] Muhamad FN, Ahmad RB, Asi SM, Murad MN. Performance analysis of Needleman-Wunsch algorithm (global) and Smith-Waterman algorithm (local) in reducing search space and time for DNA sequence alignment. In *Proceedings of 1st International Conference on Green and Sustainable Computing (ICoGeS)*, Kuching, Malaysia, November 25–27, 2017, p. 012085.

[31] Manavski SA, Valle G. CUDA compatible GPU cards as efficient hardware accelerators for Smith-Waterman sequence alignment. *BMC Bioinf.* 2008, 9(Suppl 2):S10.

[32] Rucci E, Garcia C, Botella G, De Giusti A, Naiouf M, *et al.* SWIFOLD: Smith-Waterman implementation on FPGA with OpenCL for long DNA sequences. *BMC Syst. Biol.* 2018, 12(Suppl 5):96.

[33] Gao F, Zhang C. Comparison of various algorithms for recognizing short coding sequences of human genes. *Bioinformatics* 2004, 20(5):673–681.

[34] Schwartz AS, Pachter L. Multiple alignment by sequence annealing. *Bioinformatics* 2007, 23(2):e24–e29.

[35] Benegas G, Albors C, Aw AJ, Ye C, Song YS. A DNA language model based on multispecies alignment predicts the effects of genome-wide variants. *Nat. Biotechnol.* 2025, 43:1960–1965.

[36] Lynch TJ. Sequence time coding for data compression. *Proc. IEEE.* 1966, 54(10):1490–1491.

[37] Bahl L, Jelinek F. Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition. *IEEE Trans. Inf. Theory* 1975, 21(4):404–411.

[38] Elias P. Universal codeword sets and representations of the integers. *IEEE Trans. Inf. Theory* 1975, 21(2):194–203.

[39] Wiberg N, Loeliger HA, Kotter R. Codes and iterative decoding on general graphs. *Eur. Trans. Telecommun.* 1995, 6:513–525.

[40] Sweeney P, Wesemeyer S. Iterative soft-decision decoding of linear block codes. *IEE Proc.-Commun.* 2000, 147(3):133–136.

[41] Kora AD, Saemi A, Cances JP, Meghdadi V. New list sphere decoding (LSD) and iterative synchronization algorithms for MIMO-OFDM detection with LDPC FEC. *IEEE Trans. Veh. Technol.* 2008, 57(6):3510–3524.

[42] Uchoa AGD, Healy CT, de Lamare RC. Iterative detection and decoding algorithms for MIMO systems in block-fading channels using LDPC codes. *IEEE Trans. Veh. Technol.* 2016, 65(4):2735–2741.

[43] Elias P. Coding for noisy channels. *IRE Conv. Rec.* 1955, 3:37–46.

[44] Forney GD. The viterbi algorithm. *Proc. IEEE* 1973, 61(3):268–278.

[45] Bahl L, Cocke J, Jelinek F, Raviv J. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Inf. Theory* 1974, 20(2):284–287.

[46] Hagenauer J, Hoeher P. A Viterbi algorithm with soft-decision outputs and its applications. In *Proceedings of 1989 IEEE Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond'*, Dallas, USA, November 27–30, 1989, pp. 1680–1686.

[47] Le HS, Schulz MH, McCauley BM, Hinman VF, Bar-Joseph Z. Probabilistic error correction for RNA sequencing. *Nucleic Acids Res.* 2013, 41(10):e109.

[48] Aita T, Ichihashi N, Yomo T. Probabilistic model based error correction in a set of various mutant sequences analyzed by next-generation sequencing. *Comput. Biol. Chem.* 2013, 47:221–230.

[49] Golod D, Brown DG. A tutorial of techniques for improving standard hidden Markov model algorithms. *J. Bioinf. Comput. Biol.* 2009, 7(4):737–754.

[50] Guillemot C, Siohan P. Joint source-channel decoding of variable-length codes with soft information: a survey. *EURASIP J. Adv. Signal Process.* 2005, 2005(6):632067.

[51] Altman RM. Assessing the goodness-of-fit of hidden Markov models. *Biometrics* 2004, 60(2):444–450.

[52] Matsumine T, Ochiai H. Recent advances in deep learning for channel coding: a survey. *IEEE J. Commun. Soc.* 2024, 5:6443–6481.

[53] Loeliger HA. An introduction to factor graphs. *IEEE Signal Process Mag.* 2004, 21(1):28–41.

[54] Zhou Z, Feng W, Li P, Liu Z, Xu X, *et al.* A fusion method of pedestrian dead reckoning and pseudo indoor plan based on conditional random field. *Measurement* 2023, 207:112417.

[55] Sutton C, McCallum A. An introduction to conditional random fields. *Found. Trends Mach. Learn.* 2012, 4(4):267–373.

[56] Weyns D, Gheibi O, Quin F, Van Der Donckt J. Deep learning for effective and efficient reduction of large adaptation spaces in self-adaptive systems. *Found. Trends Mach. Learn.* 2022, 17(1–2):1–42.

[57] Ma G, Jiao X, Mu J, Han H, Yang Y. Deep learning-based detection for marker codes over insertion and deletion channels. *IEEE Trans. Commun.* 2024, 72(10):5945–5959.

[58] Chang K, Kim M, Ko YJ, Kim I. Towards precise synchronization under phase distortion: signal design and deep learning. *IEEE Trans. Veh. Technol.* 2025, 74(4):6715–6720.

[59] Islam N, Shin S. Deep Learning in physical layer: review on data driven end-to-end communication systems and their enabling semantic applications. *ACM Trans. Auton. Adapt. Syst.* 2024, 5:4207–4240.

[60] Kitaev AY. Quantum computations: algorithms and error correction. *Russ. Math. Surv.* 1997, 52(6):1191–1249.

[61] Guo S, Wang Y, Zhang N, Su Z, Luan T, *et al.* A survey on semantic communication networks: architecture, security, and privacy. *IEEE Commun. Surv. Tut.* 2025, 27(5):2860–2894.

[62] Kalade S, Crockett LH, Stewart RW. Training deep filters for physical-layer frame synchronization. *ACM Trans. Auton. Adapt. Syst.* 2022, 3:1063–1075.

[63] Wei P, Lu R, Ye G, Xie S, Wang R. Channel synchronization based on deep learning. *Trans. Emerging Telecommun. Technol.* 2023, 34(1):e4656.

[64] Wu Q, Ng BK, Lam CT, Cen X, Liang Y, *et al*. Shared graph neural network for channel decoding. *Appl. Sci.* 2023, 13(23):12657.

[65] Bajpai A. Intelligent LDPC decoding using graph neural networks for beyond 5G communication systems. *J. Inf. Commun. Converg. Eng.* 2025, 23(4):227–235.

[66] Zhang W, Wu S, Meng S, He J, Zhang Q. Engineering a lightweight deep joint source-channel-coding-based semantic communication system. *IEEE Internet Things J.* 2025, 12(1):458–471.

[67] Xu J, Tung T, Ai B, Chen W, Sun Y, *et al*. Deep joint source-channel coding for semantic communications. *IEEE Commun. Mag.* 2023, 61(11):42–48.

[68] Zhang P, Liu Y, Song Y, Zhang J. Advances and challenges in semantic communications: a systematic review. *Natl. Sci. Open* 2024, 3(4):36.

[69] Li A, Wei X, Wu D, Zhou L. Cross-modal semantic communications. *IEEE Wireless Commun.* 2022, 29(6):144–151.

[70] Zheng Z, Zhong YF, Tian S, Ma A, Zhang L. ChangeMask: deep multi-task encoder-transformer-decoder architecture for semantic change detection. *ISPRS J. Photogramm. Remote Sens.* 2022, 183:228–239.

[71] Liu X, Huang Z, Zhang Y, Jia Y, Wen W. CNN and attention-based joint source channel coding for semantic communications in WSNs. *Sensors* 2024, 24(3):957.

[72] Ozates T, Koç A. Semantic communication over channels with insertions, deletions, and substitutions. *IEEE Commun. Lett.* 2025, 29(3):586–590.

[73] Bourtsoulatze E, Kurka DB, Gündüz D. Deep joint source-channel coding for wireless image transmission. *IEEE Trans. Cognit. Commun. Networking* 2019, 5(3):567–579.

[74] Zhao T, Li F, Du H, Sun L. Deep reinforcement learning- and information bottleneck-enabled task-oriented semantic communication. *IEEE J. Sel. Areas Commun.* 2025, 43(12):4198–4214.

[75] Lv S, Li X, Liu J, Shi M, Chen X. End-to-end optimization of semantic communication systems: joint source-channel-host-tasks. *IEEE Trans. Veh. Technol.* 2025, 74(5):7580–7593.

[76] Uras Kargı E, Duman TM. Symbol-level deep learning based decoders for concatenated codes over insertion/deletion channels. *IEEE Trans. Commun.* 2026, 74:4187–4208.

[77] Xia L, Sun Y, Liang C, Zhang L, Imran MA, *et al*. Generative AI for semantic communication: architecture, challenges, and outlook. *IEEE Wireless Commun.* 2025, 32(1):132–140.

[78] Bo Y, Duan Y, Shao S, Tao M. Joint coding-modulation for digital semantic communications via variational autoencoder. *IEEE Trans. Commun.* 2024, 72(9):5626–5640.

[79] Hagimara M. Introduction to quantum deletion error-correcting codes. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* 2025, 108(3):363–375.

[80] Con R, Shpilka A, Tamo I. Optimal two-dimensional Reed-Solomon codes correcting insertions and deletions. *IEEE Trans. Inf. Theory* 2024, 70(7):5012–5016.

[81] Yin Z, Liu W. Key error correction method with low complexity for satellite-to-ground quantum key distribution experiment based on polar codes (In Chinese). *Chin. J. Quantum Electron.* 2025, 42(3):354–360.

[82] Yin J, Li Y, Liao S, Zhang L, Li Y, *et al*. Entanglement-based secure quantum cryptography over 1,120 kilometres. *Nature* 2020, 582(7813):501–505.

[83] Haeupler B, Shahrasbi A. Synchronization strings and codes for insertions and deletions—a survey. *IEEE Trans. Inf. Theory* 2021, 67(6):3190–3206.

[84] Davey MC, Mackay DJC. Reliable communication over channels with insertions, deletions, and substitutions. *IEEE Trans. Inf. Theory* 2001, 47(2):687–698.

[85] Sala F, Gabrys R, Schoeny C, Dolecek L. Exact reconstruction from insertions in synchronization codes. *IEEE Trans. Inf. Theory* 2017, 63(4):2428–2445.

[86] Nachmani E, Marciano E, Lugosch L, Gross WJ, Burshtein D, *et al*. Deep learning methods for improved decoding of linear codes. *IEEE J. Sel. Top. Signal Process.* 2018, 12(1):119–131.

[87] Ouyang Y, Lai C. Linear programming bounds for approximate quantum error correction over arbitrary quantum channels. *IEEE Trans. Inf. Theory* 2022, 68(8):5234–5247.