# Research on clustering analysis of eye diagram point set of digital signal based on equivalent time sampling

Jingkai Ni *

College of Technology, Yanbian University, Yanji, China, 133002

* Corresponding Author Email: 15105774086@163.com

**Abstract.** With the continuous development and improvement of communication technology, the research and application of high-frequency signals are becoming more and more important, especially in important fields such as electronic communication, aerospace and aviation. The analysis of high-frequency signals is the most basic and most important. High-frequency digital signals are mainly obtained by equivalent time sampling and sequential sampling. This paper first analyzes the basic principles of equivalent time sampling and real-time sampling, and compares the advantages and disadvantages of the two and the limitations of each sampling method through various indicators. After that, the eye diagram point set based on equivalent time sampling is clustered and analyzed. By comparing the contour coefficients, the sum of squared errors, and the time complexity of each clustering algorithm, the clustering method of the eye diagram point set is further optimized, and the most efficient and accurate clustering algorithm is selected. The clustering algorithm is optimized by increasing the multi-selective convergence threshold based on each cluster center.After many experiments and simulations, from the perspective of various clustering indicators, there is no significant difference between the K-Means clustering algorithm and the K-Mediods clustering algorithm in the case of fewer data points. However, in the case of relatively large data points, the K-Mediods clustering algorithm is more accurate and efficient than the K-Means clustering algorithm. Moreover, compared with the original K-Mediods clustering algorithm, the clustering effect of the K-Mediods clustering algorithm after multi-selective optimization in terms of convergence threshold is reflected in both the sum of squared errors and the contour coefficient. Both have better accuracy.

**Keywords:** Equivalent time sampling, Error sum of squares, Time complexity, clustering.

## 1. Introduction

The rapid development of cloud computing, Internet of Things and other related industries has caused the continuous growth of bandwidth demand and interface transmission rate. With the continuous improvement of signal transmission frequency and rate, the demand for high-bandwidth signal precision measurement and analysis instruments is increasing in various industries, especially in the industrial field. As an important instrument for waveform analysis and measurement, oscilloscope is indispensable for the advancement of work. There are two kinds of oscilloscopes, one is real-time oscilloscope, and the other is sampling oscilloscope. The real-time oscilloscope uses a high-speed digital-to-analog converter to continuously sample and display real-time signals, and the sampling rate is determined by the digital-to-analog converter. The sampling oscilloscope adopts the equivalent sampling technology. Under the same sampling accuracy, the sampling rate of the signal is lower. The application of equivalent sampling technology not only has new achievements in bandwidth, frequency and other important parameters, but also overcomes many problems brought by analog-to-digital converter (ADC). Sampling oscilloscope has a variety of functions, mainly including eye diagram analysis, waveform analysis, eye diagram module analysis and other functions. In the eye diagram mode, it can be more intuitive and convenient to judge the quality of the channel, measure the bit error rate of the channel and the feasibility of the work.

Aiming at the transient trigger sampling of the sampler and the identification and reconstruction of the transient signal, Yang Kun realized the conditioning and transformation of the narrow pulse signal by using the method of differential impedance integral transformation and the design of multi-stage filter amplifier circuit, and verified the feasibility of the designed joint debugging circuit through experiments [1]. Liu Jingjing proposed a sparse signal reconstruction algorithm based on sequential equivalent time sampling to solve the limitations of traditional sampling on signals, which achieved higher sampling efficiency and obtained more accurate signals [2]. Du et al. proposed an equivalent time sampling based on time discrimination for the problem of low accuracy of random equivalent sampling trigger point and the next sampling point in a short time and complex equipment ,which achieved high-precision measurement results and obtained more accurate reconstructed waveforms [3].Aiming at the shortcomings of the traditional equivalent time sampling principle and method, Liu Jianbo proposed an equivalent time sampling method based on Field-Programmable Gate Array ( FPGA ) , which realized the frequency conversion equivalent time sampling in the periodic analog bandwidth [4].Aiming at the complexity of the traditional eye diagram intersection estimation method and the tediousness of the data, Liu Jie proposed an eye diagram intersection estimation algorithm based on local interpolation , which greatly reduced the amount of data required under the same accuracy [5].

The main work of this paper is to study the oscilloscope related projects based on equivalent time sampling. Firstly, the sequential equivalent time sampling method is used to restore a more complete eye diagram. Then, the basic parameters of the eye diagram signal are measured and analyzed by writing different algorithms. The exact values of the eye diagram parameters under various algorithms are gradually optimized. Finally, the most suitable algorithm is selected to complete the design of the eye diagram module.

## 2. Sampling principle analysis and the basic characteristics of eye diagram

### 2.1. Real-time sampling principle

Based on the Nyquist sampling theorem, the real-time sampling is carried out at an equal time interval with a frequency greater than twice the original signal. Sampling is performed several times in a cycle. After each sampling, the digital-to-analog converter is used to quantify and encode each sampling point into a number, and the quantization operation is repeated. Finally, the obtained data is sorted in order to restore the waveform, as shown in the figure 1.

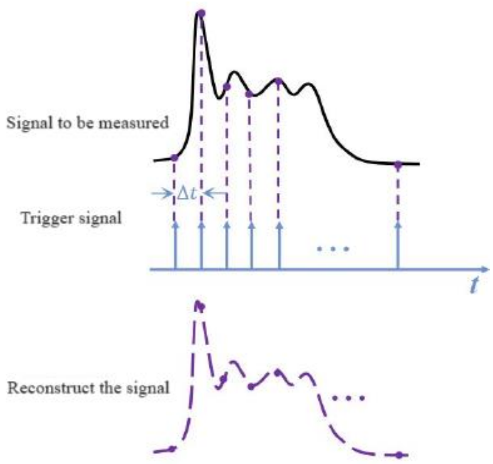


**Figure 1.** Real-time sampling waveform.

Real-time sampling requires equal interval sampling of periodic signals in a cycle to better complete the recovery of signals. The speed is fast and the accuracy is high. The collected waveform is close to the original signal in real time display. Moreover, when the sampling frequency is high enough, real-time sampling can also be applied to the sampling of aperiodic signals, so real-time sampling has a very high universality. Although the minimum sampling frequency is twice the frequency of the original signal, in order to restore the original waveform more accurately, the number of sampling points in a single cycle needs to be further increased [6]. In general, the number of sampling points in a single cycle needs to reach 10 or more.

Because the quantization of the sampled value needs to rely on the digital-to-analog converter, once the quantization time exceeds the sampling interval, the output waveform will produce serious distortion. In order to adapt to the sampling time interval, a high-speed digital-to-analog converter and a very high sampling clock are required. In the time domain signal, the sampling of high-frequency signals requires a high-precision digital-to-analog converter. The duration of a single-cycle Gaussian narrow pulse is at the nanosecond level. If you want to sample at a frequency 10 times higher than the original signal, you must make the frequency of the signal reach the GHZ order of magnitude, and the high-speed conditions will affect the conversion accuracy of the digital-to-analog converter, resulting in an increase in system cost, which is extremely unfavorable for the design of the overall experiment.

## 2.2. Equivalent time sampling

For the defects of real-time sampling in terms of rate, equivalent time sampling can be more effectively compensated. The equivalent sampling technique is the sampling of periodic signals. By triggering multiple sampling of high-frequency signals in different periods, the sampling points of different periodic signals are reorganized, and the original waveform can be reconstructed more effectively. Compared with real-time sampling, under the same sampling frequency, equivalent time sampling can obtain higher sampling rate, convert high-frequency signals into low-frequency signals, and easily expand and reconstruct signals far exceeding the Nyquist limit frequency while improving resolution. However, since the equivalent sampling is a combination reconstruction of different periodic sampling points of the signal, the equivalent sampling is suitable for repeated periodic signals. Equivalent sampling is generally divided into random sampling and sequential sampling [7]. The biggest difference between the two is the position difference of the collected signal. The sequential equivalent sampling signal can only appear after the trigger point, while the random equivalent sampling signal can appear before or after the trigger point.

Sequential equivalent sampling refers to the sampling of the input periodic signal at different positions of each cycle by using the sampling pulse with timing sequence. Through the external clock, each cycle is only sampled once. After multiple shift triggers, the periodic signal is reconstructed with the obtained signal data without the need to recover the clock [8]. The principle of sequential equivalent time sampling is shown in Figure 2.

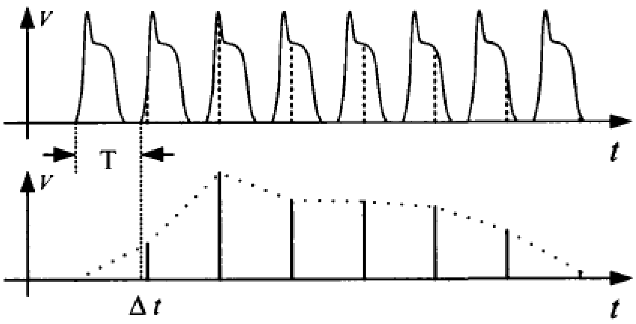


**Figure 2.** Sequential equivalent time sampling principle.

The time sequence in the sequential equivalent time sampling can be divided into step, step back, and difference. Taking the step-by-step sequential equivalent sampling as an example, the step-by-step delay pulse is relative to the signal trigger point. Each sampling point increases a precise incremental delay $\Delta t$ compared to the original sampling point, that is, each sampling point is compared with the first sampling point in turn $\Delta t, 2\Delta t, 3\Delta t \cdots n\Delta t$ delay time, so the digital-to-analog converter for sequential sampling needs to output a voltage that is proportional to the instantaneous amplitude of the signal waveform at $\Delta t, 2\Delta t, 3\Delta t \cdots n\Delta t$ in turn. It is assumed that the original signal period is T, the step delay time is $\Delta t$ , and the complete recovery of a periodic signal requires n points. The signal waveform after equivalent time sampling transformation is equivalent to the frequency drop to $1/\left(nT + n\Delta t\right)$, which realizes the conversion of high frequency signal to low frequency signal, so that it meets the Nyquist sampling law, so it has a very high sampling rate.

Random equivalent sampling is to sample periodic signals in different intervals under multiple triggers, and the time interval between each sampling point and the trigger point is random. In the random equivalent sampling method, in order to ensure that the data points sampled each time are the only accurate determination points, the trigger level needs to be set before the random equivalent sampling. When the amplitude of the measured signal reaches the trigger threshold, the data acquisition system uses the periodicity of the input signal to randomly select a series of time intervals in multiple cycles to collect a series of waveform data, and a set of sampling data is obtained by data processing. After each data acquisition system is triggered, the same number of data points are collected to form a sequence, so as to obtain multiple sampling sequences. Since the frequency and phase of the sampling clock will produce random jitter, if the trigger point position at each trigger is kept unchanged, the time interval between each trigger point and the subsequent nearest sampling point is randomly distributed within a sampling clock period, and the time intervals are not equal to each other. As shown in the figure 3, it can be seen from *t*1, *t*2, *and t3* in the figure that the size of the time interval determines the position of the sampled data obtained by this trigger in the reconstructed waveform. The final sampling point sequence can be obtained by arranging these sampling point sequences in ascending order according to the time interval.

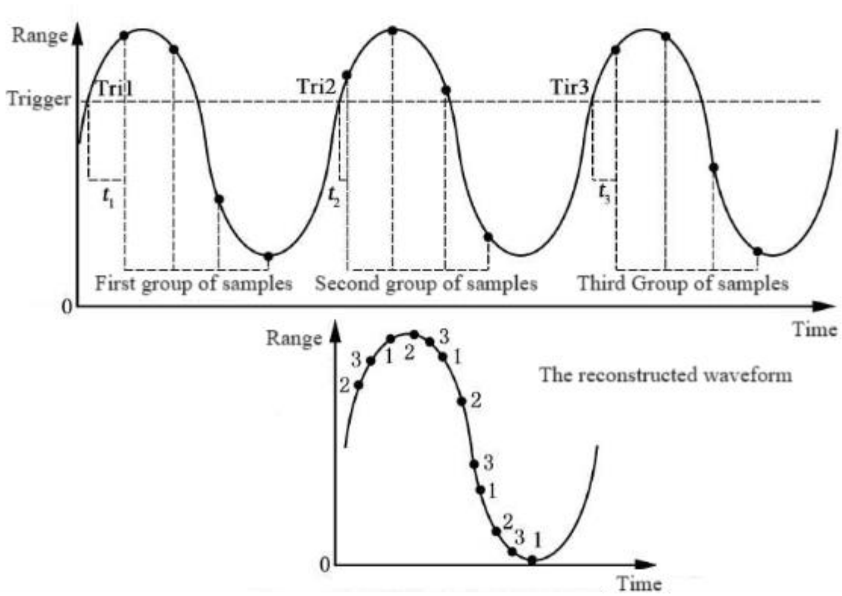


**Figure 3.** Random equivalent time sampling principle.

In the random equivalent sampling, it is necessary to accurately measure the time interval $\Delta t$ between the trigger point and its nearest sampling point. After multiple trigger sampling, the obtained multiple sets of sampling sequences are sorted by the equivalent sorting algorithm according to the time size $\Delta t$ to reconstruct a complete original signal waveform, so as to achieve a higher equivalent sampling rate at a lower sampling frequency [9]. Due to the randomness of the trigger point, the clock used in the random equivalent sampling must be internally provided.

## 2.3. Basic characteristics of eye diagram

The eye diagram is a graph formed by the superposition of different band symbols. Its shape is similar to the eye, so it is named as the eye diagram. As shown in the figure 4. Various factors of the signal itself may have a certain impact on the performance and working state of the system, such as potential noise, interference, distortion and so on. The quality of the signal, that is, inter-symbol crosstalk, jitter, noise, etc., can be visually judged by the appearance of the eye diagram, and the basic parameters of the eye diagram can also be used for quantitative analysis [10]. Therefore, the reasonable use of the high-speed digital signal global map presented by the eye diagram can more intuitively evaluate the channel quality and determine the feasibility of channel transmission of high-speed signals.

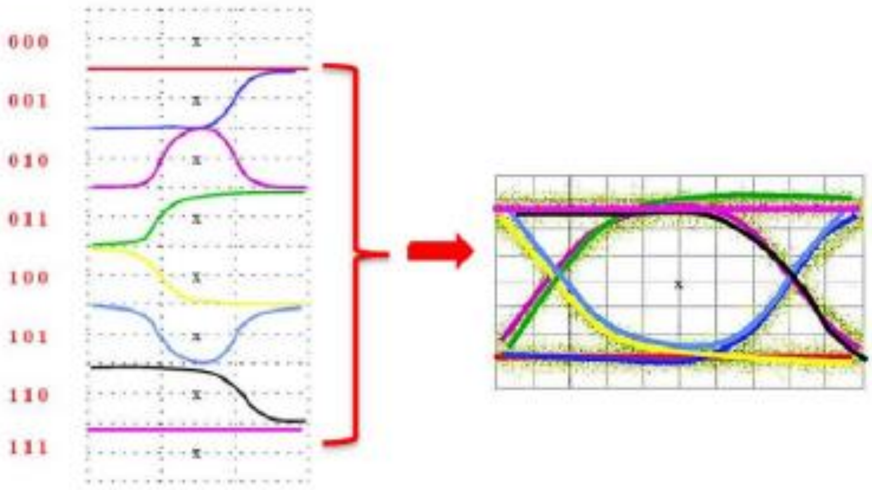


**Figure 4.** Eye diagram signal formation.

The signal sequence can be superimposed according to a certain number of symbols to form an eye diagram. According to the morphological characteristics of the eye diagram, the jitter of the eye diagram can be seen more intuitively and clearly, so as to judge the bit error rate of the signal, that is, the quality problem, as shown in the figure 5. The eye diagram covers the following parameters: the decision threshold of the signal, the optimal decision time, zero-crossing distortion, timing error sensitivity, maximum signal distortion and so on. Among them, the decision threshold level is generally the middle area of the eye diagram, and the mean value of the upper and lower eyelids can be calculated. The slope of the upper and lower lines of the eye diagram is a representation of the error sensitivity. The larger the absolute value of the slope of the line is, the higher the sensitivity is, and the more significant the jitter effect is. The maximum value of the opening amplitude of the eye diagram at the best decision time is the performance of the noise capacity. The larger the amplitude, the greater the noise tolerance; the best judgment time refers to the maximum position of the eye opening in the eye diagram. At this time, the inter-symbol interference is small and the judgment is more accurate. The variation range of the shadow part at the intersection of the upper and lower edges of the eye diagram reflects the zero-crossing distortion, and the thickness of the ' eyelid ' at the top represents the maximum distortion of the signal.

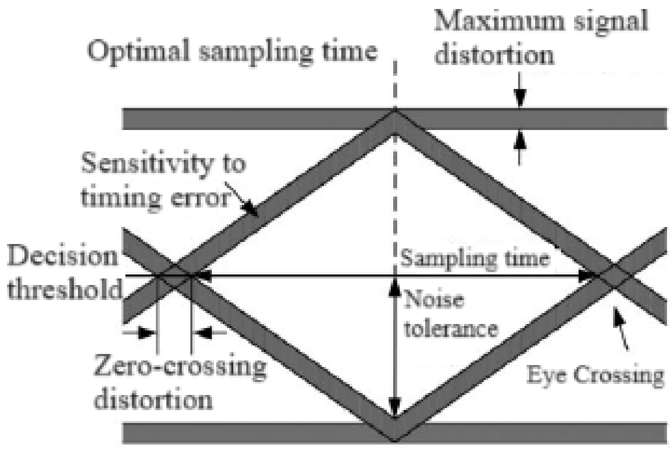


**Figure 5.** Eye diagram model.

Since the eye diagram is the superposition of the same time interval signal waveforms at different times, the synthesis of the eye diagram requires the timing synchronization clock after the signal timing synchronization. According to the timing synchronization clock, the eye diagram calculation module defines the appropriate symbol number as the unit length, divides the signal sequence into several equal parts and superimposes them in the same picture, so as to observe the statistical characteristics of the signal.

## 3. Clustering algorithm analysis of various point sets

Through the analysis of the characteristics and definitions of eye diagram parameters, the data point set of eye diagram parameters can be obtained through the following steps. Firstly, the data transmission rate is calculated, and the data point set of the single cycle of the eye diagram is intercepted. Then, the data point set of the eye diagram is divided into two parts to obtain the cross-data point set of the eye diagram. By calculating the left and right cross points, the data set of the 0 / 1 level of the eye diagram is obtained. Finally, the simulation of the clustering algorithm is carried out according to the data point set of the 0 / 1 level data set. The simulation steps and results are as follows.

### 3.1. DBSCAN clustering algorithm

DBSCAN is a density-based clustering algorithm. It defines clusters as the largest set of points connected by density, and can divide regions with sufficient density into clusters. For dense point sets in space, it is a cluster, and noise is found between different density regions. The radius and the minimum number of points (MinPts) defined by this algorithm determine the number of objects assigned to the specified cluster and the distance between objects [11]. The DBSCAN algorithm has no requirement for the preset number of clusters and is more effective in dealing with special clusters.

The following is the DBSCAN algorithm flow:

(1) Select any uncalibrated starting point;

(2) find all points within radius $\varepsilon$ with the starting point as the center;

(3) For points not less than MinPts found in the subrange, they form new clusters with adjacent points. Repeat the above steps and continue to expand the new clusters.

(4) When the maximum limit of the number of clusters is reached, the initial step is returned and the next new point is selected for operation.

(5) Repeat the above steps until all unmarked points are used.

### 3.2. K-means clustering algorithm

The K-Means clustering algorithm uses the Euclidean distance between data points as the grouping basis. According to the degree of aggregation, the scattered point set is divided into k sets, so that the distance between data points in each set reaches the minimum and the distance between data points in other non-same sets reaches the maximum [12].

The workflow of K-Means algorithm is as follows:

(1) Determine the number of clusters (K);

(2) K representative data points are randomly selected from the data set as the initial clustering center {C1, C2, C3... Ck}, $1 < k < n+1$

(3) Calculate the distance between the remaining data points and each cluster center, allocate the remaining data to each class according to the minimum distance principle, and finally obtain k clusters {S1, S2, S3... Sk}.

$$dis(X_i, C_i) = \sqrt{\sum_{i=1}^{m} (X_{it} - C_{it})^2} \tag{1}$$

Among them, $X_i$ is the *i* th object, $C_j$ is the *j* th cluster center, and $X_{it}$ is the *t* th attribute of the *i* th object ; $C_{jt}$ is the *t* th attribute of the *j* th cluster center.

(4) Recalculate the mean value of the cluster and redistribute the data set according to the mean value.

(5) Repeat 3 and 4 steps until the end of the algorithm convergence.

Compared with the DBSCAN clustering algorithm, the K-Means algorithm has a more comprehensive clustering effect and a wider range of use. And in the case of massive data points, the K-Means algorithm is more accurate.

### 3.3. K-mediods algorithm

The K-Mediods algorithm is a partition-based clustering algorithm, and it is also an optimization and improvement of the K-Means algorithm [13]. The clustering center selects the median in the cluster, which can better reduce the influence of outliers on the clustering results. Since the clustering center point is supported by actual data, the K-Mediods algorithm is less affected by outliers and noise [14]. The K-Mediods algorithm works as follows:

(1) Determine the number of clusters (K);

(2) K representative data points are randomly selected from the data set to do the initial clustering center medoids {M1, M2, M3... Mk},$1 < k < n+1$, for any medoids are corresponding to a cluster.

(3) Calculate the distance between the remaining data points and each cluster center, assign the remaining data to each class according to the minimum distance principle, and finally obtain k clusters {N1, N2, N3... Nk}.

(4)The center point of the new cluster is recalculated, and the data set is redistributed according to the above algorithm.

(5) The algorithm repeats the above steps until all medoid is equal to the previous medoid.

The flow of K-Mediods algorithm is as shown in the figure 6:

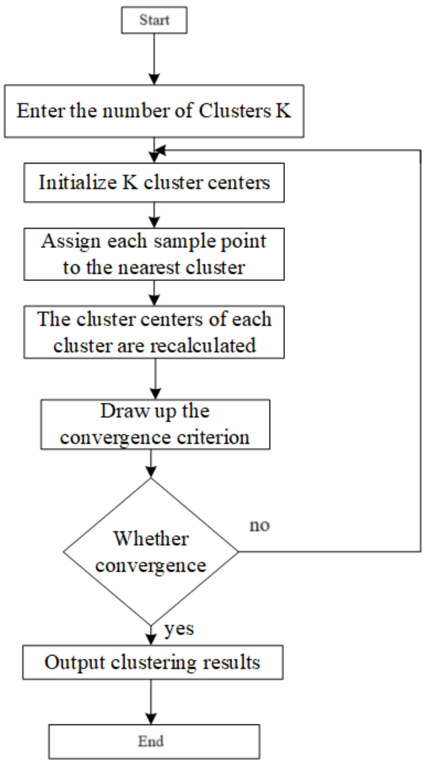


**Figure 6.** The K-Mediods algorithm flow.

Compared with the K-Means algorithm, the K-Mediods algorithm has higher complexity of cluster head selection time and cluster overlap space, and the lower outlier sensitivity of the K-Mediods algorithm has a positive effect on execution time and anti-noise performance. Compared with the K-

Mediods clustering algorithm, the K-Means algorithm is faster and more efficient [15]. However, due to the high sensitivity to outliers, the clustering center will be greatly affected by outliers, and the convergence results of clustering will also be changed to a certain extent, resulting in instability.

## 4. Simulation and experimental analysis

The simulation of the clustering algorithm selects the data point set based on the sequential equivalent time sampling. There are many evaluation indexes for the clustering results, such as the sum of squared errors, the Dunn index, the contour coefficient, etc. The Dunn index refers to the shortest distance of points in any cluster divided by the longest distance between any two clusters. The sum of squared errors is the sum of squared distances from all data points to the center. The contour coefficient is a combination of cohesion and separation, and it is also one of the important indexes to judge the clustering results. In this paper, the contour coefficient and the sum of squared errors are selected as the indexes to judge the simulation effect, so as to obtain the best eye diagram clustering algorithm.

Firstly, the sum of squared errors is used as the criterion function of K-Mediods clustering algorithm and K-Means algorithm simulation:

$$S = \sum_{i=1}^{k} \sum_{x \in X_i} \left\| x - \mu_i \right\|^2 \tag{2}$$

$K$ is the number of clusters, $X_i$ is the cluster set, $x$ is the object in the cluster set, $\mu_i$ is the cluster center point, S represents the compactness of the sample. In general, the smaller the value of S, the more compact the cluster point set, and the better the clustering effect.

K-Mediods clustering algorithm and K-Means algorithm are applied to the point set sampled by sequential equivalent time sampling. The sum of error squares obtained by each simulation is obtained by multiple simulations, and the mean value of the sum of error squares is obtained. The simulation results are as shown in the figure 7:

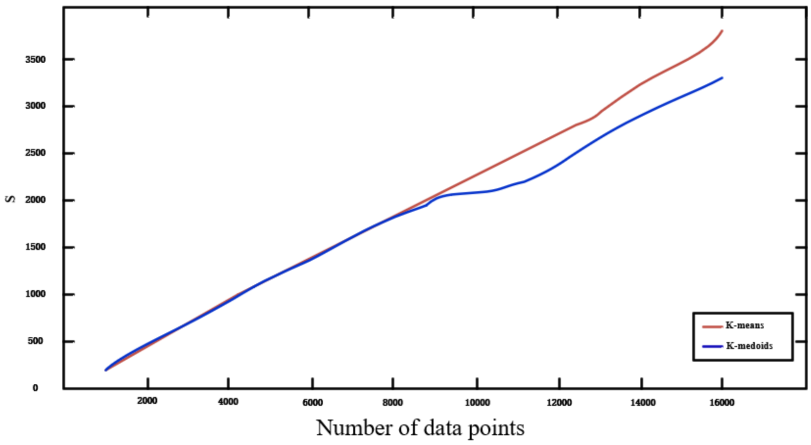


**Figure 7.** K-Mediods clustering algorithm, K-Means algorithm error sum of squares.

By comparing the error square sum of K-means algorithm and K-mediods algorithm, it is not difficult to find that in the case of less data points, the clustering results of the two algorithms are close, and there is no big difference. However, in the case of more data points, the error square sum of K-mediods algorithm is obviously smaller than the error square sum of K-means algorithm. Therefore, the processing accuracy of K-mediods clustering algorithm is better than that of K-means clustering algorithm when dealing with massive multi-source data.

In order to better evaluate the clustering effect of these two clustering algorithms, the contour coefficient is selected as the evaluation criterion. The contour calculation formula of a single sample is:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{3}$$

$a(i)$ Represents the average distance from the sample to other samples in the same cluster, and $b(i)$ represents the minimum value of the average distance from the sample to all other clusters. The size of $a(i)$ reflects the correlation between the sample and the same cluster, and $b(i)$ reflects the degree of non-correlation between the sample and the cluster. Through experimental calculation and analysis, the contour coefficient of K-mediods algorithm is 0.6838, and the contour coefficient of K-means algorithm is 0. 6415. Compared with the contour coefficient obtained by the experiment, the contour coefficient of K-mediods algorithm is higher than that of K-means algorithm, and the clustering effect of K-mediods algorithm is better again.

In order to further determine a better clustering algorithm, the time efficiency of the two types of algorithms is compared here,as shown in the figure 8 :

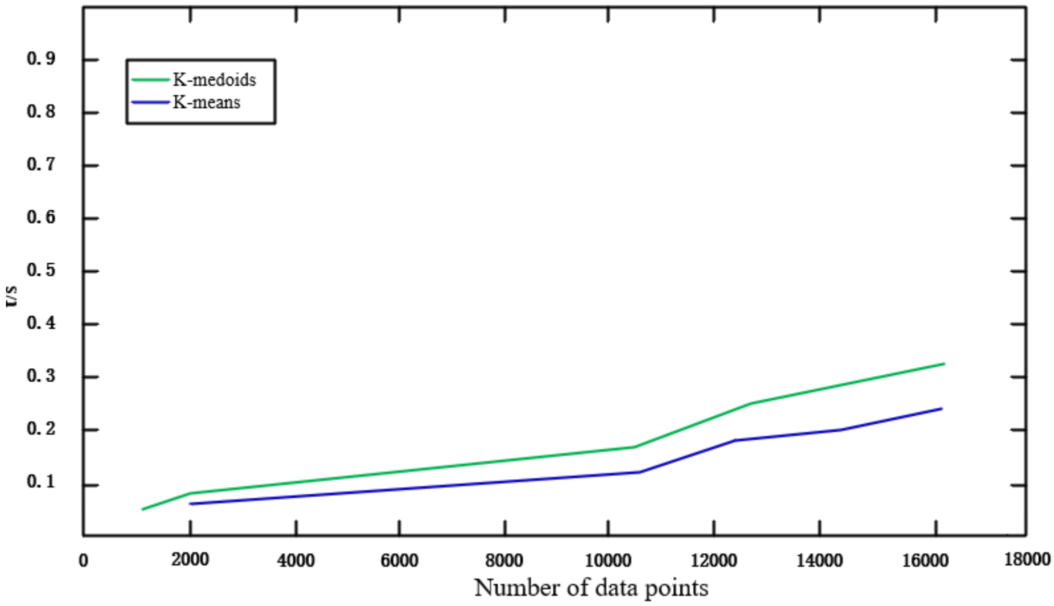


**Figure 8.** Time complexity.

The time complexity of the algorithm is mainly affected by the number of sample points, the number of clusters, and the number of iterations. Under the premise that the time complexity is similar or basically the same, the running time of the K-mediods clustering algorithm is relatively small, and in the case of the increasing number of data points, the running time difference between the K-mediods clustering algorithm and the K-means algorithm is more and more significant. Therefore, the K-mediods clustering algorithm is more efficient than the K-means algorithm.

Whether it is in the sum of squared errors, contour coefficients, or time complexity, there is not much difference between the two in the case of a small number of sample points, but when the base number of sample points is large, K-mediods clustering algorithm is superior to K-means algorithm in all aspects. On the whole, the K-mediods clustering algorithm is superior to the K-means algorithm.

## 5. Conclusion

In this paper, the principle and method of equivalent time sampling and real-time sampling are studied for the eye diagram formation method. Under the premise of sequential equivalent time sampling, the clustering method of data points is optimized. The K-mediods clustering algorithm and the K-means clustering algorithm are used to simulate the data clustering. Comparing the clustering of the two algorithms, it is found that the K-mediods clustering algorithm and the K-means clustering algorithm have little difference in various clustering indicators in the case of fewer data points, but in the case of more complex data point sets. The contour coefficient and error square of the K-mediods clustering algorithm are better than the K-means clustering algorithm, and the time complexity of the K-mediods clustering algorithm is lower. Therefore, on the whole, the K-mediods clustering algorithm is better for the clustering of eye diagram data sets based on sequential equivalent time sampling.

## References

[1] Yang Kun. Study on Transient Sampling and Signal Conditioning Circuits of Sampling Oscilloscope [D]. North Central University, 2023.

[2] Liu Jingjing. Research on random sampling model and its signal processing algorithm [D]. University of Electronic Science and Technology of China2018.

[3] Du Xiuli; li Shixin; qiu Shaoming; chen Bo. Time discrimination based random equivalent sampling short time measurement method [J]. Instrument technology and sensor.2016, (11):114-117+122.

[4] Theodore W. Hansy. Optical wave interference sampling oscilloscope [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition).2021, 33(05):699-713.

[5] Liu Jianbo, Guo Wenxiu, Zhang Jie and so on. Equivalent Time Sampling Based on FPGA [J]. Electronic Design Engineering.2015, 23(02):122-124+129.

[6] Cheng C, Desheng R, Yanlin Y, et al. TOPSIS based multi-fidelity Co-Kriging for multiple response prediction of structures with uncertainties through real-time hybrid simulation[J].Engineering Structures,2023,280

[7] Feng Xinru, Jingning, Yinziyan. High-speed chirp signal recovery based on equivalent sampling [J]. Single-chip microcomputer and embedded system applications, 2023, 23 (09): 37-40.

[8] Li Haitao, Ruan Linbo, Tian Geng. Sequential equivalent sampling method based on cascaded step delay and its implementation [J]. Automation Instrumentation, 2020, 41 (10): 74-77.

[9] Zhixin T, Dingkai Z, Shunhe H, et al.A Hierarchical Random Graph Efficient Sampling Algorithm Based on Improved MCMC Algorithm[J].Electronics,2022,11(15):2396-2396.

[10] Gao Qiang, Ma Shimin, Liu Jie. Vector network analyzer eye diagram intersection point estimation method [J]. Metrology.2023, 44(05):777-782.

[11] Zhan C, Bai K, Tu B, et al.Offshore Oil Spill Detection Based on CNN, DBSCAN, and Hyperspectral Imaging[J].Sensors,2024,24(2).

[12] Guo Kenan. Data detection technology based on fusion improved K-means clustering algorithm [J]. Electronic Design Engineering, 2024, 32 (05): 41-45.

[13] Wang W, Lou B, Li X, et al.Intelligent maintenance frameworks of large-scale grid using genetic algorithm and K-Mediods clustering methods[J].World Wide Web: Internet and Web Information Systems,2020,23(3):1177-1195.

[14] Zhang Shangyao. Research on runoff prediction and multi-model combination based on multi-feature-driven machine learning [D]. Xi 'an University of Technology, 2023.

[15] Chen X, Güttel S. Fast and explainable clustering based on sorting [J]. Pattern Recognition, 2024, 150110298-.