Real-Time D-PMU Data Compression for Edge Computing Devices in Digital Distribution Networks

Jiahui Yang, *Student Member, IEEE*, Hao Yu, *Senior Member, IEEE*, Peng Li, *Senior Member, IEEE*, Haoran Ji, *Member, IEEE*, Wei Xi, Jianzhong Wu, *Senior Member, IEEE*, and Chengshan Wang, *Senior Member, IEEE*

***Abstract*—The proliferation of distribution-level phasor measurement units (D-PMUs) with a high reporting rate brings a heavy transmission burden to communication systems of distribution networks, which necessitate efficient data compression on edge computing devices. This paper proposes a real-time D-PMU data compression algorithm, including three stages of prediction, quantization, and Bitpack. The current data frame is predicted by the adaptive normalized least mean square predictor based on the stochastic gradient descent algorithm. Then, the prediction errors are quantized as integers and Bitpack is established to extract significant bits and losslessly reduce the redundancy of the quantized data. For edge computing devices accessing multiple D-PMUs in distribution networks, a performance optimization mechanism is proposed. The spatial similarity of D-PMUs is explored to multiplex the predictor and reduce the computation burden. In addition, the compression performances can be adaptively adjusted to diminish the transmission delay in limited bandwidth conditions. Finally, the proposed method is validated and compared with the state-of-the-art methods using the field data and simulated data in normal and fault conditions of distribution networks. Moreover, the impacts and countermeasures of data quality are considered. The results demonstrate that the proposed method achieves accurate and efficient real-time compression in different scenarios.**

***Index Terms*—Data compression, digital distribution networks, D-PMU, edge computing, real-time.**

I. Introduction

THE application of advanced information technologies has significantly accelerated the digitalized transformation of distribution networks [1], [2]. Massive data and extensive links are basic characteristics of digital distribution networks, where edge computing devices are deployed to realize local measurement collection, analysis, and processing [3], [4]. Refined information obtained by edge computing devices can be further uploaded to the cloud for centralized analysis and decision-making, forming a cloud-edge collaborative architecture of distribution networks [5], [6]. In this context, the advanced sensing and measurement technologies, represented by distribution-level phasor measurement units (D-PMUs), become essential for enhancing the observability of digital distribution networks [7].

Typically, D-PMUs can provide global positioning system-synchronized data frames, such as voltage phasors, current phasors, and frequency [8]. In digital distribution networks, D-PMU data are transmitted to the distribution management system (DMS) at 30-240 frames per second (fps) in real-time [9]. A number of D-PMUs and their high reporting rate result in tremendous measurement data. As discussed in [10], 1100 PMUs with a reporting rate of 30 fps via IEEE C37.118 protocol will increase the data volume by over 700 GB per day. Therefore, data compression in the edge side becomes crucial to prevent the transmission of D-PMU data from causing unacceptable delay and blockage to the communication system. Moreover, within the cloud-edge collaborative architecture of digital distribution networks, data from multiple D-PMUs are collected by an edge computing device and then uploaded in a batch [11], further aggravating the pressure on computation and communication resources. Therefore, there is a need for efficient D-PMU data compression algorithms in edge computing devices.

Compressing D-PMU data is still a challenging task. The D-PMU data have been widely used in many applications, such as real-time state estimation (RTSE) [12], [13], topology detection [14], fault location [15], and voltage control [16], [17]. Compared with supervisory control and data acquisition (SCADA) or advanced metering infrastructure (AMI), the most prominent characteristic of D-PMUs is its synchronized high-frequency data acquisition capability. For state estimation, D-PMU provides more accurate data with higher refresh rates to promptly capture the real-time operation state of distribution networks. For fault location, the unique phase angle data collected by D-PMUs can contribute to identifying fault locations more rapidly and accurately. These applications require real-time D-PMU data of at least 50-60 fps to guarantee excellent performance. Hence, the D-PMU data compression algorithm should effectively reduce the data size while avoiding introducing too much time delay, which is referred to as "real-time data compression".

Manuscript received 12 April 2023; revised 13 October 2023; accepted 12 November 2023. Date of publication 22 November 2023; date of current version 20 June 2024. This work was supported by the National Natural Science Foundation of China under Grants U22B20114 and 52277117. Paper no. TPWRS-00524-2023. *(Corresponding author: Hao Yu.)*

Jiahui Yang, Hao Yu, Peng Li, Haoran Ji, and Chengshan Wang are with the Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China (e-mail: yjh@tju.edu.cn; tjuyh@tju.edu.cn; lip@tju.edu.cn; jihaoran@tju.edu.cn; cswang@tju.edu.cn).

Wei Xi is with the Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China, and also with the Digital Grid Research Institute of China Southern Power Grid, Guangzhou 510700, China (e-mail: xiwei@csg.cn).

Jianzhong Wu is with the Institute of Energy, School of Engineering, Cardiff University, CF24 3AA Cardiff, U.K. (e-mail: wuj5@cardiff.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2023.3335282.

Digital Object Identifier 10.1109/TPWRS.2023.3335282

In general, data compression techniques are classified into two categories: lossless compression methods and lossy compression methods [18], [19]. A lossless compression method usually employs data statistics and performs efficient bitwise encoding. Therefore, data can be reconstructed without any loss of information. Common lossless compression algorithms include Huffman coding, Lempel-Ziv coding, and Golomb coding [20], [21]. Some lossless compression methods have been employed for synchrophasor data compression. In [21], frequency-compensated difference encoding is used to reduce the phase angle data entropy. After the preprocessing stage, the data are further compressed by Golomb-Rice encoding. In [22], a delta coding and exclusive or (XOR) coding-based method is proposed for the real-time low-latency data transmission in smart grids. The method has been tested on several datasets, including the D-PMU data from the Illinois Institute of Technology Microgrid. However, the applicability of the method in different scenarios, such as faults, is yet to be explored. In [23], a compression algorithm combining slack-referenced encoding and off-the-shelf compression tools is introduced, which compares the effects of various preprocessing methods and generic compression tools in different scenarios. In [24], an improved time-series compression and its combination with a delta-difference Huffman method are proposed to compress the frequency data and the phase angle data respectively. The algorithms were tested using the field-collected D-PMU data of the FNET/GridEye and Oak Ridge National Laboratory. All in all, lossless compression algorithms focus on scenarios with high precision requirements, but the compression ratio (CR) is usually low.

Lossy compression methods sacrifice accuracy for a better CR. Some techniques have been employed in the lossy compression of D-PMU data, including principal component analysis (PCA), singular value decomposition (SVD), wavelet transform (WT), [19], [25], [26], etc. Ref. [19] proposes an adaptive subband compression method capable of providing real-time streaming of high-resolution continuous point-on-wave and PMU data. In [25], a two-stage algorithm is developed for D-PMU data compression based on PCA and discrete cosine transform. A statistical change detection technology is introduced to adjust the compression parameters and preserve the critical disturbance information. In [26], an improved SVD-based compression algorithm for D-PMU data is proposed. The orthogonal property of two sub-matrices generated by SVD is utilized to enhance the compression performances. However, the algorithm is offline and can not be applied in real-time data compression. In [27], real-time exception and swing door trending (SDT) data compression is proposed for PMUs. The SDT compression requires a certain time interval, which may cause an extra time delay. In [28], a WT-based data compression algorithm is designed, where multiresolution analysis and its properties are studied for data compression and signal denoising. Ref. [29] proposes a synchrophasor data compression method based on singular value decomposition, which was tested with the D-PMU data from École Polytechnique Fédérale de Lausanne (EPFL) campus. In [30], a lossy compression method that incorporates prediction, quantization, and entropy encoding is developed for time series data. However, it necessitates a block-sorting stage and is not suitable for real-time D-PMU data compression. Overall, most off-the-shelf lossy compression algorithms are developed for historical data and are not applicable to the real-time compression of the D-PMU data stream.

In addition, it is necessary for the compression algorithm design to consider the characteristics of edge computing devices and D-PMU data. Under the nearby branch of radial distribution networks, data from multiple D-PMUs accessing an edge computing device exhibit the spatial similarity. For example, the variation law of nodal voltage amplitude from the nearby branch is close. These characteristics may be exploited to enhance the performance of data compression. However, overall computation resources of edge computing devices are limited, which are allocated to various service functions such as monitoring, protection, control, metering, PMU, etc. The available communication bandwidth changes with the number of running services and influences the transmission delay of the compressed D-PMU data. For the applications in the cloud, the compressed D-PMU data from the edge side must preserve valuable information, meaning that the compression algorithm should be error-bounded. These problems bring new challenges to the real-time D-PMU data compression at the edge side in digital distribution networks.

This paper aims to provide a real-time compression method for edge computing devices of distribution networks. The raw D-PMU data frame can be compressed as a whole for all data types in the progressive way of prediction, quantization and Bitpack. Therefore, we refer to the proposed method as PQB. On the receiving side, exploiting the high degree of symmetry of PQB, data reconstruction can be performed by carrying out the inverse transform of the compression process in real-time. The proposed method has been verified using field data and simulated data during normal and fault conditions of distribution networks. All in all, PQB realizes efficient data compression of D-PMUs accessing edge computing devices and meets requirements of multi-scenario applications, which is real-time, error-bounded, high-precision, and lightweight. The main contributions are summarized as follows:

1) A real-time D-PMU data compression method is developed based on progressive prediction, quantization, and Bitpack. Prediction is performed for each data frame, and the prediction error is quantized as integers in real-time. Due to the temporal continuity in D-PMU data, the integer prediction errors tend to concentrate within a narrow value range and contain numerous invalid bits. A Bitpack algorithm is designed to remove these redundant bits, resulting in a significant reduction in data size.
2) A compression performance optimization mechanism is developed for edge computing devices connected with multiple D-PMUs in power systems. D-PMUs connected to the same edge computing device usually have nearby locations and share common features. The predictor can be multiplexed and dynamically adjusted according to the operational state at the edge-side, which leads to a reduction in computational burden and makes the method suitable for edge-side applications.

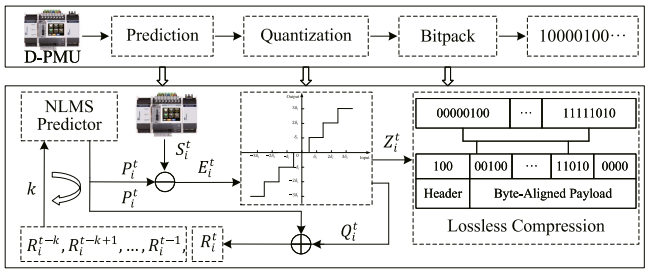

Fig. 1. Diagram of the real-time D-PMU data compression algorithm.

3) The proposed method is validated using both field data from an actual pilot project and simulated data from a standardized IEEE system. Various scenarios are considered, including normal and fault conditions in distribution networks, as well as communication and synchronization failures of D-PMUs. The method is implemented on a Raspberry Pi, serving as a simulation platform for edge computing in power systems. The results effectively demonstrate the feasibility and effectiveness of our method.

The remainder of this paper is organized as follows. Section II presents the real-time D-PMU data compression method. Section III describes the performance optimization mechanism. In Section IV, data reconstruction and evaluation metrics are introduced. In Section V, the method's performances are evaluated for different scenarios using filed data and simulated data. Finally, conclusions are given in Section VI.

II. Real-Time D-PMU Data Compression

The real-time D-PMU data compression algorithm called PQB consists of three steps: prediction, quantization, and Bitpack, as shown in Fig. 1. During the prediction stage, the NLMS predictor employs a linear combination of historical reconstruction data to make predictions of the current data, and the weights are updated based on the stochastic gradient descent algorithm. In the quantization stage, the prediction errors are fed into the uniform quantizer. In the Bitpack stage, the meaningless leading zeros or ones are eliminated, and the data redundancy is reduced further.

A. Prediction

Due to the high reporting rate, D-PMU data exhibit the temporal similarity, which means that D-PMU data change slightly over short periods. Specifically, the phase angle data changes in a linear manner when the frequency is relatively stable [9], as demonstrated in (1). Therefore, D-PMU data can be predicted by the NLMS predictor.

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = 2\pi\left(f - f_0\right) \tag{1}$$

where φ is the phase angle, f is the frequency, and f_0 is the reference frequency.

The NLMS algorithm is originally an adaptive filtering algorithm for echo cancellation [31], [32] and adaptive beamforming [33]. Ref. [30] demonstrates the use of the NLMS adaptive filtering algorithm for predicting time series data. In this paper, the NLMS algorithm is employed to predict the current D-PMU data frame.

The object of prediction is a frame of D-PMU data, including three-phase nodal voltage phasors, three-phase branch current phasors, and frequency. The core of the NLMS predictor is a linear prediction for multiple variables. For each data type, the predictor employs a linear combination of historical data of the same type to make the prediction, as depicted in (2).

$$P_i^t = \sum_{j=0}^{k-1} w_i^t(j) R_i^{t+j-k} \tag{2}$$

where $i \in [0, q-1]$, $i \in \mathbb{Z}$, q represents the number of different types of data, P_i^t is the predicted value of the i-th data at time t, R_i^{t+j-k} represents the reconstruction value of the i-th data at time $t+j-k$, and $w_i^t(j)$ is the j-th weight of the i-th data at time t. The D-PMU data frame at the next time step does not participate in the prediction of the current D-PMU frame, so the prediction process is real-time in addition to the computation delay.

The weight matrix $\boldsymbol{w}$ is adaptively updated in real-time to realize the accurate prediction based on the stochastic gradient descent algorithm. Since only the current sample is chosen to calculate the gradient of the loss function, the algorithm is lightweight. The loss function J_i^t is calculated as follows.

$$J_i^t = \frac{1}{2}\left(R_i^t - P_i^t\right)^2 \tag{3}$$

where the reconstructed value R_i^t is used, instead of the raw data S_i^t in the calculation of the loss function. This is because the raw data are discarded during the compression process, and the reconstruction values are used to ensure the symmetry of the NLMS predictor so that the prediction parameters are not unnecessary to be transmitted. The symmetry means that the NLMS predictor can be completely recovered in the reconstruction process. The update method of the weight matrix $\boldsymbol{w}$ refers to the following equations.

$$w_i^{t+1}(j) = w_i^t(j) - n\frac{\partial J_i^t}{\partial w_i^t(j)} \tag{4}$$

$$w_i^{t+1}(j) = w_i^t(j) - nR_i^{t+j-k}\left(P_i^t - R_i^t\right) \tag{5}$$

$$n = \frac{m}{\alpha + \sum_{j=0}^{k-1}\left(R_i^{t+j-k}\right)^2} \tag{6}$$

where m is the learning rate and set to the experience value 0.5. n is the normalized learning rate and changes with the data trend, and α is a constant to prevent n from being too large due to the small value of $\sum_{j=0}^{k-1}(R_i^{t+j-k})^2$.

B. Quantization

Defining the raw D-PMU data at time t is S_i^t, the prediction error can be calculated according to (7).

$$E_i^t = S_i^t - P_i^t \tag{7}$$

A uniform quantizer is utilized to convert the floating-point prediction error E_i^t to an integer, enabling the application of lossless compression methods for integers. The quantization process is described in (8).

$$Z_i^t = \text{round}\left(\frac{E_i^t}{\delta_i}\right) \tag{8}$$

$$\delta_i = 2\gamma_i \tag{9}$$

where round$(*)$ represents rounding of the data $*$, δ_i is the quantization step size of the i-th data, and Z_i^t is the quantized value of E_i^t. Equations (8) and (9) imply that the quantization process will introduce an absolute error, bounded by the maximum value γ_i between the prediction error E_i^t and its reconstruction. In this paper, γ_i is the maximum absolute error requirement of the i-th data. As the further compression for Z_i^t is lossless, the reconstruction value of E_i^t and S_i^t can be calculated by (10) and (11), respectively.

$$Q_i^t = Z_i^t \delta_i \tag{10}$$

$$R_i^t = P_i^t + Q_i^t \tag{11}$$

where Q_i^t and R_i^t are the reconstruction values of E_i^t and S_i^t, respectively.

Due to the accurate prediction of the current data frame, the prediction errors only occupy a small number of bits. Therefore, an 8-bit quantizer is employed with a quantization range of $[-127, 127]$. If Z_i^t exceeds the range, Z_i^t is set to 128, and E_i^t is saved in the compressed data.

C. Bitpack

The NLMS predictor is capable of accurately predicting D-PMU data in the normal condition of distribution networks, resulting in a small numerical value for the quantized value Z_i^t. When using an 8-bit quantizer, Z_i^t may contain meaningless leading zeros or ones. To reduce redundancy and improve CR, the lossless compression methods for integers can be employed to encode Z_i^t.

Fast methods of compressing integers are often based on bit packing [34]. The underlying concept is to use at most b bits to represent integers in the range of $[0, 2^b - 1]$ and store these bits contiguously. In this paper, an improved bit packing algorithm called Bitpack is proposed, which is specifically designed for the quantized D-PMU data.

The Bitpack process is shown as Algorithm 1. By performing a bitwise AND operation and comparing the result with the symbolic variable β, Bitpack can extract the maximum number of significant bits of a quantized D-PMU data frame, using at most b bits to represent each value. The data consist of two parts: header and payload. The header is used to store the maximum number of significant bits b and the payload is to store the lower b bits of each value. Moreover, one bit as the sign bit is required for each value. To satisfy the requirement for byte alignment, if the data size converted into bytes is not an integer, 0 will be used for padding. Fig. 2 illustrates the Bitpack process. Before Bitpack, the 8-bit quantized values are stored in contiguous memory, as illustrated in the upper blocks. In each data block, the bits on the right of the dotted line are the significant bits of Z_i^t, while the bits to the left of the dotted line are meaningless leading zeros or ones. The bottom blocks represent the data after Bitpack. In this example, the maximum number of significant bits b is 4, and the header is 100. The payload includes the sign bit, as well as the lower b bits of each Z_i^t. The total data size is 68 bits, with additional padding of 0000 to ensure byte alignment.

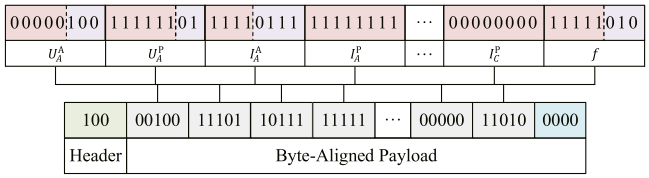

Fig. 2. Diagram of the Bitpack process.

Algorithm 1: Bitpack.

Input: A quantized D-PMU data frame
Output: A D-PMU data frame after Bitpack
Calculate the maximum significant bits b**:**
for i = 0 to q-1 **do**
  if $Z_i^t \geq 0$ **then**
    Set β = 1;
  else
    Set β = 0;
  end
  for each bit Z_g in Z_i^t **do**
    if Z_g & 1 == β **then**
      The number of significant bits for Z_i^t is set to $g + 1$;
    end
  end
end
b is the maximum number of significant bits for all Z_i^t.
Bitpack:
Write b as the header of a message;
for i = 0 to q-1 **do**
  Write the sign bit and the lower b bits of Z_i^t;
end
Execute byte alignment.

III. Performance Optimization Mechanism for Edge Computing

With the advancement of digital distribution networks, cloud-edge collaboration is becoming a prevalent computing paradigm [35], as illustrated in Fig. 3. The radial nature of distribution networks allows for their division into various areas based on branches. An edge computing device can be deployed in each area to perform functions such as local collection of measurement data from multiple D-PMUs, local analysis, and

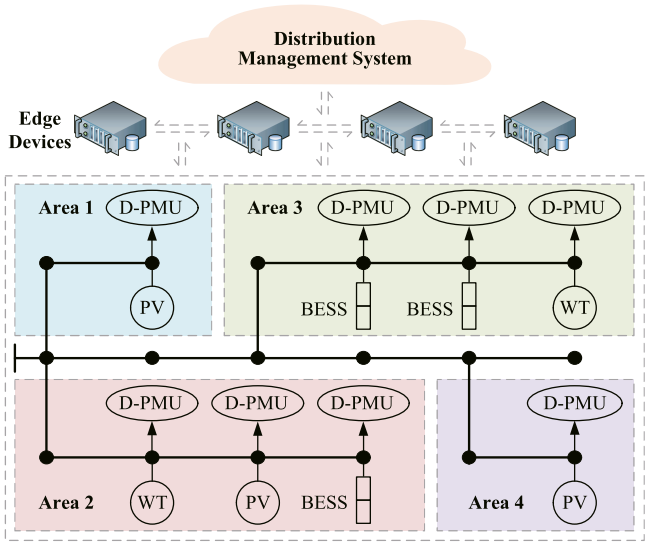

Fig. 3. Cloud-edge collaborative architecture of distribution networks.

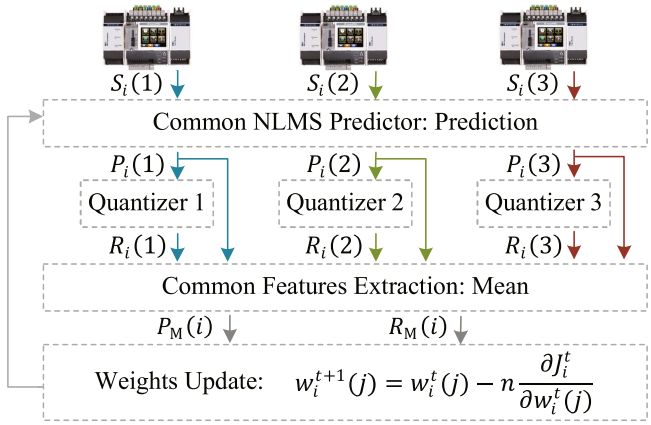

Fig. 4. Diagram of the NLMS predictor multiplexing.

local control of wind turbines (WT), photovoltaic generators (PV), battery energy storage stations (BESS) [36], [37], etc. The distribution management system, serving as the cloud, aggregates data from edge computing devices in various areas and performs centralized analysis and operation optimization. The communication system, consisting of fifth-generation technology, optical fiber communication, etc., functions as the pipeline and connects the power grid, the edge computing layer, and the cloud.

However, the computation resources of edge computing devices are limited. Moreover, with more D-PMUs accessing edge computing devices of distribution networks, every D-PMU will be allocated fewer computation resources, which may cause longer runtime. In addition, the communication resources also vary with the start and stop of different services within the edge computing device. A performance optimization mechanism is proposed to reduce the computation burden and transmission delay in limited bandwidth conditions, including the following two parts. In this paper, only common radial distribution networks topologies are considered. For different network structures, such as a ring topology, the mechanism may require specific settings according to the equivalent electric distances of D-PMUs.

A. Predictor Multiplexing on Edge Computing Devices

Section II applies the temporal similarity to address data prediction of a single D-PMU, thereby establishing the basis for subsequent compression. In Fig. 3, certain areas, such as Area 2 and Area 3, include multiple D-PMUs. For D-PMUs situated within the nearby branch of radial distribution networks, D-PMU data exhibit the spatial similarity, which suggests that the voltage amplitude or frequency data have similar changes in a short period. Therefore, the predictor multiplexing strategy is proposed to reduce the computation burden of the real-time D-PMU data compression on an edge computing device.

The NLMS predictor multiplexing includes two parts: data prediction and weights update. In the process of data prediction, the current data are predicted based on historical data from its own D-PMU and a common NLMS predictor. In addition, at the weights update stage, to enable the predictor to represent the common characteristics, the average of data for the same data type at the same moment from different D-PMUs are applied to update weights through the stochastic gradient descent algorithm. Fig. 4 describes the whole process using 3 D-PMUs as an example. The average of the data can be calculated by (12) and (13). The process for updating weights is similar to a single D-PMU.

$$P_{\mathrm{M}}^t(i) = \frac{1}{N}\sum_{m=0}^{N-1} P_i^t(m) \tag{12}$$

$$R_{\mathrm{M}}^t(i) = \frac{1}{N}\sum_{m=0}^{N-1} R_i^t(m) \tag{13}$$

where $P_{\mathrm{M}}^t(i)$ and $R_{\mathrm{M}}^t(i)$ are the average of the predicted data and the reconstructed data, respectively. N is the number of D-PMUs. The predictor multiplexing means that the D-PMU data from multiple D-PMUs located in the nearby branch is predicted by the same predictor for one data type. Thus, the computation resource usage of edge computing devices at prediction stage is reduced to about $1/N$ of that before the predictor multiplexing.

B. Adaptive Adjustment of the Compression Performances Considering Bandwidth

The communication resources can be measured by the bandwidth, which has a direct impact on the transmission delay of the compressed D-PMU data. An adaptive adjustment mechanism is proposed that takes into account the communication bandwidth to prevent unacceptable delay in limited bandwidth conditions. The mechanism works by dynamically adjusting the quantization step size matrix to reduce the size of the compressed data.

Under the assumption that the available bandwidth is known [38], [39], the maximum available bandwidth is set to b_0, and the current available width is b_{c}. The bandwidth ratio

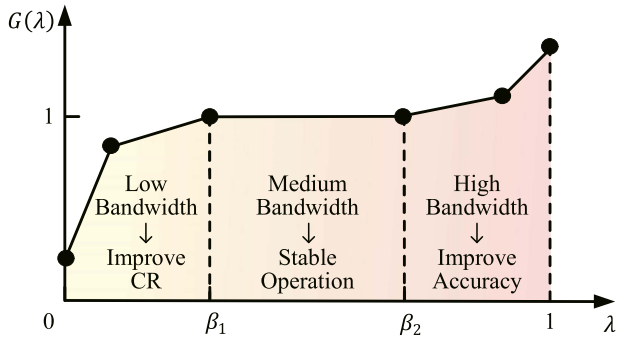

Fig. 5. Diagram of the dynamic amplitude modulation function.

λ is defined as the ratio of b_c to b_0, as derived in (14). The quantization step size matrix $\boldsymbol{\delta}$ changes according to the dynamic amplitude modulation function $G(\lambda)$, as presented in (15). $G(\lambda)$ is illustrated in Fig. 5, which is composed of 5 line segments and the slope demonstrates the sensitivity of the compressed data size to the available bandwidth. The quantization step sizes are set to $\boldsymbol{\delta_s}$ when λ is in $[\beta_1, \beta_2]$.

$$\lambda = \frac{b_c}{b_0} \tag{14}$$

$$\boldsymbol{\delta} = \frac{\boldsymbol{\delta_s}}{G(\lambda)} \tag{15}$$

IV. Data Reconstruction and Evaluation Metrics

A. Data Reconstruction

On the receiving side, to analyse the dynamic characteristics of distribution networks based on D-PMU data, it is necessary to reconstruct the compressed data to obtain the full data. In view of the symmetry of the real-time D-PMU data compression algorithm, the reconstruction is equivalent to the inverse transformation of the compression process.

Firstly, unpack the bitpacked data. Read three bits from the byte streams as the header, and the maximum number of significant bits b is obtained. Continue to read $b+1$ bits as the payload of each value, which is compromised of the sign bit and the lower b bits. Hence, Z_i^t is obtained. Next, the inverse transform of uniform quantization is performed by (10). Additionally, the historical reconstructed data, not the raw data, is applied to predict the current data. Finally, the current reconstructed data are gained with reference to (11).

B. Evaluation Metrics

To evaluate the compression performances of PQB, this paper introduces two evaluation metrics: compression ratio and reconstruction accuracy. The CR is defined as the ratio of the raw data size s_r to the compressed data size s_c for a period, as expressed in (16).

$$\xi_{CR} = \frac{s_r}{s_c} \tag{16}$$

The reconstruction accuracy consists of three aspects: maximum absolute error (MAX), mean absolute error (MAE) as defined in (17) and (18). In theory, the maximum absolute error is less than or equal to half of the quantization step size. The mean absolute error is used to assess the overall compression accuracy.

$$\xi_{MAX} = \max_{t_0 \le t \le t_1} \left| S_i^t - R_i^t \right| \tag{17}$$

$$\xi_{MAE} = \frac{T}{t_1 - t_0} \sum_{t=t_0}^{t_1} \left| S_i^t - R_i^t \right| \tag{18}$$

where T is the reporting interval, t_0 is the initial moment, and t_1 is the end moment. In this paper, MAX and MAE are calculated among data of the same type in all three phases.

V. Case Studies

In this section, the performances of the proposed method PQB are verified using both field data and simulated data under normal and fault conditions of distribution networks. The compression performances of PQB are compared with the state-of-the-art methods, including the improved SVD-based data compression (ISVD) [26], the discrete wavelet transform-based data compression (DWT) [28] and the improved exception and swing door trending data compression (IESDC) [40]. In addition, the scenarios involving edge computing devices connecting multiple D-PMUs and considering bandwidth are designed to evaluate the performance optimization mechanism. Finally, we discuss the impacts of data quality on compression performances and give our countermeasures.

IESDC reduces data volume by discarding data points with small variations [27]. Its application requires the accumulation of data points for a certain period, which introduces a delay of T_{max}. In [40], the Exponential-Golomb coding is employed to further reduce data redundancy losslessly and improve the compression ratio (CR). IESDC is also error-bounded and the maximum compression interval T_{max} is set to 200 ms, which is consistent with that in [27].

ISVD transforms the raw D-PMU data to two orthogonal matrices and one diagonal matrix, which contain the feature information of the D-PMU data. By discarding some of the smaller singular values, it is possible to reduce the data size while maintaining reconstruction accuracy. Ref. [26] introduces an additional lossless compression stage to further enhance the compression ratio, utilizing the orthogonal property of the two sub-matrices generated by ISVD.

DWT orthogonally decomposes the raw D-PMU data into scaling coefficients and wavelet coefficients, which contain the approximation and details of D-PMU data, respectively. Consequently, by discarding some of smaller wavelet coefficients, the overall data size can be reduced. In this paper, the Order 2 Daubechies wavelet and scale 5 are used as the wavelet function and the decomposition level, which is consistent with that in [28].

The field data were collected from a pilot distribution network in Guangzhou, China, where edge computing devices have been installed to enable local measurement and control of PV inverters. Fig. 6 shows the topology of the 53-node distribution networks with 14 PVs and 18 PMUs. The voltage level is 10 kV with a frequency of 50 Hz. The reporting rate of D-PMUs is

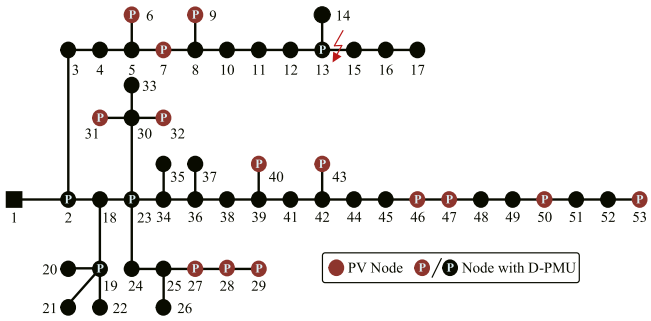

Fig. 6. Topology of actual distribution networks in Guangzhou.

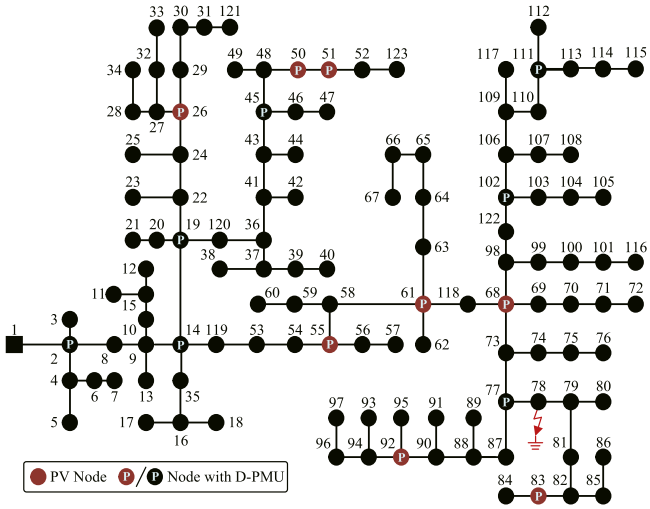

Fig. 7. Topology of the modified IEEE 123-node test case.

50 fps. The data include three-phase nodal voltage phasors, three-phase branch current phasors and frequency. These data are in floating-point format, with each data occupying 4 bytes. Two scenarios are considered: i) Data from the D-PMUs during normal operation spanning 800 s, and ii) data from the D-PMU at node 13, spanning 6 s, derived with a 0.3-second phase-to-phase short circuit experiment.

The simulated D-PMU data were generated with the modified IEEE 123-node test case using PSCAD. Fig. 7 illustrates the topology of the modified IEEE 123-node test case, which operates at a nominal voltage of 4.16 kV and a frequency of 60 Hz. The case includes 118 lines, 85 loads and 8 PVs. 15 D-PMUs with a reporting rate of 60 fps are placed on the buses with PVs and important junction nodes. The total loads are 3490+j1925 kVA. The simulation lasts for 120 s, during which two events are imposed: i) The active power output of PVs fluctuates and the total maximum active power of all PVs is about 1020 kW, and ii) a single-phase grounding fault at node 78 occurs at $t = 60$ s and lasts for 0.3 s. To mimic real-world D-PMU data, 92 dB additive white Gaussian noises are added to the simulated data [41]. The data format involved in this test case is the same as the field data.

The control parameters include the order k of the NLMS predictor, the learning rate m, the control constant α, the initial values of the weight matrix $\boldsymbol{w}$, and the quantization step size matrix $\boldsymbol{\delta}$. These parameters determine the compression performances, which are set based on the requirements of each scenario. In particular, in the first k moments, there is not enough historical data, so the reconstructed data from the last moment is used as the predicted values for the current moment, and the initial predicted values are set to 0.

TABLE I
QUANTIZATION STEP SIZES IN THE NORMAL CONDITION

Data Types	U_Φ^{A}	U_Φ^{P}	I_Φ^{A}	I_Φ^{P}	f
Quantization Step Sizes	$0.02\%U_0$	$0.02°$	$0.1\%I_0$	$0.02°$	0.0002Hz

In this paper, the Raspberry Pi 4 is chosen as the edge computing device to test the performance of PQB. The CPU is ARM Cortex-A72 @1.5 GHz, and the memory is 4 GB. All tests are conducted in the C++ programming language. The D-PMU data are sent to the compressor frame by frame to simulate the real-time generation.

A. Field Data

1) Test in Normal Condition: Similar to [27], fixed control parameters are adopted to maintain controllable accuracy throughout the entire compression process. In this case, a dataset of 40000 data frames, spanning 800 seconds, from the D-PMU located at node 23 during the normal operation of the distribution network, is utilized. The voltage level is 10 kV therefore the phase voltage amplitude reference U_0 is set to 5.774 kV. The phase current amplitude reference I_0 is set to 1000 A, which is in accordance with the rated capacity and practical operation conditions. To ensure stability and rapid response to data changes by the predictor, the order k is set to 8, and the data from the previous 160 ms are utilized for prediction. The initial value of each element for the weight matrix $\boldsymbol{w}$ is set to 1/8, and the initial predicted values are calculated as the average of historical data. The learning rate m adopts the empirical value of 0.5, and α is set to 1.

Section II confirms that the quantization step size is theoretically twice the maximum absolute error. The quantization step sizes for different data types should be relatively uniform, ensuring that the number of significant bits in Bitpack is similar. Considering the more pronounced fluctuations in the current amplitude, its accuracy level is lower than that of voltage amplitude. The phase angle difference at both ends of the line is small, and the frequency is stable in normal operation of the distribution network. Therefore, the quantization step size matrix is set in Table I, where Φ represents the phase, and A and P are abbreviations for amplitude and phase angle.

a) Comparison of reconstruction accuracy: As shown in Table II, the maximum absolute error (MAX) and mean absolute error (MAE) are used to quantitatively analyze the reconstruction accuracy of PQB, ISVD, DWT, and IESDC. It is clear that the MAX values of PQB are less than or equal to half of the quantization step sizes in Table I. As the prediction errors are approximately evenly distributed, the MAE values of PQB are about a quarter of the quantization step sizes. In comparison, the MAX and MAE values of ISVD, DWT, and IESDC are

TABLE II
COMPRESSION PERFORMANCES IN THE NORMAL CONDITION

Metrics	Data Types	PQB	ISVD	DWT	IESDC
MAX	U_{Φ}^{A}(V)	0.580	4.442	2.522	1.045
	U_{Φ}^{P}(deg)	0.010	0.049	0.064	0.036
	I_{Φ}^{A}(A)	0.500	5.472	6.545	1.228
	I_{Φ}^{P}(deg)	0.010	0.175	0.105	0.038
	f(Hz)	1e-4	0.002	0.002	0.001
MAE	U_{Φ}^{A}(V)	0.289	0.352	0.462	0.292
	U_{Φ}^{P}(deg)	0.005	0.009	0.009	0.005
	I_{Φ}^{A}(A)	0.250	0.717	0.815	0.267
	I_{Φ}^{P}(deg)	0.005	0.017	0.015	0.005
	f(Hz)	5e-5	3e-4	4e-4	1e-4
CR	-	7.249	2.274	2.611	5.211

close to those of PQB for specific data types, such as the phase angle data, but are obviously higher for other types of data.

Fig. 8 illustrates the mean absolute errors for individual data of the same type in three phases between the raw data and the reconstructed data. It is clear that the errors for these four methods are mostly quite small. For PQB, the errors are relatively stable and mostly lower than those of other methods. According to the operation logic of IESDC, some raw data may be preserved, resulting in errors close to 0 for a small number of data points. For ISVD and DWT, both methods are not error-bounded, and their accuracy is determined by the number of retained singular values or wavelet coefficients. Consequently, their error profiles exhibit occasional spikes, indicating the potential for large errors at certain moments.

b) Comparison of compression ratio: The CR measures the relative reduction in the size of data representation. Table II shows the compression ratios of the four methods at their respective accuracy levels. PQB achieves a CR of about 7, which is the best among the four methods and means that at least 85% of communication resources can be saved. The reduction in data size comes from two aspects. The 8-bit quantizer reduces the data volume to one-fourth of the original size. Bitpack is exploited to encode the compressed data efficiently, realizing a CR of about 1.8 while being lossless. IESDC achieves a CR of around 5, which is slightly lower than PQB. The reason is that PQB compresses a D-PMU data frame as a whole, while IESDC compresses data separately for different data types, and the size of indexes for all data types occupies a certain proportion. Furthermore, when voltage or power data experience significant fluctuations, the fixed parameters in the compression process prevent IESDC from achieving accurate fitting, resulting in the preservation of most of the data. However, IESDC is efficient if data exhibit a linear trend such as the phase angle data. For ISVD and DWT, the two methods transform data into another space, and data are compressed from a global perspective. To achieve the required high accuracy, it is necessary to preserve more singular values or wavelet coefficients; therefore, the CRs are relatively lower than other methods.

c) Computation efficiency: Time delay is a significant metric to evaluate a real-time compression method. The delay of

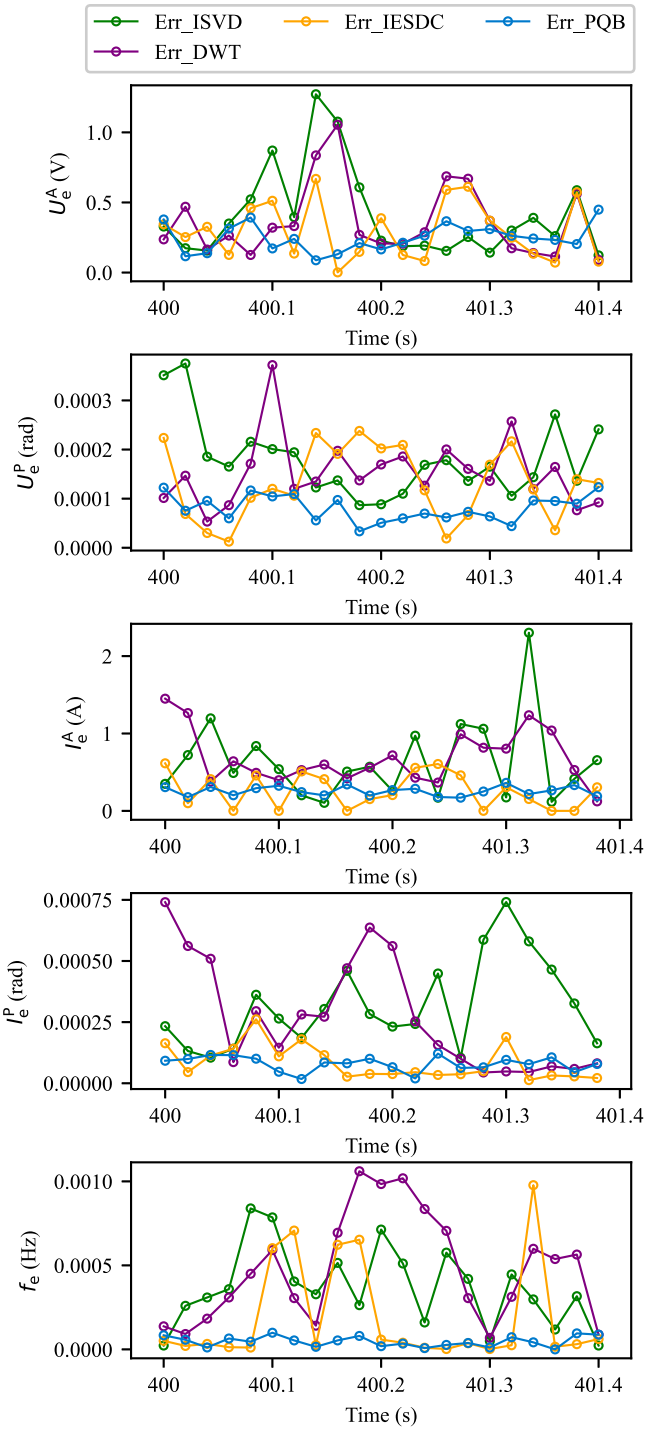

Fig. 8. Error profiles in the normal condition.

a compression algorithm consists of two parts: the constructed time window due to the algorithm operation logic, called the logic delay, and the runtime. The transmission delay is not considered in this paper. For PQB, there is no logic delay, and the time delay is only the runtime spent on the calculation of prediction, quantization and Bitpack, among which the prediction accounts for the majority. To ensure real-time performance, the runtime of each frame should be less than the reporting interval

TABLE III
COMPUTATION EFFICIENCY

Data Size	40000 Frames	One Frame
Average Runtime	1.703 s	0.043 ms

TABLE IV
COMPRESSION PERFORMANCES IN THE FAULT CONDITION

Metrics	Data Types	PQB	ISVD	DWT	IESDC
MAX	U_{Φ}^{A}(V)	0.579	2.289	3.281	1.392
	U_{Φ}^{P}(deg)	0.010	0.044	0.133	0.036
	I_{Φ}^{A}(A)	0.099	0.242	3.445	0.376
	I_{Φ}^{P}(deg)	0.010	0.086	0.090	0.033
	f(Hz)	1e-4	0.006	8e-4	0.003
MAE	U_{Φ}^{A}(V)	0.287	0.520	0.794	0.289
	U_{Φ}^{P}(deg)	0.005	0.011	0.015	0.007
	I_{Φ}^{A}(A)	0.037	0.039	0.042	0.043
	I_{Φ}^{P}(deg)	0.005	0.013	0.014	0.005
	f(Hz)	5e-5	2e-4	5e-5	7e-5
CR	-	6.202	2.505	2.964	4.615

of D-PMU, for example, 20 ms. For IESDC, the time delay includes the logic delay, which is $T_{\max}$, and the runtime. ISVD and DWT are offline compression methods and not suitable for real-time compression.

The test is repeated 20 times on Raspberry Pi 4, and the average results are shown in Table III. It is observed that the average runtime of one frame is significantly less than 20 ms, indicating that PQB can realize real-time compression on edge computing devices.

2) Test in Fault Condition: The short-circuit fault experiment in the pilot distribution network was conducted under conditions in which most critical loads have been shed, resulting in a normal current amplitude of only around 10 A. The short-circuit fault took place between phase A and phase B, with a duration of about 0.3 s. A total of 300 D-PMU data frames covering the entire short-circuit process are utilized. The quantization step size of the current amplitude data is set to 0.2 A, which is $2\%I_0$ and I_0 is 10 A. Other parameters are consistent with that in the normal condition.

a) Comparison of reconstruction accuracy: As shown in Table IV, the MAX values of PQB still do not exceed half of the quantization step sizes, proving the error-bounded property of PQB. Meanwhile, the MAX and MAE values of PQB are lower than those of other methods. Fig. 9 depicts the mean absolute errors for individual data of the same type in three phases during the fault condition. The interval within the red dashed line represents the phase-to-phase short-circuit period. The errors of PQB are concentrated at the bottom of the chart during the entire process, indicating that the accuracy of PQB is not affected by operation conditions. The overall trend of IESDC is similar to PQB, but there are a small number of data with large errors. For ISVD, the errors are small during the fault period because the retained larger singular values first preserve data information with more significant fluctuations. The data before and after the fault are relative stable, characterized by small singular values that are discarded. DWT achieves high-precision compression for frequency data but results in larger errors for other data types.

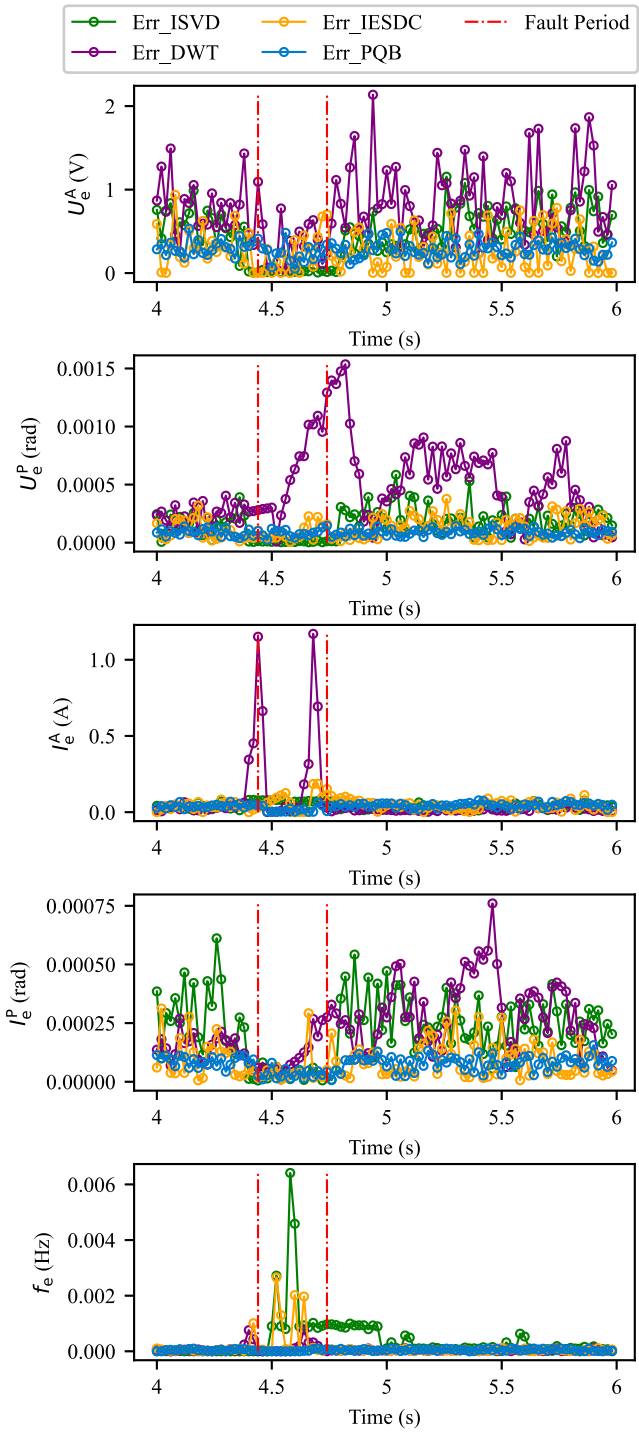

Fig. 9. Error profiles in the phase-to-phase short-circuit condition.

b) Comparison of compression ratio: Table IV compares the compression performances of the four methods. The CR of the IESDC is 4.615. As most loads have been shed in advance, the current amplitude data are stable before and after the short-circuit experiment. Therefore, IESDC can fit the data well

TABLE V
COMPRESSION PERFORMANCES OF SIMULATED DATA

Metrics	Data Types	PQB	ISVD	DWT	IESDC
MAX	U_Φ^{A}(V)	0.240	8.755	2.807	1.316
	U_Φ^{P}(deg)	0.010	0.400	0.088	0.033
	I_Φ^{A}(A)	0.250	1.422	0.700	0.798
	I_Φ^{P}(deg)	0.010	0.131	0.061	0.033
MAE	U_Φ^{A}(V)	0.120	0.587	0.498	0.149
	U_Φ^{P}(deg)	0.005	0.016	0.006	0.013
	I_Φ^{A}(A)	0.129	0.163	0.204	0.209
	I_Φ^{P}(deg)	0.005	0.041	0.015	0.014
CR	-	5.214	3.524	2.885	4.033

by exploiting the 200 ms time window, and most data points are discarded. For ISVD and IDWT, the compression performances are compromised to reduce the information loss. The CR of PQB is 6.202, which is less than that in the normal condition. The reason is that the phase-to-phase short-circuit increases prediction errors, leading to more coding bits. In addition, considering that PQB compresses the D-PMU data frame as a whole, the number of encoding bits for the current amplitude data can be influenced by another type of data even if the current amplitude data are stable. Nevertheless, the CR of PQB is still the highest.

B. Simulated Data

To evaluate the applicability of PQB in different scenarios, simulated D-PMU data are generated with the modified IEEE 123-node test case using PSCAD, in which the fluctuations of PV output and a single-phase grounding fault are considered. The total 400 data frames from the D-PMU placed at node 77 covering the entire short-circuit process are utilized. The phase voltage amplitude reference U_0 is set to 2.40 kV and the phase current amplitude reference I_0 is set to 50 A according to rated capacity. Considering significant fluctuations of the current amplitude during the fault, the quantized value is set to $1\%I_0$. The setting of other parameters is identical to that of field data in the normal condition.

1) Comparison of Reconstruction Accuracy: Table V compares the reconstruction accuracy of the four methods in the single-phase grounding fault. It is also clear that the MAX values of PQB are less than half of the quantization values. The voltage amplitude and phase angle data realize approximate accuracy as that in the normal condition. Fig. 10 illustrates the mean absolute errors for individual data of the same type in three phases during the fault. The interval within the red dashed line represents the single-phase grounding fault period. It is clear that the error profiles of PQB are below those of the other methods. The compression logic of ISVD during this fault is similar to that in the phase-to-phase short-circuit condition. IESDC and DWT achieve high-precision compression for the phase angle data, but the accuracy decreases for other data types.

2) Comparison of Compression Ratio: Table V compares the compression ratios. The CR of PQB is 5.214, which is the highest although it is still less than that of field data in the normal condition. The fluctuations of PV output and the single-phase

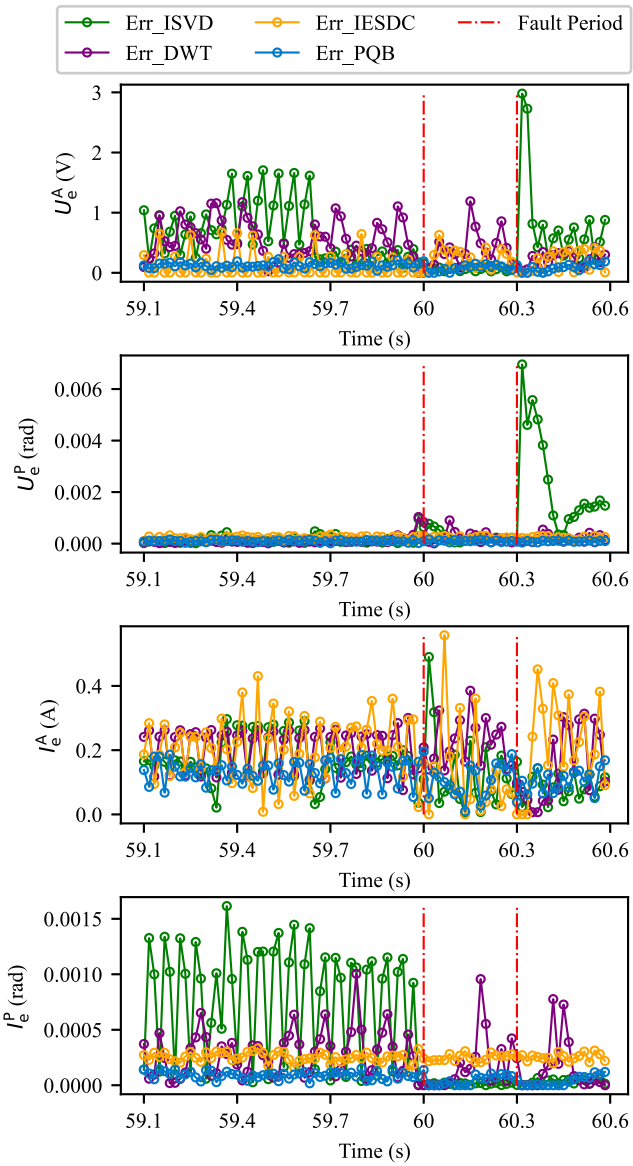

Fig. 10. Error profiles of simulated data.

grounding fault make the current amplitude data increase significantly. However, the prediction errors do not experience a high growth and the CR decreases slightly compared with that in the normal condition. For ISVD, the CR increases to about 3.5 but the accuracy decreases, which is also reflected in Fig. 10. For IESDC and DWT, the fluctuating data make it challenging to fit the raw data linearly or using a small number of wavelet coefficients. Therefore, most data and wavelet coefficients are preserved to guarantee accuracy.

C. Verification of the Performance Optimization Mechanism for Edge Computing

1) Predictor Multiplexing Strategy for Edge Computing Devices: Considering the limited computation and memory resources, the spatial similarity of multiple D-PMUs under the nearby branch of distribution networks is exploited to reduce

TABLE VI
PERFORMANCE OF THE PREDICTOR MULTIPLEXING

D-PMU ID	1	2	3	4
CR after Multiplexing	6.858	7.026	7.071	7.025
CR before Multiplexing	8.010	7.965	7.946	7.950

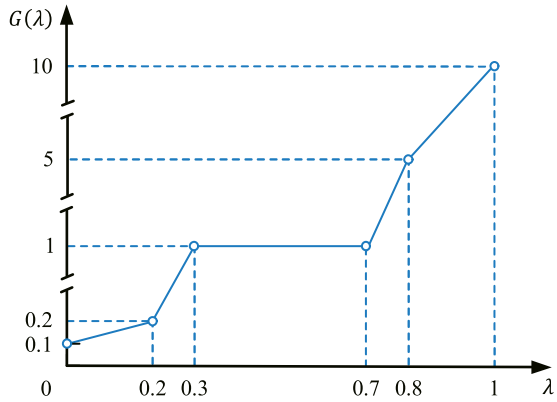

Fig. 11. Dynamic amplitude modulation function.

the computation burden of PQB. The three-phase nodal voltage amplitudes and frequency of 4 D-PMUs installed at nodes 46, 47, 50 and 53 depicted in Fig. 6, including 40000 frames for each D-PMU, are used to simulate the real-time reporting of multiple D-PMUs accessing an edge computing device. The control parameters setting is consistent with that in the normal condition of Section V-A.

The CR that involves the voltage amplitude and frequency of each D-PMU after the NLMS predictor multiplexing is shown in Table VI. The CR after multiplexing is less than that before multiplexing, but it can still reduce the amount of data by about 7 times. The average of D-PMU data, not the raw and reconstructed data, is utilized to update the weights, which increases the prediction errors. The reason is that changes in loads or distributed generators at different nodes of distribution networks cause voltage fluctuations, although the general trend is close. However, the benefits of the NLMS predictor multiplexing are also obvious. In this case, the number of D-PMUs is 4. Therefore, the number of predictors is about one quarter of that before multiplexing. The computation burden of edge computing devices can drop by about three quarters.

2) Adaptive Adjustment of Compression Performances Considering Bandwidth: The available bandwidth changes with time, which is defined in (19) and shows that the bandwidth ratio λ decreases from 1 to 0. The data from the D-PMU at node 46 and the parameters setting of the normal condition are adopted except for the quantization step size matrix. The stable operation interval is $(0.3, 0.7]$. The quantization step size matrix in the stable operation interval $\boldsymbol{\delta}_{\mathbf{s}}$ is the same as that in the normal condition of Section V-A. When the available bandwidth exceeds the stable operation interval, the amplitude modulation function $G(\lambda)$ is adjusted with reference to the requirement of the maximum absolute errors, as shown in (20) and Fig. 11. In this case, the lowest precision under the limited bandwidth is about one-tenth of that under the stable interval and the highest precision is about ten times.

TABLE VII
COMPRESSION PERFORMANCES CONSIDERING BANDWIDTH

Period	0-160 s	160-240 s	240-560 s	560-640 s	640-800 s
λ	(0.8, 1]	(0.7, 0.8]	(0.3, 0.7]	(0.2, 0.3]	(0, 0.2]
$\boldsymbol{\delta}/\boldsymbol{\delta}_{\mathbf{s}}$	[0.1, 0.2)	[0.2, 1)	1	[1, 5)	[5, 10)
CR	4.180	5.456	7.327	8.767	14.498

$$b_{\mathrm{c}} = \left(1 - \frac{t}{800}\right) b_0 \tag{19}$$

$$G(\lambda) = \begin{cases} 0.2 + 0.5(\lambda - 0.2), \lambda \in (0, 0.2] \\ 1 + 8(\lambda - 0.3), \lambda \in (0.2, 0.3] \\ 1, \lambda \in (0.3, 0.7] \\ 1 + 40(\lambda - 0.7), \lambda \in (0.7, 0.8] \\ 5 + 25(\lambda - 0.8), \lambda \in (0.8, 1] \end{cases} \tag{20}$$

The CR of each period is shown in Table VII. In the stable interval, from 240 s to 560 s, the CR is close to the normal condition. Within 0 - 240 s, the available bandwidth is abundant, and the amplitude modulation function $G(\lambda)$ is greater than 1. The result is that the accuracy increases due to the lower quantization step size matrix $\boldsymbol{\delta}$, and the bandwidth utilization is improved. From 560 s to 800 s, the available bandwidth becomes insufficient, and the quantization step size matrix is adaptively adjusted. The CR increases to around 15 to reduce the size of the compressed data and prevent the unacceptable transmission delay.

D. Influences and Countermeasures of Data Quality on Compression Performances

In actual distribution networks, D-PMU data quality may be affected by many factors, including communication failures [42], synchronization problems [43], cybersecurity risks [44], etc. In this paper, we assess the impacts of these factors on the compression performances of PQB during the normal operation of distribution networks.

If communication failures take place, D-PMU data may be missing, which would affect the prediction stage and render the prediction error E_i^t unable to be calculated correctly in (7). In this condition, the missing data can be rectified through various methods. In this paper, the missing value is filled in with the predicted value P_i^t based on the temporal similarity of D-PMU data.

Synchronization problems may be caused by the inaccurate clock [43] or GPS spoofing [45], leading to errors in the phase angle data. We consider two factors which may cause synchronization problems, including the inaccurate clock and a specific spoofing separately by Case 1 and Case 2.

Case 1: All D-PMUs receive the 1 PPS signal from the GPS to achieve the synchronization. However, the sampling clocks are not exactly synchronized during each one-second interval.

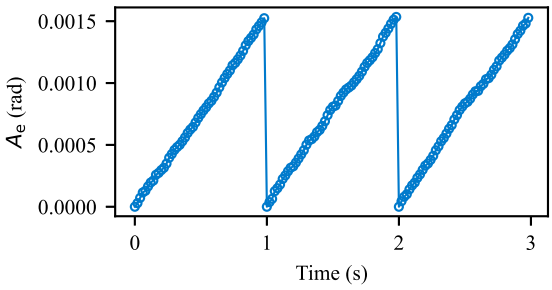

Fig. 12. Phase angle error curve in Case 1.

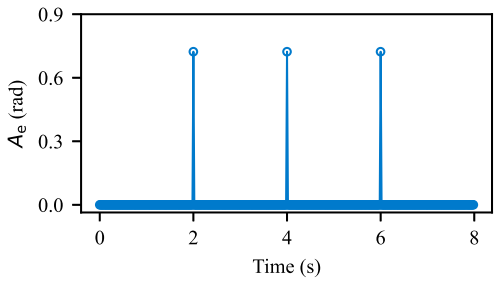

Fig. 13. Phase angle error curve in Case 2.

Therefore, the timing error originates and accumulates during the one-second interval and are cleared while the PPS signal arrives. We set that the distribution of the timing error caused by the inaccurate clock is a normal distribution with a mean of 0.1 μs and a standard deviation of 0.03 μs within a reporting interval [46]. The phase angle error A_e can be calculated by (21). Fig. 12 illustrates the phase angle error curve caused by the inaccurate clock.

$$A_e = 2\pi T_e f \tag{21}$$

where T_e is the timing error (s) and f is the frequency.

Case 2: A specific GPS spoofing is set. The consequent phase angle error may be relatively larger compared to Case 1. We consider the timing error in individual data caused by the specific spoofing attack. With reference to [45], the timing error can be up to 2.3 ms. The frequency of this attack is set at once every 2 seconds, as shown in Fig. 13.

Cybersecurity risks or unauthorized access can lead to false data injection, resulting in arbitrary false data points in any type of D-PMU data. Case 3 is set to demonstrate the performance of PQB under a typical false data injection.

Case 3: The false data are added into the three-phase voltage amplitude data, which appear as individual data point. The amplitude is $-5\%U_0$ and the frequency of this attack is set at once every 2 seconds, as depicted in Fig. 14.

Case 1–3 are tested using the same parameters as that in the normal condition of Section V-A. The test data are based on the modification of the field data. Table VIII compares the compression ratios. CR is the compression ratio based on the test data without defense. It is clear that the compression performance of Case 1 only has a slight decrease compared to that in the normal condition of Section V-A. Therefore, the influences brought about by the inaccurate clock can be overlooked. However, the

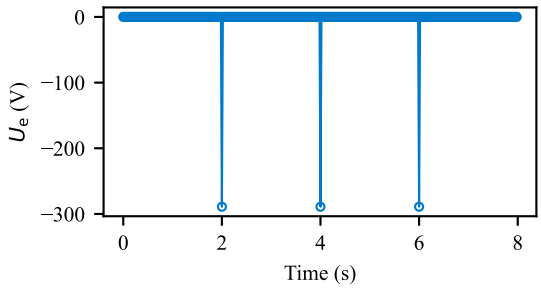

Fig. 14. False voltage amplitude data injection in Case 3.

TABLE VIII
CR COMPARISON CONSIDERING MEASUREMENT DATA QUALITIES

Scenarios	Section V-A	Case 1	Case 2	Case 3
CR	7.249	7.218	4.902	6.651
CR_d	-	-	7.254	7.250

compression performances of Case 2 and Case 3 are affected and the CRs reduce to 4.902 and 6.651. This is because of that the phase angle data error caused by the inaccurate clock is small and the prediction is still accurate. As a result, the number of significant bits for the corresponding quantized values is low. Nevertheless, the errors of individual data introduced by the GPS spoofing or cybersecurity risks are so large that the quantized values exceed the range of the quantizer. Consequently, the raw data near the time of the attack need to be preserved.

To prevent this type of attack from weakening the compression performances of PQB, we consider the current data as an attack value and replace it with the predicted value if the prediction error exceeds a reasonable threshold. CR_ds for Case 2-3 represent the compression ratio after our defense against single-point time-synchronization attacks and cybersecurity attacks. The CR_ds are 7.254 and 7.250 respectively, which are close to that in the normal condition of Section V-A. The results indicate that PQB can effectively deal with the synchronization problems caused by factors such as inaccurate clock or specific GPS spoofing, as well as the false data in individual data. Meanwhile, it should be noted that the synchronization problems may be more complicated, such as GPS signal loss [46], well-designed time-synchronization attacks or cybersecurity risks considering physical characteristics of distribution networks [47], [48], which may be difficult to be detected and require more sophisticated solutions [49], [50].

VI. CONCLUSION

D-PMUs with a high reporting rate generate massive measurement data, resulting in a huge data transmission burden for communication systems of digital distribution networks. In this paper, a real-time data compression method is proposed, including three stages: prediction, quantization and Bitpack. The current data frame is predicted by the NLMS predictor. Then, the prediction errors are quantized by the uniform quantizer to reduce the overall data size. Bitpack is employed to discard

meaningless leading bits and further reduce data redundancy as a whole for all data types. In addition, a performance optimization mechanism for edge computing devices is designed. Considering the spatial similarity of multiple D-PMUs, the NLMS predictor multiplexing is applied to reduce the computation burden. The adaptive adjustment of the compression performances considering the available bandwidth is designed to diminish transmission delay and guarantee the real-time property of D-PMU data in limited bandwidth conditions. Case studies are performed using the field data and simulated data. The results show that PQB is real-time and achieves a CR of about 5–7 with high accuracy during normal and fault conditions of distribution networks. In the future, D-PMU data compression for specific applications needs further research to improve the compression efficiency.

REFERENCES

[1] A. Marot et al., "Perspectives on future power system control centers for energy transition," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 2, pp. 328–344, 2022.

[2] C. Feng, Y. Wang, Q. Chen, Y. Ding, G. Strbac, and C. Kang, "Smart grid encounters edge computing: Opportunities and applications," *Adv. Appl. Energy*, vol. 1, 2021, Art. no. 100006.

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[4] J. Jian et al., "Supply restoration of data centers in flexible distribution networks with spatial-temporal regulation," *IEEE Trans. Smart Grid*, early access, 16 Jun. 2023, doi: 10.1109/TSG.2023.3286844.

[5] J. Li, C. Gu, Y. Xiang, and F. Li, "Edge-cloud computing systems for smart grid: State-of-the-art, architecture, and applications," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 4, pp. 805–817, 2022.

[6] J. Zhao et al., "Cloud-edge collaboration-based local voltage control for DGs with privacy preservation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 98–108, Jan. 2023.

[7] I. Kosen et al., "UPS: Unified PMU-data storage system to enhance T+D PMU data usability," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 739–748, Jan. 2020.

[8] A. V. Meier, E. Stewart, A. McEachern, M. Andersen, and L. Mehrmanesh, "Precision micro-synchrophasors for distribution systems: A summary of applications," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2926–2936, Nov. 2017.

[9] *IEEE Standard for Synchrophasor Measurements for Power Systems*, IEEE Standard C37.118.1-2011, 2011.

[10] W. Wang, C. Chen, W. Yao, K. Sun, W. Qiu, and Y. Liu, "Synchrophasor data compression under disturbance conditions via cross-entropy-based singular value decomposition," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2716–2726, Apr. 2021.

[11] L. Fiaschetti et al., "Monitoring and controlling energy distribution: Implementation of a distribution management system based on common information model," *Int. J. Elect. Power Energy Syst.*, vol. 94, pp. 67–76, 2018.

[12] Y. Liu, J. Li, and L. Wu, "State estimation of three-phase four-conductor distribution systems with real-time data from selective smart meters," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2632–2643, Jul. 2019.

[13] Z. Liu et al., "Robust state estimation of active distribution networks with multi-source measurements," *J. Mod. Power Syst. Clean Energy*, vol. 11, no. 5, pp. 1540–1552, Sep. 2023.

[14] J. Sun, Q. Chen, and M. Xia, "Data-driven detection and identification of line parameters with PMU and unsynchronized SCADA measurements in distribution grids," *CSEE J. Power Energy Syst.*, early access, May 6, 2022, doi: 10.17775/CSEEJPES.2020.06860.

[15] W. Li, D. Deka, M. Chertkov, and M. Wang, "Real-time faulted line localization and PMU placement in power systems through convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4640–4651, Nov. 2019.

[16] H. Mortazavi, H. Mehrjerdi, M. Saad, S. Lefebvre, D. Asber, and L. Lenoir, "A monitoring technique for reversed power flow detection with high PV penetration level," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2221–2232, Sep. 2015.

[17] H. Gharavi and B. Hu, "Scalable synchrophasors communication network design and implementation for real-time distributed generation grid," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2539–2550, Sep. 2015.

[18] K. Sayood, *Introduction to Data Compression*, 5th ed. San Mateo, CA, USA: Morgan Kaufmann, 2017, pp. 1–10.

[19] X. Wang, Y. Liu, and L. Tong, "Adaptive subband compression for streaming of continuous point-on-wave and PMU data," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5612–5621, Nov. 2021.

[20] L. Wen, K. Zhou, S. Yang, and L. Li, "Compression of smart meter Big Data: A survey," *Renewable Sust. Energ. Rev.*, vol. 91, pp. 59–69, 2018.

[21] J. E. Tate, "Preprocessing and Golomb–Rice encoding for lossless compression of phasor angle data," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 718–729, Mar. 2016.

[22] L. Yan, J. Han, R. Xu, and Z. Li, "Model-free lossless data compression for real-time low-latency transmission in smart grids," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2601–2610, May 2021.

[23] R. Klump, P. Agarwal, J. E. Tate, and H. Khurana, "Lossless compression of synchronized phasor measurements," in *Proc. IEEE/PES Gen. Meeting*, Minneapolis, MN, USA, 2010, pp. 1–7.

[24] C. Chen, W. Wang, H. Yin, L. Zhan, and Y. Liu, "Real-time lossless compression for ultrahigh-density synchrophasor and point-on-wave data," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 2012–2021, Feb. 2022.

[25] P. H. Gadde, M. Biswal, S. Brahma, and H. Cao, "Efficient compression of PMU data in WAMS," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2406–2413, Sep. 2016.

[26] J. Zhao et al., "Improved SVD-based data compression method for synchronous phasor measurement in distribution networks," *Int. J. Elect. Power Energy Syst.*, vol. 129, 2021, Art. no. 106877.

[27] F. Zhang, L. Cheng, X. Li, Y. Sun, W. Gao, and W. Zhao, "Application of a real-time data compression and adapted protocol technique for WAMS," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 653–662, Mar. 2015.

[28] J. Ning, J. Wang, W. Gao, and C. Liu, "A wavelet-based data compression technique for smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 212–218, Mar. 2011.

[29] R. Pourramezan, R. Hassani, H. Karimi, M. Paolone, and J. Mahseredjian, "A real-time synchrophasor data compression method using singular value decomposition," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 564–575, Jan. 2022.

[30] S. Chandak, K. Tatwawadi, C. Wen, L. Wang, J. Aparicio, and T. Weissman, "LFZip: Lossy compression of multivariate floating-point time series data via improved prediction," in *Proc. Data Compression Conf.*, 2020, pp. 342–351.

[31] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Process*, vol. 41, no. 9, pp. 2811–2825, Sep. 1993.

[32] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[33] Y. Shi, L. Huang, C. Qian, and H. C. So, "Shrinkage linear and widely linear complex-valued least mean squares algorithms for adaptive beamforming," *IEEE Trans. Signal Process*, vol. 63, no. 1, pp. 119–131, Jan. 2015.

[34] D. Blalock, S. Madden, and J. Guttag, "Sprintz: Time series compression for the Internet of Things," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 93–116, 2018.

[35] L. S. Vedantham, Y. Zhou, and J. Wu, "Information and communications technology (ICT) infrastructure supporting smart local energy systems: A review," *IET Energy Syst. Integr.*, vol. 4, no. 4, pp. 460–472, 2022.

[36] Y. Huo et al., "Data-driven coordinated voltage control method of distribution networks with high DG penetration," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1543–1557, Mar. 2023.

[37] X. Fu, X. Wu, C. Zhang, S. Fan, and N. Liu, "Planning of distributed renewable energy systems under uncertainty based on statistical machine learning," *Protection Control Mod. Power Syst.*, vol. 7, no. 41, pp. 1–27, 2022.

[38] T. Yang et al., "Elastic sketch: Adaptive and fast network-wide measurements," in *Proc. Conf. ACM Special Int. Group Data Commun.*, 2018, pp. 561–575.

[39] R. Lübben, M. Fidler, and J. Liebeherr, "Stochastic bandwidth estimation in networks with random service," *IEEE ACM Trans. Netw.*, vol. 22, no. 2, pp. 484–497, Apr. 2014.

[40] W. Xi et al., "Two-stage PMU data compression for edge computing devices of distribution networks," *Power Syst. Technol.*, vol. 47, no. 8, pp. 3184–3192, 2022.

[41] L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[42] T. Zhang, P. Yuan, Y. Du, W. Zhang, and J. Chen, "Robust distributed state estimation of active distribution networks considering communication failures," *Int. J. Elect. Power Energy Syst.*, vol. 118, 2020, Art. no. 105732.
[43] Q. Zhang, V. Vittal, G. Heydt, Y. Chakhchoukh, N. Logic, and S. Sturgill, "The time skew problem in PMU measurements," in *Proc. IEEE Power Energy Soc. Gen. Meeting.*, 2012, pp. 1–6.
[44] J. Zhao, G. Zhang, M. L. Scala, Z. Dong, C. Chen, and J. Wang, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1580–1590, Jul. 2017.
[45] X. Jiang, J. Zhang, B. J. Harding, J. J. Makela, and A. D. Domínguez-García, "Spoofing GPS receiver clock offset of phasor measurement units," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3253–3262, Aug. 2013.
[46] W. Yao et al., "Impact of GPS signal loss and its mitigation in power system synchronized measurement devices," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1141–1149, Mar. 2018.
[47] L. Che, X. Liu, Z. Li, and Y. Wen, "False data injection attacks induced sequential outages in power systems," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1513–1523, Mar. 2019.
[48] E. Shereen, M. Delcourt, S. Barreto, G. Dán, J. L. Boudec, and M. Paolone, "Feasibility of time-synchronization attacks against PMU-based state estimation," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3412–3427, Jun. 2020.
[49] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Dong, "A review of false data injection attacks against modern power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1630–1638, Jul. 2017.
[50] Z. Zhang, S. Gong, A. D. Dimitrovski, and H. Li, "Time synchronization attack in smart grid: Impact and analysis," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 87–98, Mar. 2013.

Jiahui Yang (Student Member, IEEE) received the B.S. degree in electrical engineering in 2021 from Tianjin University, Tianjin, China, where he is currently working toward the M.S. degree in electrical engineering. His research interests include edge computing in digital distribution networks and data compression.

Hao Yu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Tianjin University, Tianjin, China, in 2010 and 2015, respectively. He is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University. His research interests include the operation analysis and optimization of active distribution networks and integrated energy systems. He is also an Assistant Editor for *IET Energy Systems Integration*.

Peng Li (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Tianjin University, Tianjin, China, in 2004 and 2010, respectively. He is currently a Professor with the School of Electrical and Information Engineering, Tianjin University. His research interests include operation and planning of active distribution networks, modeling and transient simulation of power systems. Prof. Li is also an Associate Editor for IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, *CSEE Journal of Power and Energy Systems, Sustainable Energy Technologies and Assessments*, and *IET Renewable Power Generation*.

Haoran Ji (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Tianjin University, Tianjin, China, in 2014 and 2019, respectively. From 2019 to 2021, he was a Postdoctoral Research with Tianjin University. He is currently an Associate Professor with Tianjin University. His research interests include distributed generation systems and optimal operation of distribution networks. He was supported by China National Postdoctoral Program for Innovative Talents in 2019.

Wei Xi received the M.S. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2003. He is currently working toward the Ph.D. degree in electrical engineering with Tianjin University, Tianjin, China. He is currently a Professorate Senior Engineer with the Digital Power Grid Research Institute, China Southern Power Grid. His research interests include electric power chip development and digitization of power grid.

Jianzhong Wu (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Tianjin University, China, in 1999, 2002 and 2004, respectively. From 2004 to 2006, he was with Tianjin University, where he was an Associate Professor. From 2006 to 2008, he was a Research Fellow with the University of Manchester, Manchester, U.K. He is currently a Professor of multi-vector energy systems and the Head of the School of Engineering, Cardiff University, Cardif, U.K. His research interests include integrated multi-energy infrastructure and smart grid. He is the Co-Editor-in Chief of Applied Energy. He is the Co-Director of U.K. Energy Research Centre and EPSRC Supergen Energy Networks Hub.

Chengshan Wang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Tianjin University, Tianjin, China, in 1991. He is currently a Professor with the School of Electrical and Information Engineering, Tianjin University. He is also an Academician of China National Engineering Research Institute. He is the Director of the Key Laboratory of Smart Grid of Ministry of Education, Tianjin University. His research interests include distribution system analysis and planning, distributed generation system and microgrid. Prof. Wang is the Editor-in-Chief of *IET Energy Systems Integration*.