ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

AutoNanopore: An Automated Adaptive and Robust Method to Locate Translocation Events in Solid-State Nanopore Current Traces

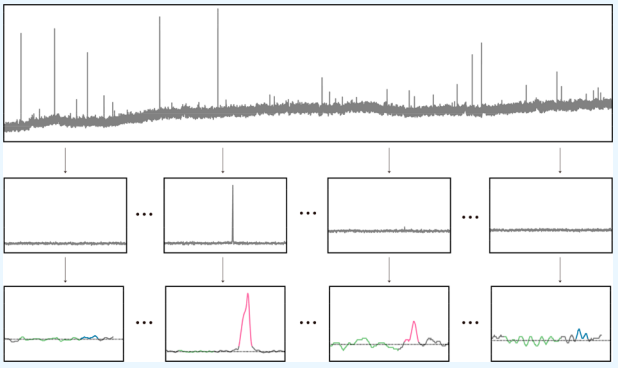
Zepeng Sun,* Xinlong Liu, Wei Liu, Jiahui Li, Jing Yang, Feng Qiao, Jianjun Ma, Jingjie Sha, Jian Li,* and Li-Qun Xu*

Cite This: *ACS Omega* 2022, 7, 37103−37111

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Solid-state nanopore sequencing has shown impressive performances in several research scenarios but is still challenging, mainly due to the ultrafast speed of DNA translocation and significant noises embedded in raw signals. Hence, event detection, aiming to locate precisely these translocation events, is the fundamental step of data analysis. However, existing event detection methods use either a user-defined global threshold or an adaptive threshold determined by the data, assuming the baseline current to be stable over time. These disadvantages limit their applications in real-world application scenarios, especially considering that the results of different methods are often inconsistent. In this study, we develop an automated adaptive method called AutoNanopore, for fast and accurate event detection in current traces. The method consists of three consecutive steps: current trace segmentation, current amplitude outlier identification by straightforward statistical analyses, and event characterization. Then we propose ideas/metrics on how to quantitatively evaluate the performance of an event detection method, followed by comparing the performance of AutoNanopore against two state-of-the-art methods, OpenNanopore and EventPro. Finally, we examine if one method can detect the overlapping events detected by the other two, demonstrating that AutoNanopore has the highest coverage ratio. Moreover, AutoNanopore also performs well in detecting challenging events: e.g., those with significantly varying baselines.



1. INTRODUCTION

Nanopore sequencing technology, which delivers ultralong reads and portable devices, is playing an increasingly important role in life sciences and molecular biology. As one of the two main categories, protein nanopore sensors have been effectively used in many sequencing scenarios, such as human whole genome sequencing,[1] SARS-CoV-2 genome analysis,[2] pathogen identification,[3] disease-causing variant identification,[4] etc. Meanwhile, during the past decade, remarkable progress has been achieved for solid-state nanopore sensors. In particular, solid-state nanopores have been used in DNA/RNA conformation detection,[5] protein fingerprinting,[6] and biomarker immunoassays.[7] Compared to protein nanopores, solid-state nanopores fabricated with SiNx, SiO_2, or MoS_2 are less vulnerable to the environment and are easier to integrate. Despite these apparent advantages, solid-state nanopore sequencing is hugely challenging, mainly due to the ultrafast speed of nucleotide translocation and significant noises in raw signals.[8−11]

Apart from the fabrication of solid-state nanopores, processing solid-state nanopore raw signals is also crucial.[12,13] So far, the processing has been mainly concentrated on the translocation event detection and the follow-up analyses of the events. Most existing event detection methods are essentially outlier identification in current values: the current variations are computed/ranked and cut off at a specific threshold. Based on the principle of how a cutoff threshold is determined, the event detection methods can simply be classified into two categories: (1) classical methods which select user-defined, and usually global, thresholds such as MiniAnalysis,[14] Easy Electrophysiology,[15] and Clampfit;[16] (2) more advanced methods which use adaptive and local thresholds determined by the data, such as OpenNanopore,[17] MOSAIC,[18] Transalyzer,[19] EasyNanopore,[20] EventPro,[21] etc. Following the event detection results, machine-learning models have been proposed to analyze the detected events: e.g., Carral et al. developed a deep-learning method to distinguish single nucleotides at high accuracies[22] and Xia et al. recently proposed a machine-learning-based method to classify signals generated by four synthetic glycosaminoglycans through

Received: May 11, 2022
Accepted: September 28, 2022
Published: October 14, 2022

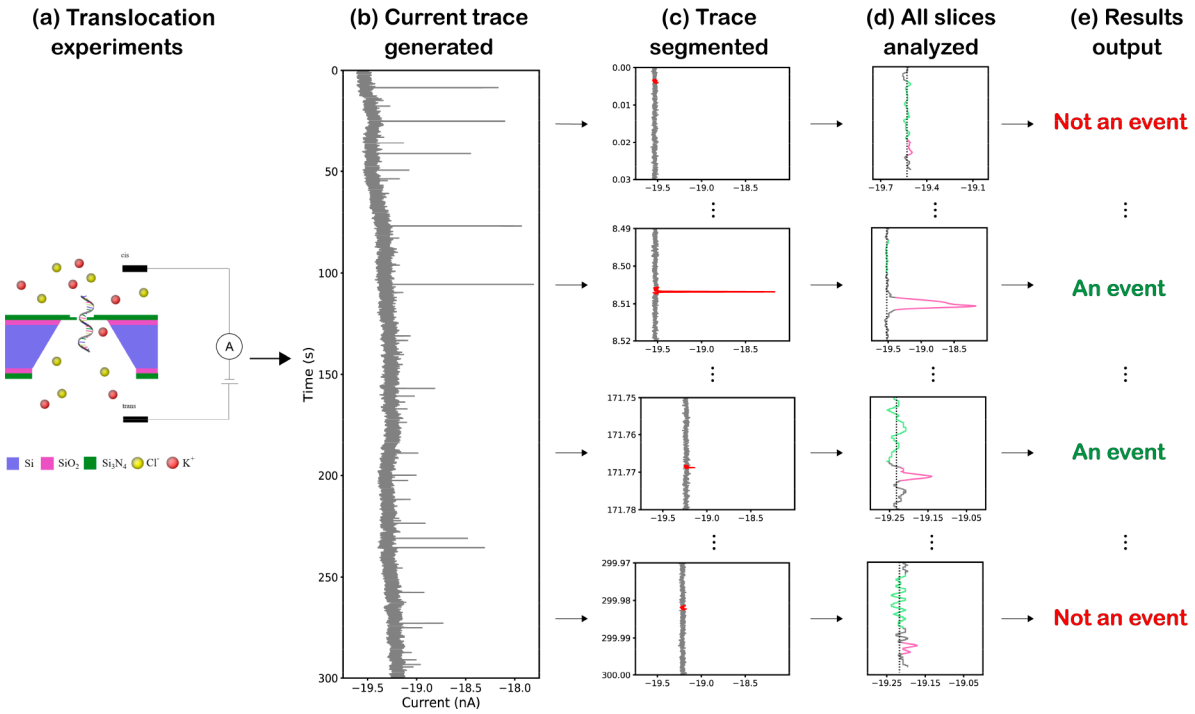


Figure 1. An exemplary close-up view of the proposed event detection method performed on data file 1.abf (refer to Figure 2 for the complete results). (a) Illustration of the translocation experiment. (b) The current traces over time, plotted in a vertical direction. (c) Four examples of the segmented slices, each being 30 ms in time. In each slice, the red part shows the range around the peak and is then analyzed. (d) The red part in each slice is enlarged and analyzed: the green part of the curve indicates a phase absent of translocation events, and the baseline current value (denoted as a horizontal dashed curve) is computed as the average of this part, and the pink curve shows the potential event range. (e) The detection results, based on the amplitude of each peak corresponding to its baseline.

solid-state nanopores.[23] There is also another study where a deep-learning method was developed for postdenoising of ionic current in a nanofluidic channel having five pairs of nanoprotrusions.[24]

Although these studies have proposed impressive machine-learning tools to effectively analyze translocation events, accurate event detection is still the first and most essential step toward analyzing the raw electrical signals. Classical event detection tools perform the task by simply comparing the amplitudes to a global threshold. Unfortunately, choosing a suitable threshold is tricky in that it requires sophisticated manual observations of the signals, which depends heavily on the expertise of researchers; hence, the task is quite labor-intensive. Of those advanced tools, OpenNanopore[17] is an excellent one because it uses a modified cumulative sums algorithm (CUSUM), where a threshold is determined by the data without additional human intervention; the tool EasyNanopore[20] has also been proposed where the filtering threshold is defined as the mean value plus a multiple of the standard deviation of all the current values before a specific data point. While these tools are impressive, they need to compute a wide range of data points in a recursive form and are thus computationally demanding, though parallel processing can accelerate the computation. More importantly, they require the baseline of the current traces to be stable.[17] To resolve the baseline variation issue, the tool EventPro[21] was proposed, where a baseline construction was performed first, and then an event was identified when either the amplitude in a fixed window size exceeded a multiple of the standard deviation of the baseline or the amplitude exceeded a user-defined threshold. This tool turns out to be capable of optimizing the event detection by avoiding the influences of baseline variations, but it requires a considerable amount of computation to perform the baseline fitting. Thus, it is urgent to develop a fast, automated, and accurate method to perform the event detection task in a robust manner. To this end, such a method should incorporate a more straightforward outlier identification strategy and concentrate on the data points near the "peaks" caused by potential translocation events rather than considering a wide range of data points.

This study aims to address the aforementioned issues and develop AutoNanopore—a fast and accurate event detection method in solid-state nanopore sequencing raw signals. It can detect the events in a vast amount of data points efficiently. The article is organized as follows. Section 2 describes the motivation and setup of the experiments and the detailed implementation steps of AutoNanopore, our proposed translocation events detection method. Then, Section 3 presents the results of AutoNanopore on a data set consisting of 23 abf files. Next, Section 4 discusses the strengths and limitations of AutoNanopore and its future research direction, and finally, the paper concludes with Section 5.

2. METHOD

This section first explains the current trace acquisition process for solid-state nanopore translocation experiments, followed by a detailed presentation of the AutoNanopore translocation event detection method.

2.1. Data Acquisition. To acquire raw solid-state nanopore current traces, we added single-strand DNA (ssDNA) molecules with a concentration of 500 pM/L into the trans-chamber. Then a positive voltage of 300 mV was applied to the trans-chamber, and the cis-chamber was electrically grounded. Two Ag/AgCl

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111

electrodes were placed in the electrolyte bath at both ends of the nanopore with a diameter of 14 nm, and the transmembrane voltage was set to generate an ionic current. The current trace was then measured by a resistive feedback amplifier (Axon MultiClamp 700B) at a bandwidth of 250 kHz and a 10 kHz low-pass filter. All single-stranded DNA in the experiments has 22 nucleotides (3′-TCAA CATC AGTC TGAT AAGC TA-5′). All of the current traces were generated during one experiment, and the files were stored every 5 min. Finally, 23 axon binary format (abf) files were generated during the experiments; each file lasts for 5 min and thus contains 75 million data points, approximately 143MB in size. All experiments were carried out in a dark Faraday cage. More details about the experiment preparation and nanopore fabrication can be found in the Supporting Information.

2.2. Event Detection. When an ssDNA with 22 nucleotides translocates through the nanopore, it gives rise to an abrupt change in the current amplitude, and such an occurrence earmarks a translocation event. The events are normally distributed in a stochastic and sparse manner in the raw traces; we aim to identify and locate them automatically. In the following, we mainly discuss the positive-going cases in which a translocation substantially increases the current value; the negative-going cases are processed in the same fashion.

Our proposed event detection method illustrated in Figure 1 includes the following five steps.

2.2.1. Step 1 (Segmentation). A single abf file is equally split into a predefined number of slices. In this study, the default window size is 30 ms, and so each file is split into 10000 slices (contains 7500 data points). Let the peak value (maximum value) found in slice i be p_i, where $i \in [1,10000]$, and the index of the peak point in slice i be k_i. Note that 30 ms per slice represents a compromise between the performance of AutoNanopore and the computational complexity for our 5 min traces. We have also tested different segmentation strategies and have concluded that, for a slice length between 6 ms and 60 ms, the performance of AutoNanopore is not significantly affected. However, for short current traces, we recommend reducing the slice's window size to favor the follow-up statistical analyses.

2.2.2. Step 2 (Baseline Search and Amplitude Computation). For each slice i, first, we need to select a suitable range for 0.5 ms (containing 125 data points) before the peak point k_i; in our experience, a range closer to the peak will better reflect the actual variation trend due to the variation of the overall baseline. To ensure the choice of a suitable range, we perform a backward search from the peak until the Z-score of the peak current value is higher than 3, considering the current values in the range $[k_i - m_i, k_i]$, where m_i is the amount of the push-back data points used for baseline search. Next, we take the average value of the range $[k_i - m_i - 125, k_i - m_i]$ as the baseline corresponding to peak p_i, denoted by b_i. Figure 1 gives an illustration of the process, where the green part depicts the selected range; the amplitude of the peak is then defined as

$$a_i = p_i - b_i$$

2.2.3. Step 3 (Amplitude Outlier Identification). A straightforward statistical analysis is performed for all 10000 amplitudes: the amplitudes are sorted, the first quantile Q1 and the third quantile Q3 are found, and the interquantile range is defined as IQR = Q3 − Q1. Next, the amplitudes that satisfy the following condition are identified:

$$a_i > \mathrm{Q3} + \theta \times \mathrm{IQR}$$

The peaks of the identified amplitudes become outliers among all amplitudes and would most probably occur due to translocation events, denoted by e_j; the corresponding peak value, baseline, and amplitude for each e_j are then denoted as p_j^e, b_j^e, and a_j^e, respectively. Note that the parameter θ determines the selection threshold, and the default value is set to 1.5. The IQR of all the amplitudes may vary significantly under different circumstances, requiring an adjustment in the value of θ to ensure accurate detection. We strongly suggest that the value should not be smaller than 1.5, since amplitudes below Q3 + 1.5 × IQR are not thought to be outliers, according to statistical principles.

2.2.4. Step 4 (Event Characterization). Since each event e_j corresponds to a local maximum current value, there must be a rising period before the peak p_j^e and a falling period after the peak. Similar to the method proposed by OpenNanopore, we set the starting point of an event as the first point before the peak whose current value is below $b_j^e + 0.1 \times a_j^e$, while the ending point is the first point after the peak whose value crosses $b_j^e + 0.1 \times a_j^e$. For each event e_j, we locate the starting and ending time by backward and a forward searches, denoted by $t_{\mathrm{start},j}$ and $t_{\mathrm{end},j}$, respectively. The orange curve indicates the time range of the event in Figure 1; the duration of e_j is obtained as $d_j = t_{\mathrm{end},j} - t_{\mathrm{start},j}$. We choose $b_j^e + 0.1 \times a_j^e$ as the threshold for starting and ending times, considering that with the ultrafast speed of the nucleotide translocation the nucleotides should not yet have entered the nanopore when the current value increases strictly to the baseline; similarly, all of the nucleotides should already have passed through the nanopore before the current value drops strictly to the baseline. In other words, the translocation only starts slightly later but finishes slightly earlier than the changes in values of the current.

2.2.5. Step 5 (Optimization). The event detection task is finished when the above four steps are done. However, some events may have a much larger duration compared to others due to the occasional clogging of the nanopores. We perform the same analysis as presented in Step 3, filter out the events having large durations that are thought to be outliers, and keep the remaining events to do further optimization. Moreover, some events may still have small amplitudes, increasing the probability of them being false detections. We score each event by examining how its amplitude is far away from the maximum amplitude: we sort all the amplitudes in ascending order, for each amplitude a_k, we compute

$$\frac{a_k - a_1}{\frac{1}{k}\sum_{i=1}^{k} a_i}$$

until its value is higher than a threshold (the default is 0.1). Here $a_i < a_j$ for all $i < j$. Finally, we select all the events with amplitudes higher than a_k to be confident events, and the others are thought to be low-confidence events. In summary, by following the five steps in our proposed AutoNanopore method, we can rapidly locate the translocation events together with their characterizations, each including the peak value, baseline, amplitude, and the starting and ending time points. Note that, in the discussion above, we present event detection in positive-going cases. However, our method is also applicable to event detection in negative-going cases, where the translocation substantially reduces the current value. To perform event detection tasks in negative-going cases, we only need to make several minor adjustments: in step 1, we pick the minimum value in each slice as the peak p_i; in step 2, the amplitude is defined as $b_i - p_i$; in step

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111

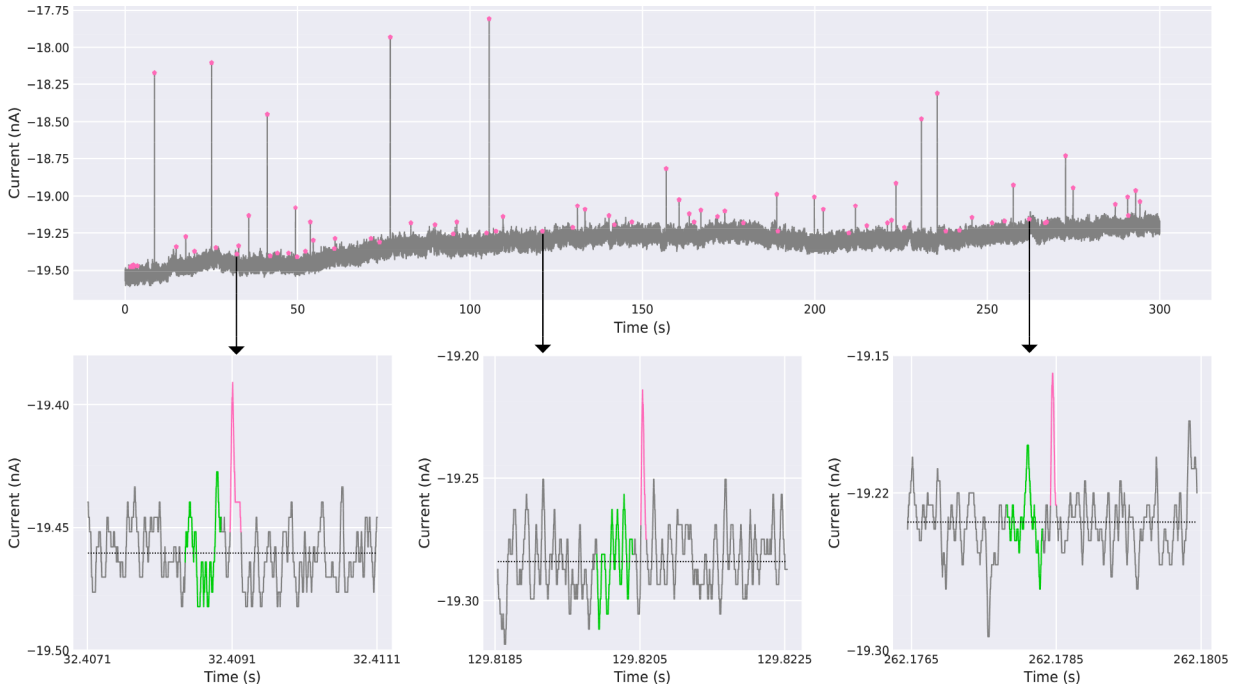


Figure 2. Illustration of the events that AutoNanopore detects, on file 1.abf. In the top panel, the gray curves show the current variation over time and the pink stars indicate the events that AutoNanopore detects; in the bottom panels, the enlarged current traces of three events (with relatively low amplitudes) are given for a better illustration.

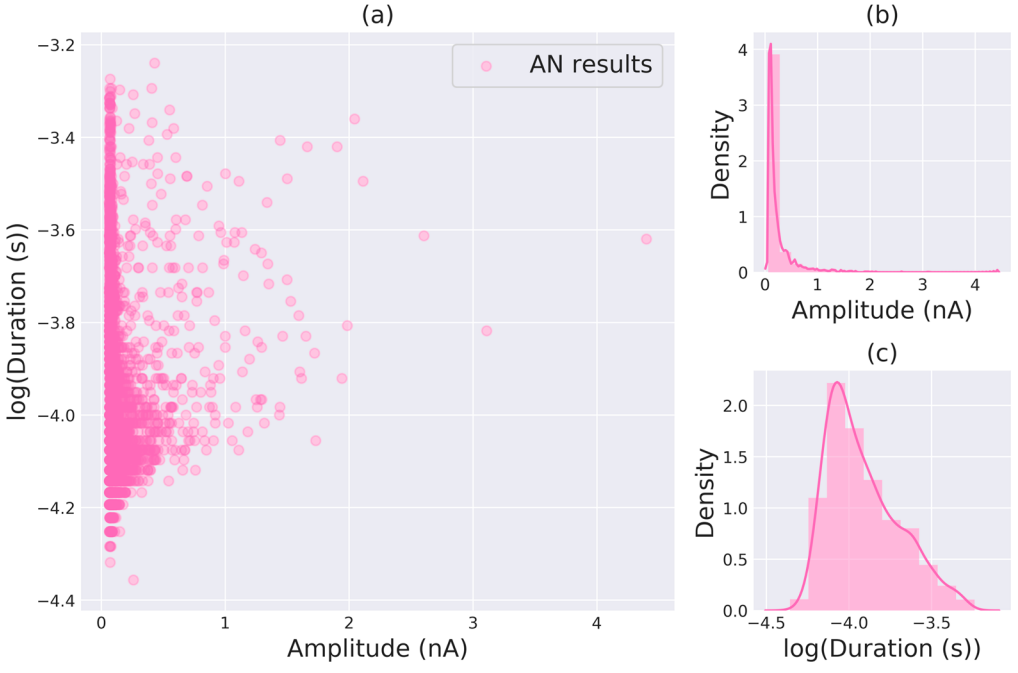


Figure 3. Characteristics of the 2111 confident events detected by AutoNanopore on all 23 abf files: (a) scatter plot of the amplitude versus duration of the events; (b) histogram of the amplitude; (c) histogram of the duration.

4, the searching for starting and ending time points is performed in the opposite direction as well.

3. RESULTS ON ALL 23 SINGLE-STRANDED DNA CURRENT TRACES

This section analyzes various results obtained when applying AutoNanopore to the acquired data sets. AutoNanopore is implemented with Python (preferably using version 3.7, a 64-bit version), and the pyabf package (version 2.3.5) is used to process the abf files. It runs efficiently on Windows, Linux, or MacOS platforms without extra specific configurations. The source code of AutoNanopore and a demo file (1.abf as shown in Figure 2) have been released for use and verification (https://github.com/bellstwohearted/AutoNanopore).

3.1. Example of the Output of AutoNanopore. Figure 2 shows the results for file 1.abf. We can observe that all translocation events result in an increase in current, with a wide range of peak amplitudes. AutoNanopore identifies 80 trans-

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111

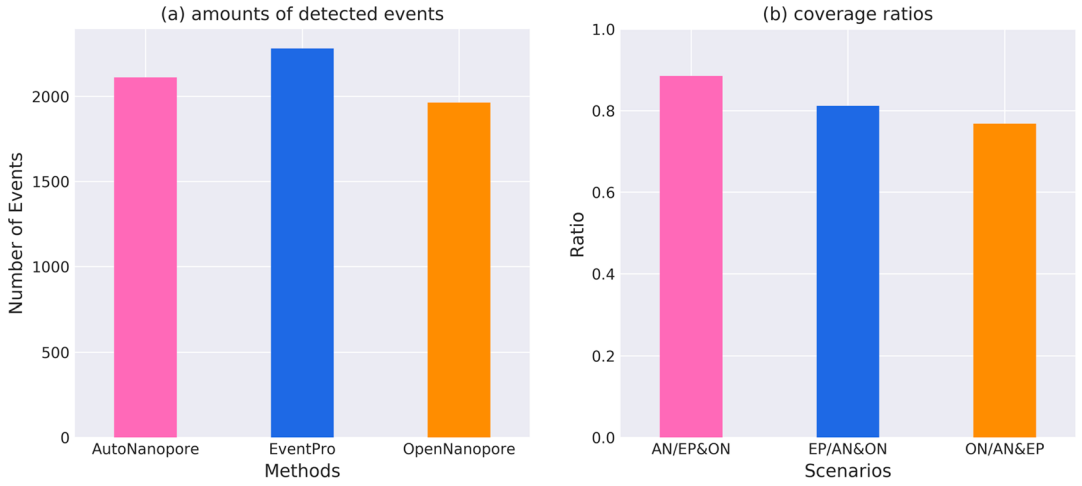


Figure 4. Comparison among the three methods. (a) The total amounts of events detected by the three methods. (b) The ratio of coverage between different comparison scenarios: e.g., AN/EP&ON means the ratio of the EP-ON-overlap events that are also detected by AN, $\frac{\#\text{of}\,(\text{AN}\cap\text{EP}\cap\text{ON})}{\#\ \text{of}(\text{EP}\cap\text{ON})}$.

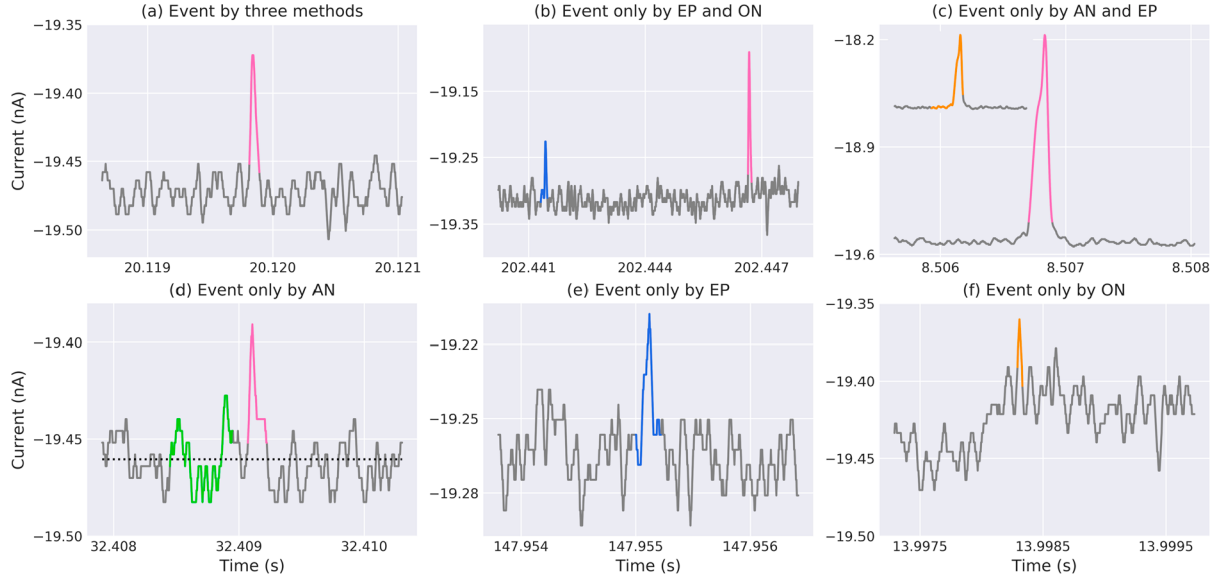


Figure 5. Representative examples of the events, as detected by different methods. (a) One event that is agreed upon by all three methods. (b) Two events that are agreed upon by both EP and ON, but only one (in pink) is detected by AN. (c) An event accurately detected by AN and EP. In the top left corner, the small figure gives the output of ON (in orange). (d–f) Examples of events detected by only one method. In all panels, pink curves show the AN-determined event ranges, blue curves shows the EP-determined ranges, and orange curves show the ON-determined ranges. In (d), the green and dashed curves have the same meanings as in Figure 1.

location events from 75 million data points in approximately 15 s, using a single CPU core. Furthermore, it also outputs the characterizations of these events, including the time of the peak, the baseline value, the amplitude of the peak, the starting and ending times, and the duration. As shown in Figure 2, 80 events are detected by AutoNanopore: some events have either large or small amplitudes, while most events have intermediate amplitudes. To better illustrate the events, especially those with relatively low amplitudes, three examples are given together with the entire current trace (bottom panels of Figure 2). It turns out that, though these events seem to be submerged in the baseline, when the range is enlarged, we can conclude that these events are correctly identified.

3.2. Overall Results of AutoNanopore on 23 abf Files. Figure 3 presents the characteristics of the events detected by AutoNanopore on the data set. In total, AutoNanopore detects 2111 events from all 23 abf files. It can be found that the amplitudes and durations of the vast majority of the events are located within a relatively small range, and a few have high amplitudes and/or longer durations (Figure 3a). Since there are a few outliers in the duration determined by AutoNanopore, to show the results in a more intuitive way, we perform a logarithmic transformation for the duration before plotting, rather than filter out the outliers further. Figure 3 also shows the distributions of the amplitudes and durations of all the events detected by AutoNanopore (Figure 3b,c). It turns out that, for such short ssDNA samples, the events' amplitudes are very concentrated, with only one peak.

3.3. Comparison with Other Methods. To examine the effectiveness and robustness of AutoNanopore, we compare the results determined by AutoNanopore to those by other two excellent methods, OpenNanopore[17] (ON) and EventPro[21] (EP); the parameter settings for using these two methods are given in the Supporting Information. We chose these two

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111

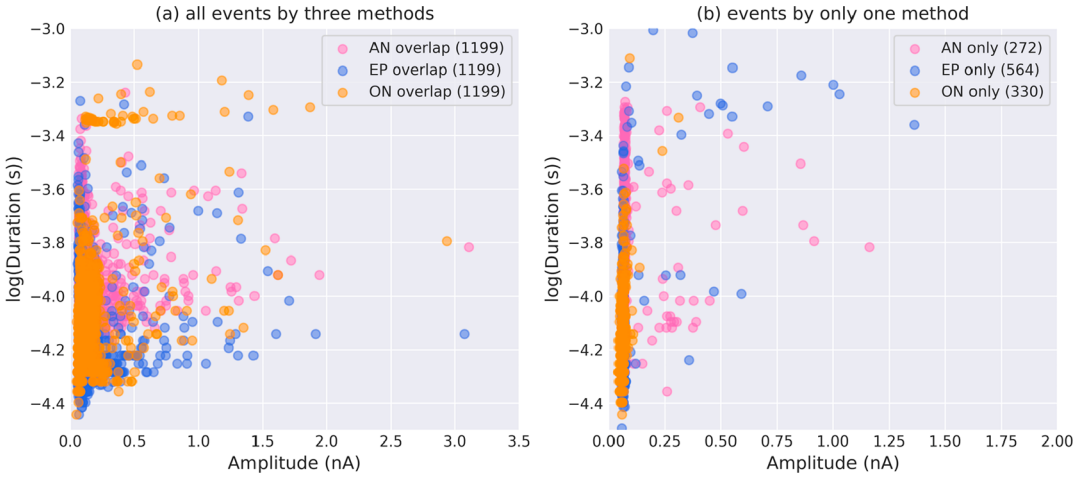


Figure 6. (a) Characteristics of the 1199 events detected by all three methods. (b) Characteristics of the AN-only, EP-only, and ON-only events.

methods for the reason that they are the same types of event detection methods that use an adaptive threshold as our method, and more importantly, they are excellent methods that have had great impacts in the field.

The comparisons are performed from two aspects: (1) whether AN is able to detect the events that are identified by both EP and ON and (2) whether the events' characteristics determined by AN are consistent with those by EP and ON. For EP, we use version 2.0, and the baseline fitting mode is "SMD and Msbackadj". This mode, according to its authors, is the most powerful. The parameter settings for EP and ON are given in the Supporting Information.

First, we analyze whether AutoNanopore is able to detect the events that are identified by other methods by investigating whether the peaks corresponding to the AN-detected events are included in the outputs of EP and ON. Since ON outputs some events that have much longer duration compared to AN and EP, before the comparison, we perform a process on the outputs of ON by filtering out the events with extreme outlier durations. This is conducted independently on each abf file, and the events with a duration longer than Q3 + 3 × IQR are filtered out. As shown in Figure 4, in total, AN detects 2111 events, EP detects 2281 events, and ON detects 1963 events. The results show that 88.5% of the EP-ON-overlap events (the events that are detected by both EP and ON) are also detected by AN, while 81.2% of AN-ON-overlap events are detected by EP and 76.8% of AN-EP-overlap events are detected by ON. Moreover, 362 events are only identified by AN and EP, 156 events are only identified by EP and ON, and 278 events are only identified by AN and ON. Separately, there are 272, 564, and 330 events that are only detected by AN, EP and ON, respectively. A detailed Venn diagram is given in the Supporting Information.

Next, we investigate the events that are not consistently detected by the three methods, and some representative examples from the results of 1.abf are shown in Figure 5. Before examination of the inconsistent events, Figure 5a shows an event that is agreed upon by all three methods. Furthermore, there are some events that are only detected by one or two methods but are missing by other methods. On the one hand, only 11.5% of EP-ON-overlap events are not detected by AN; the main reason is that, within some slices (30 ms segment of AN), there exist multiple events. However, due to the method of AN, at most one event (with the maximum peak current value) can be detected within each slice, resulting in the missing of some events (Figure 5b' AN only detects the event marked in the pink curve but misses the orange event). This can be avoided by shortening the window size of AN slices, and it is acceptable that only a few such events are missing. On the other hand, most of the AN-only events (those that are only detected by AN) have small amplitudes that do not meet the selection criteria of EP and ON, mainly due to the variation of the baseline. Figure 5d gives an example, showing that the baseline corresponding to this event varies significantly; thus, the AN strategy of searching a range close to the peak for baseline computation is to some extent advantageous.

Interestingly, Figure 5c shows an example where only AN and EP are able to detect this event. Note that ON also detects it but gives it a much longer duration (small figure on the top left corner). In such a case, ON is considered not to accurately characterize this event; actually, this event has been eliminated during the duration outlier filter for ON. Moreover, Figure 5e,f shows examples that are only detected by EP and ON, respectively. EP detects the event shown in Figure 5e, but the amplitude is too small to meet the criteria of AN and ON, for the reason that EP uses a different method to construct the baseline. ON detects the event shown in Figure 5f that is just within the range when the baseline begins to increase; looking at the current variation after the event (orange range), it is hard to judge whether it is indeed an event.

Third, we analyze whether the AN-determined characteristics are consistent with the other two methods. Figure 6 shows the distributions of the amplitude and duration of the 1199 events that are detected by all three methods and the events only detected by one method. The distribution plots corresponding to Figure 6 are given in the Supporting Information. Note that, for EP and ON, only the maximum amplitudes of each event are plotted here. It can be seen that the amplitude and duration distributions of the events agreed upon by all three methods are consistent between AN and EP (Figure 6a) and are reasonable for translocation events. This can also be seen from the distribution plot shown in the Supporting Information. Moreover, the ON-determined amplitudes/durations are more concentrated, with some events concentrating at the top boundary of the figure.

The characteristics of the events only detected by one method are shown in Figure 6b. AN-only and EP-only events have reasonable amplitudes/durations. Again, ON-only events are concentrated at the left bottom corner, indicating smaller

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111

amplitudes (and durations). Note that, though all of these results indicate that AutoNanopore performs very well in detecting the translocation events, they do not imply that AN outperforms EP and ON; all three methods have their own advantages compared to the others, due to distinct mechanisms. In particular, EP discards events based on noises; the results may be affected by noises. Summarizing, all these comparison results show that, on the one hand, the results of all three methods are consistent for high-amplitude events; on the other hand, for those low-amplitude events, results by the three methods are less consistent.

Moreover, we also test AutoNanopore and compare the results to those of EP and ON on another 5 min current trace: 48.5kb λ-DNA with a concentration of 60 pM/L, under 300 mV. The results are given in the Supporting Information. It turns out that the events driven by λ-DNA are more complex; AutoNanopore's performance is competitive and robust.

4. DISCUSSION

In this study, we present AutoNanopore as a novel and promising translocation event detection method in solid-state nanopore current traces, evaluate its performance in diverse experiments, and discuss the influence of the experimental conditions in the translocation events.

The novelty of AutoNanopore is that it detects the events by computing the quantile values and IQRs of the amplitudes to identify the outliers rather than considering the mean value and standard deviation of the current values. Thus, it combines the advantages of classic methods, such as MiniAnalysis, Easy Electrophysiology, and Clampfit, and the more advanced methods employing adaptive thresholds, such as OpenNanopore[20] and EasyNanopore.[21] Splitting the raw data into many slices first, as proposed by the classic methods, is a good strategy; since the baseline of the current traces may vary over time, as shown in Figure 2, observing the data points that are close to the peak is thus reasonable. AutoNanopore is more robust than the classic methods because it dynamically searches for a range close to the peak to compute the baseline values, which is not the case for MiniAnalysis, Easy Electrophysiology, or Clampfit (these methods usually take the fixed range before the peak). Meanwhile, each event may have distinct characteristics and so it is difficult to detect the events by a simple fixed threshold; the use of an adaptive threshold determined from neighborhood data characteristics, as proposed by OpenNanopore and EventPro, is an inevitable choice.

AutoNanopore is distinctly different from other adaptive event detection methods in several aspects. First, AutoNanopore splits the current traces into slices and analyzes the slices by concentrating only on the data points close to the peak, and a global analysis is conducted after the analyses of all slices end. Given that the baseline current may vary over time in real-world applications, the amplitudes caused by translocation events are still in a relatively stable range. Such processes enlarge the properties of local data, avoiding the negative influences caused by the variation in global data. Second, AutoNanopore uses a more straightforward outlier detection method based on the identified peaks' amplitudes, rather than considering the mean value, standard deviation, or cumulative sums of current traces. All of these new insights help improve the performance of AutoNanopore in particular in detecting those challenging events with a significantly varying baseline. Compared to EventPro, which attempts to resolve the influences of the varying baseline by several fitting methods, our method provides a more efficient way to minimize the influences of baseline variation: AutoNanopore does not require reconstructing the baseline using specific fitting methods for the reason that, no matter how the baseline varies, the amplitudes of the real events should be relatively stable. As a consequence, our method is more efficient than EventPro.

One significance of the present work is that we propose ideas and metrics for evaluation of the performances of an event detection method. Most existing methods were reported independently, without comparison or evaluation against others; this is partially due to the lack of a public, well-recognized benchmark. We suggest that such a comparison/evaluation should be necessary. Our comparison provides a useful approach for evaluating a new event detection method in the future.

AutoNanopore also has some limitations. First, AutoNanopore searches a suitable range before the peak, and this may sometimes lead to an incorrect baseline value if the current variation within that range does not reflect the actual baseline, in particular when the nanopore is clogged for a much longer time than a typical translocation event, even though the clogging is indeed the results of nucleotide translocation. In such cases, a human expert may be capable of judging the abnormal events, while AutoNanopore would not succeed. Second, AutoNanopore mainly focuses on the peak caused by the maximum current value in a segmented slice; this may sometimes result in missing of events, e.g., when multiple events indeed exist in that slice. Under such circumstances, we thus suggest adjusting the time range of baseline computation in line with the estimated event duration, which mainly depends on how many nucleotides the DNA has. Third, the current version of AutoNanopore does not incorporate the fitting of multilevel events, which are usually driven by nonlinear DNA. Basically, when an event is identified, that is, the peak of the event is located, then AutoNanopore needs to use specific methods to do changing points detection, only within the event range. Our research aims to identify specific pathogens, including three consecutive steps: the first step is event detection, followed by the identification of current traces caused by ssDNA and dsDNA (DNA-probe complex), and the third step is the identification of dsDNA corresponding to different probes using machine-learning approaches. An enhanced version of AutoNanopore will be released in the near future, together with an example of its application in a real-world scenario.

AutoNanopore has important implications in promoting the development of solid-state nanopore sequencing. Unlike sequencing by biological nanopores, where base calling can be directly performed and various machine-learning tools for data analysis have been developed,[25] the base calling in solid-state nanopore sequencing is challenging. AutoNanopore can accurately and consistently locate the events and segment the current variations in translocation events for further analysis, avoiding wrong base calling results due to processing a large portion of redundant data,[26] for the reason that these data do not output information on biomolecules we need. This is also the approach of OpenNanopore and EventPro: both tools output concatenated events containing only the data points of the identified events. It can also reduce computation time when there are plenty of current traces to be base called in real-world applications. Furthermore, the low SNR in signals is the main bottleneck of current solid-state nanopore sequencing,[14] even though some studies have reported methods of identifying four single-stranded DNA homopolymers[27,28] and even the identi-

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111

fication of single nucleotides[22,29] in solid-state nanopore sequencing. However, direct sequencing by solid-state nanopores is tricky; sequencing by hybridization[30] is thought to be a promising approach because ssDNA and dsDNA can lead to significantly different current variations due to their diameters.[7,10,31] AutoNanopore can then be applied to distinguishing the signals by ssDNA and a DNA-probe complex, revealing whether the hybridization can succeed.

5. CONCLUSIONS

We have presented an automated adaptive and robust method, AutoNanopore, for a rapid and straightforward translocation event detection in solid-state nanopore current traces. AutoNanopore has been extensively tested to perform event detection tasks on an experimental data set of 23 ssDNA current traces. The results show that it can effectively detect events from huge amounts of raw sequencing data. Additionally, we have compared our AutoNanopore favorably with two existing state-of-the-art methods, OpenNanopore and EventPro, and it exhibits high coverage against these methods and performs excellently, in particular when the events correspond to significantly varying baselines. Overall, AutoNanopore has shown its advantages in analyzing solid-state nanopore sequencing data in realistic signals and has great potential to contribute significantly to the development of solid-state nanopore sequencing.

■ ASSOCIATED CONTENT

Data Availability Statement

The source code of AutoNanopore and a demo signal file (1.abf as shown in Figures 1 and 2) are available at: https://github.com/bellstwohearted/AutoNanopore.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c02927.

Experiment preparation and nanopore fabrication, parameter settings of different methods, a Venn diagram corresponding to Figure 4, distribution plots corresponding to Figure 6, and results on a λ-DNA current trace (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Zepeng Sun − *China Mobile (Chengdu) Industrial Research Institute, Chengdu 610000, People's Republic of China;* orcid.org/0000-0002-9846-4174; Email: sunzepeng@cmii.chinamobile.com

Jian Li − *Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing 210096, People's Republic of China*; Email: jianli2014@seu.edu.cn

Li-Qun Xu − *China Mobile (Chengdu) Industrial Research Institute, Chengdu 610000, People's Republic of China*; Email: xuliqun@chinamobile.com

Authors

Xinlong Liu − *China Mobile (Chengdu) Industrial Research Institute, Chengdu 610000, People's Republic of China*

Wei Liu − *Jiangsu Key Laboratory for Design and Manufacture of Micro-Nano Biomedical Instruments, School of Mechanical Engineering, Southeast University, Nanjing 210096, People's Republic of China*

Jiahui Li − *China Mobile (Chengdu) Industrial Research Institute, Chengdu 610000, People's Republic of China*

Jing Yang − *Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing 210096, People's Republic of China*

Feng Qiao − *China Mobile (Chengdu) Industrial Research Institute, Chengdu 610000, People's Republic of China*

Jianjun Ma − *China Mobile (Chengdu) Industrial Research Institute, Chengdu 610000, People's Republic of China*

Jingjie Sha − *Jiangsu Key Laboratory for Design and Manufacture of Micro-Nano Biomedical Instruments, School of Mechanical Engineering, Southeast University, Nanjing 210096, People's Republic of China;* orcid.org/0000-0002-0797-4460

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c02927

Author Contributions

Z.S. and X.L. contributed equally to this work. L.-Q.X. and J.L. (Jian Li) conceptualized the research project. W.L. and J.S. fabricated the solid-state nanopores and carried out the solid-state nanopore translocation experiments. Z.S. designed the methodology of AutoNanopore. X.L. implemented AutoNanopore and performed the computation. J.L. (Jiahui Li) performed the comparison analyses with other tools; all authors interpreted the results. Z.S. generated the figures with inputs from X.L. and J.L. (Jiahui Li). Z.S. led the writing of the paper and wrote the first draft. L.-Q.X. examined and edited the text; all authors commented on earlier versions of the paper, and approved the final version.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was mainly supported by the Applied and Fundamental Research Program of China Mobile Communication Corp. We thank four anonymous reviewers for their constructive comments on earlier versions of this article.

■ REFERENCES

(1) Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. V.; Mikheenko, A.; Vollger, M. R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The complete sequence of a human genome. *Science* **2022**, *376*, 44−53.

(2) Bull, R. A.; Adikari, T. N.; Ferguson, J. M.; Hammond, J. M.; Stevanovski, I.; Beukers, A. G.; Naing, Z.; Yeang, M.; Verich, A.; Gamaarachchi, H.; et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **2020**, *11*, 6272.

(3) Charalampous, T.; Kay, G. L.; Richardson, H.; Aydin, A.; Baldan, R.; Jeanes, C.; Rae, D.; Grundy, S.; Turner, D. J.; Wain, J.; et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **2019**, *37*, 783−792.

(4) Goenka, S. D.; Gorzynski, J. E.; Shafin, K.; Fisk, D. G.; Pesout, T.; Jensen, T. D.; Monlong, J.; Chang, P.-C.; Baid, G.; Bernstein, J. A.; et al. Accelerated identification of disease-causing variants with ultra-rapid nanopore genome sequencing. *Nat. Biotechnol.* **2022**, *40*, 1035−1041.

(5) Kumar Sharma, R.; Agrawal, I.; Dai, L.; Doyle, P. S.; Garaj, S. Complex DNA knots detected with a nanopore sensor. *Nat. Commun.* **2019**, *10*, 4473.

(6) Raveendran, M.; Leach, A. R.; Hopes, T.; Aspden, J. L.; Actis, P. Ribosome Fingerprinting with a Solid-State Nanopore. *ACS Sens.* **2020**, *5*, 3533−3539.

(7) He, L.; Tessier, D. R.; Briggs, K.; Tsangaris, M.; Charron, M.; McConnell, E. M.; Lomovtsev, D.; Tabard-Cossa, V. Digital immuno-

assay for biomarker concentration quantification using solid-state nanopores. *Nat. Commun.* **2021**, *12*, 5348.
(8) Goto, Y.; Akahori, R.; Yanagi, I.Challenges of Single-Molecule DNA sequencing with solid-state nanopores. *Single Molecule and Single Cell Sequencing*; Springer Singapore: 2019; pp 131−142.
(9) Luan, B.; Stolovitzky, G.; Martyna, G. Slowing and controlling the translocation of DNA in a solid-state nanopore. *Nanoscale* **2012**, *4*, 1068−1077.
(10) Tabard-Cossa, V.; Trivedi, D.; Wiggin, M.; Jetha, N. N.; Marziali, A. Noise analysis and reduction in solid-state nanopores. *Nanotechnology* **2007**, *18*, 305505.
(11) Fragasso, A.; Schmid, S.; Dekker, C. Comparing current noise in biological and solid-state nanopores. *ACS Nano* **2020**, *14*, 1338−1349.
(12) Wen, C.; Dematties, D.; Zhang, S.-L. A guide to signal processing algorithms for nanopore sensors. *ACS Sens.* **2021**, *6*, 3536−3555.
(13) Das, N.; Mandal, N.; Sekhar, P. K.; RoyChaudhuri, C. Signal processing for single biomolecule identification using nanopores: A Review. *IEEE Sensors J.* **2021**, *21*, 12808.
(14) MiniAnalysis. https://scicrunch.org/resolver/SCR_002184. (MiniAnalysis was initially developed and released by Synaptosoft, its official website is now out of service.)
(15) Easy Electrophysiology. https://www.easyelectrophysiology.com/.
(16) Clampfit. https://www.moleculardevices.com/.
(17) Raillon, C.; Granjon, P.; Graf, M.; Steinbocka, L. J.; Radenovic, A. Fast and automatic processing of multi-level events in nanopore translocation experiments. *Nanoscale* **2012**, *4*, 4916.
(18) Forstater, J. H.; Briggs, K.; Robertson, J. W. F.; Ettedgui, J.; Marie-Rose, O.; Vaz, C.; Kasianowicz, J. J.; Tabard-Cossa, V.; Balijepalli, A. MOSAIC: A Modular Single Molecule Analysis Interface for Decoding Multi-state Nanopore Data. *Anal. Chem.* **2016**, *88*, 11900−11907.
(19) Plesa, C.; Dekker, C. Data analysis methods for solid-state nanopores. *Nanotechnology* **2015**, *26*, 084003.
(20) Tu, J.; Meng, H.; Wu, L.; Xi, G.; Fu, J.; Lu, Z. EasyNanopore: a ready-to-use pocessing software for translocation events in nanopore translocation experiments. *Langmuir* **2021**, *37*, 10177−10182.
(21) Bandara, Y. M.; Nuwan, D. Y.; Saharia, J.; Karawdeniya, B. I.; Kluth, P.; Kim, M. J. Nanopore data analysis: baseline construction and abrupt change-based multi-level fitting. *Anal. Chem.* **2021**, *93* (34), 11710−11718.
(22) Diaz Carral, A.; Ostertag, M.; Fyta, M. Deep learning for nanopore ionic current blockades. *J. Chem. Phys.* **2021**, *154*, 044111.
(23) Xia, K.; Hagan, J. T.; Fu, L.; Sheetzc, B. S.; Bhattacharya, S.; Zhang, F.; Dwyer, J. R.; Linhard, R. J. Synthetic heparan sulfate standards and machine learning facilitate the development of solid-state nanopore analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2022806118.
(24) Tsutsui, M.; Takaai, T.; Yokota, K.; Kawai, T.; Washio, T. Deep learning-enhanced nanopore sensing of single-nanoparticle translocation dynamics. *Small Methods* **2021**, *6*, 2100691.
(25) Cui, F.; Yue, Y.; Zhang, Y.; Zhang, Z.; Zhou, H. S. Advancing biosensors with machine learning. *ACS Sens.* **2020**, *5*, 3346−3364.
(26) Zhang, J.; Liu, X.; Ying, Y.; Gu, Z.; Meng, F.-N.; Long, Y.-T. High-bandwidth nanopore analysis by using a modified hidden Markov model. *Nanoscale* **2017**, *9*, 3458−3465.
(27) Goto, Y.; Yanagi, I.; Yokoi, T.; Takeda, K.-I. Identification of four single-stranded DNA homopolymers with a solid-state nanopore in alkaline CsCl solution. *Nanoscale* **2018**, *10*, 20844−20850.
(28) Wei, S.; Yang, H.; Sha, J.; Zhang, Y.; Chen, Y. Discrimination of single-stranded DNA homopolymers by sieving out G-quadruplex using tiny solid-state nanopores. *Electrophoresis* **2019**, *0*, 1−8.
(29) Feng, J.; Liu, K.; Bulushev, R. D.; Khlybov, S.; Dumcenco, D.; Kis, A.; Radenovic, A. Identification of single nucleotides in MoS_2 nanopores. *Nat. Biotechnol.* **2015**, *10*, 1070−1076.
(30) Drmanac, R.; Drmanac, S.; Chui, G.; Diaz, R.; Hou, A.; Jin, H.; Jin, P.; Kwon, S.; Lacy, S.; Moeur, B., et al.Sequencing by hybridization (SBH): advantages, achievements, and opportunities. *Chip Technology. Advances in Biochemical Engineering/Biotechnology*; Springer: 2002; pp 75−101.
(31) Zhang, J.; Liu, X.; Hu, Z.; Ying, Y.; Long, Y.-T. Intelligent identification of multi-level nanopore signatures for accurate detection of cancer biomarkers. *Chem. Commun.* **2017**, *53*, 10176−10179.

https://doi.org/10.1021/acsomega.2c02927
ACS Omega 2022, 7, 37103−37111