

Pore Characterization and Event Detection in Solid-State Nanopores

Garrett Wong

Advised by Christopher O'Donnell and William Dunbar
Baskin School of Engineering
UCSC

Abstract—Nanopores are used for DNA sensing. Solid-state nanopores, which are milled through a silicon-based substrate, lack the atomic-level geometric precision of biological protein-mediated pores. However, they show great promise due to their greater stability and potential for modification. We developed tools to characterize solid-state nanopores by using their resistance to infer a functional diameter estimate, and developed a rudimentary protocol for DNA translocation event detection which is robust to signal degradation.

I. INTRODUCTION

In a nanopore sensing regime, a membrane perforated by a nano-scale hole separates two regions, which we label *cis* and *trans*, all immersed in an ionic solution. When a voltage is applied across the membrane, voltage force drives a current through the pore. This current can be measured, allowing us to effectively measure the resistance of the pore. A general nanopore setup can be seen in figure 1.

This regime is useful for sensing DNA, which holds a negative charge. When DNA, with a diameter of about a nanometer, is pulled through the pore, the current through the pore is attenuated. This measured current attenuation indicates the DNA translocation event.

A. Biological nanopores

There are two primary foci of research for nanopores, separated by the chosen pore mediator: biological pores and solid-state pores. In a biological nanopore, a pore-mediating protein forms a precise hole in a lipid membrane. Biological nanopores have been created and studied for decades, and their activity is relatively well-characterized compared to solid-state nanopores [1].

One of the greatest benefits of the biological nanopore is the incredible similarity between pores: because the pore is mediated by a protein, the geometry of the pore is incredibly precise, to an atomic-scale precision. Accordingly, data from biological nanopores lacks much of the noise characteristic of solid-state nanopores, and characterization of individual pores is a trivial task beyond ensuring that the pore was properly constructed.

However, the geometric precision which characterizes these nanopores also serves as the biological nanopore's greatest weakness. The biological nanopore best characterized, α -hemolysin, is only around 1.5 nanometers across, which will admit single-stranded but not double-stranded DNA [1]. This geometric rigidity is difficult to overcome due to the difficulties associated with protein engineering, and prevents useful

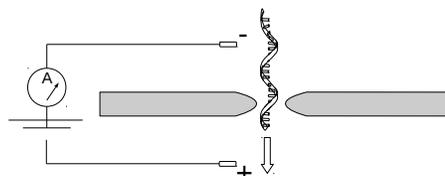


Fig. 1. A nanopore sensing setup, where a voltage differential across the membrane pulls DNA through a pore. Current through the pore is sensed, and can be used to identify and characterize DNA translocation events.

experiments like sensing of nucleosomes or other DNA-bound proteins, which would be possible with larger engineered nanopores. Additionally, biological nanopores are relatively unstable, at least in comparison to the resilience of solid-state nanopores [1].

B. Solid-state nanopores

In contrast to biological nanopores, solid-state pores are simply milled in a solid substrate, usually a silicon compound. Because the pore is not mediated by a protein, denaturation is not a concern; this leads to great promise for increased portability of devices, since temperature control is less vital to the apparatus's proper functioning [1]. Since the pores are milled, a large variety of diameters of pores can be created, allowing the pore to be customized to the sensing task: small diameter pores for sequencing, and large diameter pores for detection of bound proteins.

Again, the pore mediator leads to significant drawbacks. In solid-state pores, the pore milling process leads to pores with much less precision than the pore-mediating proteins. Because of this, there are significant differences in the signal output by each new solid-state nanopore. Accordingly, in contrast to biological nanopores, characterization of the physical properties of individual nanopores becomes an essential prerequisite to understanding the signal from that pore.

Finally, our solid-state nanopore milling process yields a nanopore diameter near 10 nanometers, which is almost an order of magnitude larger than α -hemolysin. This means that the noise in the signal is larger, and the attenuation of the current signal when 1 nm wide DNA passes through the pore is about an order of magnitude smaller.

This makes the signal processing downstream of data capture more challenging than that for biological nanopores,

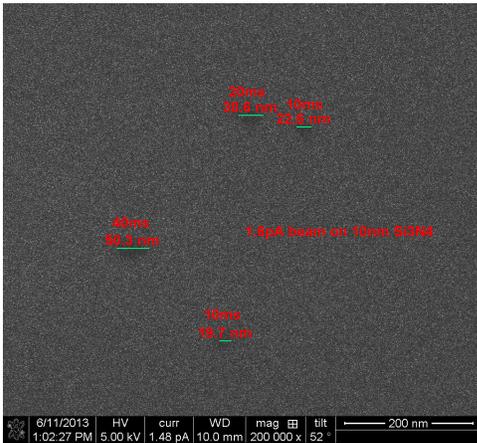


Fig. 2. A tunnelling electron microscope image of several nanopores milled in a silicon-based membrane using an electron beam from a scanning electron microscope. Labels on the figure give rough estimates for pore diameter and the time the electron beam was pulsed to mill each pore. Note that pores bored for the same amount of time differ significantly (up to 10%) in diameter.

and increases the importance of robustness to large noise and systemic signal degradation. The increased importance of the characterization of individual pores in this setting motivates our development of tools to accelerate this characterization process and diagnose problems as quickly as possible.

II. TOOLS DEVELOPED FOR PORE DIAMETER CHARACTERIZATION

Pores milled using a timed pulse of an electron beam from a scanning electron microscope vary significantly in size, as seen in figure 2. In addition, at the nano scale, traditional imaging techniques like the tunneling electron microscope image in figure 2 tell us little about the diameter and geometry of the pore.

Consequently, we need to use a different method to infer the diameter of the pore. Using the resistance of the pore to infer diameter is especially apt, as this is the functional measurement we seek, and so systemic error arising from differences between geometry and measurement are naturally arrested [2].

A. Data acquisition

In order to characterize the resistance of the pore, we take current measurements at a variety of voltages. In a typical testing regime, for example, we step the current between 200 and -200 mV in 20 mV steps, and repeat this pattern for five minutes, with current and voltage data taken 250,000 times per second. This data is imported into MATLAB for analysis. Accordingly, our tools are developed in MATLAB.

B. Measurement of resistance

We first seek to fit a line to the data in order to measure the resistance of the pore in accordance with Ohm's Law. However, because there is inevitably some capacitance in the system, we need to cull data which is not at steady-state. In the tool we developed, we filter data by dividing data points into bins of width greater than the impulse response time between

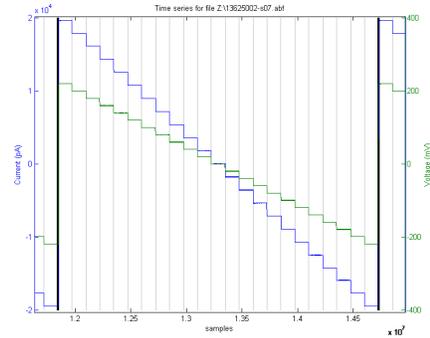


Fig. 3. A time series plot of the current and voltage over time. Data discarded for not being steady-state is demarcated with gray vertical lines. Thick black lines mark frame boundaries.

voltage change and current change and throw out all of the data in bins for which the difference between minimum and maximum values falls above an established threshold.

This method is very conservative, and ensures that data taken between the voltage jump and current response (which is locally difficult to distinguish from a true steady-state) is thrown out. However, relatively little of the data (usually around 1 percent) is culled, as seen in figure 3; additionally, we can see from figure 4 that the culling is quite successful in removing non-steady-state data. In addition, we sense the boundaries between voltage stepdowns (when the voltage jumps from -200 mV to 200 mV) and divide the data between these boundaries into frames so that we can more easily analyze differences that arise across the course of data acquisition.

Once the data has been culled and divided into frames, our tool simply plots the I-V curve. We plot data from subsequent frames, as well as the least-square best fit line for those data, in gradient colors in order to show any change throughout the course of measurement. By plotting this information, our tool allows the user to easily identify and diagnose any systemic change in the I-V curve across the period of data acquisition.

For example, in the I-V curve seen in figure 4, we can see that the current at the 160 millivolt step dropped steadily from frame to frame. This indicates that the resistance of the pore is dropping slowly over time, indicating that, for example, a pore diameter estimate taken over a greater period of time is likely to be larger than an estimate taken for only a short period of time.

Finally, we calculate the best fit line for all of the data and plot it in black, and return the slope (resistance).

C. Inference of Pore Diameter

In order to infer the pore diameter, we first need to know the experimental conditions under which the data was taken. Specifically, the user is prompted for the concentration and type of salt in the buffer so that we can estimate the conductivity of the buffer used. Because the conductivity of electrolytes in aqueous solutions is empirically derived, it is nontrivial to analytically convert this into a conductivity estimate.

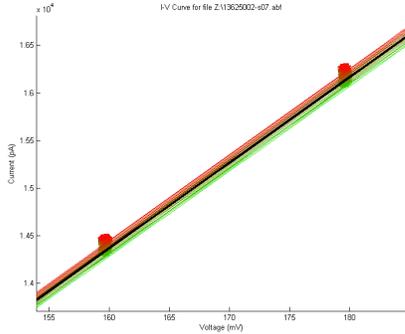


Fig. 4. A section of a typical I-V curve. The demonstrated lack of outliers indicates the success of our stringent culling protocol. Data points (plotted with “+”) and best fit lines are plotted in a red-green gradient across frames, with the first frame in red and the last frame in green. Note the drop in current over subsequent frames, indicating that the conductance of the pore slowly increased over the course of data acquisition.

In our tool, we hard-code in conductivity data at a variety of concentrations as (given by CRC Handbook) for KCl and LiCl, the two salts we foresee using. If the user enters a different salt, the tool simply prompts for the conductivity data. Often, we need to interpolate between known values of conductivity. Because conductivity curves are empirically derived, we simply use spline interpolation in these cases.

Finally, we estimate pore diameter using the method described by Kowalczyk et al [2]. In short, the model assumes a pore geometry which is hyperboloid rather than the cylindrical geometry used in past models. This pore geometry, Kowalczyk et al. show, aligns better with experimental data and with intuition about the genuine geometry of solid-state pores.

III. DNA TRANSLOCATION EVENT IDENTIFICATION

We wish to measure the duration of DNA translocation events, the current attenuation during events, and the time between events to characterize the functionality of the solid-state pore.

Ideally, a DNA translocation event manifests in the current signal as a square-wave attenuation. However, our solid-state nanopores exhibit significant noise in the signal. In addition, experimental setup, filtering, imperfections in the pore, and capacitive effects all contribute to systemic signal degradation. Due to the significantly degraded signal quality, robustness becomes a dominant factor in algorithm choice for the identification of DNA translocation events.

Several algorithms which were designed with piecewise constant data in mind were rejected on the basis of the dominant importance of robustness in our particular use case.

For example, Max Little’s methods and tools for noise removal in piecewise constant signals [3], in which a generalized functional is created which integrates many different denoising methods, and is minimized using a general solver algorithm, are particularly suited to data with events which take on a square-wave form, much like the data from a solid-state nanopore ideally would. Unfortunately, the capacitive effects in our data, which smooth the ideally instantaneous

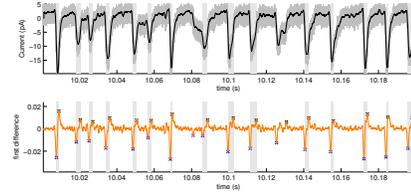


Fig. 5. Above, a time-series plot of experimental data in medium grey, with Savitsky-Golay filtered signal in black. Below, a time-series plot of Savitsky-Golay derivative, with identified extrema marked with “x”. In both plots, identified events, which occur between extrema, are marked by a light grey background.

transitions of current attenuation, violate the piecewise constant assumption too flagrantly for the algorithm to adequately cope, leading to an insufficiently denoised signal.

Likewise, Ajmami et al’s event identification algorithm is specifically designed for the identification of DNA translocation signals in nanopores [4]. This algorithm seeks to identify event boundaries as extrema in the second derivative of the signal (points of greatest convexity/concavity in the original function). However, this algorithm relies on a stable second derivative, which is difficult to obtain with the noise levels we see in our data. In addition, the systemic signal degradation we face presents a pathological case for the identification of points of greatest convexity/concavity, which are quite variable under this signal degradation.

Instead, our proposed event identification simply identifies event boundaries as extrema in the first derivative (inflection points in the original signal). Notably, in the ideal case, these should be a constant distance away from the identified event boundaries of Ajmami et al., preserving the event durations between algorithms. The pathological signal degradation that significantly skews event boundaries for Ajmami et al. has mitigated effect on these event boundaries, leading to an event identification algorithm that is very robust. Additionally, this method, due to its simplicity, leads to fast implementation and runtime.

Our implementation of this rudimentary event identification algorithm uses Savitsky-Golay to filter and differentiate; we found that this method preserves high-frequency components but also leads to a much more stable derivative estimate than, for example, a naïve first difference.

As seen in figure 5, our implementation also plots the time-series signal and shows identified events, allowing the user to fine-tune the thresholds for event detection.

IV. CONCLUSION

Solid-state nanopores, milled through a silicon substrate, lack the atomic-level geometric precision of biological, protein-mediated pores, but show great promise due to their increased stability and potential for modification. We created a flexible program which gives pore diameter estimates for a variety of experimental conditions, and allows easy extensibility to unforeseen experimental conditions. Additionally, we developed a rudimentary DNA translocation event identification algorithm which is more robust to systemic signal degradation than existing algorithms. These tools can be used to begin the

characterization of new nanopores, allowing the more accurate estimate of pore properties and the rapid diagnosis of problems with experimental setup.

ACKNOWLEDGMENT

The author would like to thank Shea Ellerson for her indispensable contribution to the project.

REFERENCES

- [1] R.D. Maitra, J. Kim and W.B. Dunbar, *Electrophoresis* 2012, 22, 3418–3428
- [2] S.W. Kowalczyk, A.Y. Grosberg, Y. Rabin, C. Dekker, *Nanotechnology* 2011, 22, 315101
- [3] M.A. Little, N.S. Jones, *Proc. R. Soc. A* 2011
- [4] N. Arjmandi, W. Van Roy, L. Lagae, G. Borghs, *Nanopores for Bioanalytical Applications: Proceedings of the International Conference, 2012*, 11