Quantitative Biology 2020, 8(3): 187–194
https://doi.org/10.1007/s40484-020-0201-x

MINI REVIEW

DNA sequencing using nanopores and kinetic proofreading

Xinsheng Sean Ling*

Department of Physics, Brown University, Providence, RI 02912, USA
* Correspondence: xinsheng_ling@brown.edu

Received December 5, 2019; Revised January 16, 2020; Accepted January 29, 2020

We discuss the feasibility of using a nanopore sandwich device to implement the principle of kinetic proofreading to discriminate incorrect hybridizing oligonucleotides on a target DNA or RNA. We propose a method of sequencing DNA or RNA using this approach. The design parameters for such a DNA sequencer are estimated from the Hopfield-Ninio theory of kinetic proofreading and Schrödinger's first-passage-time distribution function.

Keywords: DNA sequencing; nanopore sequencing; biosensing using kinetics

Author summary: A novel device concept is proposed for DNA sequencing without employing polymerases that were essential in Sanger and other DNA sequencing methods. The idea is to combine the nanoscale sensitivity of solid-state nanopores and the kinetics of oligonucleotides. The device consists of two nanopores in close proximity such that correct hybridization probes stay on the DNA during the passage, while the incorrect ones will be melted off during the transit. The proposed device contains two essential functions of a DNA polymerase, suppression of Brownian motion and discrimination based on the kinetics of the correct and incorrect Watson-Crick pairs.

INTRODUCTION

DNA sequencing, namely the determination of the sequential arrangement of the four basic nucleotides on a DNA molecule, involves at least two fundamental challenges. The first challenge is to determine the identity of the nucleic acids, *i.e.*, adenine (A), guanine (G), thymine (T), and cytosine (C). The second is to accurately determine the relative positions between the nucleotides. The challenges are immense due to the minute differences in their chemical composition and physical scales. There are two hydrogen bonds between A-T pairing, and three between G-C pairs. The nearest neighbor nucleic acids are spaced at 0.34 nm in the double helix compact form [1], at 0.7 nm when it is single stranded [2,3].

To date, the most accurate method of sequencing a DNA molecule was developed by Sanger and coworkers, known as the Sanger method [4,5], or the method of 3′ dideoxy chain termination. In the Sanger method, before the start of a DNA replication polymerase chain reaction (PCR), dideoxyribonucleoside triphosphates (ddNTPs) with proper coding (using dye or other labels) are mixed with the four deoxyribonucleoside triphosphates (dNTPs). When a ddNTP is incorporated by the DNA polymerase into the growing strand of DNA, the reaction is terminated due to the lack of an OH group at the 3′ position of the ddNTP which is required for forming a phosphodiester (covalent) bond with the next dNTP [4]. These terminated strands of DNA are melted off from the template strand and separated by electric field in a gel and then compared with the ddNTP coding [5]. The order by which the termination occurs provides the DNA sequence of the synthesized strand. In this approach, the order of the genetic sequence of A, T, G, C is determined without a direct measurement of their physical positions on the DNA. With high fidelity polymerases, the Sanger method is capable of reaching 10^{-5}–10^{-6} in the error rates per base.

Among all commercially available DNA sequencing technologies, the Sanger method is most precise, but it is inherently slow, hence costly. To speed up the process and lower the costs, new methods of sequencing-by-synthesis

(SBS) have been developed and commercialized [6,7]. These next generation sequencing (NGS) technologies still rely on DNA polymerase and Sanger's 3′ termination idea, but use greatly simplified base calling procedures. Instead of separating out the dideoxy-terminated strands from other strands of DNA by gel electrophoresis, the SBS method utilizes localized (by attaching sample ssDNA to the surfaces of beads or wells) colonies of identically created ssDNA (by PCR), and incorporates only dye-coded 3′-blocked dNTPs. After each incorporation reaction, all non-incorporated dNTPs are washed off from the sample, only those incorporated stayed on the sample. Each colony is then imaged and its dominant color determines the base that was incorporated. Next all dyes are bleached out, the 3′ blocking groups are cleaved, opening the DNA 3′ end for the next incorporation cycle. Since these tasks can be done in parallel, many fragments of DNA can be sequenced simultaneously, greatly speeding up the sequencing tasks. Similar base calling procedure is also found in the platform of pyrosequencing [8].

The enhanced speed of the NGS SBS technology comes at a price in sequencing errors. In the SBS method, typically one uses DNA polymerases with reduced 3′ to 5′ exonuclease proofreading activities to avoid the incorporated bases being chewed off [9]. Thus the incorporation error is enhanced due to the lack of proofreading capabilities, to about 10^{-3} per base. Furthermore, the residues left on the dNTP after the 3′ blocker was cleaved, and the quenched dye molecules interfere with the Watson-Crick pairing of the double-strand formation and the polymerase incorporation function, leading to more incorporation errors. Thus for *de novo* sequencing, all DNA sequences generated with SBS technologies should be viewed with caution, especially with regard to single-nucleotide mutation [10].

A more ambitious idea is to do away DNA polymerases altogether in the DNA sequencing process, drastically reducing the need for expensive reagents, and avoiding the intrinsic errors of DNA polymerases. In 1996 Kasianowicz and coworkers proposed a novel concept: nanopore sequencing [11]. Their idea was to use the nanopore ionic current as a direct sequencing mechanism. They conjectured that as a DNA passes through a nanopore in an insulating membrane (lipid or solid-state) with a diameter so small that only single-stranded DNA can pass through, and with a pore length so short that individual nucleotides passing (translocating) through can be distinguished in their signatures in the ionic current through the pore, one may be able to read out the DNA sequence directly in the current variations. This simple concept has generated tremendous interests during the past two decades [12].

From physics perspective, this idea is intriguing since the electric-field driven DNA translocation takes place at room temperature, thermal effects in the DNA translocation process are expected to be important [13]. Indeed, subsequent experiments [14,15] confirmed the importance of Brownian motion of DNA molecules during electric-field driven translocation.

Given the importance of thermal effects, a compromised solution, hybridization-assisted nanopore sequencing (HANS), was proposed [16]. This approach was based on Southern's original idea of sequence detection using hybridization probes (Southern blotting [17]), called sequencing-by-hybridization (SBH) [18]. However, it quickly became clear that SBH is a fundamentally flawed concept since if a probe binds at multiple locations on a long ssDNA, it becomes mathematically impossible to reconstruct a unique sequence based on the hybridizing probes. While the HANS platform does bring SBH from a state of flawed idea to a plausible concept, it brought a new challenge, a large number of samples that have to be analyzed. For example, for probes of oligonucleotides of n-mers, there are $N = 4^n$ possible combinations of A, T, G and C. For $n = 12$, it is a library of $N = 16{,}777{,}216$ oligonucleotides.

For years, the main thinking in SBH or Southern sequencing research was to design long, stable, hybridization probes [18]. Short oligo probes in the range of 2–5-mers have never been considered useful for SBH since their life-times on an ssDNA are too short for imaging. Inspired by the 3′ to 5′ exonuclease proofreading function of DNA polymerases, we exnoticed that it is entirely feasible to use solid-state nanopore measurements to distinguish the life times between the correct and incorrect probes in hybridization. In fact, we showed that it is possible to observe hybridized vs. de-hybridized events in nanopore experiments [19]. This observation led us to the realization [20] that by using the melting kinetics of shorter probes, it may be more efficient in discriminating against incorrect probes. As shown in Fig. 1, one simple way to discriminate an incorrect probe from a correct one is to do a second measurement after a waiting time t_w [20]. We show below that it is possible to use the Hopfield-Ninio kinetic proofreading principle [21,22] to carry out high precision *de novo* sequencing.

A NEW PRINCIPLE FROM NUMERICAL RESULTS

As shown in Fig. 2, left panel, we propose a nanopore sandwich device [23] to perform the wait-time task outlined in Fig. 1. The two nanopores are separated by a small cavity such that they form a two-resistors-in-series device. The physical parameters for the pores are designed in such a way that the two pores have different resistances. When a DNA hybridization segment passes

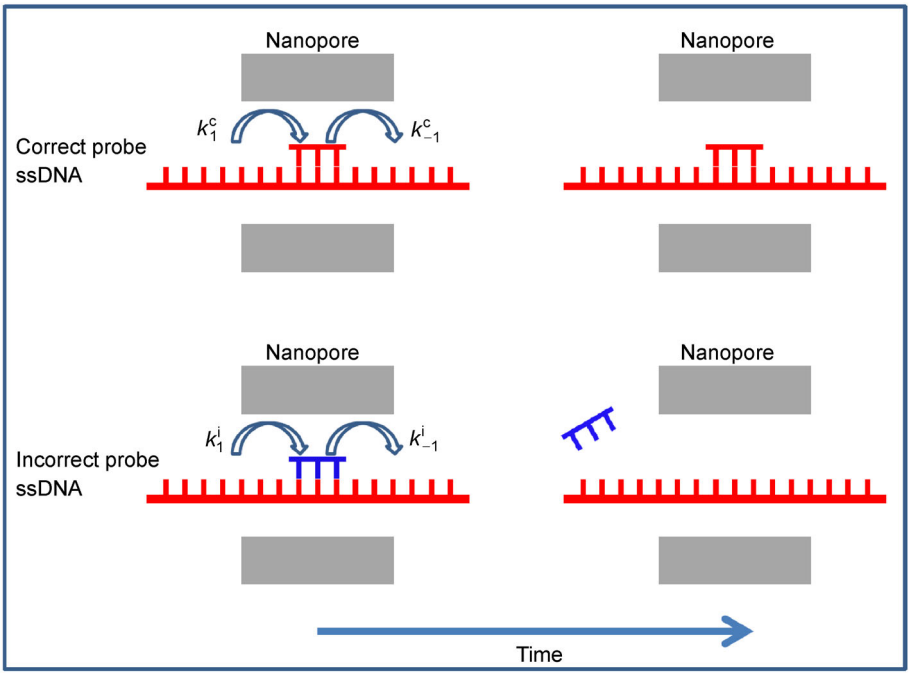

Figure 1. The basic concept of nanopore sequencing using hybridizing and melting kinetics of oligonucleotides on a target DNA. (Top left) A single-stranded DNA inside a nanopore with a hybridizing 3-mer oligo probe, where k_1^c and k_{-1}^c are the on- and off-rates of a correct probe, respectively. (Top right) The correct probe stays hybridized after a waiting time. (Lower left and right) An incorrect probe hybridizes the target ssDNA, then melts off during the waiting time. Respectively, k_1^i and k_{-1}^i are the on- and off-rates of an incorrect probe.

through these two pores, the ionic current through the whole device shows two dips that are of different characters.

An example of a correct-probe current profile, *i.e.*, the ionic current vs. the position of the probe, calculated using COMSOL is shown in the top right panel in Fig. 2. During a sequencing run, by dragging the DNA through the pore at a speed of 1 μm/s, we effectively measure the probe twice with a wait time of 20 nm/1 μm/s = 20 ms. For an incorrect probe, it will be melted away from the DNA before it reaches the second pore. Since a 3-mer oligo has a large diffusion constant, its passage through the pore will not be observable in the pore current. Thus the expected incorrect probe current profile has only one dip, as shown in the lower right in Fig. 2.

We can apply the well-known theory of kinetic proofreading [21,22] to analyze the accuracy of using the above technique for DNA sequencing. The probability density $[p{:}S]$ of probes of concentration $[p]$ binding to a target substrate $[S]$ (= 1) is governed by the rate equation [21,22]:

$$\frac{\mathrm{d}[p{:}S]}{\mathrm{d}t} = k_1[p][S] - k_{-1}[p{:}S] \qquad (1)$$

where k_1 and k_{-1} are the on- and off-rates of the probe, respectively.

The symbol $p{:}S$ represents a hybridized probe. The on-rate k_1 is determined by the concentration and the diffusion constant of the probe molecules in the buffer. It should be constant for probes of the same length. However, the off-rate k_{-1} should be sensitive to the nature of hybridization. For example, for a target site of 5′-AGT-3′, the probes of 3′-TCA-5′ will bind for a longer time than those of 3′-GCA-5′ due to the difference in the number of hydrogen bonds they form.

According to Eq. (1), the equilibrium probability density should be $[p{:}S] = (k_1/k_{-1})[p][S]$. For a correct probe, with perfect Watson-Crick pairings, $[p_c{:}S] = (k_1^c/k_{-1}^c)[p_c][S]$; for an incorrect probe, $[p_i{:}S] = (k_1^i/k_{-1}^i)[p_i][S]$. For the same applied concentration of probes, $[p_c]=[p_i]$, the error rate for sequencing based on equilibrium probability densities, $R_0 = [p_i{:}S]/[p_c{:}S] = k_{-1}^c/k_{-1}^i$ here we used $k_1^c \approx k_1^i$ (the on-rates are mainly determined by diffusion).

Since the off rates are determined by the probe binding free energies, the equilibrium error rate is expected to be determined by the difference in their free energies, $R_0 \sim e^{-\Delta G/k_B T}$, where ΔG is the free energy difference between correct and incorrect probe bindings. For a mismatch of one base on an incorrect probe, the binding free energy is weaker by an amount $\Delta G \sim 3$ kcal/mol ~ 5

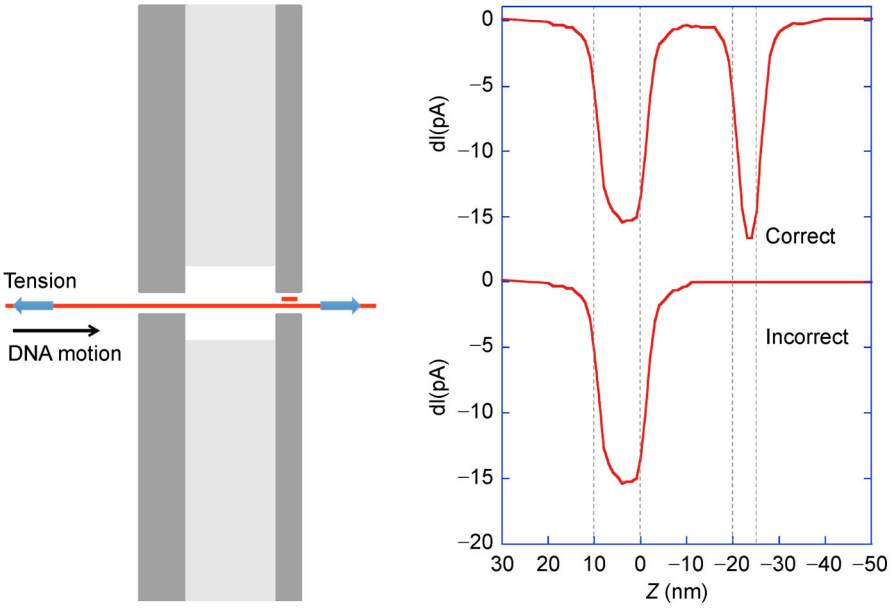

Figure 2. Sandwich nanopore device geometry and calculated signals. (Left) A device concept for a sandwich nanopore structure. The dark layers indicate a robust material such as silicon nitrides in which uniform pores can be drilled through using ion beams. The lighter material such as silicon oxides between the pores in the two dark layers is etched off using chemicals, leaving a nanoscale cavity in which the oligos that fall off the ssDNA can escape into. (Right) The calculated ionic current traces using COMSOL for a probe that stayed on the ssDNA during the entire translocation which is judged to be the correct one (top), and that fell off before arriving at the second pore which is deemed incorrect (bottom). The device parameters are: from left to right 10 nm/20 nm/5 nm, diameter of the small pores is 5 nm, the diameter of the inner cavity is 20 nm. The conductivity of 1 M KCl at room temperature is used. Applied voltage = 0.1 V. The background ionic current is subtracted.

k_BT, per molecule [24]. Thus the order of magnitude of the error rate based on equilibrium values alone is $R_0 \sim 9e^{-\Delta G/k_BT} \sim 9\mathrm{e}^{-5} \sim 9\times 0.007 = 0.063 = 6.3\%$. The multiplicity pre-factor of 9 is due to the fact that for every correct probe, there are 9 incorrect probes with one mismatched nucleotide [25]. The probes with two or three mismatches can be ignored. The basic error rate of 6.3% per base is an unacceptable starting point for a DNA sequencing technology. No matter how well the rest of the sequencing machine is constructed, this rate is determined by the fundamental kinetics of the probe binding-unbinding processes.

However, if after a probe binding event has been detected, we take a second measurement after some waiting time to determine the probability of the probe staying on the binding site, the rate equation should instead be [21,22],

$$\frac{\mathrm{d}[p{:}S]}{\mathrm{d}t} = -k_{-1}[p{:}S] \qquad (2)$$

Namely, the probability that a probe remains hybridized decays with time exponentially, $[p{:}S] \sim e^{-k_{-1}t}$. Figure 3 illustrates the time characteristics of correct and incorrect

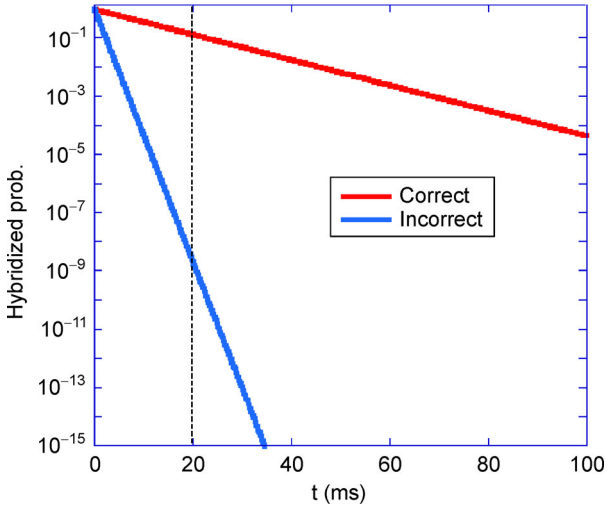

Figure 3. Discrimination using kinetics. The time dependence of the probability that 3-mer oligo probes with perfect match (correct) and with one mismatch (incorrect) stay hybridized onto the target DNA. The dashed line indicates 20 ms waiting time (the time it takes for the ssDNA to travel the distance that is equal to the thickness of a sandwich nanopore device.

probe staying (or survival) probabilities. Since $k^i_{-1} \gg k^c_{-1}$, the error rate after a waiting time t_w becomes

$$R_w \approx R_0 e^{-k^i_{-1} t_w}. \quad (3)$$

Thus in the approach outlined in Fig. 2, the error rate in base calling has an additional error-correction factor $e^{-k^i_{-1} t_w}$, which can now be controlled by experimental waiting time t_w.

For an incorrect 3-mer probe, with a single mismatch, the off-rate is estimated to be about $k^i_{-1} \sim 10^3$ Hz by applying Kramers theory of reaction kinetics [26–28]. Thus for a moderate wait time of $t_w = 20$ ms, as indicated by the dashed line in Fig. 3, one can achieve an additional factor of $\sim 10^8$ in the suppression of sequencing error.

We envision a sequencing run as follows: (1) For each 3-mer probe, a sample ssDNA is pulled through the nanopore sandwich device at a speed of 1 μm/s while the ionic current of the device is recorded continuously (at a sampling rate of 200 kHz), locate all double-dip features and assign the matching sequence of the probe to these positions; (2) Repeat step-1 for all 64 probes. (3) Align the sequences from all 64 probes based on their positional information to obtain the full sequence of the target DNA. With this nanopore sandwich device, assuming genomic DNA length of 400 μm, one can scan the whole DNA with one probe AGC in 400 seconds. If we measure each probe for 10 times to get better statistics, that would be a little over one hour. To finish all 64 probes would take less than three days. This is comparable to the current NGS SBS platform in terms of time consumption, but it is many orders of magnitude more precise, and costs significantly less since no reagents other than re-usable oligos (and electricity) will be used.

THE GRAND CHALLENGE: SUPPRESSING DIFFUSION

Brownian motion of DNA inside the nanopore has long been considered a challenge to realizing a nanopore DNA sequencing technology. This was manifested in the broadening of the translocation times as observed by Kasianowicz *et al.* [11] and Meller *et al.* [14] in alpha-hemolysin nanopores. These broadening effects were, as pointed by Lubensky and Nelson [13], a consequence of the spread in the first-passage times for biased random walks. For solid-state nanopores, with large pore openings and weaker DNA-pore interactions, the translocation times were also found to be broad [15]. Interestingly, it was found that the exact solution for the first-passage distribution for a 1D biased diffusion model first worked out by Schrödinger [29] in 1915, can be used for the analysis DNA translocation times [30]. In fact, from the exact Schrödinger distribution, one can derive a rigorous criterion for nanopore DNA sequencing using the hybridization probes [30].

As shown in the simplified model in Fig. 4, the black probes are of different sequences from those of the red ones. The basic question is whether one can distinguish the changes in the distance between two probes on a DNA to within δx using the nanopore technique. To resolve Δx (~0.4 nm), the change in the mean first-passage time $\delta\langle T\rangle$ (between red and black) must be larger than the variance of the first-passage times for each set of the probes,

$$\delta\langle T\rangle \geqslant \sqrt{\langle T^2\rangle - \langle T\rangle^2}. \quad (4)$$

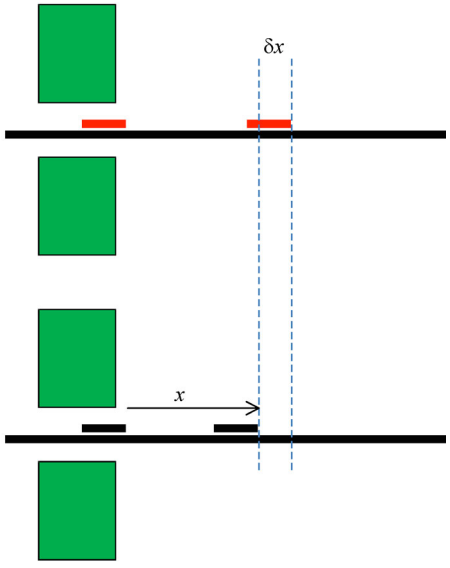

Figure 4. Requirement for a polymerase-free DNA sequencing scheme. (Upper panel) Two oligos on a target ssDNA. The DNA construct travels through the pore from right to left. (Lower panel) Two different types of oligos on the same target ssDNA. The distances between the two sets of probes differ by δx.

Using Schrödinger's theory of first-passage time distribution function to evaluate the averaged quantities in the above relation, one arrives at a rigorous DNA sequencing criterion,

$$\delta x \geqslant \sqrt{2D\frac{x}{v}}. \quad (5)$$

This result was first given in [30].

It was noted [30] that this criterion in the inequality (Eq. (5)) is identical to that used in the technique of gel electrophoresis for separating the DNA fragments in the Sanger method. In the gel electrophoresis [5], for two fragments of DNA to be spatially separated after time t, under an electric field, the diffusion length $\sqrt{2Dt}$ needs to

be smaller than their spatial separation due to their difference in drift velocities (due to the gel), $\delta x = t\delta v$. Since the diffusion length scales as $\sim\sqrt{t}$ and the migration length as $\sim t$, the spatial separation will occur at large enough t.

The Schrödinger inequality (Eq. (5)), which can be rewritten as criterion (Eq. (6)), sets the fundamental limits for the development of a nanopore sequencing technology.

$$v \geqslant x\left(\frac{2D}{\delta x^2}\right). \tag{6}$$

Figure 5 is a "phase diagram" using inequality (Eq. (6)): the four solid lines are calculated using criterion (Eq. (6)) with four different values of diffusion constants, δx = 0.4 nm, $D_0 \sim 1.8\,\mathrm{nm^2/\mu s}$ being from free DNA molecules [31] (it is expected to be even larger for the individual segments). The vertical dashed line indicates the current silicon nitride technology in producing silicon nitride membranes of 5 nm in thickness. The horizontal dashed line indicates the bandwidth, 200 kHz, of the commercial state of-the-art patch-clamp amplifier Molecular Device model Axopatch-200B.

Clearly, without suppressing the diffusion, the Schrödinger criterion (Eq. (6)) rules out most nanopore sequencing concepts [12] based on position measurements. However, it was known that by simply holding DNA under tension [32,33], one can effectively change the diffusive behavior of the DNA to that of a particle trapped in a harmonic well with effective diffusion constant suppressed. As shown in Fig. 5, by reducing the effective diffusion constant down by a factor of 10^3 can open up a parameter space for a nanopore DNA sequencing technology. The future nanopore technology development should be focused on designing mechanisms to suppress the Brownian motion of the DNA segment inside the pore.

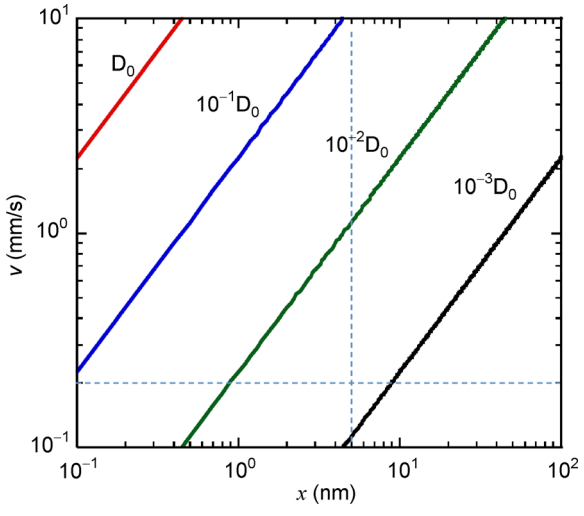

Figure 5. The parameter space for the proposed sequencer. The lines are constructed using the Schrödinger's inequality with different levels of Brownian motion as indicated by the values of the diffusion constants. The vertical dashed line indicate the current silicon nitride technology for thin film thickness, the horizontal dashed line is set by the bandwidth of the commercial patch-clamp electronics Exopatch 200B. The likely parameter space for a solid-state nanopore sequencing machine is the triangular region formed by the two dashed lines and the black solid line.

Thus a nanopore DNA sequencing machine will have to function like a DNA polymerase: it has to suppress the Brownian motion of the DNA segment that is being probed, and it has to be able to discriminate against incorrect hybridization probes beyond the thermodynamics of the hydrogen bonds.

DISCUSSION

In the proposed base-calling approach in Fig. 2 (inset), there is a mechanism for false positives: an incorrect probe that falls off during transit between the two pores finds itself re-hybridized onto the target site of the DNA.

The simple solution is to operate at low concentrations of probes such that few probe molecules are present inside the cavity. The fact that there is a bias voltage across the two-pore cavity space, implies that any probes fall off the DNA will be driven out of the space quickly, reducing the probability of re-hybridization. On the other hand, there is also the issue of false negatives at low probe concentration: if the probe density around the *cis* side of the nanopore sandwich device is too low, there may be no probe hybridization at the target site even if it has the correct Watson-Crick pairing. It is an interesting subject for future studies to find the optimal concentration of probes on the *cis* side of the device such that there are enough probe binding events for statistics while the rate of having false positives (the incorrect probe binding) is low.

It is interesting to compare our proposed method with the Sanger approach at the fundamental level. In the original Sanger method, the base termination is accomplished by DNA polymerase with 3' to 5' exonuclease proofreading activities and 3' ddNTP termination. The sequencing, *i.e.*, base-by-base calling, is accomplished after the fragments having been separated by gel electrophoresis. In our method, the base-calling is accomplished by proofreading a correct hybridization probe on a target DNA. The sequencing is done by alignments of all correct probes. In Sanger's method, one can improve precision by using high fidelity polymerases or by reducing the dNTP concentration (to slow down the DNA synthesis process). In our method, the proofreading

mechanism can be tuned to be as precise as natural polymerases with exonuclease activities, by simply slowing down the DNA movement speed, thus extending the waiting time in the exponential error reduction factor in Eq. (3).

Among the existing methods of DNA sequencing, the Sanger method is most accurate due to the fact that the 3' termination at the incorporated ddNTP is effectively 100%. The main sources of sequencing error are from the intrinsic errors of the DNA polymerase and the strand separation resolution of gel electrophoresis. Since typically high fidelity DNA polymerases are used in the Sanger method, they possess the intrinsic 3' to 5' exonuclease proofreading activities, the error rate due to polymerase can be very low, *e.g.*, $\sim 1\times10^{-6}$ per base for New England Biolab polymerase pol Q5 (thermodynamics of Watson-Crick pairing gives 10^{-3}). The sequencing error due to weak gel separation signals can be improved by making more copies of the synthesized strands, the error can be as low as $1/\sqrt{N}$ where N is the copy number of DNA fragments in each band. Thus in principle, the Sanger method can be as accurate as 10^{-6}. (In practice, the cost consideration of creating large numbers of copies of the DNA fragments may have led users to cheat, most errors in the sequencing data may not be from polymerases.) Thus the DNA sequences obtained using the Sanger method are the benchmarks in all DNA sequence data.

In most of the next-generation sequencing (NGS) platforms that use sequencing-by-synthesis (SBS) method, Sanger's idea of 3' termination is still the base-calling mechanism. In Solexa-Illuminas 3' reversible terminators approach, as well as in Roche's pyrosequencing, polymerases with reduced or suppressed 3' to 5' exonuclease proofreading capabilities are used to avoid exonuclease chewing back the incorporated bases. Thus the intrinsic errors of NGS-SBS platforms are significantly higher than that in Sanger's original method. It is a significant issue that many random mutations in DNA sequencing data can be traced to the NGS SBS methods used. One is forced to develop elaborate algorithms to deal with such errors. It is not clear if such endeavours will bear fruit in reliably removing such errors. Thus in the development of nanopore DNA sequencing, it is generally hoped [12] that a more accurate method than NGS SBS will be developed, and at the same time cost reduced. Here it is worthwhile to discuss the sources of errors that can be expected from nanopore DNA sequencing.

In our proposed approach here, we expect the fundamental error from the base calling mechanism in Fig. 3 to ravel that of Sanger sequencing with high fidelity polymerases, 10^{-6}. The biggest uncertainty is expected from the positional errors as we pull the DNA through the pore using piezo-controlled nano positioning stages. Since the anchoring mechanism cannot be too close to the nanopore chip, the DNA, being an entropic string, will exhibit positional fluctuations. Nature solved the diffusion problem in evolving DNA polymerases into a shape that effectively grabs onto the DNA. One may have to coat the nanopore inner surface with molecular brushes such that the diffusive motion of the DNA is arrested.

It should be pointed out that by simply re-sequencing the same DNA many times does not make up for the loss of proofreading. The kinetic proofreading mechanism gives an exponential factor $e^{-k^i_{-1}t_w}$ in error correction (as shown in Eq. (3)). In contrast, re-sequencing can at best give a $1/\sqrt{N}$ correction, N being times of re-sequencing.

AUTHOR CONTRIBUTIONS

The author is the original contributor of the concepts discussed in the article.

FUNDING

The patented device concept [23] described in this paper was developed while the author was spending a sabbatical year as consulting professor (visitor) at Southeast University, Nanjing, China, during which he received a partial sabbatical salary from Brown University and a stipend (living-expense) from Southeast University.

ACKNOWLEDGMENTS

The author is grateful to Professor Terence Hwa for an early helpful conversation, and to Wang Miao for carrying out the COMSOL calculation in Fig. 2, and to Alexandros Pertsinidis and Zhishan Yuan for useful discussions, and to Professors Yunfei Chen, Zuhong Lu, and Jingjie Sha for assistance, to Southeast University (Nanjing, China) for hospitality and financial support for his sabbatical year (2016–2017) in Nanjing during which the bulk of this work was done.

COMPLIANCE WITH ETHICS GUIDELINES

The author is a co-founder and consultant of the Nanjing Rhode Island Nanotech Ltd. (Nanjing, China) and NJRI Suzhou Ltd. (Suzhou, China) which acquired the patent [23] discussed in this article.

This article is a review article and does not contain any studies with human or animal subjects performed by the author.

REFERENCES

1. Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2004), Molecular Biology of the Cell (2nd Ed.), pp.176. Garland Science Publishing
2. Shusterman, R., Alon, S., Gavrinyov, T. and Krichevsky, O. (2004) Monomer dynamics in double- and single-stranded DNA polymers. Phys. Rev. Lett., 92, 048303
3. Chi, Q., Wang, G. and Jiang, J. (2013) The persistence length and length per base of single-stranded DNA obtained from fluores-

cence correlation spectroscopy measurements using mean field theory. Physica A, 392, 1072–1079
4. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, 74, 5463–5467
5. Sanger, F, Coulson, A.R. (1978) The use of thin acrylamide gels for DNA sequencing. FEBS Lett., 87, 107–110
6. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456, 53–59
7. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437, 376–380
8. Bowers, J., Mitchell, J., Beer, E., Buzby, P. R., Causey, M., Efcavitch, J. W., Jarosz, M., Krzymanska-Olejnik, E., Kung, L., Lipson, D., *et al.* (2009) Virtual terminator nucleotides for next-generation DNA sequencing. Nat. Methods, 6, 593–595
9. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. Science, 323, 133–138
10. Goldfeder, R. L., Priest, J. R., Zook, J. M., Grove, M. E., Waggott, D., Wheeler, M. T., Salit, M. and Ashley, E. A. (2016) Medical implications of technical accuracy in genome sequencing. Genome Med., 8, 24
11. Kasianowicz, J. J., Brandin, E., Branton, D. and Deamer, D. W. (1996) Characterization of individual polynucleotide molecules using a membrane channel. Proc. Natl. Acad. Sci. USA, 93, 13770–13773
12. Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., *et al.* (2008) The potential and challenges of nanopore sequencing. Nat. Biotechnol., 26, 1146–1153
13. Lubensky, D. K. and Nelson, D. R. (1999) Driven polymer translocation through a narrow pore. Biophys. J., 77, 1824–1838
14. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. and Branton, D. (2000) Rapid nanopore discrimination between single polynucleotide molecules. Proc. Natl. Acad. Sci. USA, 97, 1079–1084
15. Li, J. and Talaga, D. S. (2010) The distribution of DNA translocation times in solid-state nanopores. J. Phys. Condens. Matter, 22, 454129
16. Ling, X.S., Bready, B.; Pertsinidis, A. (2006) Hybridization assisted nanopore sequencing. WO2007041621A3, International Application No.: PCT/US2006/038748
17. Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., 98, 503–517
18. Drmanac, R., Drmanac, S., Chui, G., Diaz, R., Hou, A., Jin, H., Jin, P., Kwon, S., Lacy, S., Moeur, B., *et al.* (2002) Sequencing by hybridization (SBH): advantages, achievements, and opportunities. Adv. Biochem. Eng. Biotechnol., 77, 75–101
19. Balagurusamy, V. S., Weinger, P. and Ling, X. S. (2010) Detection of DNA hybridizations using solid-state nanopores. Nanotechnology, 21, 335102
20. Ling, X. S. (2013). Methods of sequencing nucleic acids using nanopores and active kinetic proofreading. WO/2013/119784, International Application No.: PCT/US2013/025106
21. Hopfield, J. J. (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc. Natl. Acad. Sci. USA, 71, 4135–4139
22. Ninio, J. (1975) Kinetic amplification of enzyme discrimination. Biochimie, 57, 587–595
23. Ling, X. S. and Yuan, Z. (2017) A method for fabricating bi-layer nanopore structures for sequencing DNA. China National Bureau on Intellectual Property. Patent number ZL 2016 1 0256247.9
24. Kim, D.-K., Kwon, Y.-S., Takamura, Y. and Tamiya, E. (2006) Detection of DNA hybridization properties using thermodynamic method. Jpn. J. Appl. Phys., 45, 509–512
25. Pertsinidis, A. (2018) Private Communications
26. Gunnarsson, A., Jönsson, P., Zhdanov, V. P. and Höök, F. (2009) Kinetic and thermodynamic characterization of single-mismatch discrimination using single-molecule imaging. Nucleic Acids Res., 37, e99
27. Hynes, J. T. (1985) The theory of chemical reactions in solutions. In: Theory of Chemical Reaction Dynamics, Baer, M., (ed.), Vol. IV, pp.171235. Boca Raton: CRC
28. Sorgenfrei, S., Chiu, C. Y., Gonzalez, R. L. Jr, Yu, Y.-J., Kim, P., Nuckolls, C. and Shepard, K. L. (2011) Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. Nat. Nanotechnol., 6, 126–132
29. Schrödinger, E. (1915) Zur Theorie der Fall-und Steigversuche an Teilchen mit Brownscher Bewegung. Phys. Z., 16, 289–295
30. Ling, D. Y. and Ling, X. S. (2013) On the distribution of DNA translocation times in solid-state nanopores: an analysis using Schrödinger's first-passage-time theory. J. Phys. Condens. Matter, 25, 375102
31. Robertson, R. M., Laib, S. and Smith, D. E. (2006) Diffusion of isolated DNA molecules: dependence on length and topology. Proc. Natl. Acad. Sci. USA, 103, 7310–7314
32. Smith, S. B., Cui, Y. and Bustamante, C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science, 271, 795–799
33. Peng, H. and Ling, X. S. (2009) Reverse DNA translocation through a solid-state nanopore by magnetic tweezers. Nanotechnology, 20, 185101

