A Hybrid Compression Method for Compound Power Quality Disturbance Signals in Active Distribution Networks

Xiangui Xiao, Kaicheng Li, and Chen Zhao

***Abstract*—In the compression of massive compound power quality disturbance (PQD) signals in active distribution networks, the compression ratio (CR) and reconstruction error (RE) act as a pair of contradictory indicators, and traditional compression algorithms have difficulties in simultaneously satisfying a high CR and low RE. To improve the CR and reduce the RE, a hybrid compression method that combines a strong tracking Kalman filter (STKF), sparse decomposition, Huffman coding, and run-length coding is proposed in this study. This study first uses a sparse decomposition algorithm based on a joint dictionary to separate the transient component (TC) and the steady-state component (SSC) in the PQD. The TC is then compressed by wavelet analysis and by Huffman and run-length coding algorithms. For the SSC, values that are greater than the threshold are reserved, and the compression is finally completed. In addition, the threshold of the wavelet depends on the fading factor of the STKF to obtain a high CR. Experimental results of real-life signals measured by fault recorders in a dynamic simulation laboratory show that the CR of the proposed method reaches as high as 50 and the RE is approximately 1.6%, which are better than those of competing methods. These results demonstrate the immunity of the proposed method to the interference of Gaussian noise and sampling frequency.**

***Index Terms*—Signal compression, power quality disturbance, Huffman coding, run-length coding, wavelet analysis, sparse decomposition.**

I. Introduction

POWER quality disturbance (PQD) data are useful for active distribution networks (ADNs), where different PQD waveforms in ADNs generally correspond to different ADN operations. For example, the transient oscillation signal is caused by load switching [1], and the voltage notch originates from the phase commutation of new energy resources embedded with power electronic devices in an ADN [2]. In addition, voltage sag or interruption occurs because of short-circuit faults [3], and the impulse signal derives from thunder strikes on the lines in ADNs [4]. These operations in ADNs severely distort waveforms. In addition, PQD waveforms can be used to locate disturbances in distribution networks [5]. However, as the scale of ADN increases, PQD events inevitably become more frequent. If many power grid monitoring devices are used in a distribution network, the Ethernet bandwidth will be enormous for transmitting and receiving uncompressed PQD data [6], and the disk space for storage will also be considerable, thus incurring high expense. Therefore, a highly efficient compression method for PQD data is urgently needed [7].

The main compression methods can be divided into two types: lossy and lossless. Lossless compression reduces the redundancy of PQD data by encoding methods such as Huffman coding [8] and Lempel-Ziv-Welch (LZW) coding [9]. The advantage of lossless compression is that the data can be completely restored without loss after signal recovery. Unfortunately, the redundancy of PQD data is generally low, and it is difficult to obtain a high compression ratio (CR) by encoding. Typically, engineers focus on overall PQD events instead of their details. Therefore, discarding certain details of PQD signals during compression processing is considered acceptable. Many researchers have thus developed lossy compression methods for PQD signals.

The lossy method consists of two types of compression, namely, direct and indirect compressions. Direct compression uses a wavelet to compress the PQD signals directly. Reference [10] utilized 1D wavelet analysis and improved wavelet-based methods [11] - [13] to compress 1D signals. The study obtained higher CR and lower waveform distortion than with the discrete cosine transform.

2D discrete wavelet transform (DWT) based algorithms are often applied to PQD signal compression. First, the PQD signal is reshaped into a 2D grayscale image, and the reshaped "image" is compressed through 2D DWT [14] - [16]. Although the reconstruction error is less than that of the 1D DWT under the same CR, a long data frame is necessary for data reshaping with DWT-based algorithms. In addition, singular value decomposition (SVD), which is a major image

Manuscript received: September 17, 2022; revised: December 5, 2022; accepted: February 5, 2023. Date of CrossCheck: February 5, 2023. Date of online publication: April 21, 2023.

This work was supported in part by the National Natural Science Foundation of China (No. 52077089).

X. Xiao and K. Li (corresponding author) are with the State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: xiaoxiangui@hust.edu.cn; likaicheng@hust.edu.cn).

C. Zhao is with the College of Mechanical and Electrical Engineering, Fujian Agriculture and Forest University, Fuzhou 350002, China. (e-mail: fauzhao@gmail.com).

DOI: 10.35833/MPCE.2022.000602

compression technique, has been used in PQD signal compression [17], [18]. The purpose of SVD-based compression is to discard several small values of the "image" decomposition results. Its advantage is its simplicity; its disadvantage is the long data frame required for the 1D signal to transform into a 2D image.

The goal of indirect compression is to first separate the transient component (TC) and steady-state component (SSC) in the PQD signals and then compress the TC and SSC using different schemes. Reference [19] used a Butterworth filter bank to separate the TC and SSC. The study then calculated the amplitude and angle of the SSC using interpolated FFT and compressed the TC using DWT (db4 basis). Although the high CR and low percentage of root-mean-square difference (PRD) verified the effectiveness of the algorithm, the algorithm failed to consider noise in PQD signals. Similar to [19], [20] used a Kalman filter (KF) and PQD signal compression dynamics to obtain a high CR.

Another mathematical approach for separating the TC and SSC is sparse decomposition (SD), which is based on an overcomplete hybrid dictionary (OHD) and consists of identity matrix $\boldsymbol{I}$, cosine transform matrix $\boldsymbol{C}$, and sine transform matrix $\boldsymbol{S}$ [21], [22]. Reference [23] combined $\boldsymbol{I}$ and the identity Fourier transform matrix $\boldsymbol{F}$, and [24] combined $\boldsymbol{I}$ and the identity Hartley transform matrix $\boldsymbol{H}$. In [22], the OHD [$\boldsymbol{I}$, $\boldsymbol{C}$, $\boldsymbol{S}$] was used instead of compression for PQD event classification. The number of columns in [$\boldsymbol{I}$, $\boldsymbol{C}$, $\boldsymbol{S}$] is typically large, and the solution is time-consuming. In [23], an OHD [$\boldsymbol{I}$, $\boldsymbol{F}$] was used for PQD signal compression. However, if the PQD signals contain white Gaussian noise, the CR declines significantly according to the principle of the SD approach. In addition, some orthogonal matching pursuit (OMP) based improved solutions of SD have been reported, such as regularized OMP [25], sparsity adaptive correntropy OMP [26], and adaptive group OMP [27]. However, the computation speed is limited. Therefore, in this study, the classic OMP was chosen as the solution for convex optimization.

Although various methods for PQD signal compression have been proposed in previous studies, several remaining problems require further investigation. Despite the high CR and low PRD values were achieved in [19], no Gaussian noise is apparent in the synthetic PQD signals for compression. The characteristics of synthetic PQD signals in [19] does not match the characteristics of real-life signals. When the method in [19] is applied to compress the real-life signals with Gaussian noise, the CR declines significantly. Reference [23] simply divided the number of points of PQD data by the number of non-zero values to compute CR. Although the study claimed that the CR was 10.08 when the signal-to-noise ratio (SNR) was 40 dB, the CR was actually less than 2 when the method was reproduced. Moreover, it did not compress the results using the lossless encoding method. Therefore, the results in [23] are dubious.

In [19], a wavelet was used to decompose the signals into three levels, and then the detail coefficients d_3, d_2, and d_1 were suppressed by the wavelet threshold and compressed. However, the study did not compress the approximation coefficients a_3, which resulted in a relatively low CR. If a_3 had been compressed, the CR would have been much higher.

During our experiments, we observed that CR and PRD depended on multiple wavelet thresholds. If the wavelet threshold was large, the CR would increase and the PRD would worsen, and vice versa. Thus, the parameter is significant. To the best of our knowledge, no study has been conducted on how to choose the multiple of the wavelet threshold.

Our study was conducted to overcome these shortcomings, where the main contributions are as follows.

1) Some existing studies such as [19] failed to consider noise. Although their algorithm achieved a high CR for noiseless PQD signals, it deteriorated in the presence of noisy PQD signals. In this study, PQD signals are imposed with 30, 40, and 50 dB of Gaussian noise to enable the PQD signals to match the characteristics of real-life signals in a power grid, and the PQD signals with Gaussian noise are compressed and recovered.

2) We design a novel technique for determining the wavelet threshold. Specifically, we use the fading factor (FF) of the strong tracking Kalman filter (STKF). If the FF is greater than 1, the multiple of the threshold is set to be a high value, and the CR will be high without a significant decline in the PRD. If FF is equal to 1, the multiple of the threshold is set to be a small value, and the CR will be slightly lower without a significant increase in the PRD. Thus, the CR can be automatically adjusted by the FF.

3) The TC is decomposed into detail coefficients d_3, d_2, d_1 and approximation coefficients a_3, which are compressed through the run-length and Huffman encoding methods, respectively. Because Huffman coding compresses a floating-point number that occupies 16 bits into a binary code that occupies only 3-7 bits, a_3 can be further compressed by Huffman coding. As a result, a larger CR can be obtained.

4) Small absolute values in the frequency domain, which are generally less than 0.05, correspond to the white Gaussian noise in the time domain. Therefore, all absolute values less than 0.05 are set to be zero. Only the non-zero values and positions are saved, and the sparse SSC is compressed.

The remainder of this paper is organized as follows. In Section II, the proposed method based on STKF, OHD [$\boldsymbol{I}$, $\boldsymbol{H}$], Huffman coding, and run-length coding is described in detail. Section III presents case studies, which describe how the three types of synthetic compound PQD signals in MATLAB are compressed and reconstructed. Section IV gives the experiment results based on real PQD signals. Finally, Section V draws conclusions.

II. Proposed Method

Several existing indirect compression methods perform well on PQD signals without noise, but the CR worsens when the signals contain TCs and noise. To obtain a high CR and low PRD, a novel PQD signal compression method is proposed in this study, and the flow is shown in Fig. 1.

The process flow of the proposed method shown in Fig. 1 is explained in the following three subsections. First, the PQD signal is decomposed into the TC and SSC by SD. The principle of SD is first introduced.

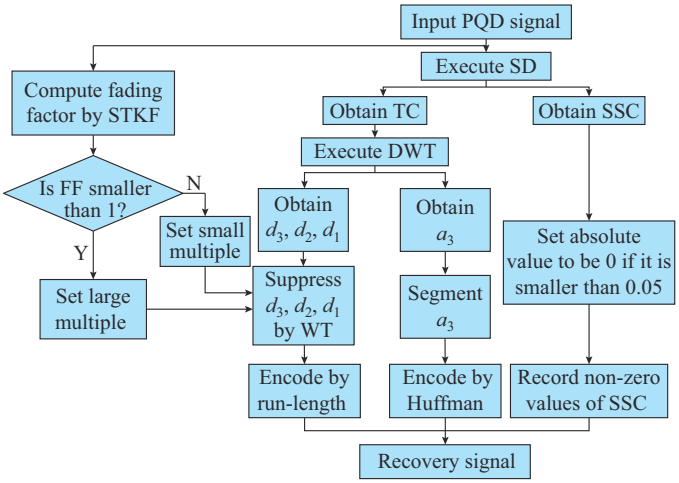

Fig. 1. Flow of proposed method.

A. SD

The SD technique is employed to separate the TC and SSC in the PQD signal for further compression.

Let $\boldsymbol{x}$ be the PQD signal of white Gaussian noise of length N. The TC and SSC can be separated by:

$$\boldsymbol{x}=\boldsymbol{A}\boldsymbol{y} \tag{1}$$

where $\boldsymbol{A}$ is an overcomplete dictionary with size $N\times M$, in which $M>N$; and $\boldsymbol{y}$ is the key to (1) with size $M\times 1$.

In this study, OHD $[\boldsymbol{I},\boldsymbol{H}]$ is used for SD because of its advantages. The disadvantage of dictionary $[\boldsymbol{I},\boldsymbol{F}]$ is that it is a complex matrix demanding double the amount of space for storage. Therefore, the storage space of $[\boldsymbol{I},\boldsymbol{H}]$ is less than that of $[\boldsymbol{I},\boldsymbol{F}]$. The disadvantage of $[\boldsymbol{I},\boldsymbol{C},\boldsymbol{S}]$ is that the frequency domain portion of the decomposition result usually has sidelobes, which impair the CR, whereas the results of $[\boldsymbol{I},\boldsymbol{H}]$ have no sidelobes. Therefore, matrix $\boldsymbol{H}$ with size $n\times n$ can be expressed as:

$$\boldsymbol{H}=\sqrt{\frac{1}{n}}\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & cas\left(\frac{2\pi}{n}\right) & \cdots & cas\left(\frac{2\pi}{n}(n-1)\right)\\ \vdots & \vdots & & \vdots\\ 1 & cas\left(\frac{2\pi}{n}(n-1)\right) & \cdots & cas\left(\frac{2\pi}{n}(n-1)^2\right)\end{bmatrix}_{n\times n} \tag{2}$$

where $cas(\cdot)=\sin(\cdot)+\cos(\cdot)$. Thus, the joint dictionary $\boldsymbol{A}$ in (1) can be expressed as:

$$\boldsymbol{A}=[\boldsymbol{I}\quad\boldsymbol{H}] \tag{3}$$

The solution to (1) is a convex optimization problem. Popular convex optimization solutions include OMP [23] and OMP-based algorithms. For details about OMP, please refer to [23].

B. Huffman Coding and Run-length Coding

After the PQD signal is decomposed into TC and SSC, the following steps are performed to compress the TC and SSC. To improve the CR, Huffman coding is applied to code the approximation coefficients obtained from wavelet analysis. As a classic lossless coding method, the process steps of Huffman coding can be described as follows.

Step 1: sort the list in ascending order from the smallest to largest.

Step 2: choose the smallest and second smallest values to construct a 2-layer binary tree with two leaf nodes and one single root node, which is the sum of the two leaf nodes. Then, delete the two values in the two leaf nodes, and add the new value of the root node to the list.

Step 3: recursively repeat this process until a complete Huffman tree is constructed.

Step 4: finally, encode all paths from the root node to the leaf nodes, with all left and right paths of the tree set to be 0 and 1, respectively.

For example, consider a list with elements 2, 3, 6, 7, 10, 19, 21, and 32 in ascending order. The encoded Huffman values of elements 2, 3, 6, 7, 10, 19, 21, and 32 are 00000, 00001, 0001, 0010, 0011, 10, 11, and 01, respectively. The numbers of bits occupied by the eight integers and encoded values are $8\times 16=128$ bits and $5\times 2+4\times 3+2\times 3=28$ bits, respectively. Therefore, the CR of the list is $128/28=4.57$.

Run-length coding is suitable for compressing sparse lists that contain many consecutive zeros. For example, consider a list with elements 0, 0, 0, 5, 0, 0, 3, 0, 0, 0, and 2, where the corresponding run-length coding result is (3, 5), (2, 3), and (3, 2). The first number in parentheses is the number of consecutive zeros, whereas the second number is the non-zero value immediately following the string of zeros. Here, the 11 integers in the list are represented by 6 integers, and the CR is $11/6=1.83$. The higher the number of consecutive zeros in the sparse list is, the higher the CR is.

C. STKF

The wavelet threshold is a key parameter that determines the CR and PRD. The FF is used to determine the multiple of the wavelet threshold, and the STKF is described as follows.

In the signal model, a typical sinusoidal waveform with harmonics $y(t)$ can be described by:

$$y(t)=\sum_{k=1}^{n}A_k\sin(k\omega t+\varphi_k) \tag{4}$$

where ω is the fundamental angular frequency; and A_k and φ_k are the amplitude and angle of the k^{th} harmonic, respectively.

The second step of STKF is establishing the state equation of the signal to facilitate subsequent processing, i.e., to calculate the FF. Let the elements of the state vector $\boldsymbol{x}(k)$ be:

$$\begin{cases}x_1(k)=A\sin(\omega kT)\\ x_2(k)=A\cos(\omega kT)\\ x_3(k)=\omega\end{cases} \tag{5}$$

where A is the amplitude; and T is the sampling interval. Thus, the nonlinear state equation is given as:

$$\boldsymbol{f}(\boldsymbol{x}(k))=\begin{bmatrix}x_1(k)\cos(x_3(k)T)-x_2(k)\sin(x_3(k)T)\\ x_1(k)\sin(x_3(k)T)+x_2(k)\cos(x_3(k)T)\\ x_3(k)\end{bmatrix} \tag{6}$$

Therefore, the state matrix $\boldsymbol{F}$ is the first-order Taylor expansion of (6) and is expressed as:

$$\boldsymbol{F}=\begin{bmatrix} F_1 & -F_2 & -x_1(k)TA-x_2(k)TB \\ F_2 & F_1 & -x_1(k)TA-x_2(k)TB \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

where $F_1=\cos(x_1(k)T)$; and $F_2=\sin(x_1(k)T)$.

The measurement matrix $\boldsymbol{H}$ is:

$$\boldsymbol{H}=[1 \quad 0 \quad 0] \tag{8}$$

After the nonlinear state and linear measurement equations are established, the parameter matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ are set to be $10^{-5}\boldsymbol{I}_n$ and 10^{-4}, respectively.

A discrete linear system can be described by state and measurement equations, which are given by (9) and (10), respectively.

$$\boldsymbol{x}(k+1)=\boldsymbol{F}\boldsymbol{x}(k)+\boldsymbol{w}(k) \tag{9}$$

$$\boldsymbol{Y}(k)=\boldsymbol{H}\boldsymbol{x}(k)+\boldsymbol{v}(k) \tag{10}$$

where $\boldsymbol{x}(k)$ is the state vector of the system at step k with size $n\times 1$; $\boldsymbol{Y}(k)$ is the measurement vector of the system at step k with size $m\times 1$; $\boldsymbol{F}$ is the state-transition matrix; $\boldsymbol{H}$ is the measurement matrix; and $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are the process and measurement noises, respectively, in which both $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ have zero means, and the variance matrices are $\boldsymbol{Q}$ and $\boldsymbol{R}$, respectively. For details about the STKF, please see [28].

III. Case Study and Analysis

To test the effectiveness of the proposed method, three types of compound PQD signals with 40 dB Gaussian noise are generated in MATLAB and then compressed and reconstructed. Because of the randomness of noise, compression and reconstruction are performed 10 times in this study, and the mean value is taken as the final result.

The sampling frequency of the PQD signal under a duration of 0.1 s is 6.4 kHz, and 640 samples are used for each signal. The two indicators of CR and PRD are used to evaluate the proposed method and are expressed in (11) and (12), respectively.

$$CR=\frac{S_{original}}{S_{compress}} \tag{11}$$

$$PRD=\sqrt{\frac{\sum_{i=1}^{N}(x(i)-\hat{x}(i))^2}{\sum_{i=1}^{N}x^2(i)}}\times 100\% \tag{12}$$

where $S_{original}$ and $S_{compress}$ are the bits occupied by the original and compressed signals, respectively; and $x(i)$ and $\hat{x}(i)$ are the original and recovered signals of the i^{th} sampling point, respectively.

A. Impulse Signal

Impulse waveforms are a common phenomenon in ADNs. Thunder strikes of transmission lines are a unique cause, and the waveform appears as a unipolar transient pulse. The increase in time of the impulse signal is typically between 5 ns and 0.1 ms, and the duration is longer than 1 ms [23]. Thus, a model of the impulse signal of the k^{th} sampling point can be given as:

$$x(k)=\begin{cases} \sin(2\pi\cdot 50k/f_s)+e^{-(k-381)} & 381\le k<388 \\ \sin(2\pi\cdot 50k/f_s)+n_{oise} & \text{others} \end{cases} \tag{13}$$

where f_s is the sampling frequency, which is 6.4 kHz; and n_{oise} is the Gaussian noise. To simulate a real-life environment, an impulse signal is imposed with 40 dB Gaussian noise.

The result decomposed by SD can be segmented into two parts in the frequency domain, i.e., a TC composed of the Gaussian noise and impulse signal and an SSC that derives from the Hartley transform, as shown in Fig. 2.

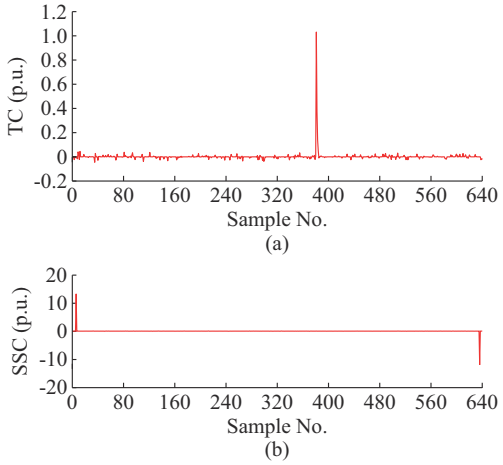

Fig. 2. Result decomposed by SD. (a) TC. (b) SSC.

The TC is first decomposed by wavelet analysis (db4 basis) into three levels to obtain the approximation coefficients a_3 and detail coefficients d_3, d_2, and d_1. Different strategies can then be used to compress the detail and approximate coefficients. The detail coefficients are suppressed by the wavelet threshold, as shown by:

$$\lambda_i=\alpha\frac{median(abs(d_i))}{0.6475}\sqrt{2\ln N} \tag{14}$$

where $i\in\{1,2,3\}$; d_i denotes the detail coefficients in which N is the length; $abs(\cdot)$ and $median(\cdot)$ are the operations for computing the absolute and median values, respectively; and α is the tunable multiple of the coefficient that determines the CR and PRD. If α is low, the threshold will be low, and numerous "burrs" in d_i will be retained, which is not conducive to a high CR. By contrast, if α is high, the "burrs" in d_i will be discarded, which leads to a relatively high CR.

The problem of choosing α must then be resolved. Because FF reflects a sudden change in the PQD waveform, we use FF as an indicator to choose the parameter. If FF is greater than 1, α is set to be 2, and if FF is 1, α is set to be 1.

Figure 3 shows that the FF of the impulse signal is greater than 1. This result indicates that a sudden change occurs in the waveform, and therefore α should be relatively high to avoid a low CR. The new detail coefficients of impulse signal suppressed according to (14) are shown in Fig. 4.

In Fig. 4, most of the values for d_3, d_2, and d_1 are set to be 0, and the new coefficients d_3', d_2', and d_1' become sparse and can be compressed by run-length coding.

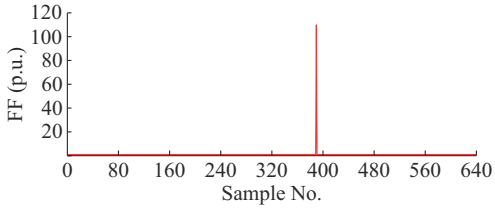

Fig. 3. FF of impulse signal.

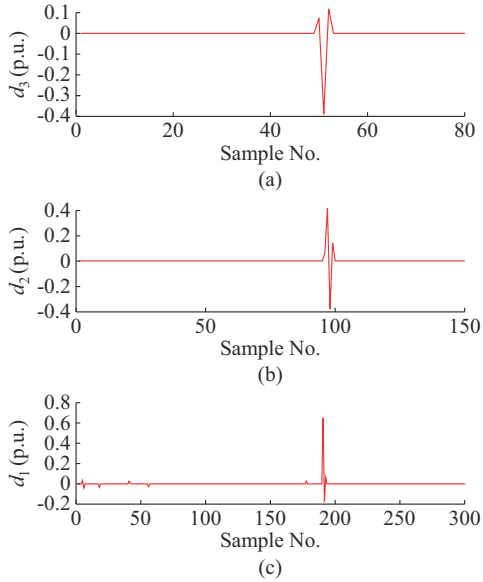

Fig. 4. New detail coefficients of impulse signal. (a) d_3. (b) d_2. (c) d_1.

With the new detail coefficient d_1' taken as an example, the run-length encoding results are (90, 0.037), (99, 0.6226), (0, –0.1758), (0, 0.0863), and 129. Note that run-length coding is previously explained in Section II. Nine floating-point numbers occupy $9 \times 32 = 288$ bits, whereas 323 numbers are included in the original detail coefficient d_1, which require $323 \times 32 = 10336$ bits. CR is thus $10336/288 = 35.89$, and therefore d_1 is efficiently compressed. Similarly, d_2 and d_3 are efficiently compressed.

The next step is to compress the approximation coefficients a_3, which is difficult because of low redundancy. To make it suitable for Huffman coding, the processing of a_3 is presented in Fig. 5.

After this processing, the new approximation coefficients a_3' have only 19 numbers. The error of a_3' and a_3 does not exceed 0.0015, whereas the redundancy is greatly improved. Huffman coding is then performed on a_3'. As a result, the least frequent value in a_3' is –0.0224, which occurs only once, and the corresponding Huffman value is 100110; the most frequent value in a_3' is 0.0046, which occurs 13 times, and the corresponding Huffman encoding value is 011. The Huffman tree for a_3' is shown in Appendix A Fig. A1. The Huffman tree can be explained as follows. Nineteen leaf nodes exist, where each block has two subblocks. The number in the lower block indicates the number in a_3', and the number in the upper block is the corresponding occurrence frequency of the number. For example, the numbers in the first block in the 4th layer indicate 11 –0.014 in a_3'.

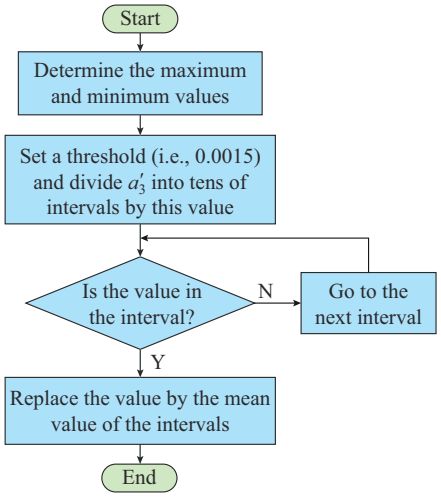

Fig. 5. Processing of approximation coefficients a_3.

The space occupied by the coefficients a_3' represents the sum of products of the frequency of each value and the number of bits of the encoded value, which is $1 \times 6 + 1 \times 6 + \ldots + 12 \times 3 + 13 \times 3 = 319$ bits. The space occupied by the original approximation coefficients a_3 is $86 \times 32 = 2752$ bits. Therefore, CR is $2752/319 = 8.63$. The original approximation coefficients a_3 are well compressed by Huffman coding after approximation.

In summary, the new approximation coefficients a_3' and new detail coefficients d_3', d_2', and d_1' occupy 319, 384, 256, and 192 bits, respectively. In addition, only two non-zero values exist at positions in the SSC, and they occupy $32 \times 2 + 16 \times 2 = 96$ bits. The original PQD signal has $640 \times 32 = 20480$ bits, and thus CR is $20480/(319 + 384 + 256 + 96) = 16.43$.

Finally, the impulse signal is recovered, and the PRD is 1.12% according to (10). The original impulse signal with 40 dB Gaussian noise, recovered signal, and error of impulse signal are shown in Fig. 6.

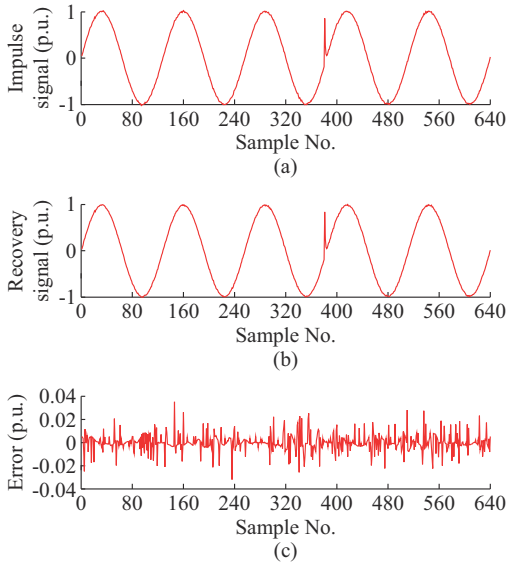

Fig. 6. Original impulse signal with 40 dB Gaussian noise, recovery signal, and error of impulse signal. (a) Original impulse signal with 40 dB Gaussian noise. (b) Recovery signal. (c) Error of impulse signal.

Considering the randomness of Gaussian noise, 10 Monte Carlo experiments are performed, where the PRD and CR values are listed in Table I.

TABLE I
COMPRESSION RESULTS OF IMPULSE SIGNALS BY DIFFERENT METHODS

Method	PRD (%)	CR
Proposed	1.19	18.86
SD	1.10	1.59
FFT + WT	1.20	5.93

Table I shows that CR is much improved under the proposed method as compared with that of the two competing methods. The CR declines significantly with the other methods due to presence of Gaussian noise in the PQD signal.

B. Decaying Harmonics

Time-varying harmonics typically derive from the arcing load in a steel plant, and the model of the harmonics under decaying amplitude is given by:

$$x(k)=\mathrm{e}^{-2\pi k/f_s}\{\sin(2\pi\cdot 50k/f_s)+0.5\sin(2\pi\cdot 150k/f_s+\pi/4)+ 0.3\sin(2\pi\cdot 250k/f_s+\pi/3)\}+n_{oise} \quad (15)$$

Because FF is 1, which suggests that no sudden change occurs in the signal, the TC and SSC can be separated by SD, as shown in Fig. 7.

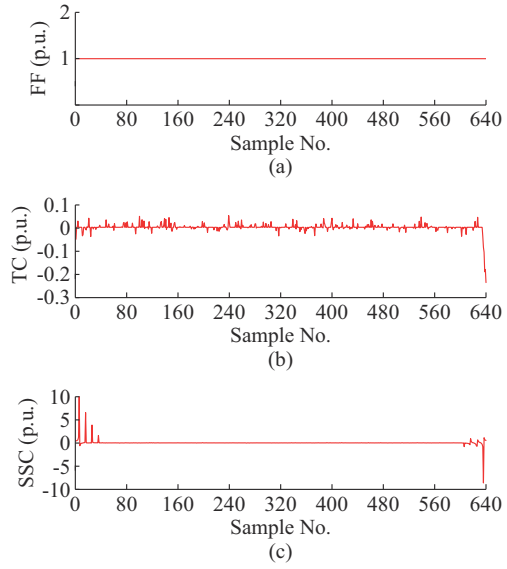

Fig. 7. FF, TC, and SSC results of decaying harmonics. (a) FF. (b) TC. (c) SSC.

The processing of the TC and SSC is identical to that described in Section III-A and is therefore not repeated here. The original decaying harmonic signal with 40 dB Gaussian noise, recovery signal, and error of decaying harmonic signal are shown in Fig. 8.

Ten experiments are conducted in this study to reduce the number of random errors, and the mean values of the 10 CRs and PRDs are 10.21 and 2.32%, respectively. The results are compared with those of SD and are listed in Table II.

The CR is much higher under the proposed method than that under the other two methods, whereas the PRD approximates that of SD, also indicating better performance by the proposed method. In addition, the CR of FFT + WT is higher than that of SD but lower than that of the proposed method because the detail coefficients d_i are compressed. Note, however, that the approximation coefficients a_3 are not compressed.

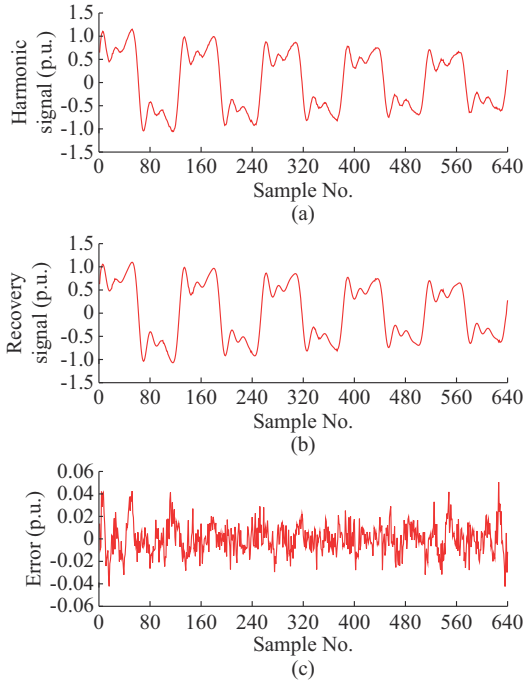

Fig. 8. Original decaying harmonic signal with 40 dB Gaussian noise, recovery signal, and error of decaying harmonic signal. (a) Original decaying harmonic signal with 40 dB Gaussian noise. (b) Recovery signal. (c) Error of decaying harmonic signal.

TABLE II
COMPRESSION RESULTS OF DECAYING HARMONIC SIGNALS BY DIFFERENT METHODS

Method	PRD (%)	CR
Proposed	2.32	10.21
SD	2.09	2.59
FFT + WT	2.20	3.74

C. Sag with Transient Oscillation

A sag with a transient oscillation signal, which is a frequent type of PQD signals in distribution networks, is caused by heavy loads that connect to the city power grid and is often accompanied by transient oscillation when the sag begins. A model of sag with oscillations can be described by:

$$x(k)=\begin{cases}0.5\sin(2\pi\cdot 50k/f_s)+ 0.8\mathrm{e}^{-0.4(k-403)}\sin(2\pi\cdot 50\cdot 18k/f_s) & 403\le k<473\\ 0.5\sin(2\pi\cdot 50k/f_s) & 473\le k<881\\ \sin(2\pi\cdot 50k/f_s) & \text{others}\end{cases} \quad (16)$$

Figure 9 shows that the FF of the sag signal is greater than 1, which suggests that two sudden changes occur in the waveform. Therefore, this waveform can be segmented into

three sections: the period before sagging, the sagging duration, and the period after sagging. The compression and reconstruction processing of these sectional waveforms is identical to that described in Section III-A. The three sectional waveforms are then connected, and the sag with the transient oscillation signal can be recovered. The waveforms of the original sag signal with 40 dB Gaussian noise, recovery signal (detail), and error of sag signal are shown in Fig. 10.

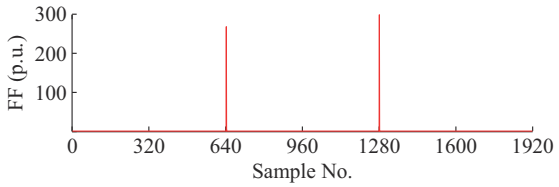

Fig. 9. FF of sag signal.

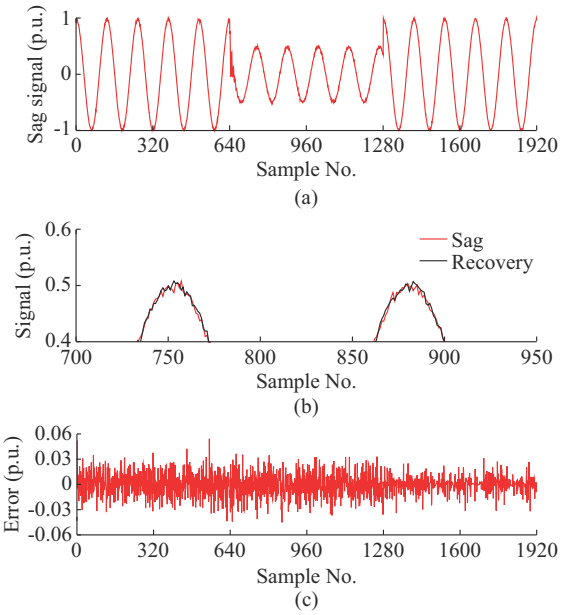

Fig. 10. Original sag signal with 40 dB Gaussian noise, recovery signal (detail), and error of sag signal. (a) Original sag signal with 40 dB Gaussian noise. (b) Recovery signal (detail). (c) Error of sag signal.

The PRD and CR results are compared with those of the SD and FFT + WT, which are listed in Table III. The CR is 27.29, which is much higher than that of the SD and FFT + WT methods, whereas the PRD approximates that of SD and FFT + WT, indicating better performance by the proposed method as compared with the two competing methods.

TABLE III
PRD AND CR RESULTS OF SAG SIGNALS WITH TRANSIENT OSCILLATION BY DIFFERENT METHODS

Method	PRD (%)	CR
Proposed	2.02	27.29
SD	1.95	2.05
FFT + WT	1.53	5.79

D. Effects of Sampling Frequency

As previously indicated, the sampling frequency is set to be 6.4 kHz, and the number of sampling points is 640. However, when high-frequency PQD signals are recorded, the sampling frequency may double to 12.8 kHz. If the sampling time remains at 0.1 s, the sampling points will increase to 1280, which requires more storage space. To analyze the effects of the sampling frequency in our evaluation of the proposed method, an indicator D_{cf} is used, which is given by:

$$D_{cf}=1000\left|\frac{CR_1-CR_2}{f_{s_1}-f_{s_2}}\right| \tag{17}$$

where f_{s_1} and f_{s_2} are 12.8 and 6.4 kHz, respectively; and CR_1 and CR_2 are the CRs under the two sampling frequencies f_{s_1} and f_{s_2}, respectively. A larger D_{cf} corresponds to a lower sensitivity to changes in f_s.

With the impulse signal taken as an example, let f_s and the total number of samples be 12.8 kHz and 1280, respectively. When the proposed method is applied to the impulse signal (α is set to be 2.3), the mean values of the CR and PRD are 29.08 and 1.20%, respectively. Thus, D_{cf} is 1.60 according to (17), which is higher than that of the competing methods, as shown in Table IV.

TABLE IV
CRS AND OVERALL EFFECTS OF DIFFERENT SAMPLING FREQUENCIES WITH DIFFERENT METHODS

Method	CR_1 (f_s=12.8 kHz)	CR_2 (f_s=6.4 kHz)	D_{cf}
Proposed	29.08	18.86	1.60
SD	7.31	5.05	0.35
FFT + WT	9.09	5.93	0.49

The results of D_{cf} reveal that the proposed method exhibits lower sensitivity to variations in f_s as compared with SD and FFT + WT. In addition, α is 2.3 and 2 when f_s is 12.8 and 6.4 kHz, respectively. The results thus also show that a slight increase in α offsets the effect of increasing f_s.

When f_s is 3.2 kHz, the CR is 23.86 as compared with 4.68 and 5.02 of SD and FFT+WT, respectively. Thus, whether f_s increases or decreases, the CR of the proposed method is higher than those of the other methods.

E. Effects of Noise

The experimental results show that the proposed method performs well when the SNR of the PQD signals is 40 dB. However, whether the method is immune to noise must be further explored.

Three types of PQD signals are imposed under 30, 40, and 50 dB SNRs. The signals are then compressed and recovered. The CR and PRD results under different SNRs are listed in Table V.

Table V indicates that the CR tends to increase, whereas PRD tends to decrease with an increase in the SNR. This can be explained as follows. When the SNR is 50 dB, the noise amplitude is relatively low, indicating that more signal details are retained during compression, and therefore CR increases while PRD decreases. However, when the SNR is 30 dB, the noise amplitude is large, indicating that more signal details of the signal are discarded during compression, and

therefore the changes in CR and PRD show inverse trends.

TABLE V
CR AND PRD RESULTS UNDER DIFFERENT SNRS

Signal type	SNR (dB)	CR	PRD (%)
Impulse signal	30	17.68	3.76
	40	18.24	1.19
	50	18.50	0.38
Sag with transient oscillation	30	22.73	3.31
	40	23.48	2.08
	50	41.21	1.83
Decaying Harmonic	30	11.56	5.47
	40	12.13	3.11
	50	12.47	2.66

F. Computational Efficiency

The computational efficiency of methods is critical. A method that is too complex results in an excessively long operational time or requires considerable storage space, and thus the method cannot be applied in practice. Therefore, the computational time cost of the proposed method is compared with that of the other two competing methods. For the comparison study, a computer is configured with 8 GB of memory and an Intel i5 2.8 GHz CPU.

The time costs of different methods are shown in Table VI. The proposed method proves to be the most time-consuming compression method. This can be explained by the fact that, in the proposed method, the coding methods (i.e., run-length and Huffman encodings) require more operational time but achieve higher CR and approximately equal errors. The added time required to achieve a higher CR is thus considered a reasonable trade-off.

TABLE VI
TIME COSTS OF DIFFERENT METHODS

Method	Time cost (s)
Proposed	5.27
SD	4.84
FFT + WT	4.76

IV. EXPERIMENT ON REAL PQD SIGNALS

To verify the feasibility of the proposed method on real-life signals, a set of PQD signals is recorded from the Dynamic Power System Simulation Laboratory at Huazhong University of Science and Technology, Wuhan, China.

The f_s of the signal is 5000 Hz, and the duration is 0.12 s, meaning that each phase signal has 600 sampling points. Because the sag is the most frequent PQD signal in a real power system, we create a voltage sag in the laboratory and use this signal to test the performance of the proposed method on real-life signals. The real signal measurement system and recorded three-phase fault waveforms are shown in Figs. 11 and 12, respectively.

The 210 V three-phase voltage signals are generated by an analog generator. They are transmitted through the transmission line, transformed to 100 V by the analog transformer, and then transmitted to the recorded signals through communication devices. Finally, the waveforms of phases A, B, and C are saved by fault recorder.

Fig. 11. Real signal measurement system.

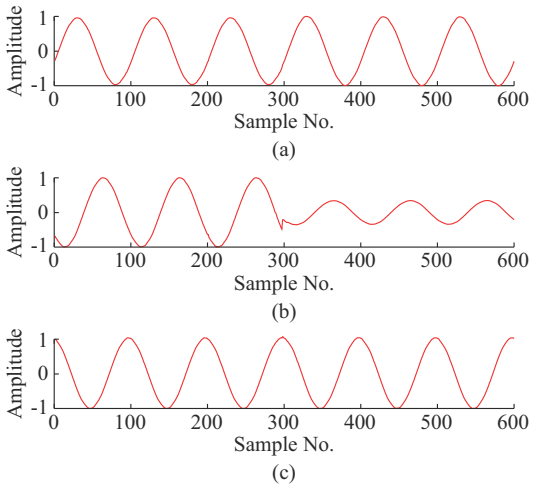

Fig. 12. Recorded fault waveforms of phases A, B, and C. (a) Phase A. (b) Phase B. (c) Phase C.

As Fig. 12 shows, the fault is set as a line-to-ground short circuit in phase B. Therefore, the recorded waveforms of phases A and C are ideal sine waves; a voltage sag occurs in phase B due to the line-to-ground short circuit. Note that the synthetic signals do not perfectly match the real-life signals, and therefore the absolute values of less than 0.005 are set to be zero to improve the CR.

The fault signals of phases A, B, and C are compressed and recovered. The CRs and PRDs of the proposed and other methods are listed in Table VII.

Table VII shows that the CR of the three-phase real signals with the proposed method is approximately 50, which is

much higher than that of SD and FFT + WT. The differences in PRD of the three methods are minimal.

TABLE VII
CRs AND PRDs OF DIFFERENT METHODS

Phase	Proposed		SD		FFT + WT	
	CR	PRD (%)	CR	PRD (%)	CR	PRD (%)
A	52.74	1.61	6.25	1.84	4.68	1.86
B	46.07	2.37	2.04	1.98	2.73	2.52
C	51.34	1.82	4.68	1.84	5.45	1.83

It should be noted that the CR of the phase B signal decreases slightly, whereas the PRD increases slightly, which explains why the error between the original and recovery signals is relatively large near the time of sag occurrence.

V. CONCLUSION

This study proposes a compression algorithm for PQD signals using an STKF, SD, and Huffman and run-length codings. The TC and SSC in the PQD signal are first separated by SD based on an overcomplete dictionary consisting of a unity matrix and a unity Hartley transform dictionary. Next, the TC is compressed by wavelet analysis, which is followed by compression of the approximation and detail coefficients by Huffman and run-length codings, respectively. The small values in the frequency domain of the SSC are then set to be zero, and the PQD signal compression is completed. Experimental results of synthetic and real-life signals verify that the proposed method can achieve the high CR and the low percentage of the root-mean-square difference. In addition, the proposed method is shown to be immune to the interference of variant sampling frequency and Gaussian noise, and the analytical results demonstrate its high robustness.

APPENDIX A

The Huffman tree of the new approximation coefficients a_3' is shown in Fig. A1.

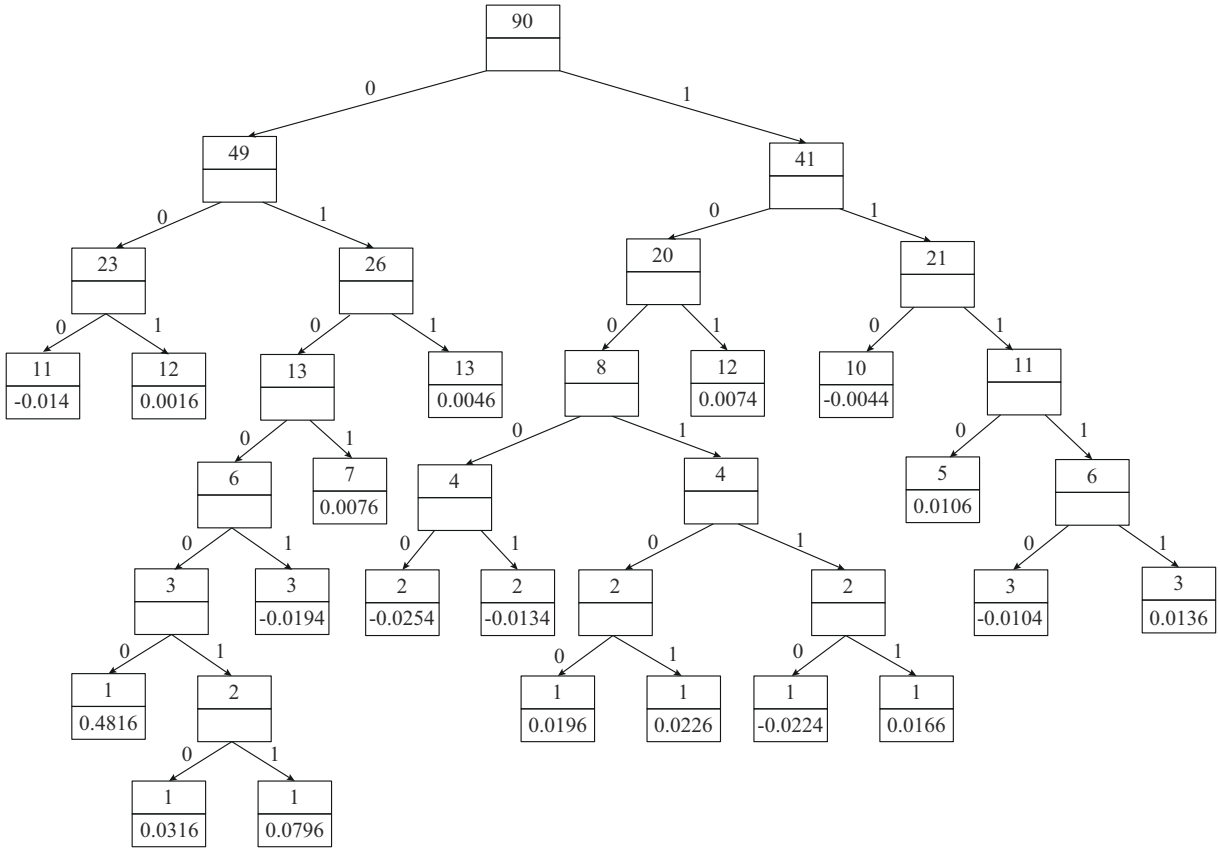

Fig. A1. Huffman tree of new approximation coefficients a_3'.

REFERENCES

[1] P. S. Pavan and S. Das, "Novel method for location of internal faults in ungrounded double wye shunt capacitor banks," *IEEE Transactions on Power Delivery*, vol. 36, no. 2, pp. 899-908, Apr. 2021.
[2] W. Wu, P. Li, B. Wang *et al*., "Integrated distribution management system: architecture, functions, and application in China," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 245-258, Mar. 2022.
[3] J. Xu, Y. Hu, H. Qian *et al*., "Delay-based phase-locked loop parameters design based on stability region of grid-connected single-phase inverter under grid voltage sags," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 11, pp. 11324-11334, Nov. 2022.
[4] *IEEE Guide for Direct Lightning Stroke Shielding of Substations*, IEEE Std 998-2012, 2013.
[5] D. Li and A. Ukil, "Fault location estimation in voltage source converter based DC system: the L location," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 11, pp. 11198-11209, Nov. 2021.
[6] M. V. Ribeiro, S. H. Park, J. M. T. Romano *et al*., "A novel MDL-based compression method for power quality applications," *IEEE*

Transactions on Power Delivery, vol. 22, no. 1, pp. 27-36, Jan. 2007.
[7] M. P. Tcheou and L. Lovisolo, "The compression of electric signal waveforms for smart grid: state of the art and future trend," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 291-304, Jan. 2014.
[8] C. Chen, W. Wang, H. Yin *et al*., "Real-time lossless compression for ultrahigh-density synchrophasor and point-on-wave data," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 2012-2021, Feb. 2022.
[9] S. Shukla, M. Ahirwar, R. Gupta *et al*., "Audio compression algorithm using discrete cosine transform (DCT) and Lempel-Ziv-Welch (LZW) encoding method," in *Proceedings of 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*, Faridabad, India, Feb. 2019, pp. 476-480.
[10] S. Santoso, E. J. Powers, W. M. Grady. "Power quality disturbance data compression using wavelet transform methods," *IEEE Transactions on Power Delivery*, vol. 12, no. 3, pp. 1250-1257, Jul. 1997.
[11] T. Gao, L. Jiang, X. Wang *et al*., "Acoustic signal compression research based on wavelet and compressed sensing," in *Proceedings of 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, Shenyang, China, Jan. 2021, pp. 777-782.
[12] J. Khan, S. M. A. Bhuiyan, G. Murphy *et al*., "Embedded-Zerotree-Wavelet-based data denoising and compression for smart grid," *IEEE Transactions on Industry Applications*, vol. 51, no. 5, pp. 4190-4200, Oct. 2015.
[13] Z. Li, C. Zheng, S. Ma. *et al*., "An improved method based on wavelet for power quality compression," in *Proceedings of 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, Hefei, China, Jun. 2016, pp. 1750-1753.
[14] O. N. Gerek and D. G. Ece, "Compression of power quality event data using 2D representation," *Electrical Power Systems Research*, vol. 78, no. 6, pp. 1047-1052, Jun. 2008.
[15] O. N. Gerek and D. G. Ece, "2-D analysis and compression of power quality event data," *IEEE Transactions on Power Delivery*, vol. 19, no. 2, pp. 791-798, Apr. 2004.
[16] D. G. Ece and O. N. Gerek, "Power quality event detection using joint 2-D-wavelet subspaces," *IEEE Transactions on Power Delivery*, vol. 53, no. 4, pp. 1040-1046, Aug. 2004.
[17] J. C. S. de Souza, T. M. L. Assis, and B. Pal, "Data compression in smart distribution systems via singular value decomposition," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 275-284, Jan. 2017.
[18] N. Hashemipour, J Aghaei, A Kavousi-Fard *et al*., "Optimal singular value decomposition based big data compression approach in smart grids," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 3296-3305, Jul.-Aug. 2021.
[19] M. Zhang, K. Li, and Y. Hu, "A high efficient compression method for power quality applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 6, pp. 1976-1985, Jul. 2011.
[20] S. He, M. Zhang, W. Tian *et al*., "A parameterization power data compress using strong trace filter and dynamics," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 10, pp. 2636-2645, Oct. 2015.
[21] M. S. Manikandan, S. R. Samantary, and I. Kamwa, "Detection and classification of power quality disturbance using sparse signal decomposition on hybrid dictionaries," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 1, pp. 27-38, Jan. 2015.
[22] M. S. Manikandan, S. R. Samantary, and I. Kamwa, "Simultaneous denoising and compression of power system disturbances using sparse representation on over hybrid dictionaries," *IET Generation, Transmission & Distribution*, vol. 9, no. 11, pp. 1077-1088, Aug. 2015.
[23] S. He, W. Tian, J. Zhang *et al*., "A high efficient approach for power quality waveform compression in the view of Heisenberg uncertainty," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2580-2591, May 2019.
[24] D. Cai, K. Li, S. He *et al*., "On the application of joint-domain dictionary mapping for power disturbance assessment," *Energies*, vol. 11, no. 2, pp. 1-17, Jan. 2018.
[25] J. Tao, C. Qi, and Y. Huang, "Regularized multipath matching pursuit for sparse channel estimation in millimeter wave massive MIMO system," *IEEE Wireless Communications*, vol. 8, no. 1, pp. 169-172, Feb. 2019.
[26] S. Zhang, X. Ma, Y. Wang *et al*., "Robust reconstruction of fluorescence molecular tomography based on sparsity adaptive correntropy matching pursuit method for stem cell distribution," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2176-2184, Oct. 2018.
[27] L. Kong, Y. An, Q. Liang *et al*., "Reconstruction for fluorescence molecular tomography via adaptive group orthogonal matching pursuit," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 9, pp. 2518-2529, Sept. 2020.
[28] Q. Ge, T. Shao, S. Chen *et al*., "Carrier tracking estimation analysis by using the extended strong tracking filtering," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1415-1424, Feb. 2017.

Xiangui Xiao received the B.E. degree from Wuhan University of Technology, Wuhan, China, in 2011, and the M.E. degree from Shanghai University of Electrical Power, Shanghai, China, in 2016. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology, Wuhan, China. His research interests include signal processing and deep learning in power systems.

Kaicheng Li received the B.Sc., M.Sc., and Ph.D. degrees all in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1983, 1988, and 1998, respectively. He is currently a Professor at the School of Electrical and Electronic Engineering, Huazhong University of Science and Technology. His research interest includes power quality analysis.

Chen Zhao received the B.E. degree in electrical engineering and the M.E. degree in agricultural mechanization engineering from Fujian Agriculture and Forest University, Fuzhou, China, in 2004 and 2008, respectively. He received the Ph.D. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently a Lecturer at the College of Mechanical and Electrical Engineering, Fujian Agriculture and Forest University. His research interests include power quality analysis and control, modern signal processing, and machine learning.