Analyst

Accepted Manuscript

Volume 143
Number 10
21 May 2018
Pages 2185-2436

Analyst

rsc.li/analyst

ISSN 0003-2654

PAPER
Huinan Yang, Chengxing Shen, Xiaoshu Cai et al
Noninvasive and prospective diagnosis of coronary heart disease
with urine using surface-enhanced Raman spectroscopy

ROYAL SOCIETY
OF CHEMISTRY

ROYAL SOCIETY
OF CHEMISTRY

rsc.li/analyst

A robust signal processing program for nanopore signals by dynamic correction threshold with compatible baseline fluctuations

Guohao Xi [1], Jinmeng Su [1,2], Jie Ma [1], Lingzhi Wu [3] and Jing Tu [1,*]

[1] State Key Laboratory of Digital Medical Engineering, School of Biological Science and Medical Engineering, Southeast University, Nanjing, 210096, China.

[2] Monash University-Southeast University Joint Research Institute, Suzhou, 215123, China.

[3] College of Science, Nanjing University of Posts and Telecommunications, Nanjing, 210046, China.

* Correspondence: jtu@seu.edu.cn (J.T.).

Abstract

Solid-state nanopores represent a powerful platform for the detection and characterization of a broader scope of biomolecules and particles, including proteins, viruses, and nanoparticles, for clinical and biochemical applications. Typically, nanopores work by measuring transient pulses of the ionic current as translocation events of molecules passing through the pore. In view of the strong noise and stochastic fluctuation of ionic current recording in nanopore experiments, the signal processing based on the statistical analysis of massive translocation events remains a crucial issue for nanopore sensing. Based on parallel computational processing and efficient memory management, we have developed a novel signal processing procedure for translocation events to improve the signal identification performance of solid-state nanopores in the presence of baseline oscillation interference. Obviously, by means of an adaptive threshold in a sliding window, we can correct the baseline determination process in real time. As a result, the features of the translocation event signals can be identified more accurately, especially for the intermittent occurrence of high-density complex signals, and the program also shows good signal differentiation. As a ready-to-use software, the data program is more efficient and compatible with diverse nanopore signals for more complex nanopore applications.

Analyst Accepted Manuscript

INTRODUCTION

In recent years, nanopore sensing has been proposed for third-generation sequencing technologies and multiple screening applications in precision medicine[1-3]. Nanopores are inspired by the transmembrane transport within cells in biology as a fundamental process in life activities[4, 5]. As nanoscale pores embedded in biological or solid membranes[6, 7], the delicate structures with high-precision detection capabilities have been applied to various fields, including the detection of nucleic acids[8-10], proteins[11-14], and nanoparticles[15, 16]. Conceptually, single analyte molecule passing through a nanoscale channel is measured by the instantaneous fluctuation of transmembrane ionic flow under the voltage and current recording with high-bandwidth sampling, which are involved the accumulated noise power at full frequency and unpredictably small-signal frequency response. Currently, the biological nanopores at fixed sizes are prevalently employed in sequencing due to the balance of the stable signal output and controlled noise amplitude[17]. Likewise, the solid-state nanopores with larger size range and customizable pore shapes are still subject to the signal acquisition and recognition with competitive noise performance, although they have been widely applied for enzymes[18, 19], viruses[20], nanoparticles[21]. Then, the current drift and unavoidable background noise are more frequent in solid-state nanopores, making it more challenging to recognize signals from solid-state nanopores[22]. Moreover, considering the transient and random of current events, the signal processing is based on the statistical analysis of a large number of nanopore events, hence, enhanced throughput and automate processes is required for the development of nanopore technology[23].

Over the decades of studies, some typical nanopore signal processing programs have been developed, such as AutoNanopore, Open Nanopore, Cavro Nanopore Sensing, Transalyzer, NanoAnalyzer, NanoPlex[24], EventPro[25] and so on[23, 26-29]. AutoNanopore and Open Nanopore employ adaptive threshold methods to detect low SNR events, but their performance may be limited in high-noise or temporally attenuating signal conditions. Cavro Nanopore Sensing and NanoAnalyzer offer high throughput and parallel analysis capabilities, yet they may struggle with noise interference in complex signal environments, especially under high salt concentrations. Transalyzer and MOSAIC have high resolution for signal precision, but may encounter challenges in event recognition accuracy when dealing with signal attenuation or significant noise effects. NanoPlex is characterized by good noise suppression and adaptability to low SNR events; however, its moderate flexibility limits its application in diverse experimental setups. EasyNanopore is known for its simplicity and low hardware requirements, but it faces challenges in handling baseline drift and complex signal environments. In contrast, our "Dynamic Correction Method" dynamically adjusts thresholds and corrects baselines in real-time, significantly improving the detection of low SNR and complex signals. Especially in solid-state nanopore applications, our method effectively mitigates noise and overcomes the limitations of traditional methods, enhancing signal detection accuracy and reliability. These methods mainly search outliers in the current traces to achieve the event recognition and information extraction of nanopore signals with the baseline and threshold algorithms. Obviously, the threshold needs to be selected by calculating and truncating the current signal changes according to the professional experience of the individual processing the data. And the fluctuation and drift of the baseline current over time is even more hard to determine a uniform global standard. On this basis, the emerging machine learning methods are used for nanopore signal

recognition and statistical analysis[30-32]. Generally, a set of machine learning-based algorithms including Hidden Markov models, fuzzy C-means, and Support Vector Machines can be efficiently to improve the recognition, feature extraction and cluster analysis of nanopore signals. Meanwhile, Neural network algorithm based on deep learning have been employed to continuously optimize the prediction results of signal processing[33-35]. However, these methods necessitate a specifically configured operational environment and training database, which may not be user-friendly for non-professionals engaged in software development.

To solve the massive signal processing, a configuration-free translocation event detection software named "Easynanopore" has been developed in our research group[29]. The method employs an adaptive thresholding approach based on low-frequency variance (utilizing local mean and local variance) to define the commencement and conclusion of an event. The Dynamic Correction Method is particularly effective in detecting low signal-to-noise ratio (SNR) events, which are often challenging to identify using traditional static threshold methods. By dynamically adjusting based on the characteristics of the signal, it is able to more accurately capture events that might be overlooked by fixed-threshold methods, especially in cases where noise or signal variations are subtle. Furthermore, when considering events that exhibit temporal attenuation or decay, the Dynamic Correction Method offers significant advantages. Temporal attenuation refers to the gradual decrease in signal amplitude over time, which can be an important characteristic of certain events, such as the translocation of larger molecules through nanopores. In such cases, the Dynamic Correction Method can more effectively track and detect events that experience a slow decay in signal, ensuring that these events are not missed. This capability is crucial for maintaining the accuracy and reliability of signal detection in experiments with complex signal profiles, such as those involving nanopores.

To better illustrate the strengths of the proposed method, we have compared its performance with that of several commonly used nanopore signal detection platforms. The following table summarizes key performance metrics, including sensitivity, accuracy, computational speed, and hardware requirements. This comparison highlights the advantages of the proposed method, particularly in terms of its efficiency and suitability for use with limited computational resources.

Table 1 : Comparison of Signal Detection Methods

Platform	Noise Management	Low SNR Event	Baseline Drift Handling	Flexibility	High Salt Conditions	Complex Signals	Hardware Requirements
AutoNanopore	Moderate	Moderate	Moderate	Good	Low	Moderate	Good
NanoAnalyzer	Good	Moderate	Moderate	Low	Moderate	Good	Moderate
Cavro Nanopore Sensing	Low	Low	Moderate	High	Good	Low	Low
Kleiner Lab Software	Low	Moderate	Moderate	Moderate	Low	Moderate	Moderate
EasyNanopore	Moderate	Moderate	Moderate	Good	Moderate	Moderate	Moderate

NanoPlex	Good	Good	Moderate	Moderate	Good	Good	Moderate
EventPro	Good	Moderate	Moderate	Good	Good	Moderate	Moderate
Dynamic Correction Method	Moderate	Good	Good	Good	Moderate	Good	Good

Additionally, a multi-threaded algorithm is employed to partition the file and adapt to low-end CPU configurations. The parallel computation method for file partitioning enhances the speed of event detection during the recognition process. However, the parameters of baseline and thresholds are determined based on the local mean variance of each point in this algorithm. As a cumulative calculation model, this model is undeniably accurate, whereas a similar error cumulative pattern will become more serious when baseline calculation deviation occurs. Especially for solid-state nanopores, where the signal output is more diverse and complex, and the baseline fluctuation situation is more drastic, persistent deviations tend to occur silently. This effect is especially pronounced in the case of dense signal fragments and mixed signal fragments. This is a principal factor contributing to the significant discrepancies and lack of reproducibility observed in the signal data obtained from solid-state nanopores.

Therefore, an improved signal processing program in our present study has been proposed based on a novel baseline and threshold correction computation model in real time during the document recognition process. This method integrates the determination of the threshold and baseline with the current recognition area of the signal, and corrective measures are implemented in accordance with the baseline conditions in the vicinity of individual signals, thereby effectively mitigating the impact of baseline fluctuations and dense signal areas. Furthermore, the conventional baseline scanning mode has been retained, allowing the user to select it freely according to the type of signal file in question. This effective corrective measure markedly enhances the identification and precision of solid-state nanopore signals, constituting a valuable contribution to the efficient and accurate classification and deployment of nanopore signals.

We refer to this signal processing program as the Dynamic Correction Method. Among various nanopore signal detection platforms, Dynamic Correction Method stands out with several unique advantages, especially in key aspects such as noise management, low signal-to-noise ratio (SNR) event detection, baseline drift handling, flexibility, high salt concentration adaptability, complex signal processing, and hardware requirements (Table 1). Compared to other platforms, Dynamic Correction Method offers significant benefits in these critical areas.

Firstly, in terms of noise management, Dynamic Correction Method effectively suppresses noise, ensuring the reliability of signals, particularly when the signal environment is complex or the noise level is high. While other platforms (such as AutoNanopore and NanoAnalyzer) also employ noise management techniques, these platforms often face challenges in environments with high salt concentrations or other complex factors, which can lead to misdetection or missed events in complicated signal backgrounds. In contrast, Dynamic Correction Method dynamically adjusts the threshold and baseline in real time, making it more resilient in noisy conditions and low-SNR environments.

While EventPro and NanoPlex employ adaptive baseline options to mitigate baseline fluctuations, their reliance on global fitting or fixed time-window updates may lead to delayed baseline adaptation and misclassification of weak signals in low SNR environments. In contrast, our Dynamic Correction Method introduces event-driven baseline updates and dynamic threshold adjustments, allowing it to track rapid fluctuations more effectively and enhance the accuracy of signal detection even under severe noise conditions.

For low-SNR event detection, the Dynamic Correction Method demonstrates a clear advantage over traditional approaches that rely on fixed thresholds. By dynamically adjusting the threshold based on real-time signal variations, our method can accurately capture low-SNR events that might otherwise be overlooked. This is particularly crucial in cases where signal fluctuations are subtle, ensuring the precise identification of weak signals while minimizing false positives.

To comprehensively evaluate the performance of different methods under these conditions, we compared several common signal processing approaches. The table 2 presents the performance of NanoPlex, EasyNanopore, NanoAnalyzer, and Dynamic Correction Method in key metrics such as baseline noise, event detection rate, and signal integrity. Other methods were not included in the comparison mainly because their performance under low SNR conditions is either similar to that of the methods included in this study or due to resource and testing limitations. Additionally, we focused on methods optimized for solid-state pore signal characteristics, ensuring the relevance of the comparison results. This comparison highlights the advantages of the Dynamic Correction Method, particularly in terms of event detection rate and signal integrity, in complex signal environments.

Table 2: Comparison of Signal Processing Methods for Low SNR

Metric	RMS Noise (pA)	Peak Noise (pA)	Event Detection Rate (%)	Signal Integrity (R)
NanoPlex	15.3±1.2	50.2±2.8	85.4±2.1	0.86±0.02
EasyNanopore	18.7±1.5	65.4±3.0	78.6±3.0	0.82±0.03
NanoAnalyzer	16.5±1.4	55.3±2.5	83.1±2.7	0.84±0.02
Dynamic Correction Method	12.8±1.0	40.5±2.3	92.8±1.5	0.91±0.01

In baseline drift handling, many platforms, such as AutoNanopore, use fixed baseline correction methods. While effective in some cases, these methods struggle when the signal exhibits significant fluctuations or complex baseline drift. Dynamic Correction Method, on the other hand, uses a dynamic baseline correction algorithm, which adjusts the baseline in real time based on the signal's characteristics, effectively managing complex signal fluctuations and drift and avoiding misjudgments caused by baseline shifts in traditional methods.

When it comes to high salt concentration conditions, particularly during nanoparticle detection, Dynamic Correction Method excels. High salt concentrations often lead to aggregation of analytes, introducing noise and affecting experimental results. Compared to traditional methods, Dynamic Correction Method maintains high sensitivity even under high salt conditions, minimizing the impact of noise on the experimental outcome.

In complex signal processing, Dynamic Correction Method demonstrates strong flexibility,

capable of handling a wide variety of signal types and complex signal patterns. In comparison, some platforms, such as Kleiner Lab Software, perform well with single-type signals but may struggle with overlapping signals or complex backgrounds. Dynamic Correction Method, however, is able to maintain stable performance under various complex signal conditions, ensuring accurate recognition of all target signals.

Finally, Dynamic Correction Method requires low hardware specifications. In contrast to other platforms, such as Cavro Nanopore Sensing, which typically require higher-end equipment, Dynamic Correction Method runs efficiently on low-end devices, reducing the experimental costs and technical barriers, making it suitable for a wider range of laboratory environments.

In summary, Dynamic Correction Method outperforms existing platforms in several areas, including noise management, low-SNR event detection, baseline drift handling, complex signal processing, high salt concentration adaptability, and hardware requirements. It demonstrates unique advantages, particularly in noisy environments and complex experimental conditions, ensuring high signal recognition accuracy and stability.

EXPERIMENT SECTIONS

Nanopore experiment

Nanoscale channels were created on typical SiNx membranes via piercing the nanopores with an electron beam. Chips were prepared and cleaned in a piranha solution (concentrated sulfuric acid with hydrogen peroxide in a volume ratio of 3:1) at 80°C for 30 minutes to enhance their surface hydrophilicity. The nanopore chips were firmly sealed with rubber pads and assembled into polydimethylsiloxane (PDMS) microfluidic channels.

The gold nanoparticles used in the experiment were obtained by the reduction reaction of sodium citrate with chloroauric acid. The polymerases used in the experiments were ordered from Sangon Biotech (Shanghai) Co. For nanopore sensing, silver chloride electrodes with bias voltages were placed on both sides of the device, and analogue current signals captured using the Axopatch 200B patch clamp (Molecular Devices, Inc. Sunnyvale, CA), filtered with a low-pass Bessel filter with a corner frequency of 10 kHz (To effectively suppress noise and preserve the signal, we chose a 10 kHz low-pass filter. This filter is suitable for both the polymerase and gold nanoparticle signals, effectively removing high-frequency noise while retaining key signal features.Since the nanopores are fabricated through dielectric breakdown, which can result in higher noise levels compared to other fabrication methods, the 10 kHz low-pass filter is employed to reduce baseline noise and minimize the interference of high-frequency noise, thereby improving the efficiency of signal extraction while maintaining the integrity of the signal characteristics.), and then digitized with a Digidata 1550B converter at a sampling frequency of 100 kHz. The translocation of analytes across the nanopore was primarily driven by an applied transmembrane voltage, which generates an electric field that induces the movement of charged particles through the pore.Data were recorded by using the PClamp software.

Conductance Calculation

To offer a more comprehensive analysis, we utilized a composite model that takes both bulk conductance and double-layer conductance into account. The model allows us to consider both the ionic conductivity of the bulk solution and the contribution of the electric double layer near the pore surface, offering a more thorough understanding of the total conductance.

The bulk conductance G_{bulk} primarily comes from the ion concentration and ion mobility in the

Analyst Accepted Manuscript

solution, σ_{bulk} is the conductivity of the solution in S/m (4.2 S/m for 1M LiCl); r is the radius of the pore in meters (25 nm); L is the length of the nanopore in meters (30 nm). The formula is:

$$G_{bulk} = \sigma_{bulk} \cdot \frac{\pi r^2}{L}$$

The double layer conductance G_{DL} is related to the charge distribution and interaction between the surface of the nanopore and the electrolyte solution, σ_{DL} is the conductivity of the double layer, typically dependent on surface charge density, ion type, and solution conditions; L_{DL} is the effective thickness of the double layer, which for a 1M LiCl solution is around 0.77 nm. It is calculated as:

$$G_{DL} = \sigma_{DL} \cdot \frac{\pi r^2}{L_{DL}}$$

After calculating both contributions, we found that the bulk conductance is significantly larger than the double-layer conductance. The bulk conductance, calculated as 2.6×10^{-8} S, is the dominant factor in determining the total conductance of the nanopore. While the double-layer conductance does play a role, its contribution is comparatively smaller, with a calculated value of 4.16×10^{-11} S.

$$G_{total} \approx G_{bulk} + G_{DL} = \sigma_{bulk} \cdot \frac{\pi r^2}{L} + \sigma_{DL} \cdot \frac{\pi r^2}{L_{DL}}$$

By considering both factors in our model, we gained a more complete picture of the overall conductance, confirming that the primary influence comes from the bulk ionic conductivity, while the effect of the double layer is minimal, particularly at high salt concentrations such as 1M LiCl. This comprehensive approach ensures that the model used in our analysis is more robust and accurate in predicting the total conductance of the system.

Signal file acquisition suggestions

The resolution of the signal files has a certain impact on data extraction performance, and this limitation primarily arises from the combination of sampling rate and filter settings. While these settings optimize noise suppression and signal clarity, they limit the resolution required for ultra-fast event detection. The following strategies can be employed to improve temporal resolution and capture faster translocation events:

Appropriate Voltage: By reducing the voltage applied across the nanopore, the speed of molecules passing through the pore can be slowed, thereby increasing the duration of the events. This allows more complete signal features to be captured at higher temporal resolution.

Nanopore Modification: Modifying the nanopore surface can further adjust the interaction between molecules and the nanopore, controlling the speed at which molecules pass through. Appropriate modifications help slow down the molecules' flow, enabling translocation events to be captured and analyzed at higher temporal resolution.

Optimizing acquisition parameters: By adjusting filter cutoff frequencies and increasing the sampling rate, we can improve temporal resolution while maintaining an acceptable signal-to-noise ratio. This will enhance the detection of fast translocation events.

These optimization strategies can improve the detection of fast translocation events, providing more complete and reliable raw data for the signal extraction algorithm, thereby enhancing the accuracy and effectiveness of the algorithm.

Program coding

Qt designer platform was used to design the visual interface of the software, Python was used to develop the back-end function of the software, and the key libraries used by the software included pyabf, pyqt5, matplotlib, etc.

RESULTS AND DISCUSSION

Signal model and program algorithm

In this paper, the many aspects of nanopore data analysis has been described in our nanopore signal model. During the recognition of the signal, the event detection thresholds (start and end thresholds determine the start and end points of the event, respectively) are determined based on the magnitude of the change in the current pulse with respect to the baseline; in other words, the accuracy of the value of the baseline determines the outcome of the nanopore signal recognition. To identify the translocation event, the signal recording is segmented in slidable windows, and the local mean and local variance for each window are computed as a means of dynamically assessing baseline drift and determining the event detection threshold. Thus, a nonlinear threshold curve is generated according to the statistical characteristics of the signal in the window, which can flexibly adapt to the dynamic changes of the signal. Especially in the case of large fluctuations of baseline current and mass dense pulse signals. Consequently, the baseline and threshold in real time can be dynamically corrected to resist the disturbance effect to improve the accuracy of event detection. The procedure is split into a series of successive stages, with each stage utilizing parameters determined in the previous stage, as shown in Figure 1.
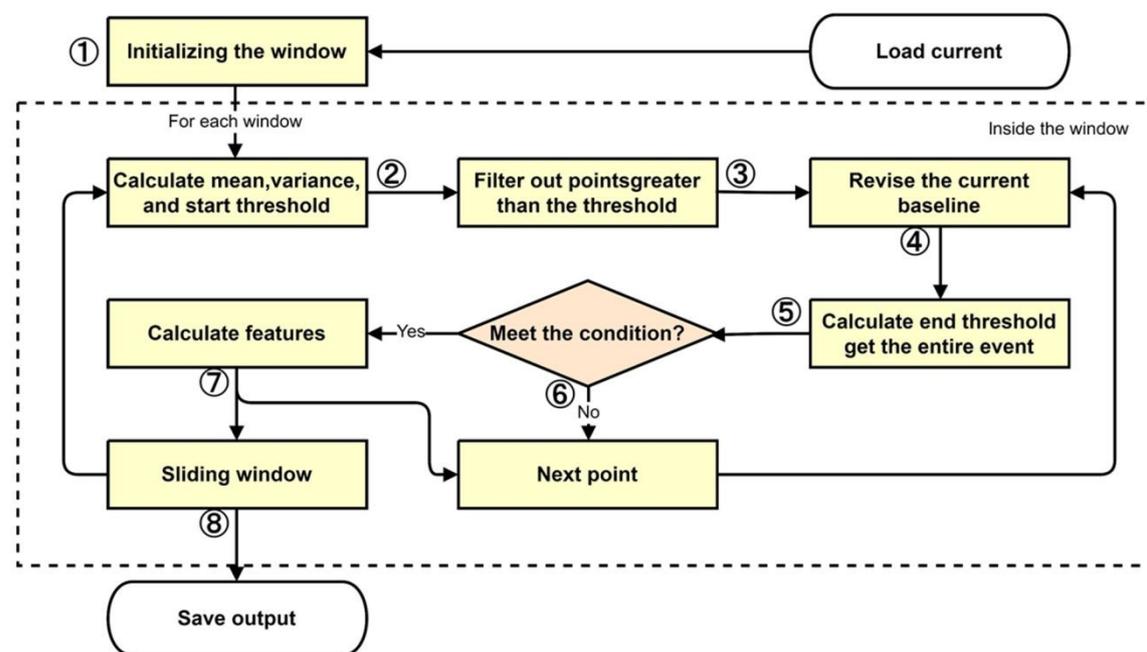

Figure 1. The flow chart of Dynamic correction method.

Step 1: Window Initialization. First, three concepts are clarified. Window size (W) refers to the number of sample points considered at one time when processing the signal. Sampling frequency (f) refers to the number of sample points collected per second in Hz, and the sampling frequency

determines the time resolution of the signals. Time (t) is the actual time covered by the window, which can be calculated from the window size and sampling frequency. The mathematical relationship between the three parameters is expressed as:

$$t = \frac{W}{f} \ (1)$$

Next steps are to set the appropriate window size (W), step size (S) and buffer size (B) and scan the entire signal file by sliding the window. During each movement of batch windows, the data within the window and (the contextual components in the front and rear areas) the buffer before and after it are further adjusted and processed based on cascade feedback. Set the signal as x and the length as N, then a sliding window function $f(k)$ can be defined with k denoting the index of the window and i denoting the signal point index as follows.

$$f(k) = \{x[i] \mid (k-1)S <= i < kS + W\}, for \ k = 1,2,...,N/S \ (2)$$

For each index k, the sliding window function $f(k)$ returns a collection of consecutive data points of number W that are signals x *from* position *(k-1) S* to *kS+V-1*. This formula assumes that the step size S is less than or equal to the window size W, ensuring the overlap between the windows. If the step size S is greater than the window size W, there will be no overlap between the testing windows. This formula also assumes that the signal length N is an integer multiple of the step size S, which ensures that the last data point of the signal is exactly at the end of a window. If the signal length N is not an integer multiple of the step size S, then the last window may contain fewer data points, or the signal may need to be filled or truncated accordingly. The window is then scrolled through the entire signal according to the selected step size. After each sweep, the working data within the window and the connected data near to the windows are ready for next processing.

Step 2: Preliminary Threshold Initialization. The mean variance within the window is the primary part of calculating the preliminary threshold. Here a first-order digital low-pass filter is used to calculate the local mean and variance at each point of the original signal with the following formula.

$$ml(0) = \frac{1}{N}\sum_{i=0}^{N-1} S_{raw}(i) \ (3)$$

$$ml(i) = \alpha * ml(i-1) + (1-\alpha)S_{raw}(i), i = 1,2,... \ (4)$$

$$vl(0) = \frac{1}{N}\sum_{i=0}^{N-1}[S_{raw}(i) - ml(0)]^2 \ (5)$$

$$vl(i) = \alpha * vl(i-1) + (1-\alpha)[S_{raw}(i) - ml(i)]^2, i = 1,2,... \ (6)$$

$$T_{us}(i) = ml(i) + \beta_s * \sqrt{vl(i)} \ (7)$$

$$T_{ds}(i) = ml(i) - \beta_s * \sqrt{vl(i)} \ (8)$$

Here $S_{raw}(i)$ is the i-th point of the original signal. $ml(i)$ and $vl(i)$ are the local mean and local variance of the original signal at the i-th point, respectively. $T_{us}(i)$ and $T_{ds}(i)$ are the global mean and the global variance of the original signal, respectively, and α are the coefficients of the filter.

$T_{us}(i)$ and $T_{ds}(i)$ represent the start thresholds for amplitude increase and decrease, respectively.

The parameter β_s is used to calculate a preliminary threshold for event detection by setting a distance between a signal point and the local mean (baseline level), where the value of this distance is s times the local standard deviation, and signal data with current fluctuations greater than this distance are filtered. the larger the value of s, the higher the preliminary threshold for event detection and the less sensitive the algorithm is to the event.

Step 3: Data Filtering. Filter out some points that are significantly below the current threshold. The filtering action of signals will traverse the entire window and the new thresholds derived from step 2 are constantly set depending on the correction process. If the sudden-change of the current data are more than the specified threshold, an event is considered to be identified and the starting points are recorded in a list.

Step 4: Event Detection Threshold Correction. The Dynamic correction method (Figure 2a) calculates the mean and variance within a localized window and obtains a baseline (red dashed line) and an event detection threshold (blue dashed line). A user-configurable yellow pane is displayed within the signal file. Upon the detection of a recordable event, the window initiates a baseline correction. This involves replacing the mean and variance values within the baseline calculation for the specified area with the mean and variance values at a number of points located prior to and subsequent to the event's start and end points. The number of points included in this replacement is also adjustable. This dynamic adjustment process brings the event detection thresholds of the pane-formatted file closer to the baseline, thereby ensuring greater accuracy in the values of each feature of each recorded event. Concurrently, this process is tantamount to circumventing the interference occasioned by the fluctuating current of the event in question with each instance of its documentation, thus ensuring that the event detection threshold remains unimpaired. In contrast, the baseline scanning method illustrated in Figure 2(b). As events are detected, the cumulative changes in the mean and variance of the entire window can shift the event detection threshold, which may affect signal recognition accuracy. This shift can lead to small signals being masked or the characteristics of detected signals becoming inaccurate. In contrast, the Dynamic Correction Method dynamically adjusts the threshold in real time, mitigating the impact of such shifts. By localizing the threshold change curve, this method better fits the local signal characteristics, enhancing detection sensitivity. Additionally, it allows for more precise extraction of event features, such as duration and amplitude, improving the accuracy and reliability of signal detection, especially in complex or low signal-to-noise ratio environments.
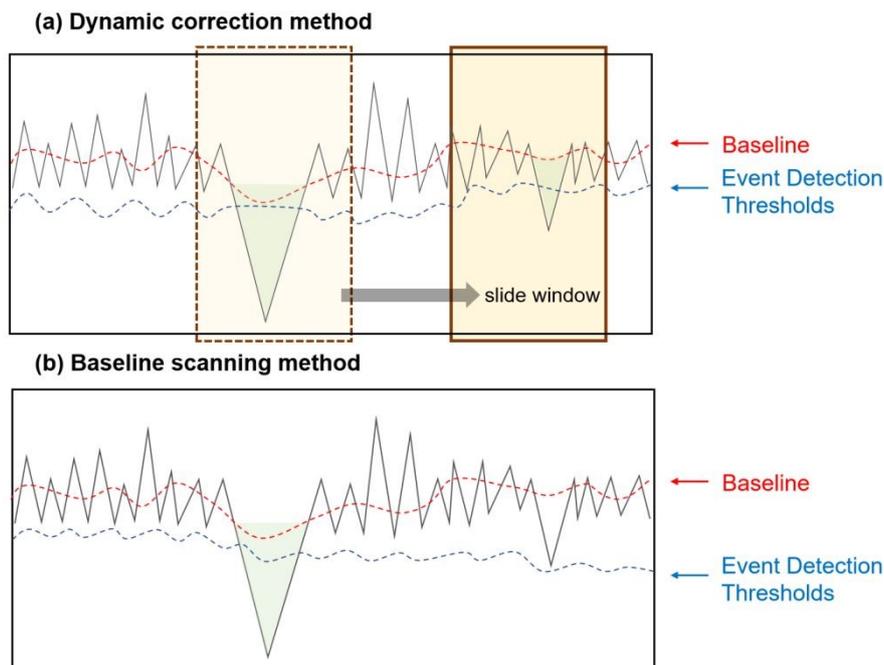
Figure 2: The identification of transition events by different methods. (a) Dynamic correction method; (b) Baseline scanning method.

Step 5: Event identification. As above, the end threshold will be determined based on the local mean and local variance to detect the events in the current traces. While the data between the start and the end of the events are checked back and forth according to the upward and downward threshold to reaches its optimal results. The formula for the event end threshold is shown as below.

$$T_{ud}(i) = ml(i) + \beta_e * \sqrt{vl(i)} \quad (9)$$

$$T_{de}(i) = ml(i) - \beta_e * \sqrt{vl(i)} \quad (10)$$

$T_{ud}(i)$ and $T_{de}(i)$ represent the thresholds with upward and downward amplitude for independent event recognition, respectively. β_e is a parameter used to compute the end threshold.

It controls the distance between the end threshold and the local mean, which is e times the local standard deviation. The larger the value of e, the lower the end threshold, and the more stringent the algorithm is in determining the end of the event. By adjusting the values of s and e, it is possible to regulate the sensitivity of the algorithm to events, as well as its requirement for the duration of events. The application of larger values of s will result in a reduction of the number of events that are identified, although this may result in the omission of some true events. Conversely, the application of larger values of e will increase the duration of events, yet this may result in the misclassification of certain noise as part of the event. In practice, the values of s and e should be set according to the specific characteristics of the signal and the requirements of the application in order to achieve the optimal results in event detection.

Step 6: Screening of Events. The detected events are subjected to screening according to preset conditions with the objective of eliminating possible false positives.

Step 7: Feature extraction. Save the current event information and calculate its features. Output the relevant information of the filtered events into the result array.

Step 8: Slide the window to subsequent data processing. Perform the detection of the translocation event in the next window.

Event characterization for nanopore sensing for polymerase

In order to assess the efficacy of the Dynamic correction method, we conducted experiments using polymerase as the target for detection. The concentration of polymerase used was 50 μM, and the experiments were performed on solid-state nanopores with a pore size of 25 nm and an applied voltage of 400 mV. The nanopore blockage current signal of these experiments are presented in Figure 3 a.

To evaluate the improvement brought about by our algorithm, we analyzed the signal distributions by applying the Dynamic correction method(counts：865)and Baseline scanning method(counts：841).The signal amplitude and width distributions were fitted using a Gaussian function model, and the fitted curves are depicted as dashed lines in the Figure 3. The goodness of fit was assessed using the R-Square values, which were found to be 0.827, 0.877, 0.905, and 0.988 for Figures 3 b-e, respectively. Our analysis revealed that the time of the signals in relation to the bar graph of the blockage current of the half-peak width became narrower after applying the improved algorithm. This indicates that the data became more concentrated. Additionally, the fitted peaks in Figures 3 b-e were determined to be 0.135ms, 0.161ms, 8991.4pA, and 9003.2pA, respectively. The consistency of these peaks suggests that our algorithm optimizes the identification process while maintaining the accuracy of the data. Figure 3f compares the signal data obtained using Dynamic correction method(red) with the data obtained using Baseline scanning method (black). The data points derived from our algorithm are concentrated in the center region, which is crucial for the classification of solid-state nanopore signals. This observation suggests that we can obtain various features of the sample using a smaller sample size. Overall, our findings demonstrate the effectiveness of our algorithm in improving the analysis of solid-state nanopore signals.
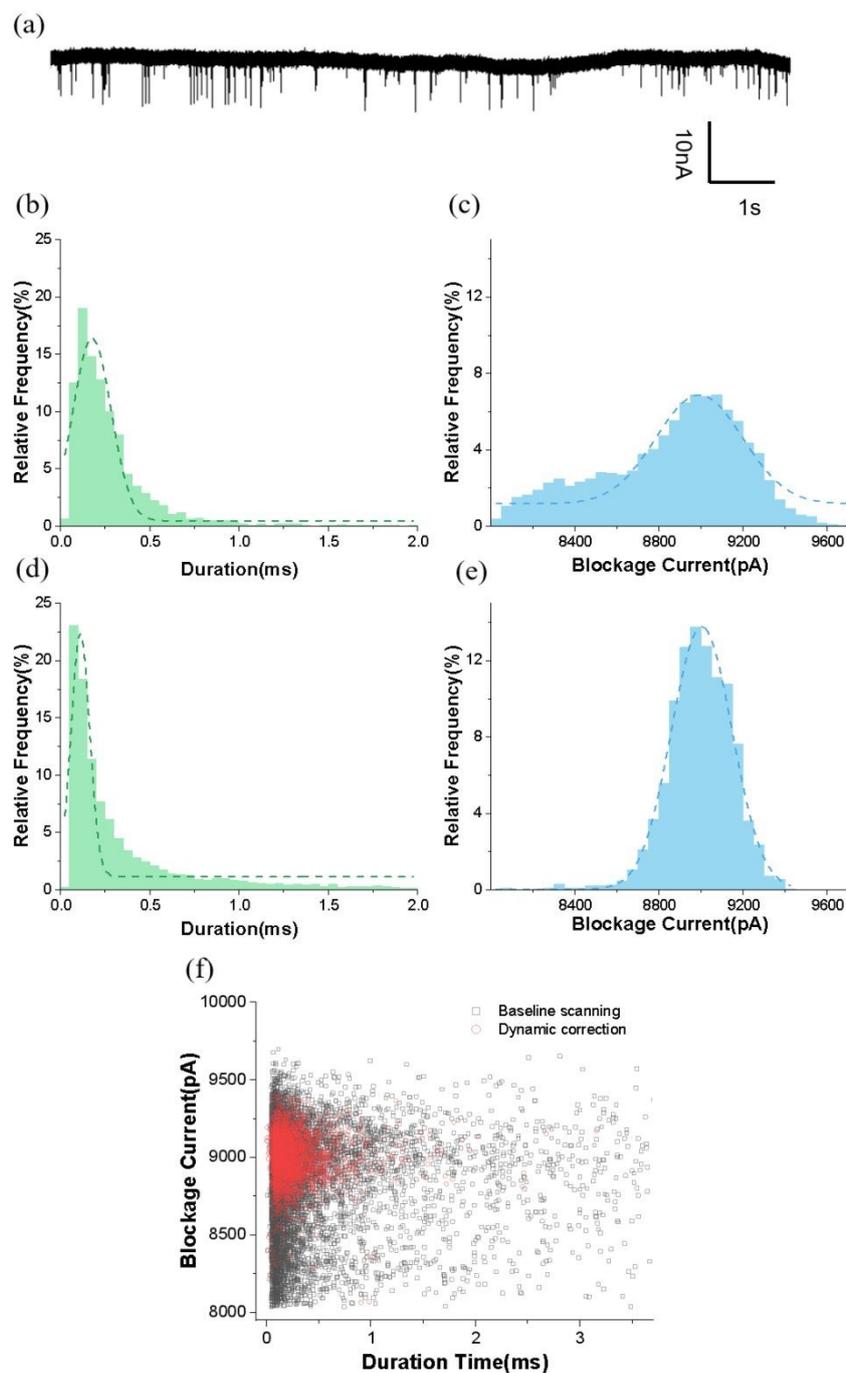
Analyst Accepted Manuscript

Figure 3. Enhancement of data accuracy after using the Dynamic correction method and Baseline scanning method. (a)The nanopore blockage current signal graph of polymerase. (b) Duration distribution of the nanopore signal by Baseline scanning method; (c)Blockage current distribution of the nanopore signal by Baseline scanning method; (d) Duration distribution of the nanopore signal by Dynamic correction method; (e) Blockage current distribution of the nanopore signal by Dynamic correction method; (f) Scatter plot of the signals by two methods.

Dynamic correction method for classification of signals from gold nanoparticles

Due to the limitations of the micro- and nanofabrication process, the baseline current of solid-state nanopores often fluctuates to a large extent, which can affect the accuracy of signal recognition. At the same time, the extent of such fluctuations is also related to the substance to be measured. For

example, proteins are more likely to cause baseline fluctuations compared to DNA, and the larger the spatial volume of the substance, the more likely it is to form a collision process during the pore entry process, thus causing baseline fluctuations. The more rigid metal nanoparticles are more likely to cause irregular oscillations of the nanopore baseline currents because they tend to form clusters in the salt solution. As shown in Figure 4(a), we detected two kinds of gold nanoparticles with particle sizes of 10 nm versus 15 nm (both with a concentration of 20 nM) using a solid nanopore with a pore size of 20 nm, and the mixed samples of metal particles made the nanopore baseline fluctuation very unstable.In this study, the term "ionic current fluctuations" refers to variations in the ionic current signal caused by multiple factors. These fluctuations can be classified into two types: slow fluctuations and fast fluctuations. Slow fluctuations primarily arise from electrochemical reactions occurring near the nanopore orifice, leading to gradual changes in the baseline current. In contrast, fast fluctuations are typically caused by the dynamic oscillatory motions of particles at the nanopore orifice, resulting in rapid and short-lived changes in the ionic current. Notably, the nanopore chips used in this study were fabricated using a dielectric breakdown method, which differs from traditional approaches involving high-energy electron or ion beam drilling. This method offers advantages such as low cost and rapid pore formation. The process involves initially creating a small pore via breakdown and then gradually enlarging it, resulting in a more loosely configured pore structure. This structural characteristic makes ionic current fluctuations, particularly those caused by particle motion, more likely to occur. Figure 4(b) shows the signal counting using Baseline scanning method to count the signals, and the number of signals for the two kinds of particles is 128(red) and 405(green), respectively, and Figure 4(c) shows the signal counting graph after using Dynamic correction method, and the number of signals for the two kinds of particles is elevated to 418 (red) and 691 (green), respectively, and it can be clearly seen from the graphs that, at the same time of elevating the number of signals recognized, the aggregation and distribution of signals are basically have not changed significantly. In particular, it should be noted that the concentration of 10nm and 15nm gold nanoparticles is 20nM, and theoretically the capture probability of the two is not much different, but the capture frequency ratio under the Baseline scanning method is only 31% (128/405). This means that baseline fluctuations and signal complexity will affect the acquisition of small signals. However, when we use the Dynamic correction method to process the same batch of files, the capture frequency ratio increases to 60% (418/691), which reflects the improvement of threshold adjustment for complex sample signal acquisition, especially when the baseline fluctuation is obvious.
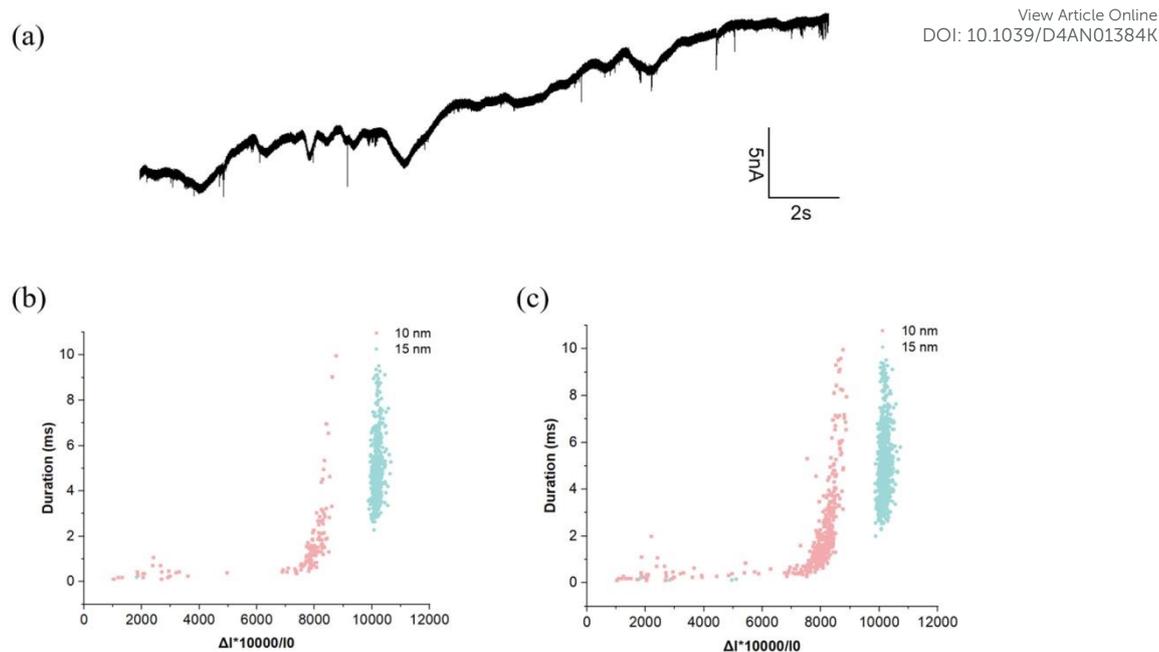
Figure 4. (a) Signal recognition of nanopore signals at fluctuating baseline (test sample is a mixture of 10nm and 15nm gold nanoparticles); (b) Distribution of nanopore signals detected by Baseline scanning method; (c) Distribution of nanopore signals detected by Dynamic correction method.

We have counted the signal distributions obtained by the Baseline scanning method and Dynamic correction method, and the results are shown in Figure 5, where (a)(b) shows the distribution of blocking current versus pore time for 10 nm gold nanoparticles, and (c)(d) shows the distribution of blocking current versus pore time for 15 nm gold nanoparticles. The most significant difference is 0.041 ms in Figure 5(b) (1.028 ms and 1.069 ms for the Baseline scanning method and Dynamic correction method, respectively), while the difference in blocking current is much smaller, with 86.45 pA and 53.70 pA for 10 nm and 15 nm gold nanoparticles, respectively. This demonstrates that Dynamic correction method can significantly improve the signal recognition rate while maintaining the signal characteristics, especially for smaller current signals that are missed due to baseline fluctuations.
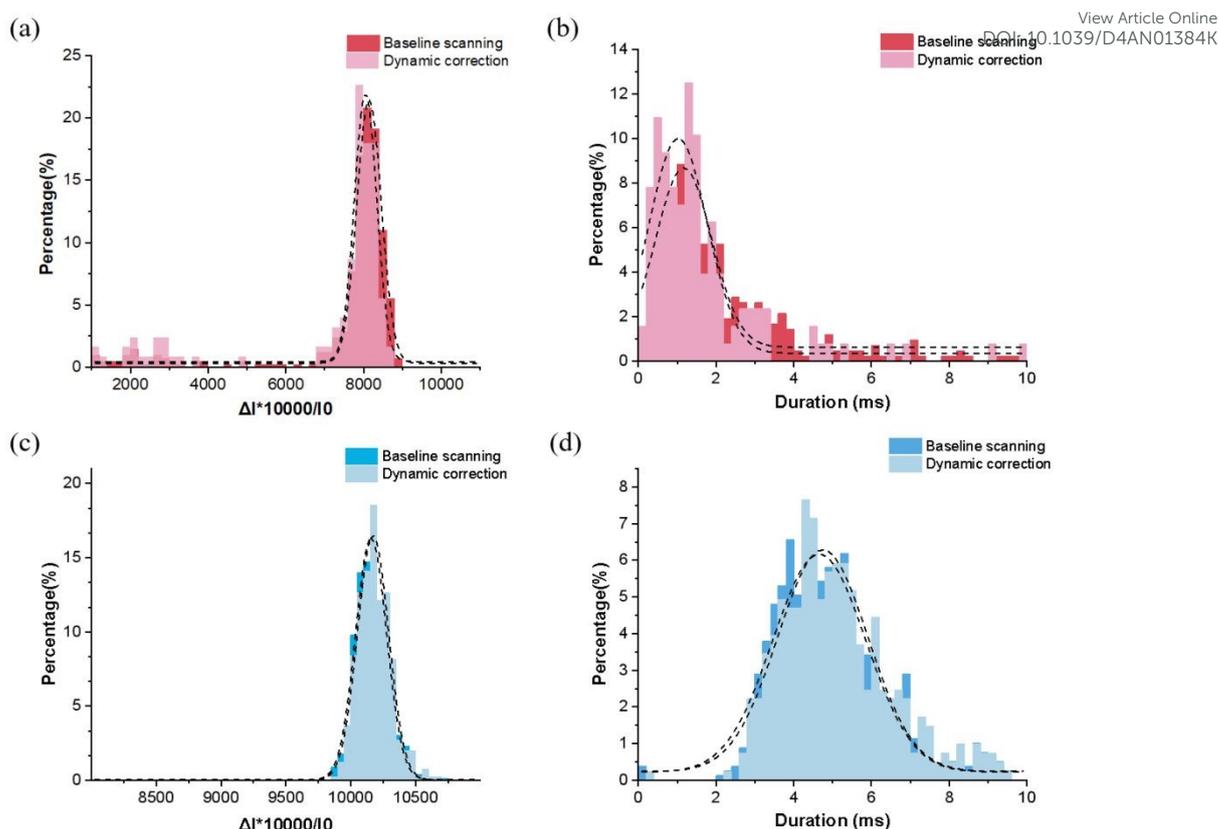
Figure 5 : Signal distributions obtained by Baseline scanning method and Dynamic correction method, (a)(b) for current versus transit time for 10 nm gold nanoparticles, (c)(d) for current versus transit time for 15 nm gold nanoparticles.

CONCLUSION

Overall, we developed a software program to perform baseline correction in real time using Dynamic correction method, overcoming the interference of baseline fluctuations of solid-state nanopores and solving the problem of difficult detection of small and medium signals in signal-dense regions. This improves the acquisition rate and accuracy of nanopore signals. The data obtained from biomolecules, including enzymes and nanoparticles, after the application of our dynamic correction method to solid-state nanopores demonstrates effective signal recognition. This study encompasses a range of topics related to biochemical experiments, signal processing, and data analysis. Furthermore, it offers valuable insights into solid-state nanopore signal research by employing the technique of dynamic correction of event detection thresholds to address the current challenges in solid-state nanopore signal analysis.

ASSOCIATED CONTENT

The clients for Windows, Linux, and MacOS operation systems are available and can be downloaded from *https://github.com/eventdetector/Event-Detector*.
The source code can be downloaded from *https://github.com/eventdetector/Event-Detector-Code*.

AUTHOR CONTRIBUTIONS

GX conceived and designed the study, developed the program, conducted the experiments, and

drafted the original manuscript. JS was responsible for module construction. JM performed data analysis. LW revised the manuscript. JT supervised the study, managed the project administration, and finalized the manuscript.

ACKNOWLEDGEMENT

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

References

(1) Dekker, A. S. C. S. J. C. H. Z. J. J. C. Fast DNA Translocation through a Solid-State Nanopore. *Nano Lett.* 2005, *5* (7), 1193-1197.

(2) Dekker; Nanotechnology, C. J. N. Solid-state nanopores. *Nature Nanotechnology* 2007, *2* (4), 209-215.

(3) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. J. N. B. The potential and challenges of nanopore sequencing. *Nature Biotechnology* 2008, *26* (10), 1146-1153.

(4) Howorka, S.; Siwy, Z. J. C. S. R. Nanopore analytics: sensing of single molecules. *Chemical Society Reviews* 2009, *38* (8), 2360-2384.

(5) Cao; Chan; Long; Yi-Tao. Biological Nanopores: Confined Spaces for Electrochemical Single-Molecule Analysis. *Accounts of Chemical Research* 2018, *51* (2), 331-341.

(6) Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Akeson, M. Automated Forward and Reverse Ratcheting of DNA in a Nanopore at Five Angstrom Precision. *Nature Biotechnology* 2012, *30* (4), 344-348.

(7) Rang, F. J.; Kloosterman, W. P.; Jeroen, D. R. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology* 2018, *19* (1),

90.

(8) Kyle; Briggs; Harold; Kwok; Vincent; Tabard-Cossa. Automated Fabrication of 2-nm Solid-State Nanopores for Nucleic Acid Analysis. *Small* **2014**, *10* (10), 2077-2086.

(9) De Zoysa, R. S. S.; Jayawardhana, D. A.; Zhao, Q.; Wang, D.; Armstrong, D. W.; Guan, X. Slowing DNA translocation through nanopores using a solution containing organic salts. *Journal of Physical Chemistry B* **2009**, *113* (40), 13332-13336.

(10) Rozevsky, Y.; Gilboa, T.; Kooten, X. F. V.; Kobelt, D.; Meller, A. Quantification of mRNA Expression Using Single-Molecule Nanopore Sensing. *ACS Nano* **2020**, *14* (10), 13964-13974.

(11) Diego, C.; Nicoletta, G.; Agne, K.; Pierre-Eugene, C.; Benoit, C.; Jean-Marc, J.; Sebastien, B. Unexpected Hard Protein Behavior of BSA on Gold Nanoparticle Caused by Resveratrol. *Langmuir : the ACS journal of surfaces and colloids* **2018**, *34* (30), 8866-8874.

(12) Arima, A.; Harlisa, I. H.; Yoshida, T.; Tsutsui, M.; Tanaka, M.; Yokota, K.; Tonomura, W.; Yasuda, J.; Taniguchi, M.; Washio, T. Identifying Single Viruses Using Biorecognition Solid-State Nanopores. *Journal of the American Chemical Society* **2018**, *140* (48), 16834-16841.

(13) Chae, H.; Kwak, D. K.; Lee, M. K.; Chi, S. W.; Kim, K. B. Solid-state nanopore analysis on conformation change of p53TAD–MDM2 fusion protein induced by protein–protein interaction. *Nanoscale* **2018**, *10* (36), 17227-17235.

(14) Kowalczyk, S. W.; Hall, A. R.; Dekker, C. Detection of Local Protein Structures along DNA Using Solid-State Nanopores. *Nano Lett.* **2010**, *10* (1), 324–328.

(15) Coglitore, D.; Coulon, P. E.; Janot, J. M.; Balme, S. Revealing the Nanoparticle-Protein Corona with a Solid-State Nanopore. *Materials* **2019**, *12* (21).

(16) Raveendran, M.; Leach, A. R.; Hopes, T.; Aspden, J. L.; Actis, P. Ribosome Fingerprinting

with a Solid-State Nanopore. *ACS Sensors* **2020**, *5* (11), 3533-3539.

(17) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. The potential and challenges of nanopore sequencing. *Nature Biotechnology* **2008**, *26* (10), 1146-1153.

(18) Yusko, E. C.; Bruhn, B. R.; Eggenberger, O. M.; Houghtaling, J.; Rollings, R. C.; Walsh, N. C.; Nandivada, S.; Pindrus, M.; Hall, A. R.; Sept, D. Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nature Nanotechnology* **2017**, *12* (4), 360-367.

(19) Yang, W.; Laura.Restrepo-Pérez; Bengtson, M.; Heerema, S. J.; Birnie, A.; Torre, J. V. D.; Dekker, C. Detection of CRISPR-dCas9 on DNA with Solid-State Nanopores. *Nano Lett.* **2018**, *18* (10), 6469-6474.

(20) Wang, M.; Fu, A.; Hu, B.; Tong, Y.; Liu, T. Virus Detection: Nanopore Targeted Sequencing for the Accurate and Comprehensive Detection of SARS‐CoV‐and Other Respiratory Viruses (Small 32/2020). *Small* **2020**, *16* (32), 2002169.

(21) Tang, H.; Wang, H.; Yang, C.; Zhao, D.; Li, Y. A novel strategy for nanopore-based selective detection of single carcinoembryonic antigen (CEA) molecules. *Analytical Chemistry* **2020**, *92*, 3042-3049.

(22) Tabard-Cossa, V.; Wiggin, M.; Trivedi, D.; Jetha, N. N.; Marziali, A. Single-Molecule Bonds Characterized by Solid-State Nanopore Force Spectroscopy. *ACS Nano* **2009**, *3* (10), 3009-3014.

(23) Sun, Z.; Liu, X.; Liu, W.; Li, J.; Yang, J.; Qiao, F.; Ma, J.; Sha, J.; Li, J.; Xu, L. Q. AutoNanopore: An Automated Adaptive and Robust Method to Locate Translocation Events in

Analyst Accepted Manuscript

Solid-State Nanopore Current Traces. *ACS omega* **2022**, *7* (42), 37103-37111.

(24) Y. M. N. D. Y. Bandara; Shankar Dutt; Buddini I. Karawdeniya; Jugal Saharia; Patrick Kluth; Tricoli, A. A Robust Parallel Computing Data Extraction Framework forNanopore Experiments. *Small Methods* **2024**, *8* (12), 240045.

(25) Kluth, Y. M. N. D. Y. B. S. I. K.; Kim, M. J. Nanopore Data Analysis: Baseline Construction and Abrupt Change-Based Multilevel Fitting. *Analytical Chemistry* **2021**, *93* (34), 11710-11718.

(26) Raillon, C.; Granjon, P.; Graf, M.; Steinbock, L. J.; Radenovic, A. Fast and automatic processing of multi-level events in nanopore translocation experiments. *Nanoscale* **2012**, *4* (16), 4916-4924.

(27) Plesa, C.; Dekker, C. Data analysis methods for solid-state nanopores. *Nanotechnology* **2015**, *26* (8), 084003.

(28) Forstater, J. H.; Briggs, K.; Robertson, J.; Ettedgui, J.; Marie-Rose, O.; Vaz, C.; Kasianowicz, J. J.; Tabard-Cossa, V.; Balijepalli, A. J. A. C. MOSAIC: A Modular Single-Molecule Analysis Interface for Decoding Multistate Nanopore Data. *Analytical Chemistry* **2017**, *88* (23), 11900-11907.

(29) Tu, J.; Meng, H.; Wu, L.; Xi, G.; Fu, J.; Lu, Z. EasyNanopore: A Ready-to-Use Processing Software for Translocation Events in Nanopore Translocation Experiments. *Langmuir : the ACS journal of surfaces and colloids* **2021**, *37* (33), 10177-10182.

(30) Zhang, J.; Liu, X.; Ying, Y. L.; Gu, Z.; Meng, F. N.; Long, Y. T. High-bandwidth nanopore data analysis by using a modified hidden Markov model. *Nanoscale* **2017**, *9* (10), 3458-3465.

(31) Tsutsui, M.; Takaai, T.; Yokota, K.; Kawai, T.; Washio, T. Deep Learning-Enhanced Nanopore Sensing of Single-Nanoparticle Translocation Dynamics. *Small Methods* **2021**, *5* (7),

e2100191.

(32) Wei, Z. X.; Ying, Y. L.; Li, M. Y.; Yang, J.; Long, Y. T. Learning shapelets for improving the single-molecule nanopore sensing. *Analytical Chemistry* **2019**, *91* (15), 10033-10039.

(33) Dematties, D.; Wen, C.; Pérez, M. D.; Zhou, D.; Zhang, S. L. Deep learning of nanopore sensing signals using a bi-path network. *ACS nano* **2021**, *15* (9), 14419-14429.

(34) Dematties, D.; Wen, C.; Zhang, S. L. A Generalized Transformer-Based Pulse Detection Algorithm. *ACS sensors* **2022**, *7* (9), 2710-2720.

(35) Misiunas; Karolis; Ermann; Niklas; Keyser; Ulrich, F. QuipuNet: Convolutional Neural Network for Single-Molecule Nanopore Sensing. *Nano Lett.* **2018**, *18* (6), 4040-4045.

The clients for Windows, Linux, and MacOS operation systems are available and can be downloaded from *https://github.com/sujinmeng/Event-Detector*.

The source code can be downloaded from *https://github.com/sujinmeng/Event-Detector-Code*.

All data included in this article is available from the authors on request.