

PAPER

A signal analysis and identification scheme for an online multiphase micron-sized particle analyzer system

To cite this article: Fuhai Wang *et al* 2021 *Meas. Sci. Technol.* **32** 085111

View the [article online](#) for updates and enhancements.

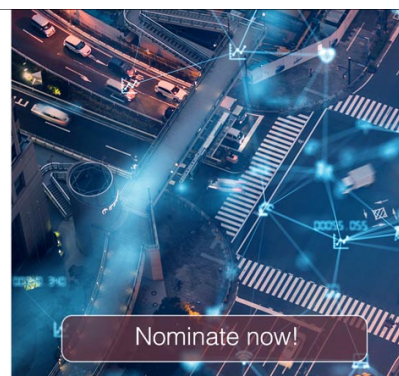


The Electrochemical Society
Advancing solid state & electrochemical science & technology

The ECS is seeking candidates to serve as the
Founding Editor-in-Chief (EIC) of ECS Sensors Plus,
a journal in the process of being launched in 2021

The goal of ECS Sensors Plus, as a one-stop shop journal for sensors, is to advance the fundamental science and understanding of sensors and detection technologies for efficient monitoring and control of industrial processes and the environment, and improving quality of life and human health.

Nomination submission begins: May 18, 2021



A signal analysis and identification scheme for an online multiphase micron-sized particle analyzer system

Fuhai Wang¹ , Hongjian Cao^{1,2,3}, Pingping Wang^{1,2}, Xiaokang Guo¹, Jinlong Han³, Haifeng Dong⁴, Xiangping Zhang⁴ and Xiaodong Wang¹

¹ Center of Materials Science and Optoelectronics Engineering, College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

² Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

³ Guang Wei (Zhongshan) Intelligent Technology Co, Zhongshan 528437, People's Republic of China

⁴ Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

E-mail: xiaodong.wang@ucas.ac.cn

Received 29 October 2020, revised 30 January 2021

Accepted for publication 17 February 2021

Published 25 May 2021



CrossMark

Abstract

Online microparticle detection is of utmost importance for industrial production. This paper proposes a signal processing and feature identification strategy to achieve particle size statistics for online measurement in a three-phase stirred tank reactor based on the electrical sensing zone (ESZ) method. Signal denoising and de-interference are achieved using the wavelet soft threshold method combined with mathematical morphological filtering. Pulse selection is implemented using pulse width limiting conditions. The key features that distinguish the pulse waveforms are defined based on the differences in the motion characteristics of the different types of particles through the aperture. Finally, the unsupervised classification algorithm balanced iterative reducing and clustering using hierarchies clustering is employed to distinguish the pulsed features between hard particles and bubbles. The results show that the particle size distribution identified by this strategy agrees with offline measurements indicating the effectiveness of the scheme. The effects of electromagnetic noise and the interference of small bubbles that approximate the particle size in solution in the online, *in-situ* measurement task are solved. This study scheme has a guiding and facilitating role in applying the ESZ principle to the industrial online measurement environment.

Keywords: electrical sensing zone, multiphase microparticle detection, wavelet threshold denoising, morphological filtering, feature recognition, hierarchical clustering

(Some figures may appear in colour only in the online journal)

1. Introduction

In industrial applications, the online detection of the size and spatial distribution of microparticles in suspensions is of paramount importance. For example, solids suspension properties are examined to study the optimum parameters for three-phase reactors in process industries such as various

hydrogenations and oxidations, ammonolysis, evaporation crystallization, mineral processing, and slurry preparation [1–4]. In bioprocesses, the condition of cells in biological reactors is usually tested to study the potential fluid-mechanical cell damage, as well as the cell growth rate, and maximum cell density [5–8]. In addition, microparticle testing is also often used to measure the condition of water quality

[9, 10]. The online detection of microparticles can facilitate the understanding of the monitoring of non-uniform characteristics [11] in an industrial application, as well as the feedback, adjustment, and optimization of processes based on particle size variations.

In recent decades, the electrical sensing zone (ESZ) method has been widely used, as one of the offline particle size measurement techniques, in chemical sciences [12], life sciences [13], environmental applications [14], pharmaceutical sciences [15] and other fields due to its good resolution [16] and other characteristics. To obtain a high signal-to-noise ratio, high particle size resolution, and stable signals, extensive research studies have been carried out on microporous design [17] and simulation, flow channel structure design [18], signal processing [19] and identification algorithms.

Initially, researchers used hardware filtering for signal denoising under laboratory conditions [20–22]. The ESZ signals are non-stationary [23], and traditional analog and digital filtering methods often result in an obvious distortion of the signal waveform [24–26]. Therefore, denoising and detrending should avoid causing distortions in the pulse waveform [27]. Most researchers then set a single fixed threshold [28] to perform the division between the pulse amplitude and noise to achieve peak localization. This approach is highly demanding for upfront detrending and denoising. To study the detection of different kinds of particles, researchers have conducted considerable numerical simulation work around the phenomena of the particle density, shape, size, and off-axis of particles passing through the micropore [29–31]. Li *et al* suggested by using numerical simulations the differences in the motion of particles of different densities and sizes passing through a small aperture [31–33]. Our team [34, 35] carried out numerical studies on the particle properties and the transient deformation of deformable particles such as bubbles. The findings of the above studies have guided the recognition of particle pulses in reality. Researchers have extracted typical pulse features such as the amplitude, pulse width, and slope [36, 37] and used models such as support vector machines for classification [38]. In recent years, Honrado *et al* [39] used self-designed microfluidic impedance chips and relied on deep learning to extract particle information from the pulsed signals. Machine learning and deep learning models [20, 40, 41] provide more an additional means of exploration for the characterization of pulsed signals in the ESZ.

However, there are challenges with respect to signal processing and waveform identification with the direct online measurement method. These challenges are related to electromagnetic noise in the measurement environment, which can weaken or even cause errors in signal recognition, and doped bubbles that approximate the size of hard particles can also create pseudo-pulses through the aperture, all of which can lead to deviations in the particle statistics. To date, although the online ESZ method has been developed [42], in-depth research on signal processing and identification algorithms are still not sufficient in industrial processes.

To address the challenges of online measurements, this paper proposes an optimization strategy and feature identification scheme for signal processing. This scheme focuses

on solving the denoising and de-interference problems of the online ESZ method. The multiphase microparticle pulse identification and classification problem are highlighted. An optimization scheme is proposed to remove high-frequency noise using wavelet multi-resolution analysis and the wavelet soft threshold method, and to remove low-frequency interferences, such as the signal baseline drift and power frequency interference, using morphological filtering. The key features of the pulse waveform are extracted based on the motion characteristics of particles with different characteristics through the small aperture, and a dataset of the pulse waveform between the bubbles and hard particles is constructed. The balanced iterative reducing and clustering using hierarchies (BIRCH) clustering algorithm is proposed to conduct the classification of pulse features between bubbles and hard particles. The statistics of hard particles in multiphase microparticles are finally completed effectively.

The remainder of the paper is organized as follows. In section 2, the framework for the signal detection theories of the ESZ method is introduced. In section 3, the measurement device and environmental setup are described. In section 4, the principles and results of online signal denoising and de-interference are analyzed. Section 5 presents a discussion of the feature extraction and classification methods for hard particles and bubble waveforms. Finally, a conclusion is presented in section 6.

2. The signal detection theories of the ESZ method

2.1. Basic principles of the ESZ method

The ESZ method, also known as the Coulter principle [43, 44], is an effective method for detecting the number and size of microparticles in a suspension, as illustrated in figure 1.

The impedance changes collected using the ESZ method are recorded by the current or voltage. From Maxwell's equations, the amount of voltage change can be simplified and expressed as follows [44] when the particle diameter is much smaller than the length of the small aperture

$$\Delta V = 4\rho_e Id^3 / \pi D^4 \quad (1)$$

where ρ_e is the resistivity of the liquid, I is the current flowing through the small aperture, d is the nominal size of the microparticle and D is the aperture diameter.

The characteristics of the pulses can reflect the particle characteristics [45], flow rate [39], and channel shape [46]. Furthermore, the statistical characteristics of the pulses can reflect information such as the concentration of particles in the solution [47].

2.2. Traditional signal processing solutions for the ESZ method

The identification of the pulse waveform signal is an effective means to discriminate particles with different physical properties (size, deformation, conductivity, density, etc) in the solution. The key steps of signal processing by the ESZ method are

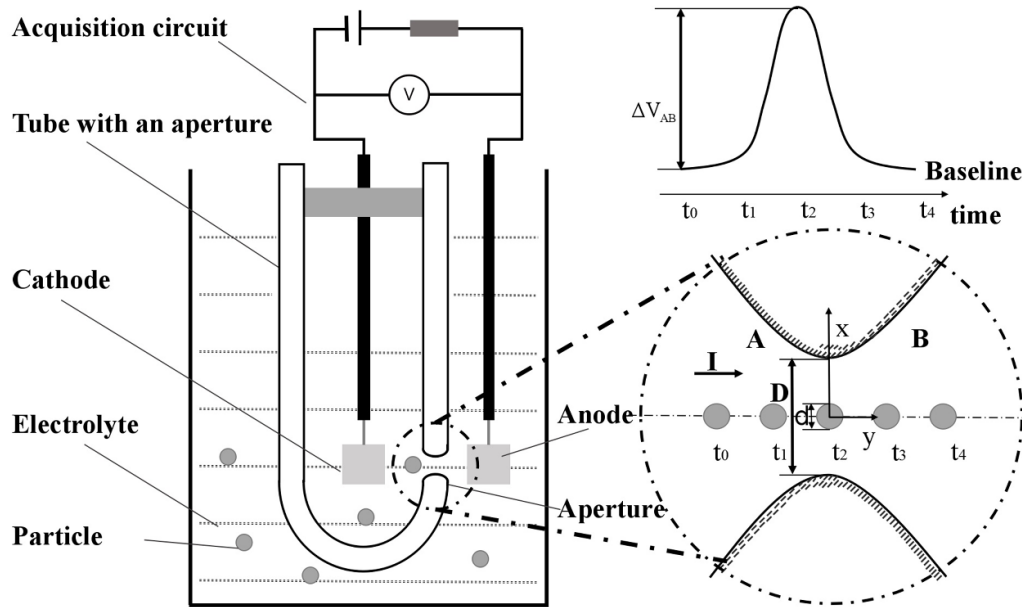


Figure 1. Schematic diagram of the ESZ method. The particle passing through the micropore will create a change in the resistance, which responds to the voltage signal, as shown in the upper right corner of the figure.

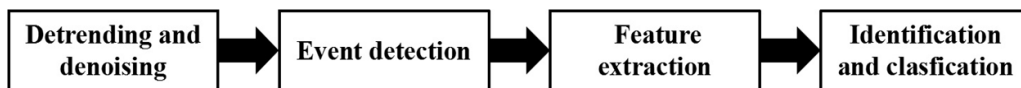


Figure 2. The signal processing flow for the ESZ.

shown in figure 2 and can be divided into four steps: detrending and denoising, event detection, feature extraction, and identification and classification.

2.3. Optimal signal processing identification scheme for the ESZ method

For the classification of denoising and the pulse waveform recognition of signals in the online ESZ method, the overall denoise and recognition design solution flow is shown in figure 3.

3. Materials and methods

In order to simulate a detection scenario in a multi-phase environment, the signal acquisition of microparticles was performed in a chemical three-phase stirred tank reactor using online micron-sized particle analyzer developed by ourselves. The processing of the signal data, including denoising, pulse waveform capture, feature extraction, and data set construction, was implemented using MATLAB. The algorithms for pulse identification and particle size statistics were done using Python. All computations were run on an Intel Core i7-9700F CPU@3.00 GHz processor with 16 GB RAM.

The concentration of the solution was 0.9% saline, and the test tube micropore diameter was 200 μm . The platinum wire electrodes were placed on both sides of the micropore.

The experiments were performed using a nano-air disc to simulate the bubbles present during gas-phase agitation. Polystyrene standard particles of Class α (nominal size 57.9 μm , CV = 4.15%) and Class β (nominal size 80 μm , CV = 5.54%) were tested. The experimental setup is shown in figure 4. The following three types of data were collected at a sampling frequency of 200 kHz.

- Class A data are for the stirred tank operating in the liquid–solid phase with both Class α and Class β particles.
- Class B data are for the stirred tank during gas–liquid two-phase operation with only bubbles.
- Class C data are for the stirred tank operating in a three-phase system and containing both Class α particles, Class β particles and bubbles.

4. Signal denoising

According to the characteristics of the acquired signal, noise can be divided into three categories. The first is Gaussian white noise, the second is the signal base drift caused by solution fluctuations, and the third is the power frequency and its harmonic interference. The amplitude of the signal noise fluctuation and interference collected by the ESZ is comparable to the amplitude of particles with diameters from 30 μm to 50 μm , which causes the worst situation with respect to the reliability of identification counting.

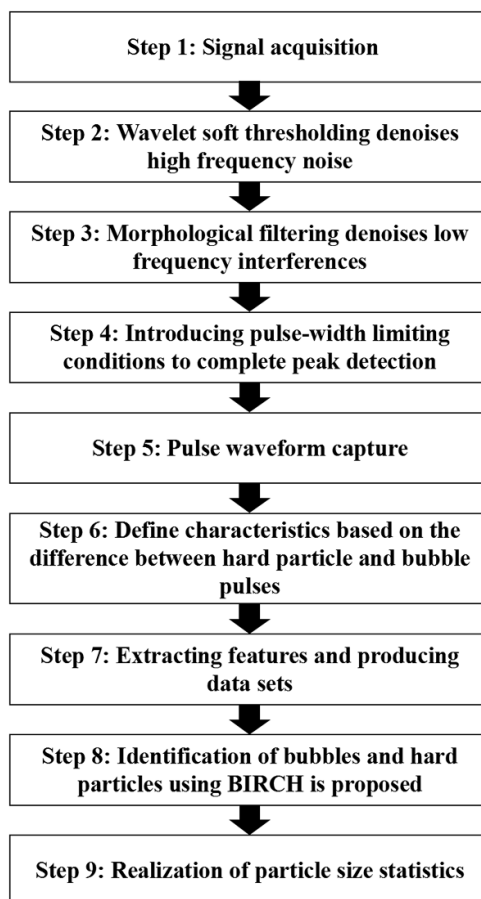


Figure 3. Overall flow chart of denoising recognition.

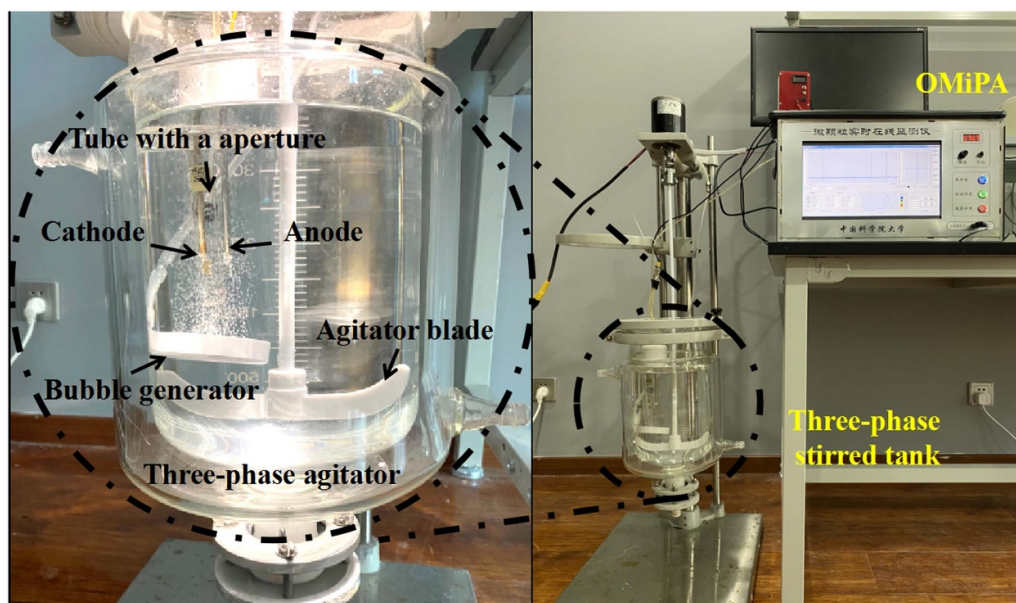


Figure 4. Experimental setup.

Therefore, the scheme is first pre-processed to remove high-frequency noise using the wavelet soft threshold method, and then morphological filtering is used to remove low-frequency

interferences, such as signal base fluctuations and power frequency interference. The advantage of this scheme is that the true waveform of the pulse is maintained maximally, which

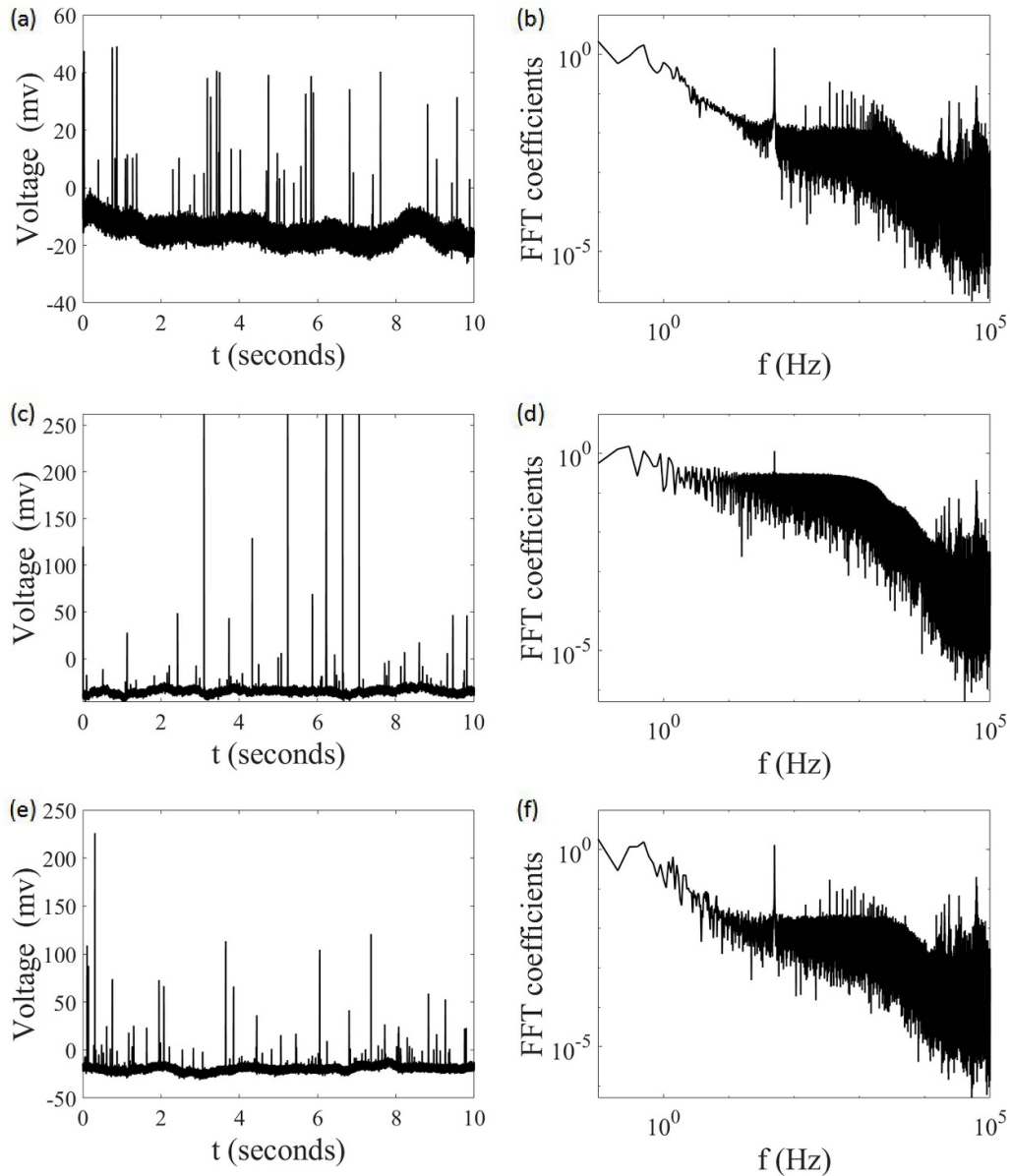


Figure 5. Raw signals of categories A, B, and C: (a), (c) and (e) time-domain signals of categories A, B, and C and (b), (d) and (f) amplitude–frequency characteristics of categories A, B, and C. From the frequency domain, with low-frequency interference, such as basal fluctuations, 50 Hz power frequency and high harmonics, and high-frequency noise interference.

does not cause errors for subsequent feature extraction and identification.

4.1. Raw signal analysis

The collected raw signal can be expressed as $y[n]$:

$$y[n] = x[n] + f[n] + p[n] + \varepsilon[n], 0 < n < N - 1 \quad (2)$$

where $x[n]$ is the useful signal containing pulses, $f[n]$ is the base fluctuation signal, $p[n]$ is the power frequency and harmonic interference, and $\varepsilon[n]$ is the Gaussian white noise in the environment. The time-domain signal and frequency domain characteristics of the three types of data A, B, and C, are given in figure 5.

4.2. Wavelet soft threshold method to remove high-frequency noise

At present, modern digital signal processing methods have been effectively applied in the signal denoising of pulses in the ESZ method, such as the treatment of high-frequency noise in a signal using adaptive filtering [48], Hilbert–Huang transform denoising [49], and wavelet [19, 27] denoising. Considering challenge task of the present online measurement mode, taking the advantages of wavelet in non-stationary signal analysis, wavelet soft threshold denoising is introduced to pre-process the high-frequency denoising of the online ESZ method, while also maintaining the true waveform of the pulse.

In the wavelet transform domain, the signal can be decomposed into an approximate sub-band of the highest layer and a detailed partial representation of each decomposed

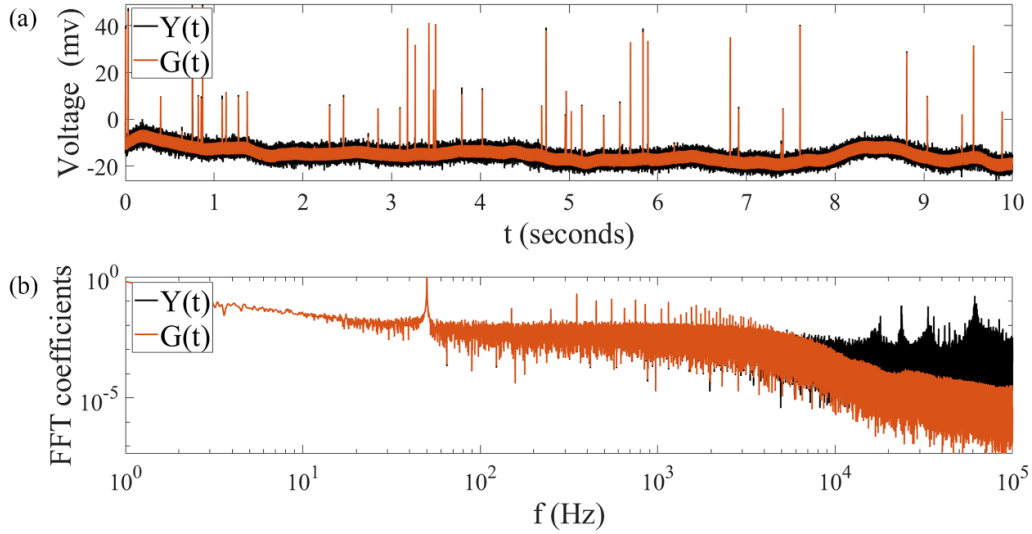


Figure 6. Time-frequency domain of signal $Y(t)$ and signal $G(t)$ after wavelet threshold noise reduction: (a) raw signal $Y(t)$ and reduced time-domain signal $G(t)$ in data A and (b) frequency domain features corresponding to $Y(t)$ and $G(t)$.

layer. To remove noise, the detail sub-bands can be treated by setting the thresholds. The steps of wavelet thresholding for noise removal can be divided into wavelet decomposition, threshold function construction, threshold selection, and wavelet reconstruction. In particular, the threshold function selection and threshold estimation steps play a key role in denoising. Threshold selection often uses universal threshold (Visushrink thresholds) [27].

The number of detail component layers after the wavelet decomposition is reasonably selected to denoise the soft threshold method with the mode-squared threshold function. The mode-squared thresholding function is formulated as

$$\hat{w}_{j,k} = \begin{cases} 0, & |w_{j,k}| \leq \lambda_j \\ \text{sgn}(w_{j,k}) \cdot \sqrt{w_{j,k}^2 - \lambda_j^2}, & |w_{j,k}| > \lambda_j \end{cases} \quad (3)$$

where λ_j is the universal threshold for the layer j wavelet, $\hat{w}_{j,k}$ is the wavelet coefficient after threshold processing, $w_{j,k}$ is the original wavelet coefficient, sgn is the symbolic function, and the expression $\text{sgn} = \begin{cases} 1, & w_{j,k} \geq 0 \\ -1, & w_{j,k} < 0. \end{cases}$

In this article, we use the db10 wavelet basis function to implement a five-level wavelet decomposition. The signal $G(t)$ is obtained by denoising the first four layers of the detail components by the mode-squared threshold method. For processing a data stream collected in 10 s, wavelet denoising takes about 0.5 s. Figure 6 shows the signal after wavelet denoising. The signal-to-noise ratio is significantly improved after wavelet soft threshold denoising, which effectively suppresses the high-frequency noise interference and ensures the integrity of the pulse waveform without distortion.

4.3. Morphological filtering to remove low-frequency interference such as base fluctuations and power frequencies

For online measurement mode, we must solve disturbances and strong external electromagnetic interference, and the signal can suffer from basal drift or fluctuations due to the power frequency and harmonic interferences. To reduce the effect of signal base fluctuations on pulse recognition, researchers have proposed methods such as smooth fitting [50], high-pass filtering [38], and wavelet de-baselining [27]. In this section, we introduce a mathematical morphology filter method to effectively remove a variety of low-frequency interferences, such as base drift and power frequency interference, while keeping the waveform characteristics of the pulse realistic.

Morphological filtering is a nonlinear signal processing method that is first applied in two-dimensional (2D) image processing. The basic idea of mathematical morphology in one-dimensional (1D) signal processing is to view a discrete section of the signal as a collection of data and then perform a special kind of ensemble operations with the set of structural elements to obtain the signal components of interest [51] finally. These special set operations are called morphological operations. Let $f(n)$ be the discrete 1D signal of the definition domain $F = \{0, 1, \dots, N-1\}$ and $g(n)$ be the structural element on the definition domain $G = \{0, 1, \dots, M-1\}$. The basic morphological operators [52] include erosion, dilation, opening, and closing.

A flexible combination of the opening and closing operations can attenuate positive or negative shocks in the measurement signal. By setting the size and shape of the structural element $g(n)$, the low-frequency fluctuations and interference can be removed. Since the pulse waveform in the ESZ is a positive shock, in order to obtain a complete forward shock waveform, the open method is used for subsequent low-frequency interference removal.

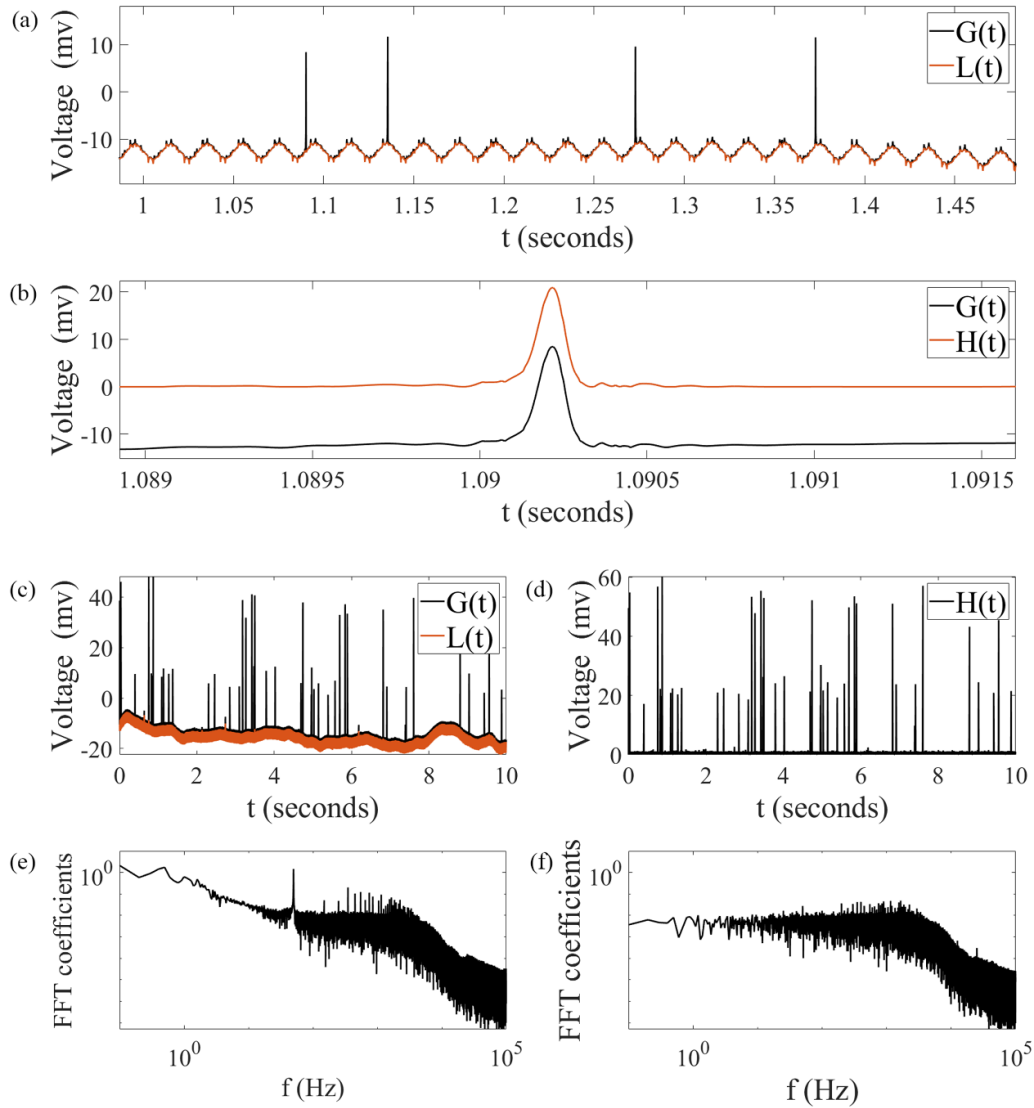


Figure 7. (a) Comparison of the details between $G(t)$ and the low-frequency interference $L(t)$ and (b) comparison of the baseline changes in $G(t)$ and $H(t)$; (c), (d) comparison of $G(t)$ and $H(t)$ after removing low-frequency interference and (e), (f) comparison of amplitude frequency characteristics of $G(t)$ and $H(t)$.

After the morphological filter operation is determined, the selected structural element is the main factor that affects the filter output. $g(n)$ is selected as the linear structure element for filtering. The width L of $g(n)$ is set to 100 according to the pulse width, which is within 200 sampling points.

After the open operation, the low-frequency interference signal $L(t)$ is obtained, and the low-frequency interference $L(t)$ is subtracted from the signal $G(t)$ to obtain the signal $H(t)$. $H(t)$ effectively eliminates low-frequency interference and ensures that the signal reference line is zero. Additionally, the nonlinear operation of $H(t)$ also maintains the true pulse waveform. Morphological filtering takes an average of 1.5 s to process a 10 s stream of data. The results of the signal processing after morphological filtering are shown in figure 7. The baseline of $H(t)$ obtained by morphological filtering becomes zero, and the baseline is smoother while maintaining the true shape of the waveform to the maximum extent.

4.4. Peak positioning and pulse interception

The peak search algorithm is first used to achieve the localization of all peaks in the signal, and peak selection is performed according to the dual-threshold method. Simultaneously, the criterion that the distance between two adjacent peak points is not less than one pulse width is introduced, which effectively eliminates the interference of spurious and overlapping peaks [20, 53].

For Class A data, the solution contains only two particles of Class α and Class β . By setting the upper and lower thresholds to 120 mv and 12 mv, respectively, the peak points of the signal pulses for Class α and Class β are searched. Since Class B and Class C contain bubbles, all pulses of the bubbles and particles with an amplitude in the range of 12–120 mv are selected by the upper and lower thresholds, respectively. This process takes a very small amount of time around 0.15 s. Later on, we use the feature recognition method to classify and remove

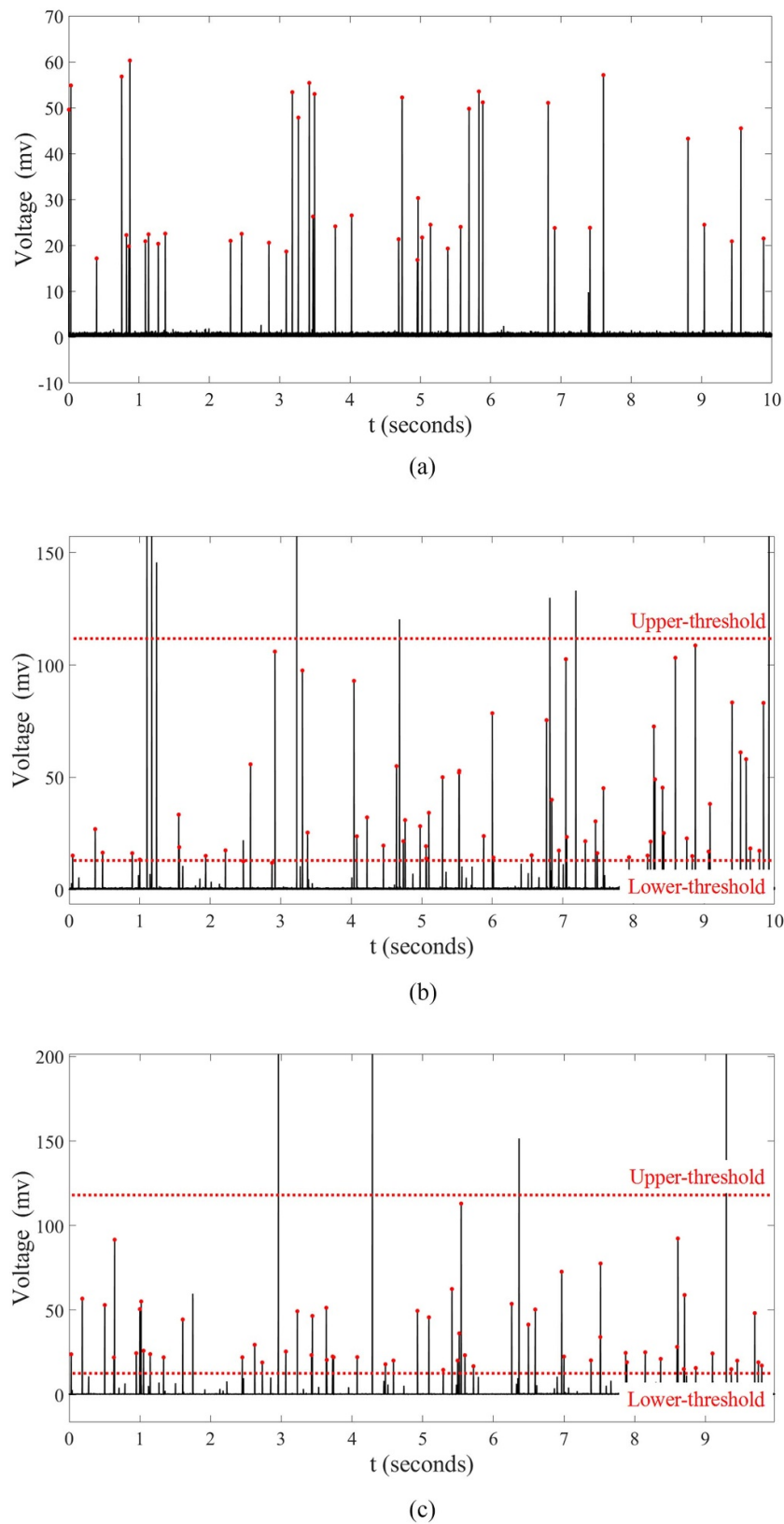


Figure 8. (a)–(c) Peak positioning of the pulses in data A, B and C, respectively.

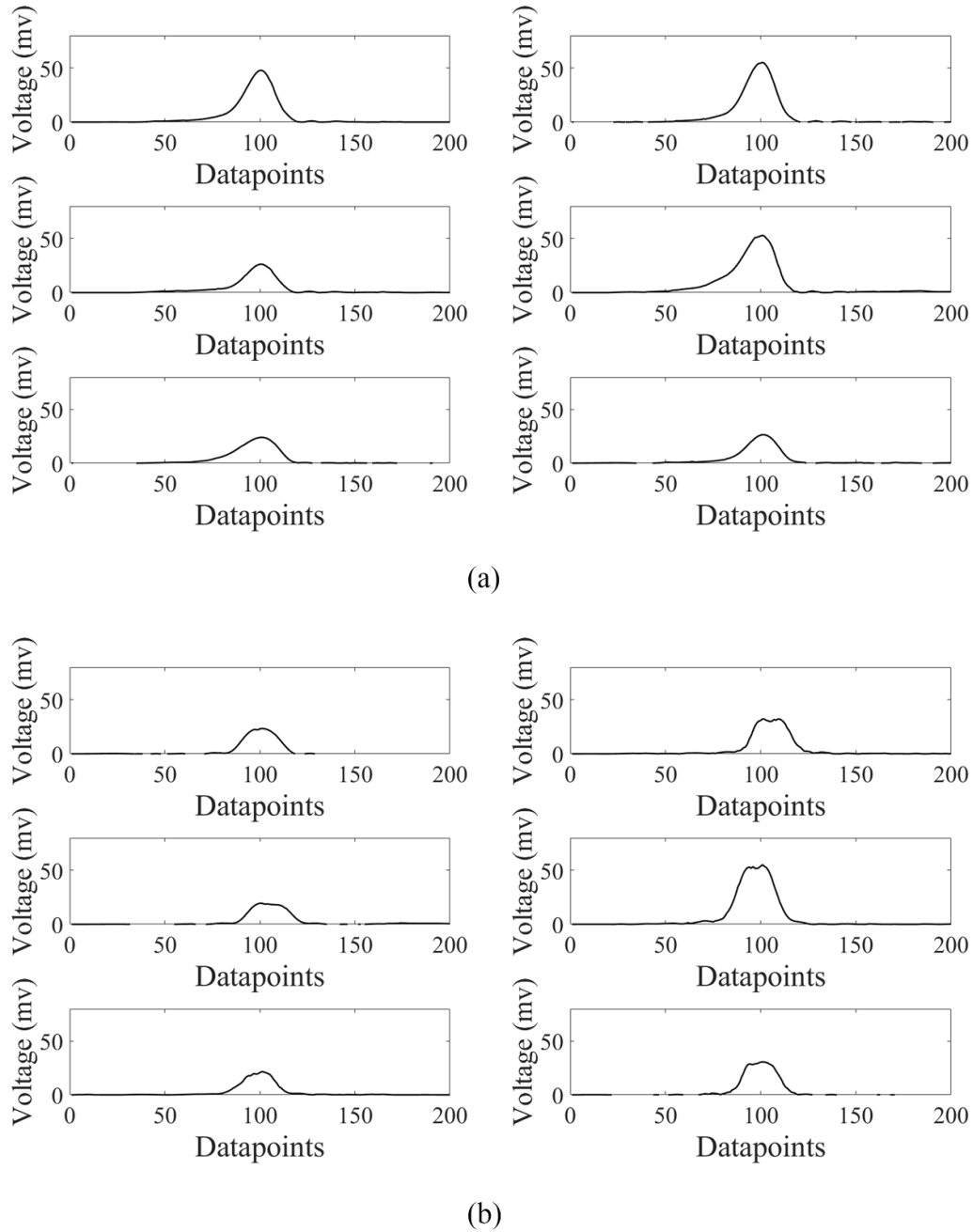


Figure 9. Intercepted pulse waveform: (a) six particle pulse waveforms randomly selected from Class A data and (b) six bubble pulse waveforms randomly selected from Class B data.

the bubbles to complete the particle size statistics, as shown in figure 8.

After the peak positioning, N data points before and after the peak point are extracted to obtain the pulse waveform. Based on the sampling frequency to determine N , 200 sampling points before and after the peak point are selected in this section. The identified and extracted pulses are shown in figure 9. Compared to hard particles, there are more small fluctuations in the signal waveform of bubbles. The bubble waveform has a tendency to show fluctuation and downward concavity at the peak, which is closely related to the geometry of the bubble as it passes through the cylindrical aperture, which provides the possibility for subsequent feature identification.

5. Feature extraction and classification of hard particles and bubble waveforms

5.1. Pulse waveform characterization

Researchers often choose the pulse amplitude as a typical feature to convert the amplitude into the particle size. However, it is difficult to distinguish between particles of the same size with different substances, such as small bubbles and hard particles, based on the amplitude alone. Besides, it is also difficult to obtain ideal waveforms after signal denoising treatment in real industrial environments.

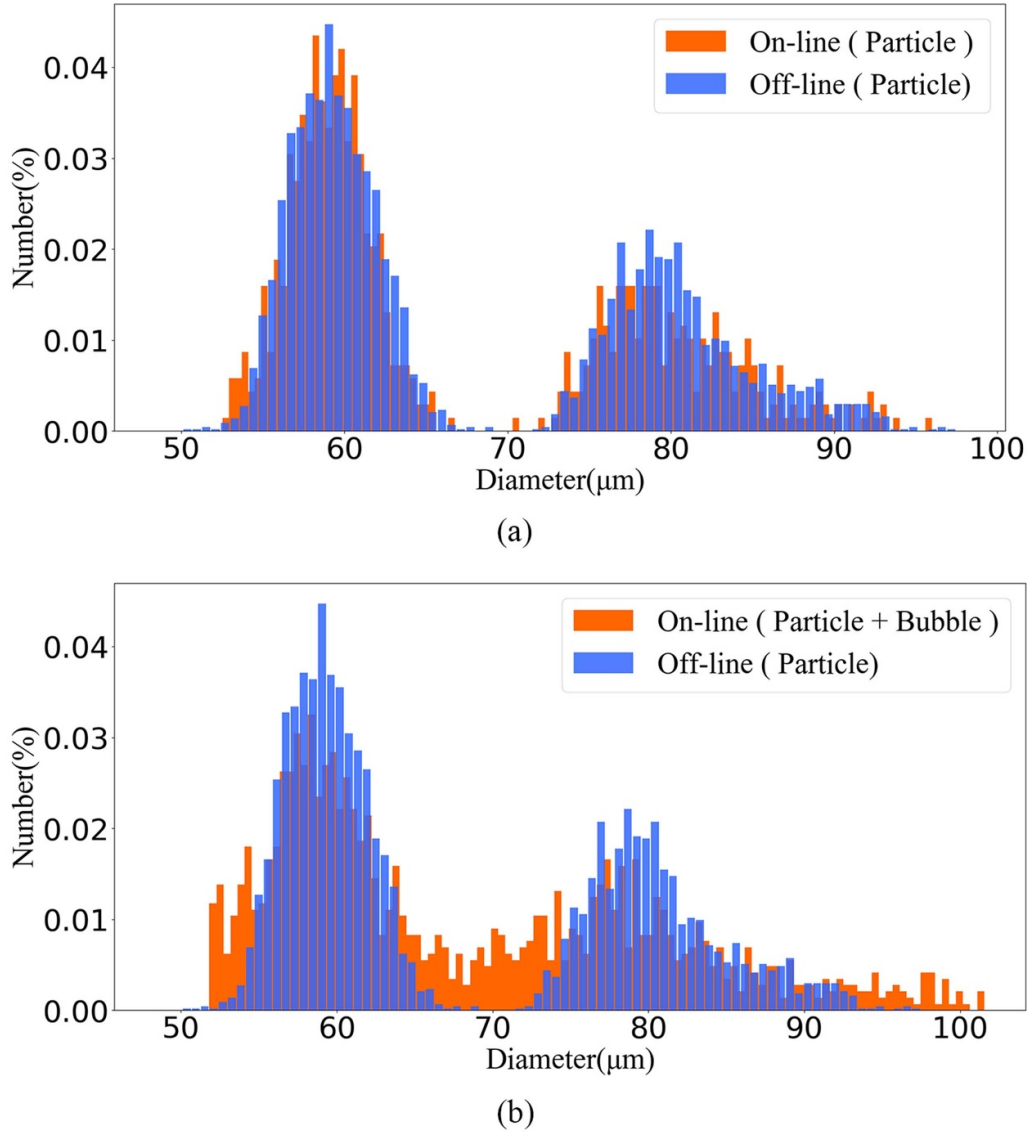


Figure 10. Particle size distribution adopting online and offline measurement modes, respectively: (a) comparison of the particle size distribution of Class A data with offline measurements; (b) comparison of particle size distribution of Class C data (with bubbles) with offline measurements.

Related numerical simulation studies have been carried out by our team to elaborate and compare the transient motion of hard particles and various deformable particles with different capillary numbers, Reynolds numbers and confinement ratios [35]. For deformable and non-deformable particles, there are differences in the transient motion processes between them. When passing through the aperture, the bubbles undergo very complex geometrical deformations, such as spherical cap, dimpled ellipsoidal cap or skirt shapes, due to a mixture of shear and extensional flows. Differences in the perturbation of the electric field by these shapes can cause differences in the pulse signal according to standardized methods in this paper. In addition, the transit time varies depending on the particle density and flow resistance. Therefore, the following features are defined in the morphological, geometrical, and time–frequency domain [36, 37] in conjunction with the actual

pulse waveforms collected to achieve the accurate depiction of hard particles and bubbles waveforms. The single pulse signal sequence is represented as $P = \{p_1, p_2, \dots, p_N\}$.

- (a) Amplitude. The maximum value of the pulse waveform is selected as the amplitude value to complete the conversion relationship between the amplitude and particle size:

$$I_1 = H_{\text{peak}} = \max \{p_i\}. \quad (4)$$

- (b) Pulse width. The full width at half maximum. The defined pulse width can be used as a basis for discrimination when positioning the pulse:

$$I_2 = \text{card} \{i | |p_i| > 0.5p_{\text{max}}\} \quad (5)$$

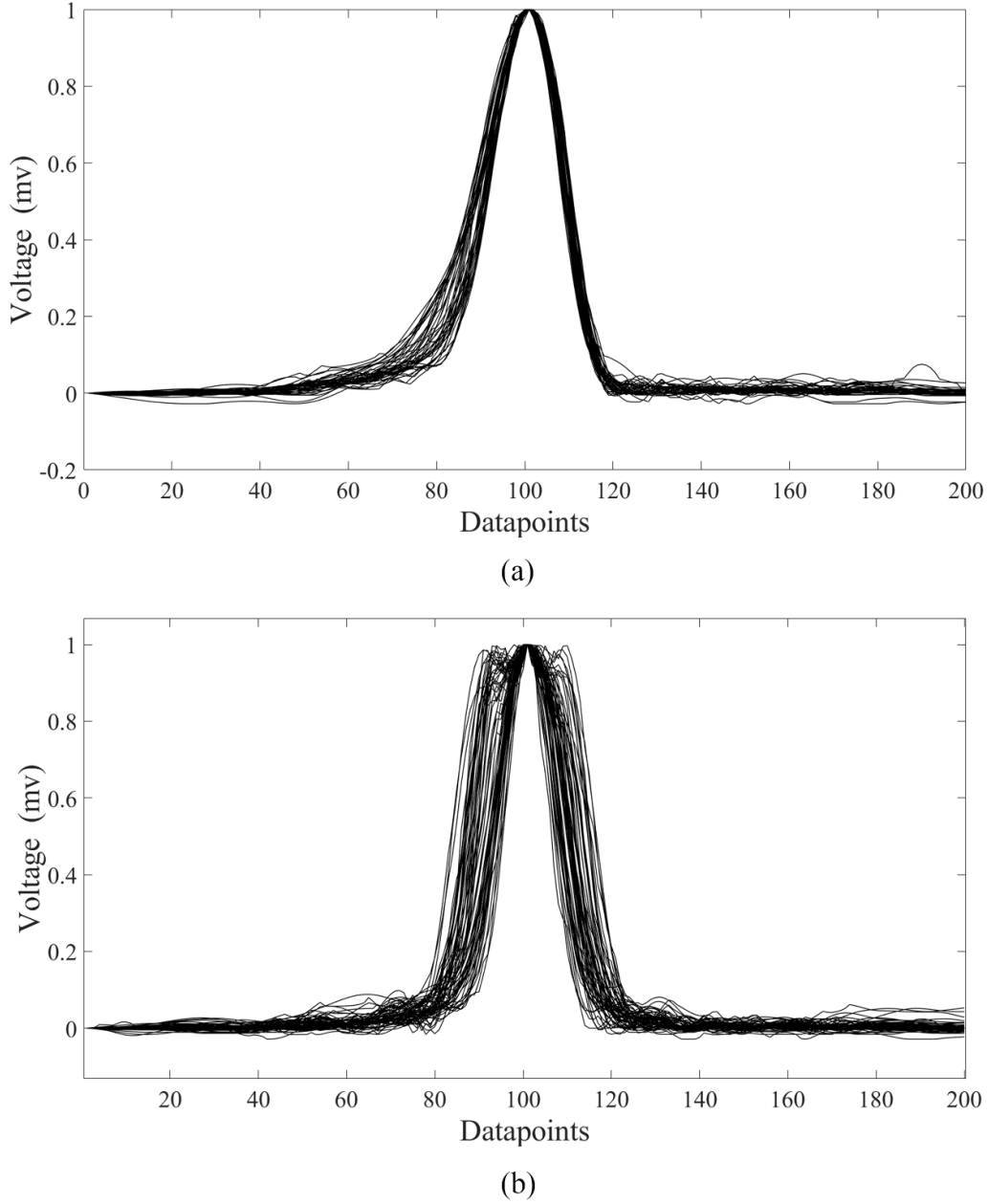


Figure 11. Pulsed signals after standardization: (a) standardized particle pulses in data A; (b) standardized bubble pulses in data B.

where card is a function that counts the number of elements in a set.

- (c) 20% of the amplitude pulse width. Define the pulse width that exceeds the 20% height of the amplitude:

$$I_3 = \text{card} \{i \mid |p_i| > 0.2p_{\max}\}. \quad (6)$$

- (d) Rise time. Defined as the time it takes for a pulse to rise from 10% of peak height to 90% of the peak height I_4 .

- (e) Maximum rising slope:

$$I_5 = \max \{p_{i+1} - p_i\}, i \in (1, N-1). \quad (7)$$

- (f) The skewness of the pulse obtained by fast Fourier transform (FFT):

$$I_6 = S = \frac{\frac{1}{M} \sum_{i=1}^M (p_{Ti} - \mu_{\text{tran}})^3}{\left(\frac{1}{M} \sum_{i=1}^M (p_{Ti} - \mu_{\text{tran}})^2 \right)^{\frac{3}{2}}}. \quad (8)$$

$\mu_{\text{tran}} = \frac{\sum_{i=0}^M p_{Ti}}{M}$ is the average amplitude of pulses obtained by FFT. p_{Ti} is indicated as the frequency domain signal after FFT.

- (g) Low-frequency component energy in the third level of the wavelet domain:

Table 1. Classification results for different parameters of BIRCH.

T	Branch factor	Number of misclassified samples	CH	Accuracy	Recall	Specificity	Precision	Time(s)
0.2	200	34	1319.0	0.980	0.983	0.977	0.969	0.0628
0.2	100	1009	769.6	0.390	0.001	0.667	0.003	0.0668
0.2	60	35	1315.8	0.978	0.985	0.974	0.964	0.0708
0.1	200	30	1314.4	0.982	0.990	0.976	0.967	0.0967
0.1	100	35	1326.9	0.979	0.994	0.968	0.957	0.1037
0.1	60	31	1313.5	0.981	0.993	0.973	0.963	0.1287
0.08	200	30	1314.4	0.982	0.990	0.976	0.967	0.1087
0.08	100	30	1314.4	0.982	0.990	0.976	0.967	0.1107
0.08	60	30	1314.4	0.982	0.990	0.976	0.967	0.1157
0.06	200	30	1314.4	0.982	0.990	0.976	0.967	0.1147
0.06	100	30	1314.4	0.982	0.990	0.976	0.967	0.1177
0.06	60	31	1313.5	0.981	0.993	0.973	0.963	0.1307
0.02	200	35	1310.5	0.979	0.993	0.969	0.958	0.1306
0.02	100	35	1310.5	0.979	0.993	0.969	0.958	0.1266
0.02	60	35	1310.5	0.979	0.993	0.969	0.958	0.1346

Note: The bold values are the final selected parameters of BIRCH, which means better parameters for classification.

Table 2. Classification results for dataset 2.

	Actual class: particle	Actual class: bubble	Total
BIRCH predicted class: particle	682	23	705
BIRCH predicted class: bubble	7	941	948
Total	689	964	1653

Note: The bold values are the final selected parameters of BIRCH, which means better parameters for classification.

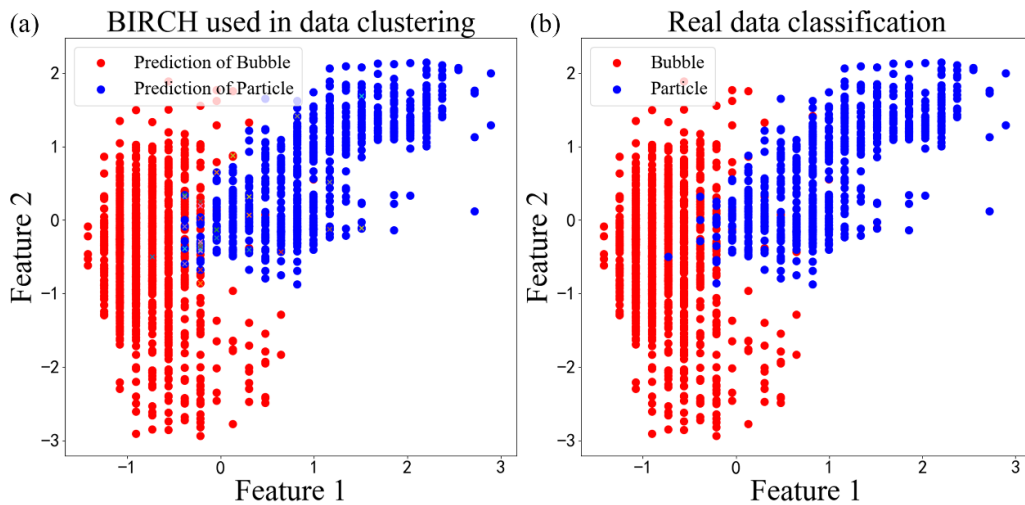


Figure 12. BIRCH clustering results: (a) predicted distribution of particle and bubble classification; (b) two-dimensional distribution of the real sample labels.

$$I_7 = E = \sum_{k=0}^{N-1} |w_{j=3,k}|^2 \quad (9)$$

where $w_{j=3,k}$ is the k -point wavelet coefficient for the third asymptotic component of the wavelet base function of ‘coif3’ after wavelet decomposition.

The amplitude and pulse width are defined for effective signal localization and extraction. The features $I_3 \sim I_5$ reflect the differences in the velocity and acceleration of materials of different densities as they pass through the ESZ, respectively.

The features I_6 and I_7 in the frequency and wavelet domains can reflect the effect of the deformation of bubbles and hard particles on the fluctuation of the signal as they pass through the aperture. These five features $I_3 \sim I_7$ characterize the motion state and morphology of hard particles and bubbles as they pass through the aperture, which is also a key link to achieve subsequent effective cluster classification.

Based on the definition of the above features, the pulse amplitudes from the Class A and C data are extracted in this paper. The amplitude-to-size [44] conversion can give the particle size distribution graph. The particle size results

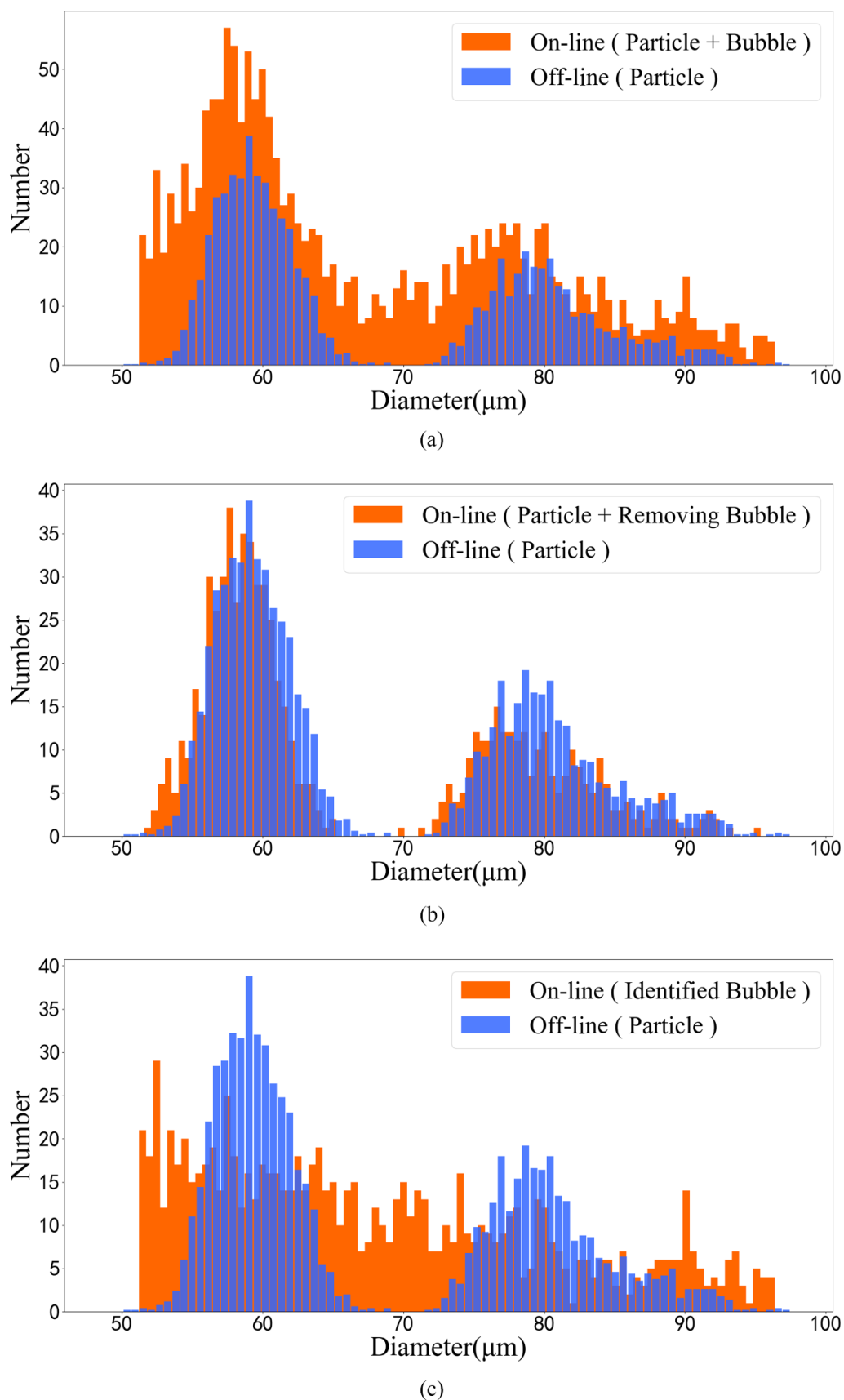


Figure 13. Particle size distribution in data 2: (a) online particle size distribution and offline particle size distribution; (b) online particle size distribution identified by BIRCH clustering and offline particle size distribution; (c) online bubble interference identified by BIRCH clustering and offline particle size distribution.

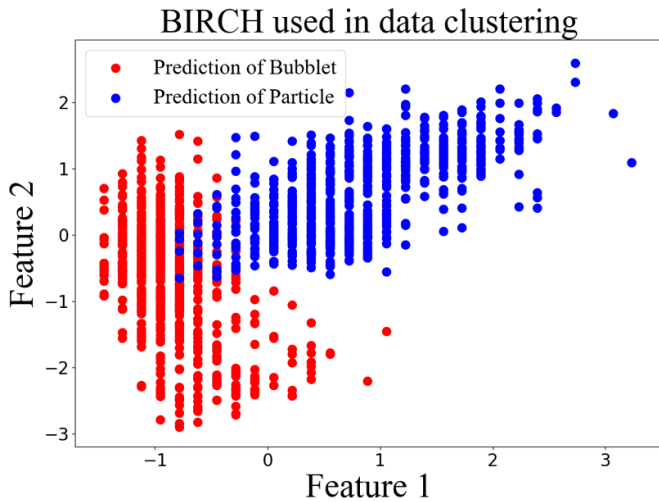


Figure 14. Online bubble and particle classification results under continuous acquisition of dataset 1.

measured by the MultisizerTM 4e COULTER COUNTER[®] Particle Analyzer produced by Beckman Coulter, Inc. in offline mode are presented too. After the treatment with the proposed signal denoising and de-interference strategy, the accuracy of the particle size statistics in Class B data can reach the offline measurement level. However, due to the interference of bubbles in the Class C data, accurate particle size statistics cannot be achieved from the amplitude directly. This task requires subsequent algorithms to identify and remove small bubbles to achieve the actual statistical result of hard particles, as shown in figure 10.

5.2. Data set construction

In this paper, the pulse waveforms of the particles and bubbles are standardized by linearly deflating to a uniform amplitude. This step eliminates the influence of the amplitude on the other dimensional features. The waveforms after pulse standardization are shown in figure 11.

After pulse standardization, the features of each pulse waveform are extracted according to the features defined in section 4.1. The five typical features are the 20% amplitude pulse width I_3 , rise time I_4 , maximum rising slope I_5 , frequency domain bias I_6 and the energy I_7 of the approximate component of the 3rd layer of the wavelet decomposition. The definition and extraction of features was implemented on MATLAB in an average time of 0.12 s for a 10 s signal data stream.

Firstly, the data of Class C are processed to obtain dataset 1, containing a total of 1445 particles and bubbles. Secondly, according to the above criteria, the data of Class A and Class B are processed. After extracting the normalized impulse features, a sample set containing labels that can be used to verify the classification results is produced. A mixed validation dataset 2 of labeled particles and bubbles is produced. Dataset 2 contains 689 samples of particles and 964 samples of bubbles.

5.3. BIRCH hierarchical clustering algorithm

Multi-feature extraction combined with machine learning provides an effective solution for pulse waveform recognition. By recognizing the features of extracted particles of different substances that generate pulses, we can effectively distinguish particle types and achieve tasks such as determining the size of classifying particles. In this paper, the BIRCH, a clustering algorithm is developed for hard particle and bubble classification identification. This algorithm was proposed by Tian Zhang in 1996 to cluster and order data using a hierarchical approach [54]. BIRCH allows fast clustering using a tree structure. Each node of the tree structure is composed of several clustering features (CFs), which have a good linear relationship and can be updated efficiently in the CF tree structure. There are three key parameters for the CF tree structure, namely, the maximum number of CFs in the inner node B , the maximum number of CFs in each leaf node L and the maximum sample radius threshold T of CFs in each leaf node. The BIRCH has the advantages of memory saving and fast clustering [54].

The BIRCH hierarchical clustering method can clearly discover the hierarchical relationship of classes, and the reasonable setting of the three parameters of CF tree structure can effectively achieve the effect of hierarchical clustering. The algorithm can also realize the detection of outliers. Although BIRCH and K-means both use distance to achieve class measurement, and are applicable to partition the dataset of hyperspherical shape distribution clusters, the former can be more flexible with regard to achieve the clustering of other shape clusters. Compared to the K-means method, BIRCH is more applicable to the distribution of features due to noise and other interference in the present scheme.

5.4. Results and discussion

The unsupervised classification of particles and bubbles was attained using the BIRCH clustering algorithm for the characteristics of the pulses, and the accuracy, recall, specificity and precision evaluation of the classification results were verified using labels [55].

BIRCH is used to classify and discriminate the sample sets of five classes of features from dataset 2. In this section, the maximum number of CFs in the inner node B is set to the same maximum number of CFs in each leaf node L and uniformly named as branch factor. Therefore, only the threshold T and branch factor need to be adjusted using Python software to obtain the clustering results. The Calinski–Harabasz (CH) index is used to evaluate the clustering results for different combinations of parameters, with a larger CH index representing better clustering results [56]. Accuracy, recall, specificity and precision are also introduced as the judging criteria. We compared the classification effects of different parameters and summarized some typical parameters in table 1. The final parameters are selected as $T = 0.08$ and branch factor = 100. With these parameters, it takes about 0.11 s to process the clustering classification of 1653 samples, and the classification results are shown in tables 2.

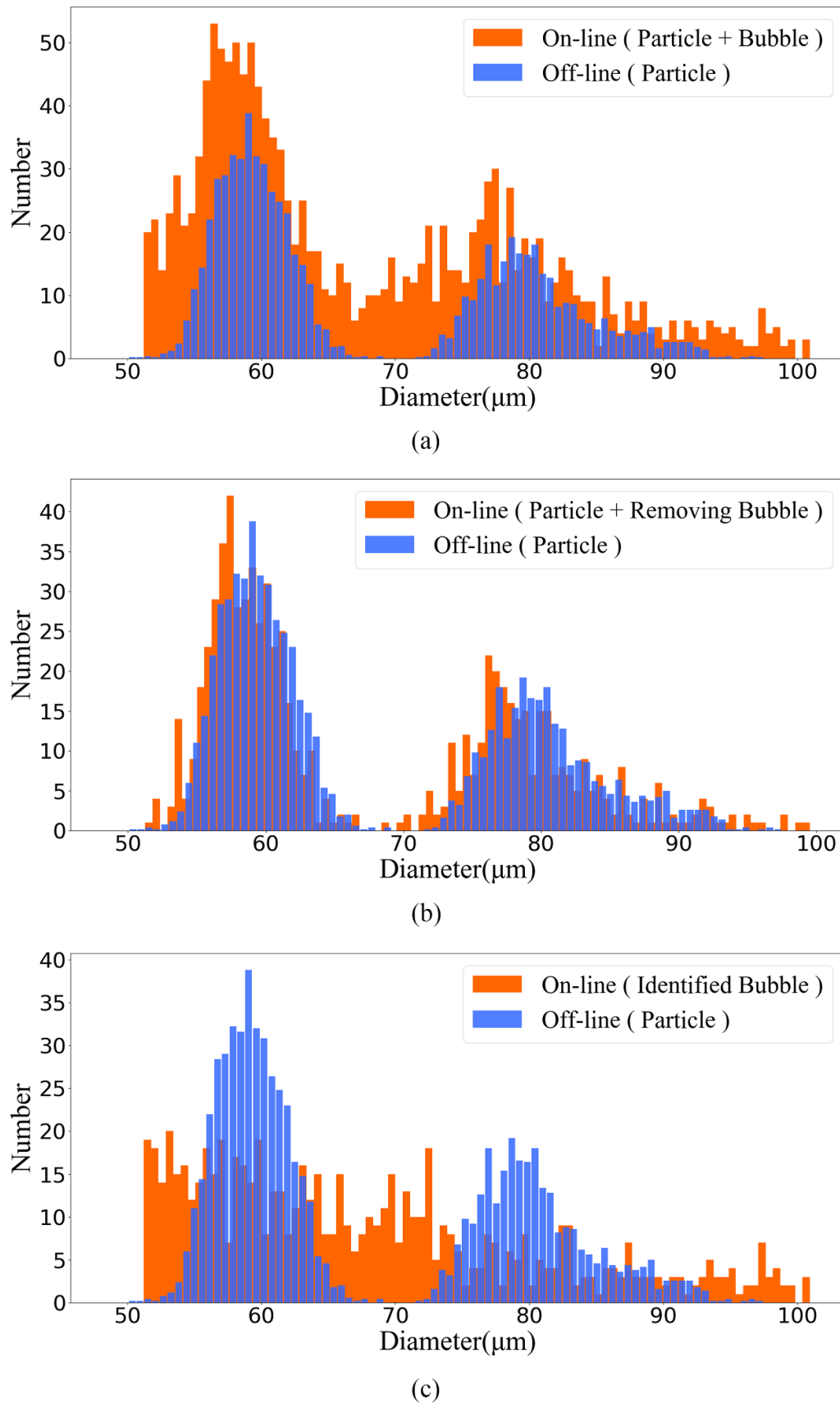


Figure 15. Particle size distribution under online continuous measurement in data 1: (a) respective online and offline particle size distribution; (b) online particle size distribution identified by clustering and offline particle size distribution; (c) online bubble interference identified by clustering and offline particle size distribution.

The results were validated by labeling with an overall sample classification accuracy of over 0.98, a recall of 0.988, a specificity of 0.976 and a precision of over 0.96. The clustering results are shown in figure 12, and the data labels are used to validate the classification results. Only the rise time I_4 and the skewness of the pulse obtained by FFT I_6 are selected to show the 2D distribution of clustering. BIRCH conforms to the expected classification profile.

This section combines five typical features extracted from the transient motions of hard particles and bubbles. By clustering, hard particles and bubbles can be effectively distinguished. By analyzing the difference between the waveforms of bubbles and hard particles in the actual collected signals, the bubbles have a shorter transit time. This is due to the lower density of the bubbles and the lower resistance to flow, resulting in higher acceleration and shorter transit times. The liquid flows through the aperture to form a jet and the acceleration of hard particles and bubbles is higher before entering the ESZ, resulting in a more prominent difference in the first half of the crossing time between particles and bubbles. In addition, there is peak attenuation and undulation of the bubble waveform. This is due to the low surface tension of the bubble, which is subjected to the extensional flow and undergoes ellipsoidal deformation. This has a relatively low perturbation to the electric field, leading to a decrease in resistance [35]. In addition, the bubbles are susceptible to shape changes. Random shape changes can also produce changes in the resistance and cause fluctuations in the waveform. The characteristic differences in the 2D distribution can also be seen, as shown in figure 12.

The size distribution of the particles and bubbles after identification of the delineated particles and bubbles is matched to the offline results of the MultisizerTM 4e COULTER COUNTER[®] Particle Analyzer as shown in figure 13. The results of the particle size distribution after removing the bubbles are in agreement with the offline results. The validity of the feature construction and the feasibility of the BIRCH clustering algorithm to achieve bubble and particle clustering classification are verified. The above scheme achieves the purpose of particle identification.

In this paper, the same operation was performed on dataset 1, and the classification results are shown in figure 14. The particle size distribution of the identified particles and bubbles are shown in figure 15. The time taken for clustering is 0.134 s. The identified particle size distributions are agreement with the results of the MultisizerTM 4e COULTER COUNTER[®] Particle Analyzer.

In this section, five classes of features are proposed to be defined and extracted based on the differences in the characteristics of the particle and bubble pulses collected by the ESZ. Particles and bubbles are effectively classified into two classes by using BIRCH. Removing bubbles from the particles yields particle size distribution results that are consistent with the offline measurements. This scheme effectively solves the problem of confusing the identification of particle pulses with different properties yielding the same amplitude range. The characterization and classification of signal pulses

in the ESZ are achieved, which makes it possible to study the particle size distribution of hard particles and small bubbles in the gas–liquid–solid three-phase chemical, medical or other environment in a more effective way.

This section extracts the pulse waveform features of hard particle and bubble based on the features and differences between the instantaneous motions of the two particles, and verifies the feasibility of the classification and identification by experimental means. For a signal with a sampling frequency of 200 kHz and a sampling duration of 10 s, the average time taken for denoising is less than 2.5 s. Simultaneous noise reduction and feature extraction while acquiring data is fully achievable. For classification of thousands of samples, the clustering algorithm takes less than 200 ms. The solution is potentially suited for real-time development.

6. Conclusions

In this study, we designed a signal analysis and identification scheme for an online multiphase micron-sized particle analyzer system in chemical and other industrial environments. Online ESZ measurements were carried out from signal analysis and feature recognition to achieve particle size statistics for particles with different attributes. The results demonstrate that the algorithm is potentially real-time. The wavelet soft threshold method is used to remove the noise at higher frequencies. For the first time, morphological filtering was used to deal with fundamental fluctuations, power frequency interference, and other low-frequency disturbances in the ESZ signal. The improved strategy of signal processing and pattern recognition can be more suitable for online measurement tasks. This approach effectively solves the effects of high-frequency noise and low-frequency fluctuations in the online measurement without distorting the waveform and demonstrates a good pulse waveform feature preservation performance. The peak detection algorithm was optimized based on the pulse width limit to overcome the effects of pseudo-pulses and overlapping peaks. By analyzing the differences between hard particle and bubble transient processes, five distinctly different features between hard particles and bubble pulses were effectively defined, and feature datasets of particles and bubbles were established. The BIRCH clustering algorithm was used to achieve the two categories of hard particles and bubbles, and the overall accuracy of particle identification was over 98%. The validity of denoising and feature recognition was verified in terms of both dataset validation and particle size distribution.

The online ESZ method has the advantages of a high detection size, accurate statistical counting and a direct measurement. Online measurements of the signal analysis and character recognition classification of research on the industrial process particle size distribution, local particle distribution, or the understanding of particles and particle group motion and other nonuniform characteristics of the monitoring, and the optimization of the process have a realistic significance.

Acknowledgments

This study was supported by the Chinese Academy of Sciences (CAS) Project (Grant No. YJKKYQ20200053), the ‘Double First-Class’ Construction Fund (Grant No. 111800XX62), the Industrial Strong Foundation Project (Grant No. DTCC28EE190929) and the Mechanical Engineering Discipline Construction Fund (Grant No. 111800M000).

ORCID iD

Fuhai Wang  <https://orcid.org/0000-0001-9044-6855>

References

- [1] Paglianti A and Pintus S 2001 An impedance probe for the measurements of liquid hold-up and mixing time in two/three-phase stirred tank reactors *Exp. Fluids* **31** 417–27
- [2] Dohi N, Takahashi T, Minekawa K and Kawase Y 2004 Power consumption and solid suspension performance of large-scale impellers in gas–liquid–solid three-phase stirred tank reactors *Chem. Eng. J.* **97** 103–14
- [3] Ahmed N and Jameson G 1985 The effect of bubble size on the rate of flotation of fine particles *Int. J. Miner. Process.* **14** 195–215
- [4] Bao Y, Hao Z, Gao Z, Shi L and Smith J 2005 Suspension of buoyant particles in a three phase stirred tank *Chem. Eng. Sci.* **60** 2283–92
- [5] Collins P C, Nielsen L K, Patel S D, Papoutsakis E T and Miller W M 1998 Characterization of hematopoietic cell expansion, oxygen uptake, and glycolysis in a controlled, stirred-tank bioreactor system *Biotechnol. Prog.* **14** 466–72
- [6] Garcia-Ochoa F and Gomez E 2009 Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview *Biotechnol. Adv.* **27** 153–76
- [7] Kunas K T and Papoutsakis E T 1990 Damage mechanisms of suspended animal cells in agitated bioreactors with and without bubble entrainment *Biotechnol. Bioeng.* **36** 476–83
- [8] Mitchell-Logean C and Murhammer D W 1997 Bioreactor headspace purging reduces dissolved carbon dioxide accumulation in insect cell cultures and enhances cell growth *Biotechnol. Prog.* **13** 875–7
- [9] O’hern T, d’agostino L and Acosta A 1988 Comparison of holographic and Coulter counter measurements of cavitation nuclei in the ocean *J. Fluid Eng.* **110** 200–7
- [10] Scardina P, Letterman R D and Edwards M 2006 Particle count and online turbidity interference from bubble formation *J. Am. Water Works Assoc.* **98** 97–109
- [11] Yang L X, Liang W H, Xin F, Jie C, Sha M Z, Zheng L G and Chao Y 2019 Progresses in measurement technologies of heterogeneous characteristics in multiphase reactors *Chem. Ind. Eng. Prog.* **38** 45–71
- [12] Li Y, Liu F, Dong H F, Zhang X P, Chen J R, Liu R C, Jin S, Wu Z P and Wang X D 2019 Application of micro particle real-time online analyzer in solid-liquid system measurement *CIESC J.* **070** 181–90
- [13] Wang J J, Jie Y, Lun Y Y and Tao L Y 2019 Nanopore-based confined spaces for single-molecular analysis *Chem. Asian J.* **14** 389–97
- [14] Kang J H, Hwang H, Hong S B and Hur S D 2014 Particle size distribution analysis of mineral dust in polar snow using a coulter counter *Ocean Polar Res.* **36** 319–26
- [15] Simões S, Sousa A and Figueiredo M 1996 Dissolution rate studies of pharmaceutical multisized powders—a practical approach using the coulter method *Int. J. Pharm.* **127** 283–91
- [16] Terejanský P, Makra I, Fűrjes P and Gyurcsányi R E 2014 Calibration-less sizing and quantitation of polymeric nanoparticles and viruses with quartz nanopipets *Anal. Chem.* **86** 4688–97
- [17] Yang L and Yamamoto T 2016 Quantification of virus particles using nanopore-based resistive-pulse sensing techniques *Front. Microbiol.* **7** 1500
- [18] Brazey B, Cottet J, Bolopion A, Lintel H V, Renaud P and Gauthier M 2018 Impedance-based real-time position sensor for lab-on-a-chip devices *Lab Chip* **18** 818–31
- [19] Shekar S, Chien C-C, Hartel A, Ong P, Clarke O B, Marks A, Drndic M and Shepard K L 2019 Wavelet denoising of high-bandwidth nanopore and ion channel signals *Nano Lett.* **19** 1091–7
- [20] Stewart R, Gideoni I, Zhu Y and Yang N 2011 *Systems and Computational Biology—Bioinformatics and Computational Modeling* (Rijeka: InTech)
- [21] Bezrukov S M 2000 Ion channels as molecular Coulter counters to probe metabolite transport *J. Membr. Biol.* **174** 1–13
- [22] Colquhoun D and Sigworth F 1995 Fitting and statistical analysis of single-channel records *Single-Channel Recording* (Berlin: Springer) pp 483–587
- [23] Sattigeri P, Thiagarajan J J, Ramamurthy N K, Joshi P, Spanias A, Goryll M and Thornton T 2010 Analysis of Coulter counting data from nanopores using clustering *Sensor Signal Processing for Defence (SSPD 2010)* (<https://doi.org/10.1049/ic.2010.0236>)
- [24] Naivar M A and Galbraith D W 2015 *Current Protocols in Cytometry* (New York: Wiley)
- [25] Pedone D, Firnkes M and Rant U 2009 Data analysis of translocation events in nanopore experiments *Anal. Chem.* **81** 9689–94
- [26] Arjmandi N, Roy W V, Lagae L and Borghs G 2012 [1207.2319](https://doi.org/10.1002/anie.201202319)
- [27] Jagtiani A V, Sawant R, Carletta J and Zhe J 2008 Wavelet transform-based methods for denoising of Coulter counter signals *Meas. Sci. Technol.* **19** 065102
- [28] Hanif M, Hafeez A, Suleman Y, Mustafa Rafique M, Butt A R and Iqbal S M 2016 An accelerated framework for the classification of biological targets from solid-state micropore data *Comput. Methods Programs Biomed.* **134** 53–67
- [29] Qin Z P, Zhe J and Wang G X 2011 Effects of particle’s off-axis position, shape, orientation and entry position on resistance changes of micro Coulter counting devices *Meas. Sci. Technol.* **22** 045804
- [30] Garboczi E J 2017 The influence of particle shape on the results of the electrical sensing zone method as explained by the particle intrinsic conductivity *Powder Technol.* **322** 32–40
- [31] Li M, Carozza C and Guthrie R I L 2013 Particle discrimination in water based LiMCA (liquid metal cleanliness analyzer) system *Can. Metall. Q.* **39** 325–38
- [32] Guthrie R I L and Li M 2001 *In situ* detection of inclusions in liquid metals: part II. Metallurgical applications of LiMCA systems *Metall. Mater. Trans. B* **32** 1081–93
- [33] Guthrie R I L and Li M 2001 *In situ* detection of inclusions in liquid metals: part I. Mathematical modeling of the behavior of particles traversing the electric sensing zone *Metall. Mater. Trans. B* **32** 1067–79
- [34] Wang X, Isac M and Guthrie R I L 2009 Numerical studies on the *in-situ* measurement of inclusions in liquid steel using the E.S.Z. or LiMCA technique *ISIJ Int.* **49** 975–84
- [35] Guo X K, Liu R C, Wang X D, Zhang Z Y, Mihaela I and Roderick G 2020 Two-phase flow simulation for distinguishing deformable particles with a LiMCA system *Appl. Math. Modelling* **88** 106–21

- [36] Raillon C, Che J, Thill S, Duchamp M, Desbiolles B X, Millet A, Sollier E and Renaud P 2019 Toward microfluidic label-free isolation and enumeration of circulating tumor cells from blood samples *Cytometry A* **95** 1085–95
- [37] Guo J, Chen Z, Ban Y L and Kang Y 2017 Precise enumeration of circulating tumor cells using support vector machine algorithm on a microfluidic sensor *IEEE Trans. Emerg. Topics Comput.* **5** 518–25
- [38] Cao X 2015 Detrending and denoising of impedance cytometry data MSc Thesis (Rutgers University-Graduate School-New Brunswick)
- [39] Honrado C, Mcgrath J S, Reale R, Bisegna P, Swami N S and Caselli F 2020 A neural network approach for real-time particle/cell characterization in microfluidic impedance cytometry *Anal. Bioanal. Chem.* **412** 3835–45
- [40] Shen Y, Yalikun Y and Tanaka Y 2019 Recent advances in microfluidic cell sorting systems *Sensors Actuators B* **282** 268–81
- [41] Wang N, Liu R, Asmare N, Chu C-H and Sarioglu A F 2019 Processing code-multiplexed Coulter signals via deep convolutional neural networks *Lab Chip* **19** 3292–304
- [42] Barnett M, Sims E and Lines R 1976 Laboratory evaluation of the Coulter on-line particle monitor *Powder Technol.* **14** 125–30
- [43] Coulter W H 1953 Means for counting particles suspended in a fluid *US Patent Specification* 2656508
- [44] Deblois R W and Bean C P 1970 Counting and sizing of submicron particles by the resistive pulse technique *Rev. Sci. Instrum.* **41** 909–16
- [45] Ahuja K, Rather G M, Tian L Z, Ye S J, Fei X P, Tuan L, Bertino J R and Javanmard M 2019 Toward point-of-care assessment of patient response: a portable tool for rapidly assessing cancer drug efficacy using multifrequency impedance cytometry and supervised machine learning *Microsyst. Nanoeng.* **5** 34
- [46] Hampson S M, Pollard M, Hauer P, Salway H, Christie S D R and Platt M 2019 Additively manufactured flow-resistive pulse sensors *Anal. Chem.* **91** 2947–54
- [47] Khodaparastasarabad N, Mohebbi A and Falamaki C 2018 A novel microfluidic high-throughput resistive pulse sensing device for cells analysis *Microsyst. Technol.* **25** 3643–53
- [48] Sun T, Berkel C V, Green N G and Morgan H 2009 Digital signal processing methods for impedance microfluidic cytometry *Microfluid. Nanofluid.* **6** 179–87
- [49] Siwal D et al 2014 Pulse shape analysis of a two fold clover detector with an EMD based new algorithm: a comparison *Nucl. Instrum. Methods Phys. Res. A* **741** 108–16
- [50] Gu Z, Ying Y L, Cao C, He P and Long Y T 2015 Accurate data process for nanopore analysis *Anal. Chem.* **87** 907–13
- [51] Maragos P and Schafer R W 1987 Morphological filters—part I: their set-theoretic analysis and relations to linear shift-invariant filters *IEEE Trans. Acoust. Speech Signal Process.* **35** 1153–69
- [52] Wang J, Xu G, Zhang Q and Liang L 2009 Application of improved morphological filter to the extraction of impulsive attenuation signals *Mech. Syst. Signal Process.* **23** 236–45
- [53] Bocaz-Beneventi G, Latorre R, Farková M and Havel J 2001 Artificial neural networks for quantification in unresolved capillary electrophoresis peaks *Anal. Chim. Acta* **452** 47–63
- [54] Zhang T, Ramakrishnan R and Livny M 1999 BIRCH: an efficient data clustering method for very large databases *ACM SIGMOD Rec.* **25** 103–14
- [55] Powers D M 2020 (arXiv:2010.16061)
- [56] Calinski T and Harabasz J 1974 A dendrite method for cluster analysis *Commun. Stat.—Theory Methods* **3** 1–27