THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2018, 9, 5718−5725

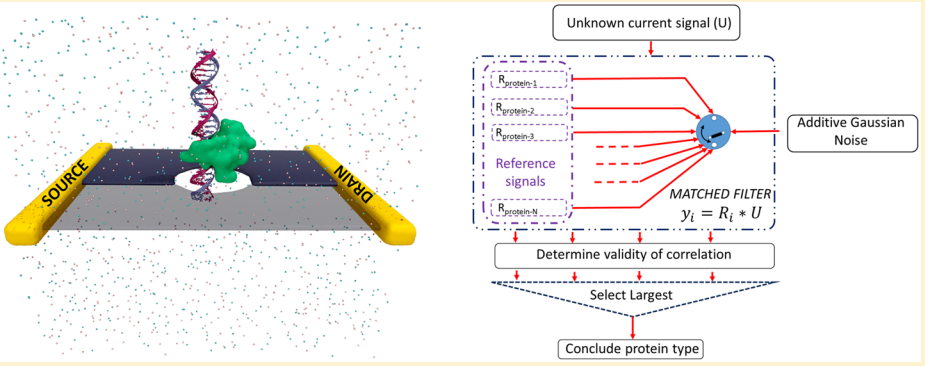
pubs.acs.org/JPCL

Classification of Epigenetic Biomarkers with Atomically Thin Nanopores

Aditya Sarathy,*[,†,‡] Nagendra B. Athreya,*[,†] Lav R. Varshney,*[,†,§,‡] and Jean-Pierre Leburton*[,†,‡,∥]

[†]Beckman Institute for Advanced Science and Technology, [‡]Department of Electrical and Computer Engineering, [§]Coordinated Science Laboratory, [∥]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

Supporting Information



ABSTRACT: We use the electronic properties of 2D solid-state nanopore materials to propose a versatile and generally applicable biosensor technology by using a combination of molecular dynamics, nanoscale device simulations, and statistical signal processing algorithms. As a case study, we explore the classification of three epigenetic biomarkers, the methyl-CpG binding domain 1 (MBD-1), MeCP2, and γ-cyclodextrin, attached to double-stranded DNA to identify regions of hyper- or hypomethylations by utilizing a matched filter. We assess the sensing ability of the nanopore device to identify the biomarkers based on their characteristic electronic current signatures. Such a matched filter-based classifier enables real-time identification of the biomarkers that can be easily implemented on chip. This integration of a sensor with signal processing architectures could pave the way toward the development of a multipurpose technology for early disease detection.

The search for a low-cost, fast, and reliable method to access, assess, and decode the human genome and epigenome is a great technological challenge in modern medicine,[1,2] which conventional biochemical sequencing processes[3] may not be able to meet. One possible alternative uses a very thin membrane with a nanoscale pore,[4] through which DNA molecules are threaded to identify not only the nucleotide sequence but also traits of DNA such as methylation.[5,6] Indeed, methylation may be as crucial as the sequence itself for the diagnosis and identification of epigenetic diseases such as cancer,[7,8] through its role in silencing key cancer-related genes. Nowadays commercially available nanopore sequencers that have been used to sequence DNA[9] and RNA[10] would be unable to identify epigenetic markers attached to methylated sites owing to size discrepancy between the DNA−marker complex and the nanopore. Therefore, solid-state nanopores may be the only viable solution as a versatile and general sensor technology for detecting methylation patterns.

It is well known that the utilization of 2D materials in solid-state nanopores offers the highest detection resolution. In this regard, Girdhar et al. proposed the utilization of solid-state multilayer nanopore membranes within multifunctional electronic devices to increase their detection sensitivity.[11] Qiu et al. further demonstrated that methylated cytosines labeled by methyl−CpG binding domain proteins can be detected with solid-state nanopores composed of 2D materials such as graphene or molybdenum disulfide (MoS_2).[12,13] This detection was based on two simultaneous signals: ionic current variations through the pore and electronic current changes along the 2D membrane. The two signals together yield higher protein identification and classification accuracy than a single signal alone. Whereas the ionic currents provide information regarding the size of the translocating biomarker via the current blockade and dwell time, the electronic sheet current variations are specifically dependent on the charge distribution within each biomarker, thereby providing unique signatures for each marker. It was also shown that the electronic detection of these labeled proteins offers a higher resolution in measurement as compared with ionic current-based detection.[14] Indeed, the resolution of detection of these labeled sites via electronic current is limited by the sizes of the labeled proteins rather than by the electronic measurement quality itself. Aside

Received: July 13, 2018
Accepted: September 13, 2018
Published: September 18, 2018

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

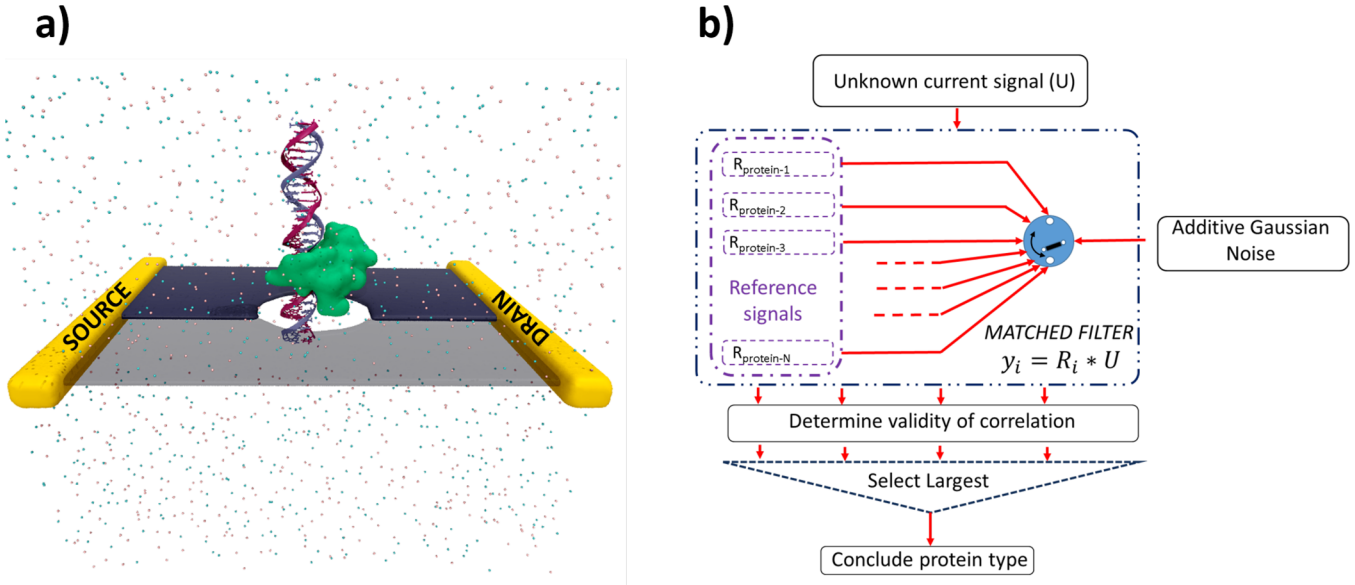


Figure 1. (a) Schematic of a multifunctional 2D semiconductor material (usually graphene or MoS_2) utilized for the detection of methylated sites along the DNA. Distinct peaks in the transverse conductance and dips in ionic current are indicative of methylated site translocation. (b) Schematic of a matched filter-based detection workflow whereby the type of marker protein on a DNA is concluded using correlation between the noisy input signal (test signal) and a dictionary of precomputed reference current signatures from "noise-free" translocation of DNA−marker complexes.

from recent works,[14,15] little attention has been paid to explore the utility of solid-state nanopores to identify epigenetic biomarkers along a double-stranded DNA that are indicative of intricate mechanisms of gene regulation.[16] Also, efforts to detect, identify, and map DNA methylation patterns using solid-state nanopores have been unsuccessful because of the significant noise introduced in the measured signal due to the conformational stochastic fluctuations of DNA inside the pore. In this Letter, we propose an integrated approach that combines electronic simulation based on device physics with statistical signal processing techniques to characterize the resolution limit of solid-state nanopore sensing and further to develop algorithms for epigenetic marker classification. Our novel sensor technology is capable of detecting and mapping regions of hyper- and hypomethylations across the genome by utilizing bulky biomarkers. These biomarkers are further classified using electronic sheet currents resulting from electrically active 2D nanopore membranes because each marker produces a current signature unique to its structure and spatial charge distribution.

Among bulky groups to label methylated cytosines along a double-stranded DNA, we utilize either methyl-CpG binding domain (MBD-1) protein or methyl CpG binding protein 2 (MeCP2) to identify regions of hypermethylation. In humans, these two proteins bind to regions of hypermethylation along the DNA and are thought to repress transcription from methylated gene promoters.[17] Abnormal levels of MBD protein and their polymorphisms have been associated with the overall risk of lung cancer.[18,19] Furthermore, MeCP2 mutations are thought to be responsible for Rett syndrome, a severe neurodevelopmental disorder. The expression of MeCP2 in the brain is mostly in mature neurons and therefore in the identification of neurological diseases. Analogously, to identify regions of hypomethylation, we consider the detection of unmethylated CpGs marked by γ-cyclodextrin (GCD). This synthetic biomarker can be used to identify hypomethylated sequences, similar to the approach described by Gilboa et al.[15]

Figure 1a illustrates the proposed model setup utilized to obtain the electronic current signatures for epigenetic marker proteins. The setup consists of a 2D material (graphene, MoS_2, or other transition-metal dichalcogenide membranes) connected between two electrodes, that is, the source and drain to enable the flow of current through the membrane under an applied bias. The detection sensitivity of the membrane is controlled via a gate electrode separated from the membrane by a high-κ dielectric (not shown). A circularly shaped nanopore, chosen to have a diameter of 5 nm, allows the translocation of the biomolecule through the membrane. This dimension is the smallest pore diameter through which the DNA−marker complex can translocate without any hindrance. The whole setup is immersed in water containing a 1 M electrolyte of potassium chloride (KCl). A DNA strand, complexed with a marker protein either at the methylated cytosine site (for hypermethylation) or at unmethylated adenine (for hypomethylation), is translocated through the nanopore using an applied bias across the cis and trans chambers (V_{TC}). Modulation in current flowing through the membrane (sheet current) enables calibration of the local electrostatic potential distribution within the nanopore at a given time instant.

For our statistical analysis, we need signal references that are obtained from frozen DNA current signatures, where the biomolecule is artificially translocated through the pore in the absence of all-atom molecular dynamics (MD) simulations, as previously performed by Girdhar et al.[11] The noisy test signal is obtained from a computation scheme involving MD simulations coupled to semiconductor device models.[14] A detailed description of MD system setup, simulation methodology, and electronic transport calculations are provided in the Supporting Information. In the noise-free reference signal, the observed current from the DNA−marker complexes will arise solely from the charge distribution across the proteins, which are unique to the protein structures themselves. The set of these noise-free signals will comprise a set of unique reference

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

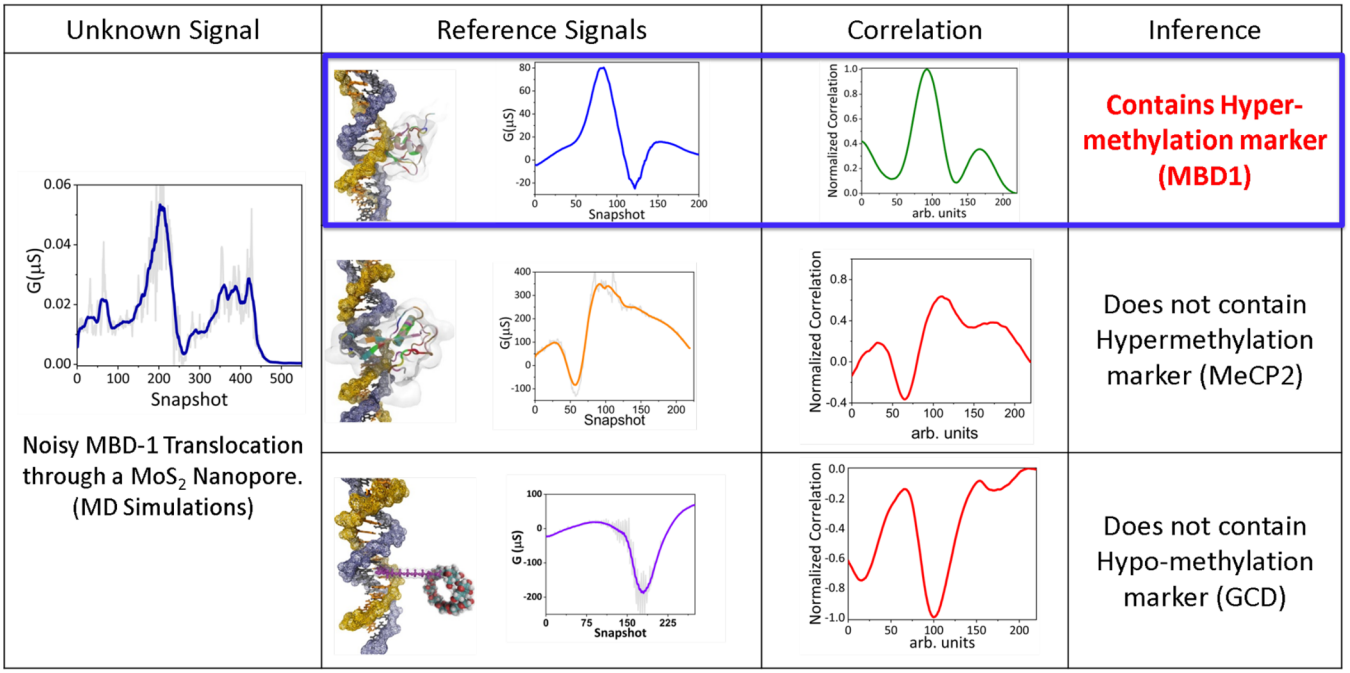


Figure 2. Classification epigenetic biomarkers: We first extend the set of reference signals by computing the sheet current signatures corresponding to translocations of frozen DNA−MBD, DNA−MeCP2, and DNA−GCD complex through a 5 nm MoS_2 nanopore (shown in the middle panel). A noisy current trace for the nonideal translocation of a DNA−MBD complex containing one MBD1 proteins (as shown in the leftmost panel) is correlated with each of the entries in the dictionary. We observe a peak when correlated with the frozen signature of a DNA−MBD complex with one protein (middle panel, rightmost column) and no discernible peaks when correlated with the other current signatures. The corresponding Q factors calculated at $\rho = 0.7$ are maximum for the correlation of the unfrozen DNA−MBD complex with the current signature of the MBD protein, indicating the presence of a MBD1 protein along the methylated DNA.

current signatures for epigenetic markers. Once this reference set is built, it can be used to identify the type or number of proteins by a statistical signal processing algorithm such as a bank of matched filters, as outlined in Figure 1b. The reference signals for each of the marker proteins are denoted as $R_i(t)$, where i is the marker protein. An unknown noisy signal (denoted as U), obtained by MD simulations, where the marker protein is unknown, is input into the filter bank that classifies the marker-protein type depending on correlations between the unknown current signal and reference signals.

In this context, it is well known that the optimal filter for detecting pulses in the presence of additive white Gaussian noise is the matched filter[20−22] (a similar model describing ionic currents in the presence of wide-band Gaussian noise was used for translocation event detection[23]). It can further be shown that the probability of detecting a weak signal in the presence of noise is largest when the signal-to-noise ratio is also largest.[24] Because the matched filter output is just the correlation with the reference signal, circuit implementations[22,25] in nanoscale computing technologies are simple.

In the presence of additive colored noise, for example, $1/f$ noise rather than white noise, one would have a whitening prefilter followed by designing the matched filter for the whitened reference signal. Our present noise model considers only the low-frequency regime of detection, that is, <100 kHz. When the sampling rate is increased toward mid- and high-frequency regimes, different noise models need to be considered and the corresponding matched filter implementations will be modified. Midfrequency will mainly consist of thermal noise, whereas the high-frequency regime will be dominated by capacitance effects. Power spectral densities for the different frequency regions are outlined by Parkin et al.[26]

Using our matched filter-based detection method, we can develop a unified framework to detect, classify, and count multiple marker proteins along the DNA and build upon our dictionary of reference signals. The reference signatures for each of the biomarkers were calculated for noise-free trajectories on graphene and MoS_2 quantum point contact nanoribbons. Additionally, another important aspect of the respective current signature is the shape of the trace because the magnitude depends on the stochastic fluctuation of the complex and its spatial orientation within the pore. In this regard, we previously showed that during the translocation of a methylated DNA complexed with either one or two MBD-1 protein complexes, depending on the number of methylated sites, the conductance square deviation is drastically different due to the strong dependence of the sheet conductance on the angular position of the marker protein within the pore.[13]

Once the reference current signatures are obtained, a correlation of a current signal consisting of an unknown marker with each of the reference signals will identify the type of marker protein. Given a reference signal $r_i(t)$ and a signal of an unknown marker $u(t)$, the cross-correlation is given as

$$\rho(t) = \int_{-\infty}^{\infty} r_i(\tau)u(t+\tau)\,d\tau \tag{1}$$

To capture just the shape of a signal and compare signals due to different marker proteins across various orientations, we normalize the correlation signals to range between −1 and 1. Given a correlation trace $\rho(t_i)$, where t_i is the sampled time instant, the normalized current trace is given by

$$\rho_{norm}(t_i) = \frac{\rho(t_i) - \rho(t_{open})}{\max(\rho(t)) - \min(\rho(t))} \tag{2}$$

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

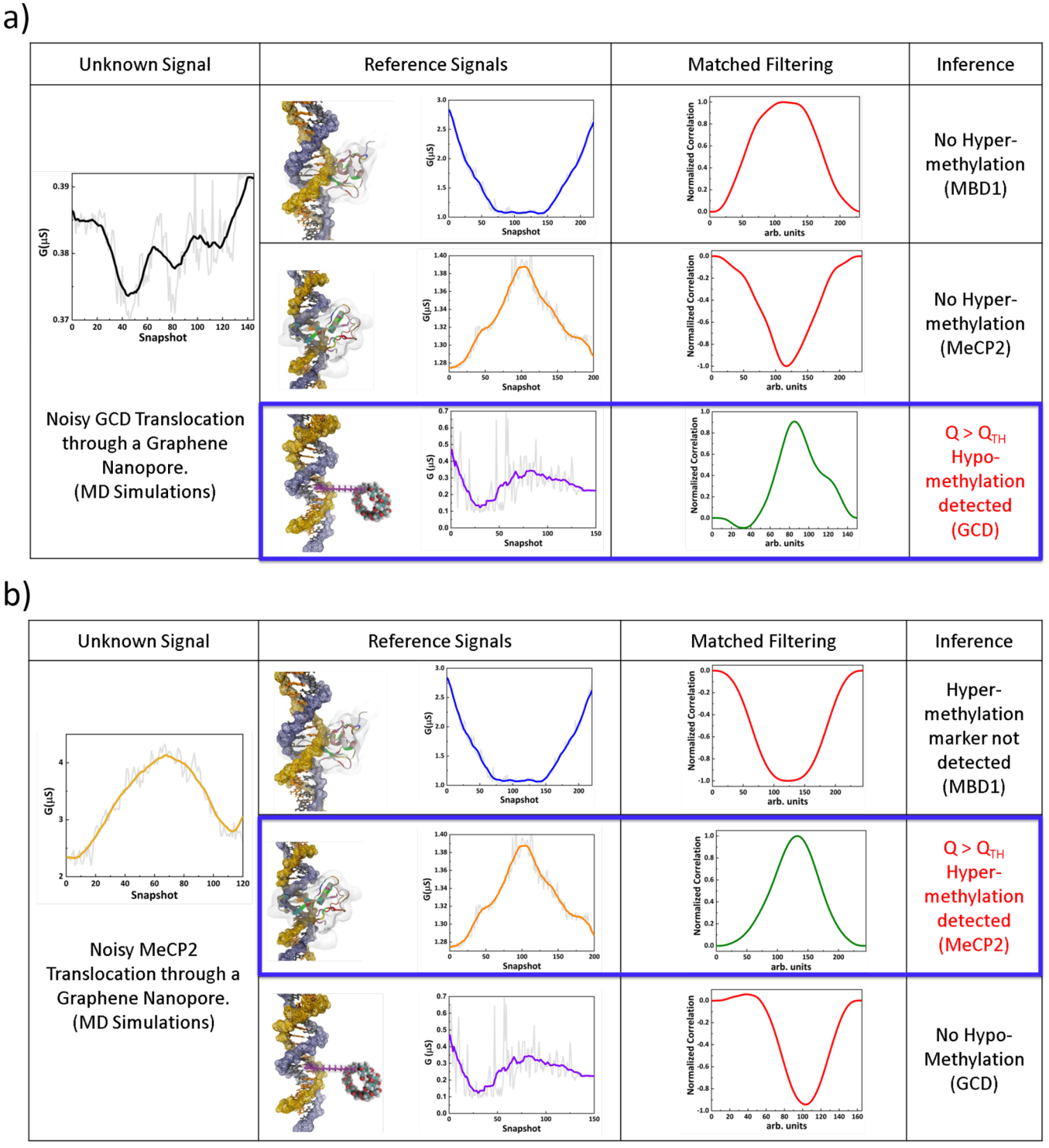


Figure 3. Matched filter operations to classify regions of hyper- and hypomethylation biomarkers along a DNA translocating through graphene nanopores

Here $\max(\rho(t))$ and $\min(\rho(t))$ denote the maximum and minimum values of the calculated trace during the translocation period during which the signal was acquired. This normalization allows us to compare the different correlated signals irrespective of their angular position during translocation.

Specifically, the criterion used to infer the type of marker protein from the correlated signal is the Q factor defined as

$$Q = \frac{1}{BW_{\text{corr}}} \tag{3}$$

where BW_{corr} is the bandwidth of correlation between the test and a particular dictionary signal. The value of correlation chosen to estimate the bandwidth is a hyperparameter (i.e., chosen by the user, based on statistics of different calculations, usually ranging from 0.6 to 0.8, as indicative from our calculations). In these sets of simulations, we choose to

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

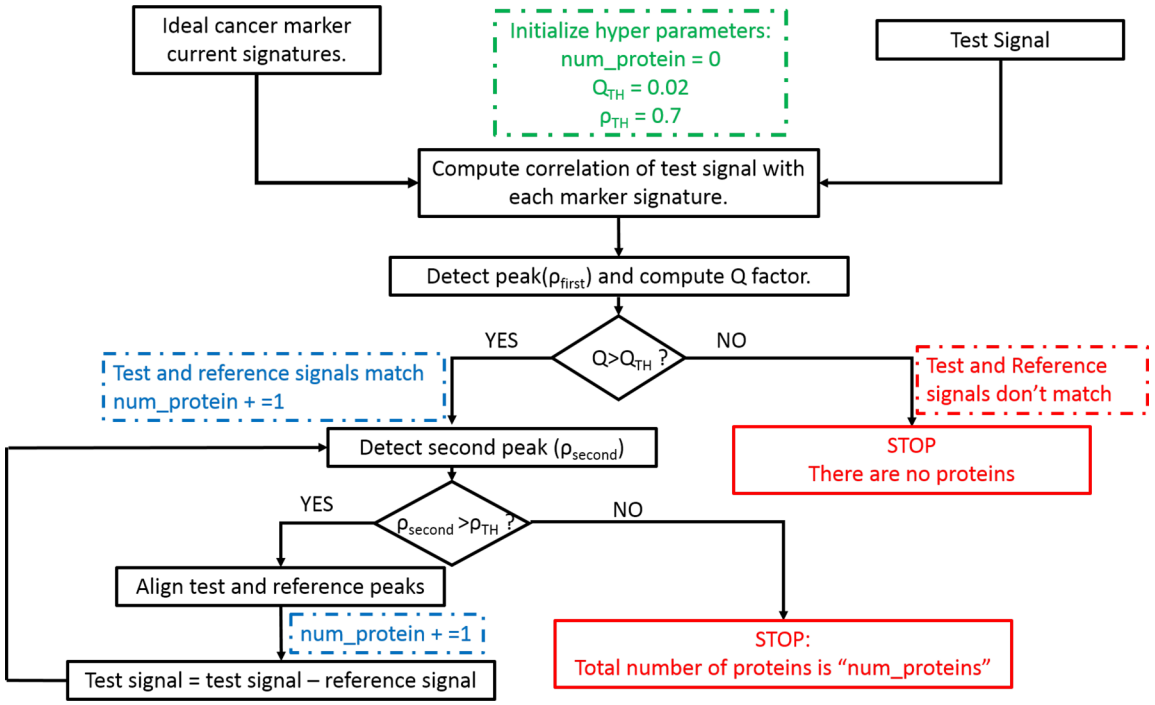


Figure 4. Flowchart illustrating the algorithm to detect the type of marker protein and the count the number of markers in the vicinity of the protein recursively. The hyperparameters chosen are the threshold Q factor and the threshold correlation value that decide the hypothesis of whether a particular marker and vicinity markers are present, respectively. Furthermore, this procedure of detecting and counting the marker proteins can be computed recursively and implemented in hardware.

calculate the BW_{corr} at ρ_i = 0.70, where ρ_i denotes the correlation between the test and reference signal of protein i (MBD, GCD, MeCP2). These hyperparameters are non-physical quantities that can be fine-tuned by obtaining statistics of multiple translocations of the proteins with different configurations and initial conditions. Essentially, once the correlations between the test and dictionary signals have been computed, the protein whose current signature provides the maximum Q factor is inferred to be present along the DNA.

Figure 2 illustrates the utility of the Q factor as a metric to infer the presence or absence of a particular protein along a translocating DNA. In these simulations, the test signal is obtained from translocating a 30 bp long DNA complexed with a single MBD1 protein at a CpG site. This test signal is noisy because it is obtained from calculating the current trace along the MoS_2 membrane from the resulting trajectory of an all-atom MD simulation. As shown in Figure 2, the first (leftmost) panel indicates the noisy current signature of the translocating MBD1 protein, whereas the second panel displays the normalized current signatures from MBD−DNA, GCD−DNA, and MeCP2−DNA complexes, respectively. The noisy current signal (on the leftmost panel) was obtained from a previous work of two of the authors (A.S. and J.-P.L.).[14] The second panel consists of the calculated current traces during a noise-free translocation of a DNA complexed with a MBD1, GCD, and MeCP2 marker protein, respectively. The third panel illustrates the matched filter operation between the noisy signal and the respective reference signals normalized to the interval [−1, 1]. The Q factor is calculated at ρ = 0.7, resulting in a value of 0.023 for the MBD1 reference signal. For the MeCP2 and GCD correlations, the Q factor is 0 because the maximum value of the cross-correlation signal is <0.7. It is therefore evident that the Q factor of the correlated outcome between noisy DNA−MBD and that of current signature of MBD is the highest, indicating the presence of the MBD protein along the DNA and the corresponding absence of the GCD or MeCP2 markers.

To illustrate the generality of this approach, we further illustrate the classification of the marker groups with unknown and reference signals calculated from translocations in graphene nanopores. Figure 3a,b illustrates the normalized correlations obtained between a noisy signal of a DNA complexed with hypomethylated (GCD) and hypermethylated (MeCP2) epigenetic biomarkers, respectively. The reference transverse current signatures for MBD1, MeCP2, and GCD are shown in the second panel, whereas the corresponding normalized correlations between the noisy unknown signal and the reference signals are shown in the third panel. As in the case of hypomethylation detection, one can clearly see that the correlation of the noisy signal corresponding to the GCD biomarker with the GCD reference signal gives the sharpest peak and greatest Q factor at ρ = 0.7 of ∼0.04, whereas the Q factors corresponding to correlations with the MBD1 and MeCP2 markers are ∼0.017 and 0, respectively. Similarly, a noisy translocation of a DNA−MeCP2 complex yields the maximum Q factor of ∼0.018 at ρ = 0.7 when correlated with the reference signal of a MeCP2 biomarker, as shown in Figure 3b. Therefore, these results indicate the versatility of our approach in developing a set of reference signals and classifying the unknown epigenetic biomarker using the matched filter.

The approach described above illustrates the use of the matched filter algorithm to classify particular epigenetic markers. This matched filter can be applied in any setting, but it can be proven to be the optimal linear detector in the presence of additive noise. Additionally, the Q factor alone can be used as a metric to infer the hypothesis of whether the particular type of protein is present or absent. However, to simultaneously detect, infer, and count the type and number of

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

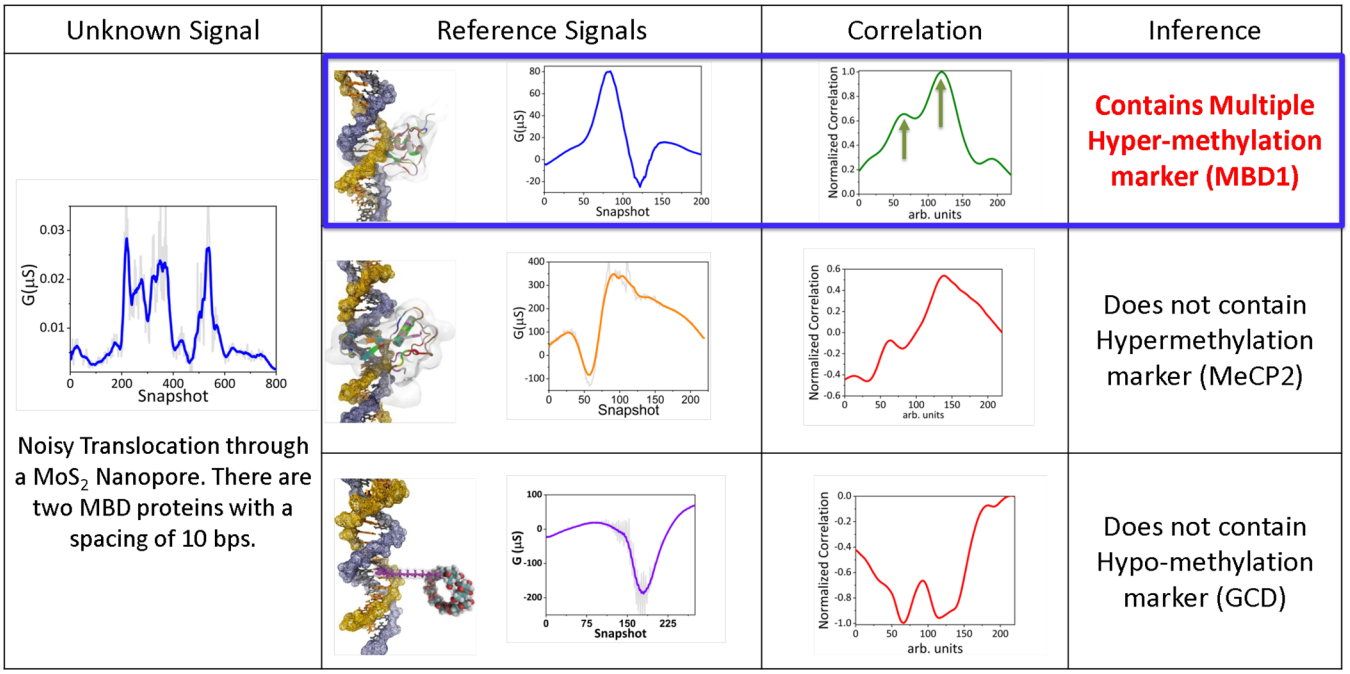


Figure 5. Detecting multiple proteins: A noisy current trace of a DNA strand containing two MBD1 proteins (as shown in the leftmost panel) is correlated with each of the entries in the dictionary. Two peaks are observed when correlated with the frozen DNA signature of a DNA−MBD complex with one protein (middle panel, rightmost column). The absence of such a peak from the correlation of the noisy signal with the other frozen DNA−marker complex signals indicates the presence of an MBD protein in the measured noisy signal.

proteins, we illustrate an algorithm that is capable of automatically deciding the validity of a marker and also counting the number of surrounding markers in its vicinity recursively. This unified algorithm is shown in Figure 4. In this algorithm, we utilize two hyperparameters denoting the threshold value (Q_{TH}) of the Q factor, indicating the validity of the particular hypothesis, that is, to make the decision of whether the vicinity protein is present or absent, and a threshold correlation coefficient (ρ_{TH}), indicative of the presence of the second protein in the vicinity.

The algorithm has two inputs: the dictionary of signals from the various epigenetic cancer markers (in our case, MBD1, MeCP2, and GCD) and a test signal from a DNA complex with an unknown marker protein. Initially, the normalized correlation and corresponding Q factors between the test and dictionary current signatures are calculated, the maximum of which indicates the identity of the marker protein. The second peak in the calculated correlations is also simultaneously monitored to infer the presence of another marker in the vicinity. If the value of the second correlation peak (ρ_{second}) is greater than ρ_{TH}, then the correlation is indicative of a second marker protein of the same type present in the vicinity. We next consider the normalized correlations between the frozen single DNA-marker complex and the noisy signals from the translocation of the DNA with a single complex (known as reference correlation) and multiple complexes (of the same type such as MBD/GCD). When these correlation peaks are aligned, a difference between them will result in the presence of a peak. From this stage onward, we recursively monitor the value of the value of second peak while subtracting the reference correlation at each iteration. The value of ρ_{second} is used to determine the presence or absence of a marker-protein in the vicinity depending on a hyperparameter threshold value (ρ_{TH}). This process of detecting and determining the value of the second peak from the subtracted signal can be performed recursively until the hypothesis is no longer valid.

We can utilize the algorithm illustrated in Figure 4 to detect and count multiple proteins complexed to a DNA, as shown in Figure 5. In this simulation scenario, we consider a 60 base pair (bp) long DNA strand that consists of two methylated CpG sites that are separated by 10 bp's. Each of these individual CpG sites are complexed by MBD1 proteins. The reason that the spacing of 10 bps chosen is due to the physical dimensions of the label protein MBD1 being 10 bp's wide, making 10 bp's the minimum possible distance that the two labeled sites can be present without mutual steric hindrance from the labels. This unconstrained DNA−MBD complex is translocated to obtain the noisy signal, as shown in the leftmost column of Figure 5. This normalized current signature of two MBD1 proteins located 10 bp's apart was previously obtained by us.[14]

The plot of the correlation between the noisy two-protein signal and the reference signals for DNA−GCD or DNA−MeCP2 complexes does not display a discernible peak, indicating a lack of similarity between the dictionary signal entry and the measured noisy signal. On the contrary, correlating the noisy two-protein signal with the frozen single-protein signal (second panel, second row) yields two peaks that could correspond to the similarity of features between each of the individual protein signatures and the single protein dictionary entry. The second peak of the correlated signal is greater than the threshold, indicating the possible presence of the second protein.

The validity of the presence of the second protein can also be determined according to the recursive matched filter algorithm shown in Figure 4. To count the number of proteins, we utilize the correlations as shown in Figure 5, where the normalized values of the frozen single-protein current signature with the noisy two-protein signal (red) and the noisy single-protein signal (blue, reference signal, i.e., second row, third

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

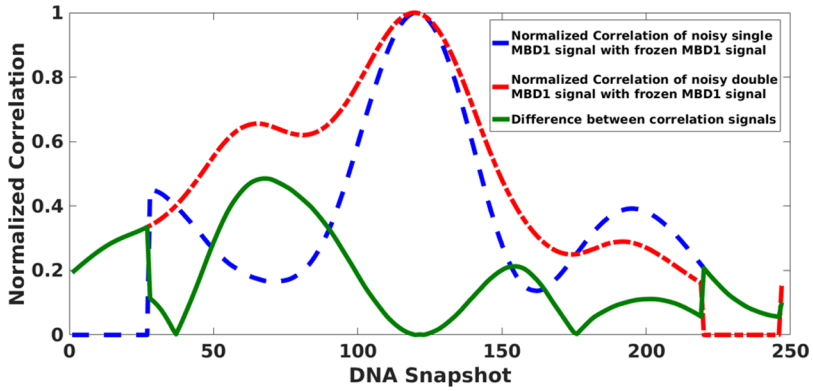


Figure 6. Counting the number of proteins: To count the number of proteins, we consider the normalized correlations between the frozen single MBD−DNA complex and the noisy signals from the translocation of the DNA with a single-MBD (blue) and multiple-MBD (red). A difference (green) between the two aligned correlation peaks results in the presence of a peak, the value of which can be used to determine the presence or absence of the second protein depending on a preset threshold value.

panel of Figure 2) are plotted. As shown in the algorithm, we monitor the value of the second peak to determine the presence of the second protein. Because $\rho_{second} > \rho_{TH}$ peaks between the two normalized signals (correlated and reference correlation signal) are aligned and subtracted, we can examine the height of ρ_{second} in the resulting curve (green curve), as shown in Figure 6. Because the second peak of the green curve is less than ρ_{TH}, we conclude that only two proteins are present. This algorithm can be performed recursively for counting multiple proteins in the vicinity. However, the real-time detection and classification of epigenetic biomarkers from a time series of current data would involve a combination of event detection similar to the approaches utilized by nanopolish in Oxford Nanopore's MinION basecaller,[27] applied to solid-state nanopores. However, a detailed implementation of this algorithm remains to be tackled as a future problem. Additionally, this approach can currently count the quantity of marker proteins in the vicinity only if they are of the same type. We intend to generalize this algorithm in future work to detect different marker proteins within the resolution limit by checking if ρ_{second} correlates with the dictionary signals. We would like to note here that this approach is sensitive to the choice of hyperparameters, which need to chosen specifically depending on the dictionary signals obtained and can be fine-tuned depending on the set of proteins signatures available.

In summary, considering signals from solid-state nanopore devices, we have outlined an algorithm to determine the type of marker proteins and simultaneously identify the possible epigenetic markers in the vicinity. This approach has been illustrated to detect and identify the presence of different biomarkers corresponding to hypo- and hypermethylation within a limited set of dictionary signals. It has not yet been tested on experimental traces due to the technical difficulties in the fabrication and electric measurements in the nanopore devices.[28] However, we strongly believe that it is general enough to be expanded to include an exhaustive set of current signatures for various epigenetic markers calculated from different sensing materials. This approach can also be generalized to incorporate signals from different noise models in the matched filter algorithm. In fact, our algorithm of using matched filter banks can be efficiently implemented in hardware,[24,25] which can enable the development of a DNA sensor chip consisting of a highly dense array of nanopores[29] with sensing and inference logic realized on the same wafer.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.8b02200.

Detailed description of the molecular dynamics simulations protocol to obtain noisy current signatures, followed by the electronic-transport model description for graphene and MoS_2 nanopore membranes to obtain the sheet currents of the biomarkers translocating through the nanopores. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*A.S.: E-mail: sarathy2@illinois.edu.
*N.B.A.: E-mail: nathreya@illinois.edu.
*L.R.V.: E-mail: varshney@illinois.edu.
*J.-P.L.: E-mail: jleburto@illinois.edu.

ORCID

Jean-Pierre Leburton: 0000-0002-8183-5581

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Prof. Amit Meller for his insightful suggestion of utilizing γ-cyclodextrin for hypomethylation detection. This work was supported by grants from Illinois-Proof of Concept (i-POC) fund and the CompGen fellowship program at the University of Illinois at Urbana−Champaign. We also acknowledge supercomputer time provided through the Extreme Science and Engineering Discovery Environment (XSEDE) Grant TG-MCB170052 and Blue Waters, which is supported by the NSF (awards OCI-0725070 and ACI-1238993) and the State of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana−Champaign and its National Center for Supercomputing Applications.

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725

REFERENCES

(1) Bayley, H. Nanotechnology: holes with an edge. *Nature* **2010**, *467*, 164−165.
(2) Ehrlich, M. DNA methylation in cancer: too much, but also too little. *Oncogene* **2002**, *21*, 5400−5413.
(3) Kilianski, A.; Haas, J. L.; Corriveau, E. J.; Liem, A. T.; Willis, K. L.; Kadavy, D. R.; Rosenzweig, C. N.; Minot, S. S. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *GigaScience* **2015**, *4*, 12.
(4) Gracheva, M. E.; Xiong, A.; Aksimentiev, A.; Schulten, K.; Timp, G.; Leburton, J.-P. Simulation of the electric response of DNA translocation through a semiconductor nanopore capacitor. *Nanotechnology* **2006**, *17*, 622−633.
(5) Gracheva, M. E.; Aksimentiev, A.; Leburton, J.-P. Electrical signatures of single-stranded DNA with single base mutations in a nanopore capacitor. *Nanotechnology* **2006**, *17*, 3160−3165.
(6) Heerema, S. J.; Dekker, C. Graphene nanodevices for DNA sequencing. *Nat. Nanotechnol.* **2016**, *11*, 127−136.
(7) Laszlo, A. H.; Derrington, I. M.; Brinkerhoff, H.; Langford, K. W.; Nova, I. C.; Samson, J. M.; Bartlett, J. J.; Pavlenok, M.; Gundlach, J. H. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18904−18909.
(8) Shim, J.; Kim, Y.; Humphreys, G. I.; Nardulli, A. M.; Kosari, F.; Vasmatzis, G.; Taylor, W. R.; Ahlquist, D. A.; Myong, S.; Bashir, R. Nanopore-based assay for detection of methylation in double-stranded DNA fragments. *ACS Nano* **2015**, *9*, 290−300.
(9) Jain, M.; Koren, S.; Miga, K. H.; Quick, J.; Rand, A. C.; Sasani, T. A.; Tyson, J. R.; Beggs, A. D.; Dilthey, A. T.; Fiddes, I. T.; et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **2018**, *36*, 338−345.
(10) Garalde, D. R.; Snell, E. A.; Jachimowicz, D.; Sipos, B.; Lloyd, J. H.; Bruce, M.; Pantic, N.; Admassu, T.; James, P.; Warland, A.; et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **2018**, *15*, 201−206.
(11) Girdhar, A.; Sathe, C.; Schulten, K.; Leburton, J.-P. Graphene quantum point contact transistor for DNA sensing. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16748−16753.
(12) Qiu, H.; Sarathy, A.; Leburton, J.-P.; Schulten, K. Intrinsic stepwise translocation of stretched ssDNA in graphene nanopores. *Nano Lett.* **2015**, *15*, 8322−8330.
(13) Sarathy, A.; Qiu, H.; Leburton, J.-P. Graphene nanopores for electronic recognition of DNA methylation. *J. Phys. Chem. B* **2017**, *121*, 3757−3763.
(14) Qiu, H.; Sarathy, A.; Schulten, K.; Leburton, J.-P. Detection and mapping of DNA methylation with 2D material nanopores. *npj 2D Mater. Appl.* **2017**, *1*, 3.
(15) Gilboa, T.; Torfstein, C.; Juhasz, M.; Grunwald, A.; Ebenstein, Y.; Weinhold, E.; Meller, A. Single-Molecule DNA Methylation Quantification Using Electro-optical Sensing in Solid-State Nanopores. *ACS Nano* **2016**, *10*, 8861−8870.
(16) Parry, L.; Clarke, A. R. The roles of the methyl-CpG binding proteins in cancer. *Genes Cancer* **2011**, *2*, 618−630.
(17) Du, Q.; Luu, P.-L.; Stirzaker, C.; Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **2015**, *7*, 1051−1073.
(18) Lopez-Serra, L.; Esteller, M. Proteins that bind methylated DNA and human cancer: reading the wrong words. *Br. J. Cancer* **2008**, *98*, 1881−1885.
(19) Jang, J.-S.; Lee, S. J.; Choi, J. E.; Cha, S. I.; Lee, E. B.; Park, T. I.; Kim, C. H.; Lee, W. K.; Kam, S.; Choi, J.-Y.; et al. Methyl-CpG binding domain 1 gene polymorphisms and risk of primary lung cancer. *Cancer Epidemiol. Prev. Biomarkers* **2005**, *14*, 2474−2480.
(20) North, D. O. An Analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems. *Proc. IEEE* **1963**, *51*, 1016−1027.
(21) Woodward, P. M. *Probability and Information Theory, With Applications to Radar*; McGraw-Hill: New York, 1953.
(22) Turin, G. An introduction to matched filters. *IRE Trans. Inf. Theory* **1960**, *IT-6*, 311−329.
(23) Raillon, C.; Granjon, P.; Graf, M.; Steinbock, L.; Radenovic, A. Fast and automatic processing of multi-level events in nanopore translocation experiments. *Nanoscale* **2012**, *4*, 4916−4924.
(24) Kotelnikov, V. A. *The Theory of Optimum Noise Immunity*; McGraw-Hill: New York, 1959.
(25) Furth, P. M.; Andreou, A. G. A design framework for low power analog filter banks. *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.* **1995**, *42*, 966−971.
(26) Parkin, W. M.; Drndić, M. Signal and Noise in FET-Nanopore Devices. *ACS Sensors* **2018**, *3*, 313−319.
(27) Simpson, J. T.; Workman, R. E.; Zuzarte, P.; David, M.; Dursi, L.; Timp, W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **2017**, *14*, 407−410.
(28) Heerema, S. J.; Vicarelli, L.; Pud, S.; Schouten, R. N.; Zandbergen, H. W.; Dekker, C. Probing DNA translocations with inplane current signals in a graphene nanoribbon with a nanopore. *ACS Nano* **2018**, *12*, 2623−2633.
(29) Athreya, N. B. M.; Sarathy, A.; Leburton, J.-P. Large Scale Parallel DNA Detection by Two-Dimensional Solid-State Multipore Systems. *ACS Sensors* **2018**, *3*, 1032−1039.

DOI: 10.1021/acs.jpclett.8b02200
J. Phys. Chem. Lett. 2018, 9, 5718−5725