Analysis Method of the Ion Current–time Waveform Obtained from Low Aspect Ratio Solid-state Nanopores

Masateru TANIGUCHI

* The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

† To whom correspondence should be addressed.

E-mail: taniguti@sanken.osaka-u.ac.jp
Abstract

Low aspect ratio nanopores are expected to be applied to the detection of viruses and bacteria because of their high spatial resolution. Multiphysics simulations have revealed that the ion current–time waveform obtained from low aspect ratio nanopores contains information on not only the volume of viruses and bacteria but also the structure, surface charge, and flow dynamics. The analysis using machine learning extracts information about these analytes from the ion current–time waveform. The combination of low aspect ratio nanopores, multiphysics simulation, and machine learning has made it possible to distinguish different types of viruses and bacteria with high accuracy.

**Keywords:** Low aspect nanopores, Solid-state nanopores, Virus, Bacteria, Multiphysics, Machine learning.
1. **Introduction**

The control and prevention of infectious diseases caused by viruses and bacteria is a common issue worldwide. Even with the rapid advancement of science and technology, it remains difficult to control and prevent pandemics due to viruses such as influenza\(^1\text{–}^5\) and Zika\(^6\text{–}^{10}\). Diagnosing infectious diseases quickly, with high accuracy and at low cost, is paramount in the effort to control infectious diseases. In order to prevent the spread of infections, early diagnosis in the state where symptoms are not present, immediately after infection, is most effective. Currently, the RT-PCR method\(^11\text{–}^{13}\) that examines the genomes of viruses and bacteria and the ELISA technique\(^14\text{–}^{16}\) that utilizes antigen–antibody reactions are widely used for diagnosis after the symptoms of infection become evident. These methods require optical detection with sophisticated microscopy by dye labeling used as a monitor. Considering that smartphones are common and widely distributed throughout the world, it makes sense to explore electrical detection methods that can be implemented with smartphones and are easy for users. A diagnostic method that can be connected to a smartphone is one of the Internet of Things.\(^17\text{,}^{18}\) By analyzing the measurement data sent from a smartphone with a data server, it is possible to monitor infectious diseases around the world. This monitoring will allow us to find out the source of the infection and elucidate how the viruses and bacteria spread. On the other hand, in order to realize early diagnosis, it is necessary to have a low-cost, highly accurate, rapid method of detection that can function when the viral or bacterial load in the body is low. A promising method that would satisfy both rapid and early diagnosis of infectious diseases as well as be used by people all over the world is electrical detection of viruses and bacteria at a low concentration with high speed, high accuracy, and low cost. One of sensing platforms for such a method employs a solid-state nanopore\(^19\text{–}^{24}\).

Solid-state nanoporous materials have through-holes with a diameter of several \(\mu\)m to several nm in the solid membrane (Fig. 1a). Two electrodes are placed on each side of the membrane. The
membrane and nanopore are filled with an aqueous solution containing ions such as potassium and chloride. When a voltage is applied between the electrodes, ions pass through the nanopore, causing an ion current to flow between the electrodes (Fig. 1a). Once viruses and bacteria with negative surface charges are introduced into the solution, these analytes migrate through the nanopore to the positive electrode side by electrophoresis. When a virus or bacterium passes through the nanopore, the ion current is reduced because the flow of ions is blocked. The ion current returns to its original value after the virus or bacterium passes through the nanopore.

SiN is a representative material for membranes that make solid-state nanopores. The film thickness of many SiNs produced on a silicon substrate is around 50 nm. Solid-state nanopores have been created to develop next-generation DNA sequencers that determine the base sequence of DNA. Much of the research toward DNA sequencers has resulted in the need for thin membranes in order to achieve single-base molecule resolution. The thin membrane was thought to be able to detect substances passing through the nanopore with higher spatial resolution. Solid-state nanopores using single-layer graphene and single-layer MoS$_2$ have been developed as ideal membranes. Although single-base molecule resolution has not been obtained yet, identification of DNA oligomers composed of several identical base molecules has been realized. In the course of this development, solid-state nanopores have been classified into two types. When the thickness and diameter of a solid-state nanopore are $L$ and $D$, respectively, $L > D$ is classified as a high aspect ratio nanopore and $L < D$ is classified as a low aspect ratio nanopore (Fig. 1b). Coulter counters, from which the principles of solid-state nanopores originated, are classified as high aspect ratio nanopores. Graphene and MoS$_2$ are classified as low aspect ratio nanopores.

The ion current–time waveform generated by the translocation of a nanoscale object obtained using solid-state nanopores is often analyzed with three values: peak current value ($I_p$), current duration ($t_d$), and signal frequency ($f$) (Fig. 2a). When an ion current–time waveform is similar to
the schematic current trace in Fig. 2b, the analyte likely collides with the nanopore first before it translocates. In practice, it is also necessary that \( t_i \) not be too low in order to avoid clogging. In order to identify two or more types of analytes, an \( I_p \) histogram, a \( t_d \) histogram, and an \( I_p-t_d \) two-dimensional map are used.\(^{21, 24, 27, 37}\) When the histograms or the heat maps of two analytic types display two peaks without overlapping or two clearly separated clusters, the analytic types can be distinguished (Figs. 2c and 2d). On the other hands, when such peaks or clusters overlap, they cannot completely distinguish the analytic types. The analysis based on this histogram assumes that the change in the ion current correlates with the volume of the analytic. In particular, the histogram of \( I_p \) used in most analyses is a typical analysis method used with high aspect ratio nanopores to calculate the volume of a specimen. However, the flow dynamics of the analytic in high and low aspect ratio nanopores differ as described by different models. The major difference is the enhanced effect of electroosmotic flow on high aspect ratio nanopores that occurs at the interface between the pore wall and the liquid.\(^{38-40}\)

The ion current–time waveform obtained with a low aspect ratio nanopore provides more information than that of a high aspect ratio pore, such as the volume, surface structure, surface charge, and flow dynamics of the analytic.\(^{41-45}\) The analysis of these can complementary be made by the development of multiphysics simulation techniques with finite element methods.\(^{46, 47}\) In order to extract these four types of information from the waveform, it is necessary to analyze the vector quantities representing the waveform, not the scalar quantities of \( I_p \) or \( t_d \). One useful method of analysis is machine learning. Even an analytic that is difficult to identify by analysis using the histogram of \( I_p \) or \( t_d \) can be identified with high accuracy by using machine learning. Because the conventional analysis based on \( I_p \) and \( t_d \) utilizes a histogram that consists of statistical data of \( I_p \) and \( t_d \), it cannot be applied for a single ion current–time waveform. However, machine learning enables the analysis of individual ion current–time waveforms. For instance,
when analyzing the waveforms produced by two types of viruses with the machine learning algorithm, it is possible to identify the type for each waveform with high accuracy.\textsuperscript{42, 45} There are not many examples of analysis using machine learning, but those that do exist have demonstrated capability for high-precision discrimination between two types of viruses or two types of bacteria having almost the same volume.\textsuperscript{42-45}

In this review, in order to compare high aspect ratio nanopores and low aspect ratio nanopores, we focus on viruses and bacteria that have measurement examples in both classes of solid-state nanopores. First, we outline the flow dynamics that form the basis for analyzing the ion current–time waveform obtained from both types of solid-state nanopore. Graphene\textsuperscript{26, 28-31} and MoS\textsubscript{2}\textsuperscript{32-34} are typical low aspect ratio nanopores and are expected to be applied to DNA sequencing. The flow dynamics of single-stranded DNA with a diameter of 1 nm can be described by molecular dynamics.\textsuperscript{48} At present, DNA flow dynamics simulation on an experimental time scale is difficult, so this review does not deal with DNA. This is because the purpose herein is to consider the reasons why high accuracy can be obtained by analysis by machine learning based on multiphysics simulation.\textsuperscript{42-45}

In order to use solid-state nanopores practically, a large number of waveforms can be obtained in a short time, i.e., a high signal frequency is required. Knowing what determines the signal frequency is essential to controlling the signal frequency. The signal frequency increases as the probability that the analyte is captured by the solid-state nanopore increases.\textsuperscript{38} We will outline experimentally evaluable capture probabilities. In addition, the phenomenon in which the analyte clogs the solid-state nanopore has been observed in an actual experiment. The cause of clogging is rarely investigated, but in practice, it is necessary to prevent clogging. We will, therefore, outline the methodology for how to prevent clogging.\textsuperscript{49-53}

Finally, an example of analyzing the ion current–time waveform using machine learning will
be introduced. The analysis using machine learning indicates that viruses and bacteria, unidentified by conventional analysis using a histogram, can be identified with high accuracy via machine learning.\cite{42-45} It also shows that the physical interpretation of features used in machine learning is possible by combining virus and bacteria multiphysics simulation and machine learning. There are many excellent reviews of solid-state nanopores.\cite{21, 24, 27, 37} For an overview of solid-state nanopores, see Dekker's review.\cite{21} There are also many reviews focused on solid-state nanopore fabrication methods\cite{21, 24, 37} and reviews focused on the fluid dynamics of specimens\cite{40, 54, 55}. In this review, we will focus on an overview of the flow dynamics of high and low aspect ratio nanopores and the machine learning analysis of ion current–time waveform data obtained from the low aspect ratio nanopores.

2. **Ion Current–time Waveform and Flow Dynamics of an Analyte**

This section will provide an overview of the basics for analyzing ion current–time waveforms obtained using solid-state nanopores. In particular, the flow dynamics of the analyte differs between the high aspect ratio nanopore and the low aspect ratio nanopore. First, we will introduce ion current–time waveforms obtained by measuring viruses\cite{42, 45, 56-58}, liposomes\cite{59}, and bacteria\cite{43, 44} using solid-state nanopores. After reviewing the flow dynamics, we will outline the frequency with which analytes are captured by solid-state nanopores.\cite{38, 40} Finally, we will review the clogging phenomenon in solid-state nanopores, which is an important practical issue.\cite{49-53}

2. **1. Ion Current–time Waveforms of Single Viruses and Bacteria**

Genomic analysis using RT-PCR\cite{11-13} or DNA sequencing\cite{60-63} is the mainstream approach for testing viruses and bacteria. The volume and structure of these microorganisms are diverse, even if the organisms are of the same type. Solid-state nanopores are used to detect the volume of the
analyte. Thus, the utility of solid-state nanopores in the study of viral and bacterial identification may be somewhat limited. Here, we introduce some identification results of viruses and bacteria using solid-state nanopores.

Negatively charged filament viruses were measured using solid-state nanopores with a SiN thickness and nanopore diameter of 20 nm and 26 nm, respectively.\textsuperscript{57} The nearly linear filament virus was 6.6 nm in diameter and 880 nm in length (Fig. 3a). In the ion current–time waveform, passing events with an average $I_p$ of 0.12 nA and an average $t_d$ of 1.2 ms were observed. In addition, waveforms corresponding to the event of the virus hitting the nanopore entrance were observed, showing a smaller average $I_p$ and average $t_d$. These two events were distinguished by a two-dimensional map of $I_p$ and $t_d$ when the voltage and KCl concentration were below 100 mV and 300 mM, respectively. However, above these thresholds, these two events could not be distinguished.

A negatively charged tobacco mosaic virus with a diameter of 18 nm and a length of 300 nm was measured (Fig. 3b).\textsuperscript{58} Nanopores with diameters of 25, 30, and 50 nm were used on a 30-nm-thick SiN membrane. In this experiment, four characteristic ion current–time waveforms were obtained. A typical waveform of the virus passing through the nanopore was obtained (Type A in Fig. 3b). The waveform with small $I_p$ and $t_d$ (Type D in Fig. 3b; Fig. 2b) is the signal indicating that the virus collided with the entrance of the nanopore. Coarse-grained Langevin dynamics simulations\textsuperscript{64-66} revealed that the Type B signal indicates that the virus fluctuates at the nanopore entrance and passes through the nanopore. Type C represents the dynamics in which the virus fluctuates thermally while passing through the nanopore. Thus, one ion current–time waveform represents the flow dynamics of one virus. For the measurements of the filament virus and tobacco mosaic virus, a high aspect ratio nanopore was used, i.e., $L/D > 1$.

Spherical negatively charged influenza viruses were measured using low aspect ratio nanopores
Nanopores 300 nm in diameter were fabricated on a 50 nm-thick SiN membrane. The diameters of influenza A (H1N1), influenza A subtype (H3N2), and influenza B range from 80 to 150 nm, in that order. The differences between these three viruses lie in the proteins that make up the virus; these differences result in different distributions of surface charges. However, the resulting ion current–time waveforms were similar for the three viruses. Therefore, the analysis using the $I_p$ and $t_d$ histograms failed to distinguish between the three viruses. As will be described later, when machine learning was used to analyze the ion current–time waveform, the three viruses were identified with high accuracy.

Negatively charged liposomes with a diameter of 80 nm were measured with $L/D = 0.8$ solid-state nanopores with a diameter of 250 nm and a thickness of 200 nm (Fig. 3d). For hard polystyrene particles, $I_p$ increases with increasing bias voltage. However, in liposomes, $I_p$ decreased when the bias voltage was increased from 100 to 600 mV. This is because liposomes are soft nanoparticles with a surface charge and are deformed by a strong electric field. Multiphysics simulations estimated that the electric field at the center of the nanopore was 14 kV cm$^{-1}$ when the bias voltage was 600 mV. Such deformation of particles by a strong electric field has been reported in giant vesicles. The research group that conducted this experiment reported that human immunodeficiency virus with a diameter of 100 nm was also deformed by a strong electric field. As described above, the softness (or stiffness) of the analyte can be examined by changing the bias voltage.

The typical bacteria, *Escherichia coli* and *Bacillus subtilis*, were also measured. Solid-state nanopores with a diameter of 3 μm were used on a 40-nm-thick SiN membrane. The diameter and length of *E. coli* and *B. subtilis* are approximately 1 and 3 μm, respectively (Fig. 3e). The zeta potentials of *E. coli* and *B. subtilis* were $-37$ and $-38$ mV. There was no significant difference in volume or surface charge, so no significant difference in the ion current–time waveform was
observed. For this reason, these two bacteria could not be distinguished with high accuracy in the histograms of \( I_p \) and \( t_d \). However, since the curvatures of *E. coli* and *B. subtilis* are different, a difference between the entry dynamics into the nanopore and the passing dynamics would be expected. An analysis of the ion current–time waveform by machine learning produced high-precision discrimination between these two bacteria. In this experiment, streptococci with a rosary shape were also measured. The measurement results of streptococci and the results of waveform simulation based on multiphysics revealed that the waveform contains information about the bacterial structure.

There are numerous types of viruses and bacteria, but only a few have been measured with solid-state nanopores. In particular, discrimination between two viruses or bacteria has not been performed, except with low aspect ratio nanopores. In the next section, we will focus on the flow dynamics of viruses and bacteria in high aspect ratio nanopores.

2. Flow Dynamics of High Aspect Ratio Nanopores

Analyses of solid-state nanopores are typically conducted using \( I_p \) and \( t_d \) histograms. The following equation is used to analyze the ion current flowing through the solid-state nanopores.\(^{72}\)

\[
I = \frac{V}{R_{\text{acc}} + R_{\text{pore}}} \tag{1}
\]

where \( R_{\text{acc}} \) and \( R_{\text{pore}} \) are the access resistance and the resistance of the solid-state nanopore, respectively. When the electrical conductivity of the electrolyte is \( \kappa \),

\[
R = 2 \times \frac{1}{2\kappa D} = \frac{1}{\kappa D} \tag{2}
\]

\[
R = \frac{4L}{\pi\kappa D^2} \tag{3}
\]

These equations describe the base current of a high aspect ratio nanopore where the electric field
strength in the nanopore is constant (Fig. 4). However, in the low aspect ratio nanopore, since the electric field strength in the nanopore is not constant, the experimental value deviates from the estimated value of the base current. In particular, in the low aspect ratio nanopore, the electric field is concentrated on the nanopore edge, and the electric field strength at the center of the nanopore is smaller than that in the high aspect ratio nanopore. In addition, this equation does not take into account the electroosmotic flow that occurs at the interface between the nanopore and the liquid. Thus, the widely used formulas (1) to (3) are suitable to estimate of the physical quantities for the high aspect ratio nanopores.

High aspect ratio nanopores are available to researchers with a product called qNano. This measurement system dynamically adjusts the diameter of the through-hole made in the polymer. With high aspect ratio nanopores, the particle diameter and zeta potential of the analyte are obtained by utilizing the fact that \( I_p \) and \( I_d \) have information on the volume, particle velocity, and surface charge of the analyte.

2. **3. Flow Dynamics of Low Aspect Ratio Nanopores**

In both the high aspect ratio nanopore and the low aspect ratio nanopore, the flow dynamics of the analyte is governed by three forces: electrophoresis, electroosmotic flow, and fluid resistance (Fig. 5). In addition, flow dynamics depends on the gradients of pressure, temperature, and salt concentration. For this discussion, the influence of these gradients and the intermolecular interaction between the nanopore surface and the analyte will not be considered. At this time, the flow dynamics of the negatively charged analyte will be classified into the three separate cases shown in Fig. 5.

The negatively charged analyte receives a force in the direction of the positive electrode by electrophoresis. For materials with negatively charged surfaces, such as SiN, cations accumulate
at the solid–liquid interface. The cations flow in the opposite direction to electrophoresis; this is the electroosmotic flow. As a result, the analyte receives a force in the opposite direction to the electrophoresis. When electrophoresis > electroosmotic flow + fluid resistance, the analyte flows in the $Z$ direction (Fig. 5a). In this case, it experiences fluid resistance in the $-Z$ direction. That is, the nanopore passage speed of the analyte is reduced by the electroosmotic flow. When the surface charge density of the analyte is low, or the surface charge density of the nanopore is high, electrophoresis < electroosmotic flow + fluid resistance may occur (Fig. 5b). In this case, the analyte cannot enter the nanopore, and an ion current–time waveform cannot be obtained. On the other hand, when the surface charge is positive, such as with Al$_2$O$_3$ solid-state nanopores, anions accumulate at the nanopore–liquid interface. As a result, the electroosmotic flow in the $Z$ direction increases the nanopore translocation speed of the analyte (Fig. 5c). Thus, control of the polarity and density of the charge on the surface of the nanopore grants control of the translocation speed of the analyte in the nanopore.

Multiphysics simulation simulates the flow dynamics of analytes with diameters of several hundred nanometers or more, that is, the translocation of viruses and bacteria through solid-state nanopores. Typically, a COMSOL program package is used. The flow dynamics of viruses and bacteria in the nanopores is described by a composite of three equations: the Poisson equation describing electrostatics in the nanopore, the Navier–Stokes equation describing fluid flow, and the Nernst–Planck equation describing ion transport.

In the nanopore, the ion flow of the solution contributes to the ion current as well as the electrostatic charge at the nanopore–liquid interface. When the ion concentration in the equilibrium state at the interface, the Poisson equation is described by the Boltzmann distribution.
where $n_i^0$, $e$, and $V$ represent the far-field concentration of the $i$-th species, the elementary charge, and the electrical potential, respectively; $z_i$, $k$, and $T$ represent the corresponding valency, Boltzmann constant, and temperature, respectively. $\varepsilon_f$ is permittivity of the fluid. For NaCl, $z_i = 1$ and $z_i = -1$ correspond to Na$^+$ and Cl$^-$, respectively.

The motion of ions in the solution is described by the following Nernst–Planck equation, wherein $u_i$, $D_i$, and $u$ are the electrophoretic mobility, diffusion coefficient, and flow velocity of the $i$-th ion.\(^{40,46,47,54}\)

\[
\nabla J_i = \nabla \cdot \left( -\text{sgn}(z_i) \mu_i n_i \nabla V - D_i \nabla n_i + n_i u \right) \quad (6)
\]

The first term represents electrical migration, the second term represents density diffusion, and the third term represents convection. $\text{sgn}(z_i)$ is a sign function, $\text{sgn}(z_i) = 1$ if $z_i > 1$, $\text{sgn}(z_i) = 0$ if $z_i = 0$, and $\text{sgn}(z_i) = -1$ if $z_i < 1$. In the nanopore, $J_i = 0$ in the wall direction of the nanopore.

Fluid flow is described by the following Navier–Stokes equation, wherein $p$, $\eta$, and $f$ are the pressure, fluid viscosity, and electrical body force exerted on the fluid due to the local charges by the unbalanced ions, respectively.\(^{46,47,40,54}\)

\[
\rho \left( \frac{\partial u}{\partial t} + u \cdot \nabla u \right) = -\nabla p + \eta \nabla^2 u + f \quad (7)
\]

\[
\nabla \cdot u = 0 \quad (8)
\]

Moreover, electrical body force is given by the following formula.

\[
f = \left( -\nabla V \right) e \sum_i z_i n_i \quad (9)
\]

In a steady state, $\partial u/\partial t = 0$ and the flow velocity in the axial direction of the nanopore does not change, so $\partial u/\partial z = 0$. Also, $u_y = u_z = 0$. The above equations (4)–(9) are combined and solved to
find $n_i$, $E_z$, and $u_z$. The ion current in the solid-state nanopore can be calculated by the following equation when density diffusion is ignored.\(^{40,46,47,54}\)

$$J_z = e \sum_i z_i (-\text{sgn}(z_i) \mu_i n_i E_z + n_i \mu_z) \quad (10)$$

Since this simulation combines a plurality of physical models, it is called multiphysics simulation. In many cases, COMSOL is used as the calculation package.\(^90\) When simulating experimental results, we make assumptions about the structure and surface charge density of the solid nanopores and analytes. The time evolution of equations (4)–(10) is calculated on a large scale. For this reason, in practice, equations (4) to (10) are calculated in a steady state in which the analyte is placed in the nanopore to obtain the ion current value.\(^{46,47}\) The analyte is moved to obtain the ion current–position relationship. The simplest and the lowest time cost is the method of obtaining the ion current by changing the position and matching the ion current–time waveform with the waveform obtained in the experiment. If this method is not used, then a position–velocity plot is created from the electric field distribution and flow field. We integrate the reciprocal of the velocity at the position with respect to time. Finally, we plot the ion current–time waveform. This method is used to reliably obtain an ion current–time waveform.

Multiphysics simulation has been used to reproduce the ion current–time waveform of low aspect ratio nanopores.\(^{41-45}\) The results of streptococci with a connected form of cocci are illustrated in Fig. 6.\(^43\) The size of one coccus is 1.1 μm in length and 0.8 μm in diameter (Fig. 6a). Bacteria consisting of linked cocci were counted using low aspect ratio nanopores with $L/D = 0.04$. The ion current–time waveform shows a shape with two peaks (Fig. 6b). One streptococcus produced one peak, and streptococci with a shape in which $N (N = 2, 3, 4)$ cocci were linked showed $N$ peaks (Fig. 6c). Multiphysics simulation was used successfully to reproduce the experimental results (Fig. 6d). In addition, by making assumptions about the structure of the specimen and the surface charge, fitting a simulation waveform to the ion current–time waveform
obtained in the experiment enables structural analysis of a single streptococcus from the ion current waveform of the experiment. This shows that the tomography of a single virus or bacterium using ion current–time waveforms is possible.\textsuperscript{43} The simulations also showed that the ion current–time waveform depends on the direction in which the analyte enters the nanopore. This multiphysics simulation shows that the ion current–time waveform has information on structure, surface charge, and flow dynamics, as well as the volume of viruses and bacteria. However, it is difficult to extract this information from the $I_p$ and $t_d$ histograms. The problem is how to extract this information from the ion current–time waveform. One technique is analysis by machine learning. Before we outline the analysis by machine learning, let us discuss the capture frequency and clogging of solid-state nanopores.

### 2. 4. Capture Frequency of Analyte

The capture of viruses and bacteria in solid-state nanopores occurs approximately via the following two processes.\textsuperscript{91, 92} At first, before entering the electrical capture sphere around the nanopore entrance, analytes move toward the pore entrance in a density diffusive manner. Upon getting into the capture sphere, the analyte becomes majorly driven by an electrophoretic force. In the above physical picturing of the capture process, the radius of the electrical capture sphere $r^*$ is defined as the distance at which the analyte diffusion speed is equal to the electrical migration speed.\textsuperscript{38} $r^*$ gives a hemisphere with radius $r^*$ from the mouth of the nanopore. When the zeta potentials of the nanopore and the specimen are $\zeta_{\text{pore}}$ and $\zeta_{\text{ana}}$, respectively, the capture sphere is given by the following equation.\textsuperscript{38}

\begin{equation}
 r^* = \frac{\varepsilon D^2}{8\eta \delta} \left| \zeta_{\text{pore}} - |\zeta_{\text{ana}}| \right| \tag{11}
\end{equation}

where $\eta$, $\delta$, and $\varepsilon$ denote the viscosity of water, the diffusion coefficient of the specimen, and the
product of the permittivity of free-space and the dielectric constant of water, respectively. This analytical formula does not match the experimental results.\textsuperscript{38} This is thought to be due to poor estimates of approximation and zeta potential. However, in common with the capture sphere obtained by other derivation methods, the capture sphere is proportional to $D^2 / (4L + \pi D)$. In other words, the low aspect ratio nanopore has a large capture sphere and can capture an analyte far from the nanopore. Using Equation 10, the capture frequency is given by\textsuperscript{38}

$$ f = 2\pi N \delta \tilde{r}^* (12) $$

where $N$ indicates the number of analytes per unit volume. $f$ should be proportional to $1/t_f$ in Figure 1, but does not agree with the experimental results. In the future, it will be useful to estimate $\tilde{r}^*$ using multiphysics simulations.

5. Analyte Clogging

The ion current–time waveform and trapping frequency have been energetically studied to establish the basic principles of solid-state nanopores.\textsuperscript{40, 46, 47, 54} When using solid-state nanopores as a practical device, it is necessary that the nanopores not be clogged. Ideally, the nanopores would remain unclogged until enough data are available for analysis.

The chemical modification of solid-state nanopores using viruses and bacteria as specimens has been applied in the detection of influenza viruses A and B.\textsuperscript{45} The surface of a SiN low aspect ratio nanopore having a diameter of 300 nm and a thickness of 95 nm was modified with a peptide that recognizes hemagglutinin on the surface of influenza virus A. The purpose of this study was to improve the accuracy of distinguishing influenza viruses A and B by using the peptide's molecular recognition ability. Efforts toward the chemical modification of the solid-state nanopore surface for detecting viruses and bacteria have just begun, and it is expected that chemical modifications according to the types of viruses and bacteria will develop in the future.
Although this review focuses on viruses and bacteria, it is also important to look at the work that has been done in DNA and protein detection.\textsuperscript{49, 50} The main causes of clogging are 1) a substance larger than the diameter of the nanopore, 2) nonspecific adsorption of the substance near the mouth of the nanopore, and 3) nonspecific adsorption of the substance to the nanopore wall. When an amyloid-beta (A\textsubscript{\beta}) peptide was measured using a solid-state nanopore with a diameter of 33 nm and a thickness of 22 nm, the nanopore became clogged within a few minutes (Fig. 7a).\textsuperscript{49} This is a result of A\textsubscript{\beta} peptide aggregation in the nanopore or the vicinity of the mouth of the nanopore in the unmodified nanoporous material. However, when the surface of the solid-state nanopore was modified with a lipid-bilayer, the nanopore did not clog for at least 40 min (Fig. 7b). Lipid bilayers have the effect of preventing the non-specific adsorption of A\textsubscript{\beta} peptides.

DsDNA and ssDNA were measured using a SiN nanopore with a diameter of 14 nm and a thickness of 30 nm coated with Tween20.\textsuperscript{50} Tween20 is an amphiphilic nonionic surfactant molecule.\textsuperscript{93, 94} The hydrophobic part interacts with SiN, and the hydrophilic part becomes the “surface.” Thus, the Tween20 coating makes the surface uncharged and hydrophilic. As a result of preventing the non-specific adsorption of DNA to the hydrophobic SiN surface, the modified nanopore was able to resist clogging for more than 100 min. The unmodified nanopore has a complex ion current–time waveform, whereas the modified nanopore showed a Gaussian waveform.

Hydrophobic surfaces cause the nonspecific adsorption of biomolecules.\textsuperscript{95, 96} In addition, since the surfaces of many viruses and bacteria are negatively charged, they interact electrostatically with solid-state nanopores that have surface charges. Thus, charged and hydrophobic surfaces render solid-state nanopores susceptible to clogging. In contrast, uncharged and hydrophilic solid-state nanopores are much less likely to become clogged. Recall that SiN has a negative surface charge and is hydrophobic; however, solid-state nanopores modified with Tween20 have
uncharged and hydrophilic surfaces.\textsuperscript{50} Such chemical modification of the solid-state nanopore surface produces an uncharged and hydrophilic surface but also has the effect of reducing the diameter of the nanopore.\textsuperscript{49, 50} In addition, the modifying molecule may selectively interact with a biomolecular analyte. Therefore, the development of a method for controlling the properties of the nanopore surface without chemical modification would be highly desirable.

3. Analysis Method Using Machine Learning

Multiphysics simulations revealed that the ion current–time waveform has information on the volume, structure, surface charge, and dynamics of viruses and bacteria.\textsuperscript{42-45} If this information or composite information is extracted from the waveform and used, individual viruses and bacteria should be identified with a single waveform. Here, we show that individual viruses and bacteria can be identified by analyzing the ion current–time waveform using machine learning.

3. 1. Analysis Method Based on Machine Learning

The goal of analysis using machine learning is to identify which virus (bacteria) is producing a given ion current–time waveform with high accuracy.\textsuperscript{42, 45} Machine learning includes a method using supervised data and a method not using supervised data. Here, an analysis method based on machine learning using training data is introduced.

We consider the case of distinguishing between virus A and virus B. First, a classifier is learned that characterizes the ion current–time waveform of virus A alone or virus B alone (Fig. 8a). In the learning process, the features are defined by the analyst. Depending on the combination of features, there are cases in which classification is not performed (features $I_p$ and $t_d$) and those in which classification (Feature 1 and Feature 2) is performed. Generally, when there are $N$ features, a feature distribution is created in an $N$-dimensional space. The characteristic quantities often used
to describe nanopores are $I_p$ and $t_d$. The feature includes not only the measurement value itself but also a feature generated from the measurement amount, for example, the sharpness of the waveform. The feature is not only a scalar quantity but also a vector quantity, such as a current vector obtained by dividing a single waveform into $N$ waveforms in the time axis direction. Therefore, even if the histograms using $I_p$ and $t_d$ overlap, there is a possibility that virus A and virus B can be distinguished. Next, an inference model for discriminating between virus A and virus B is created by determining the feature quantity and classifier that gives the highest discrimination accuracy in the learning process. Next, a reasoning model is used to classify whether the ion current–time waveform obtained in the experiment was produced by virus A or virus B. Finally, a high-accuracy classification is made as to whether the given ion current–time waveform was produced by virus A or virus B. The accuracy is evaluated by the $F$-measure: the larger the $F$-measure, the higher the accuracy of classification.97

A specific data analysis method workflow is depicted in Fig. 8b.42-45 Waveforms (pulses) corresponding to viruses and bacteria are extracted from the ion current–time data ($I$–$t$ data) obtained with solid-state nanopores using filters. This extraction process requires low current noise. 27, 98-101 The current noise of the solid-state nanopore mainly depends on the parasitic capacitance of the solid-state nanopore. In order to increase the signal-to-noise ratio, optimization of the structure and materials of the solid-state device is essential. Next, a plurality of features is defined, and features of the extracted pulse data are extracted. In the fourth step, those features are combined. Next, undersampling is performed to match the number of data with the newer, smaller number of feature data; this is because the discrepancy in the number of data would lead to over-learning, i.e., incorrect learning. In the sixth step, the classifier is trained using the feature data set. When the number of virus features is $K_1$ and the number of classifiers is $K_2$, $K_1 \times K_2$ learning can be performed. Many types of classifiers have already been developed and published.
A free software WEKA (The Waikato Environment for Knowledge Analysis) program with many classifiers is often used. The identification accuracy is evaluated for all combinations of features and classifiers. A frequently used evaluation method is \( N \)-fold cross-validation (Fig. 9).

As an example, a 5-fold cross-validation method is as follows. When a dataset of the features of virus A and virus B is given, the data are divided into five sets. Four sets of data are used for learning, and one is used for verification. With this operation, the \( F \)-measure is obtained. The same operation is performed four times, and the average of the five \( F \)-measure values is obtained. The accuracy of identification is evaluated by the \( F \)-measure value. The \( F \)-measure is defined as follows:

\[
F\text{-measure} = 2 \frac{precision \times recall}{precision + recall} \tag{13}
\]

Assume that the number of pulses obtained in virus A and virus B measurements is shown in Figure 10. It shows that there are five data incorrectly identifying virus A as virus B. These five data are false negatives for virus A. Ten data are false positives for virus A. This matrix is called a confusion matrix. Precision and recall are obtained from the confusion matrix. Precision = \( \frac{150}{150 + 10} = 0.94 \) and recall = \( \frac{150}{150 + 5} = 0.97 \) for virus A.

When the \( I_p \) and \( t_d \) histograms obtained from solid-state nanopores overlap, the two viruses cannot be distinguished with high accuracy. Analysis by machine learning is similar to searching for non-overlapping histograms using features different from \( I_p \) and \( t_d \). However, even if high accuracy is obtained, it is difficult to elucidate the obtained cause physically or chemically. In other words, at present, it is difficult to physically or chemically interpret features that give high accuracy. One way to approach this is via multiphysics simulation. In the next section, examples of viruses and bacteria that have actually been analyzed using machine learning will be introduced.

### 3.2 Analysis of Viruses and Bacteria
Influenza A (H1N1), influenza B, and influenza A subtype (H3N2) were measured using SiN nanopores with \( L/D = 50 \text{ nm}/300 \text{ nm} \). The diameters of the three viruses range from 80 to 120 nm. The difference between the three viruses lies in the proteins that make up the virus. When a bias voltage of 0.1 V was applied, spiked ion current–time waveforms were obtained with the three viruses. The ion current–time waveform obtained in the multiphysics simulation of influenza with a diameter of 110 nm was consistent with the experimental results. Therefore, although there was no difference in the \( I_p \) histograms of the three viruses, it is expected that differences will occur in the charge and flow dynamics. In fact, 10 features, including \( I_p \) and \( t_d \), were extracted from the waveform (Fig. 11a). \( J_{\text{curr}} \), \( t_d \), \( r \), \( \theta \), and \( J_{\text{time}} \) mainly reflect the transit time, i.e., the surface charge and flow dynamics of the virus. \( I_p \), \( S \), and \( S_r \) mainly reflect the current change, i.e., the volume of the virus. The \( \beta \) value reflects the structure of the virus. Using the WEKA workbench with many classifiers, machine learning was performed by combining 60 feature vectors and 71 classifiers (Fig. 11b). \( F \)-measure values for all combinations (4260 ways) were obtained, and it was found that Rotation Forest was the classifier that gave the highest accuracy. A 10-fold cross-validation method was used as the evaluation method. The \( F \)-measure values for distinguishing A (H3N2) from B, A (H1N1) from A subtype (H3N2), and A (H1N1) from B were 0.61, 0.68, and 0.72, respectively (Fig. 11c). These \( F \)-measure values apply when a single ion current–time waveform is examined. In considering an actual case, we will make a definitive diagnosis of influenza type based on several waveforms. In distinguishing A (H3N2) from B, high accuracy of 0.94 was obtained when 11 or more pulses were used (Fig. 11d). Using the \( P_{\text{rec}} \) value given by each feature as an index, it is estimated that there are differences in the flow dynamics among the three influenza viruses (Fig. 11e).
volume, and the feature amount $\beta$ reflecting the structure gave the smallest $P_{\text{rec}}$ value.

$E. \ coli$ and $B. \ subtilis$ were measured with low aspect ratio nanopores with $L/D = 40 \text{ nm}/3000 \text{ nm}$.43 $E. \ coli$ and $B. \ subtilis$ have similar surface potentials of $-37 \text{ mV}$ and $-38 \text{ mV}$. Moreover, although they have the same volume, the curvature of their structures is different. In the $I_p$ and $t_d$ histograms, the two bacteria could not be distinguished with high accuracy. Multiphysics simulations suggested that the ion current–time waveform rises differently when the bacterial structure is different. Therefore, in addition to $I_p$ and $t_d$, feature values, including the feature value $\theta$ reflecting the curvature of the edge of the structure, were extracted, and 60 feature vectors were generated (Fig. 11f). $F$-measures were obtained by combining 67 classifiers, including Rotation Forest104 obtained with the WEKA workbench102 and features (Fig. 11g). The evaluation was performed by the 10-fold cross-validation method. When $L = 1490 \text{ nm}$, the two bacteria were identified with high accuracy ($F$-measure = 0.90). Similar to the virus, the characteristic amount and the $F$-measure were analyzed, and a $P_{\text{rec}}$ with high $I_p$, $A$, and $A_{\text{long,i}}$ reflecting the volume and structure was obtained. In contrast, $I_m$ and $t_{ds}$, which reflect surface charge and flow dynamics, gave small $P_{\text{rec}}$ values. Therefore, it is thought that high identification accuracy was obtained by identifying the curvature of the bacterial structure.

The $F$-measure at each aspect ratio was obtained by the same analysis method by changing the aspect ratio of the solid-state nanopore (Fig. 11h).43 When $L/D = 40 \text{ nm}/3000 \text{ nm} = 0.013$, the $F$-measure was 0.90. As $L$ was increased, the $F$-measure decreased. This is thought to be because the spatial resolution for examining the structure of bacteria decreases as the nanopore thickness increases. In nanopores with a large thickness, the $F$-measure was predicted to decrease, but the $F$-measure increased at thicknesses greater than $L = 500 \text{ nm}$. Even in the case of solid nanopores with $L/D = 1500 \text{ nm}/3000 \text{ nm}$, the same $F$-measure as that of solid-state nanopores with $L/D = 0.13$ was obtained. This result was obtained for the first time by analysis based on machine
learning. Thus, analysis based on machine learning is also effective for optimizing the device structure.

4. Outlook

Low aspect ratio nanopores can be used to measure viruses and bacteria with high spatial resolution. Multiphysics simulations support this fact and show that ion current–time waveforms have information on viral and bacterial volume, structure, surface charge, and flow dynamics. Until now, the analysis of the ion current–time waveform obtained from the nanopore has been analyzed in terms of the maximum change current value ($I_p$) and the current duration ($t_d$). For this reason, it was quite difficult to distinguish between viruses and bacteria having the same volume. Machine learning, using features that distinguish the volume, structure, surface charge, and flow dynamics of viruses and bacteria from ion current–time waveforms, has made it possible to accurately identify viruses and bacteria. The combination of solid-state nanopores, multiphysics simulation, and machine learning allows the accurate identification of a virus from a single waveform. Similarly, it is possible to identify which bacteria produce a given waveform.

Machine learning makes it possible to classify labeled data with high accuracy. The solid-state nanopore systems can be used to measure a single virus or bacterium based on the label information contained in the resulting waveform. Therefore, machine learning and solid-state nanopores make a good combination, allowing the labeling of a viral or bacterial species for each waveform. Since the number of waveforms is equal to the number of viral and bacterial species, the combination of techniques allows for quantitative analysis. It is a natural way to count the number of viruses and bacteria of the same type on a unit basis. Even if the number is large, an integral number of virus species and bacteria species can be obtained. Thus, the combination of solid-state nanopores and machine learning creates digital analytical chemistry that can
distinguish between integer multiples of viral and bacterial species.

*Digital analytical chemistry* enables the analysis of viruses and bacteria that could not otherwise be identified or quantitatively analyzed. If training data for *N* viruses or *N* bacteria are acquired, *N* types of identifications can be made. Although this new analytical chemistry enables multi-item analysis, there are two major challenges. One is a technology for transporting viruses and bacteria to solid-state nanopores. Solid-state nanopores output an ion current–time waveform that can be analyzed if viruses or bacteria pass through the nanopore. Ideally, it would be desirable to develop a flow channel technology that separates and extracts only the viruses and bacteria that are to be detected and transports them to the solid-state nanopores.\(^\text{105}\) However, it is difficult to isolate, extract, and transport only the viruses and bacteria desired. Therefore, another problem is in the high-accuracy identification of viruses and bacteria in the presence of contaminants. The identification of single molecules in the contaminants (noise) has been demonstrated.\(^\text{106}\) The analytical system measures and learns on a sample that does not contain an analyte, in which case the resulting signal is labeled as noise. When the solution containing the analyte is measured, the signals obtained include both the noise and the analyte signal. Since the noise has already been learned, the sample signal can be extracted. This is a type of machine learning called the Positive Unlabeled Classifier (PUC) method.\(^\text{107}\) In the ideal usage of the PUC method, the target virus or bacteria can be identified in blood with high accuracy by training the system on blood that does not contain the analyte. If this can be achieved through the combination of machine learning and solid-state nanopores, the separation and extraction of impurities will ultimately become unnecessary.

**Acknowledgements**

This research was supported by KAKENHI Grant No. 19H00852.
References

1. J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, L. Brilliant, *Nature* 2009, 457, 1012.

2. G. J. D. Smith, D. Vijaykrishna, J. Bahl, S. J. Lycett, M. Worobey, O. G. Pybus, S. K. Ma, C. L. Cheung, J. Raghwani, S. Bhatt, J. S. M. Peiris, Y. Guan, A. Rambaut, *Nature* 2009, 459, 1122.

3. C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, T. Jombart, W. R. Hinsley, N. C. Grassly, F. Balloux, A. C. Ghani, N. M. Ferguson, A. Rambaut, O. G. Pybus, H. Lopez-Gatell, C. M. Alpuche-Aranda, I. B. Chapela, E. P. Zavala, D. M. E. Guevara, F. Checchi, E. Garcia, S. Hugonnet, C. Roth, W. R. P. A. Coll, *Science* 2009, 324, 1557.

4. G. Neumann, T. Noda, Y. Kawaoka, *Nature* 2009, 459, 931.

5. S. Jain, L. Kamimoto, A. M. Bramley, A. M. Schmitz, S. R. Benoit, J. Louie, D. E. Sugerman, J. K. Druckenmiller, K. A. Ritger, R. Chugh, S. Jasuja, M. Deutscher, S. Chen, J. D. Walker, J. S. Duchin, S. Lett, S. Soliva, E. V. Wells, D. Swerdlow, T. M. Uyeki, A. E. Fiore, S. J. Olsen, A. M. Fry, C. B. Bridges, L. Finelli, P. I. H. N. Virus, *New Engl. J. Med.* 2009, 361, 1935.

6. K. Pardee, A. A. Green, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O'Connor, L. Gehrke, J. J. Collins, *Cell* 2016, 165, 1255.

7. N. Pardi, M. J. Hogan, R. S. Pelc, H. Muramatsu, H. Andersen, C. R. DeMaso, K. A. Dowd, L. L. Sutherland, R. M. Searce, R. Parks, W. Wagner, A. Granados, J. Greenhouse, M. Walker, E. Willis, J. S. Yu, C. E. Megee, G. D. Sempowski, B. L. Mui, Y. K. Tam, Y. J. Huang, D. Vanlandingham, V. M. Holmes, H. Balachandran, S. Sahu, M. Lifton, S. Higgs, S. E.
Hensley, T. D. Madden, M. J. Hope, K. Kariko, S. Santra, B. S. Graham, M. G. Lewis, T. C. Pierson, B. F. Haynes, D. Weissman, *Nature* 2017, 543, 248.

8. D. Gatherer, A. Kohl, *J. Gen. Virol.* 2016, 97, 269.

9. J. J. Waggoner, B. A. Pinsky, *J. Clin. Microbiol.* 2016, 54, 860.

10. J. Z. Song, M. G. Mauk, B. A. Hackett, S. Cherry, H. H. Bau, C. C. Liu, *Anal. Chem.* 2016, 88, 7289.

11. U. E. M. Gibson, C. A. Heid, P. M. Williams, *Genome Res.* 1996, 6, 995.

12. M. W. Pfaffl, *Nucleic Acids Res.* 2001, 29, e45.

13. J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, *Genome Biol* 2002, 3, research0034.1.

14. E. Engvall, *Med. Biol.* 1977, 55, 193.

15. A. Voller, A. Bartlett, D. E. Bidwell, *J. Clin. Pathol.* 1978, 31, 507.

16. M. F. Clark, R. M. Lister, M. Barjoseph, *Method Enzymol.* 1986, 118, 742.

17. M. Elhoseny, G. Ramirez-Gonzalez, O. M. Abu-Elnasr, S. A. Shawkat, N. Arunkumar, A. Farouk, *IEEE Access* 2018, 6, 20596.

18. G. Rateni, P. Dario, F. Cavallo, *Sensors-Basel* 2017, 17, 1453.

19. A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen, C. Dekker, *Nat. Mater.* 2003, 2, 537.

20. J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, J. A. Golovchenko, *Nature* 2001, 412, 166.

21. C. Dekker, *Nat. Nanotechnol.* 2007, 2, 209.

22. F. Haque, J. H. Li, H. C. Wu, X. J. Liang, P. X. Guo, *Nano Today* 2013, 8, 56.

23. M. Taniguchi, *Anal. Chem.* 2015, 87, 188.

24. B. N. Miles, A. P. Ivanov, K. A. Wilson, F. Dogan, D. Japrung, J. B. Edel, *Chem. Soc. Rev.* 2013, 42, 15.
25. D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. H. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. A. Schloss, *Nat. Biotechnol.* **2008**, *26*, 1146.

26. S. J. Heerema, C. Dekker, *Nat. Nanotechnol.* **2016**, *11*, 127.

27. M. Wanunu, *Phys. Life Rev.* **2012**, *9*, 125.

28. S. Garaj, W. Hubbard, A. Reina, J. Kong, D. Branton, J. A. Golovchenko, *Nature* **2010**, *467*, 190.

29. Y. X. Liu, X. C. Dong, P. Chen, *Chem. Soc. Rev.* **2012**, *41*, 2283.

30. G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, C. Dekker, *Nano Lett.* **2010**, *10*, 3163.

31. C. A. Merchant, K. Healy, M. Wanunu, V. Ray, N. Peterman, J. Bartel, M. D. Fischbein, K. Venta, Z. T. Luo, A. T. C. Johnson, M. Drndic, *Nano Lett.* **2010**, *10*, 2915.

32. J. D. Feng, K. Liu, R. D. Bulushev, S. Khlybov, D. Dumcenco, A. Kis, A. Radenovic, *Nat. Nanotechnol.* **2015**, *10*, 1070.

33. K. Liu, J. D. Feng, A. Kis, A. Radenovic, *ACS Nano* **2014**, *8*, 2504.

34. A. B. Farimani, K. Min, N. R. Aluru, *ACS Nano* **2014**, *8*, 7914.

35. B. S. Bull, M. A. Schneiderman, G. Brecher, *Am. J. Clin. Pathol.* **1965**, *44*, 678.

36. R. W. Deblois, C. P. Bean, *Rev. Sci. Instrum.* **1970**, *41*, 909.

37. X. Hou, W. Guo, L. Jiang, *Chem. Soc. Rev.* **2011**, *40*, 2385.

38. M. Davenport, K. Healy, M. Pevarnik, N. Teslich, S. Cabrini, A. P. Morrison, Z. S. Siwy, S. E. Letant, *ACS Nano* **2012**, *6*, 8366.

39. C. T. A. Wong, M. Muthukumar, *J. Chem. Phys.* **2007**, *126*, 164903.

40. S. Ghosal, J. D. Sherwood, H. C. Chang, *Biomicrofluidics* **2019**, *13*, 011301.
41. S. Ryuzaki, M. Tsutsui, Y. H. He, K. Yokota, A. Arima, T. Morikawa, M. Taniguchi, T. Kawai, 
   *Nanotechnology* **2017**, *28*, 155501.

42. A. Arima, M. Tsutsui, I. H. Harlisa, T. Yoshida, M. Tanaka, K. Yokota, W. Tonomura, M. 
   Taniguchi, M. Okochi, T. Washio, T. Kawai, *Sci. Rep.* **2018**, *8*, 16305.

43. M. Tsutsui, T. Yoshida, K. Yokota, H. Yasaki, T. Yasui, A. Arima, W. Tonomura, K. 
   Nagashima, T. Yanagida, N. Kaji, M. Taniguchi, T. Washio, Y. Baba, T. Kawai, *Sci. Rep.* **2017**, 
   *7*, 17371.

44. M. Tsutsui, M. Tanaka, T. Marui, K. Yokota, T. Yoshida, A. Arima, W. Tonomura, M. 
   Taniguchi, T. Washio, M. Okochi, T. Kawai, *Anal. Chem.* **2018**, *90*, 1511.

45. A. Arima, I. H. Harlisa, T. Yoshida, M. Tsutsui, M. Tanaka, K. Yokota, W. Tonomura, J. 
   Yasuda, M. Taniguchi, T. Washio, M. Okochi, T. Kawai, *J. Am. Chem. Soc.* **2018**, *140*, 16834.

46. Y. H. He, M. Tsutsui, C. Fan, M. Taniguchi, T. Kawai, *ACS Nano* **2011**, *5*, 8391.

47. Y. H. He, M. Tsutsui, C. Fan, M. Taniguchi, T. Kawai, *ACS Nano* **2011**, *5*, 5509.

48. G. W. Slater, C. Holm, M. V. Chubynsky, H. W. de Haan, A. Dube, K. Grass, O. A. Hickey, 
   C. Kingsburry, D. Sean, T. N. Shendruk, L. X. Nhan, *Electrophoresis* **2009**, *30*, 792.

49. E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. L. Li, J. Yang, M. Mayer, 
   *Nat. Nanotechnol.* **2011**, *6*, 253.

50. X. Q. Li, R. Hu, J. Li, X. Tong, J. J. Diao, D. P. Yu, Q. Zhao, *Appl. Phys. Lett.* **2016**, *109*, 
    143105.

51. Z. P. Tang, B. Lu, Q. Zhao, J. J. Wang, K. F. Luo, D. P. Yu, *Small* **2014**, *10*, 4332.

52. G. Ando, C. Hyun, J. L. Li, T. Mitsui, *ACS Nano* **2012**, *6*, 10090.

53. G. F. Schneider, Q. Xu, S. Hage, S. Luik, J. N. H. Spoor, S. Malladi, H. Zandbergen, C. 
   Dekker, *Nat. Commun.* **2013**, *4*, 2619.

54. R. B. Schoch, J. Y. Han, P. Renaud, *Rev. Mod. Phys.* **2008**, *80*, 839.
55. P. Abgrall, N. T. Nguyen, *Anal. Chem.* **2008**, *80*, 2326.

56. A. Darvish, J. S. Lee, B. Peng, J. Saharia, R. V. K. Sundaram, G. Goyal, N. Bandara, C. W. Ahn, J. Kim, P. Dutta, I. Chaiken, M. J. Kim, *Electrophoresis* **2019**, *40*, 776.

57. A. McMullen, H. W. de Haan, J. X. Tang, D. Stein, *Nat. Commun.* **2014**, *5*, 4171.

58. H. W. Wu, Y. H. Chen, Q. Z. Zhou, R. L. Wang, B. C. Xia, D. J. Ma, K. F. Luo, Q. J. Liu, *Anal. Chem.* **2016**, *88*, 2502.

59. G. Goyal, A. Darvish, M. J. Kim, *Analyst* **2015**, *140*, 4865.

60. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **2000**, *405*, 299.

61. A. J. Davison, J. E. Scott, *J. Gen. Virol.* **1986**, *67*, 1759.

62. E. R. Mardis, *Trends Genet.* **2008**, *24*, 133.

63. W. J. Ansorge, *New Biotechnol.* **2009**, *25*, 195.

64. K. H. Zhang, K. F. Luo, *Soft Matter* **2013**, *9*, 2069.

65. K. Luo, T. Ala-Nissila, S. C. Ying, A. Bhattacharya, *Phys. Rev. Lett.* **2007**, *99*, 148102.

66. K. F. Luo, I. Huopaniemi, T. Ala-Nissila, S. C. Ying, *J. Chem. Phys.* **2006**, *124*, 114704.

67. S. J. Gamblin, J. J. Skehel, *J. Biol. Chem.* **2010**, *285*, 28403.

68. S. J. Gamblin, L. F. Haire, R. J. Russell, D. J. Stevens, B. Xiao, Y. Ha, N. Vasisht, D. A. Steinhauer, R. S. Daniels, A. Elliot, D. C. Wiley, J. J. Skehel, *Science* **2004**, *303*, 1838.

69. H. Yang, P. J. Carney, V. P. Mishin, Z. Guo, J. C. Chang, D. E. Wentworth, L. V. Gubareva, J. Stevens, *J. Virol.* **2016**, *90*, 5770.

70. K. A. Riske, R. Dimova, *Biophys. J.* **2006**, *91*, 1778–.

71. M. M. Sadik, J. B. Li, J. W. Shan, D. I. Shreiber, H. Lin, *Phys. Rev. E* **2011**, *83*, 066316.

72. J. E. Hall, *J. Gen. Physiol.* **1975**, *66*, 531.

73. E. Weatherall, G. R. Willmott, *J. Phys. Chem. B* **2015**, *119*, 5328.

74. E. Weatherall, G. R. Willmott, *Analyst* **2015**, *140*, 3318.
75. P. Hauer, E. C. Le Ru, G. R. Willmott, *Biomicrofluidics* 2015, 9, 014110.

76. J. A. Somerville, G. R. Willmott, J. Eldridge, M. Griffiths, K. M. McGrath, *J. Colloid. Interf. Sci.* 2013, 394, 243.

77. R. Vogel, W. Anderson, J. Eldridge, B. Glossop, G. Willmott, *Anal. Chem.* 2012, 84, 3125.

78. B. Lu, D. P. Hoogerheide, Q. Zhao, H. B. Zhang, T. P. Zhipeng, D. P. Yu, J. A. Goloychenko, *Nano Lett* 2013, 13, 3048.

79. Y. H. He, M. Tsutsui, R. H. Scheicher, F. Bai, M. Taniguchi, T. Kawai, *ACS Nano* 2013, 7, 538.

80. S. Duhr, D. Braun, *Proc. Natl. Acad. Sci. USA* 2006, 103, 19678.

81. P. Reineck, C. J. Wienken, D. Braun, *Electrophoresis* 2010, 31, 279.

82. Y. H. He, M. Tsutsui, R. H. Scheicher, C. Fan, M. Taniguchi, T. Kawai, *Biophys. J.* 2013, 105, 776.

83. R. M. M. Smeets, U. F. Keyser, D. Krapf, M. Y. Wu, N. H. Dekker, C. Dekker, *Nano Lett* 2006, 6, 89.

84. M. Wanunu, W. Morrison, Y. Rabin, A. Y. Grosberg, A. Meller, *Nat. Nanotechnol.* 2010, 5, 160.

85. S. H. Behrens, D. G. Grier, *J. Chem. Phys.* 2001, 115, 6716.

86. D. Stein, M. Kruithof, C. Dekker, *Phys. Rev. Lett.* 2004, 93, 035901.

87. P. Chen, T. Mitsui, D. B. Farmer, J. Golovchenko, R. G. Gordon, D. Branton, *Nano Lett* 2004, 4, 1333.

88. B. M. Venkatesan, A. B. Shah, J. M. Zuo, R. Bashir, *Adv. Funct. Mater.* 2010, 20, 1266.

89. S. Alamiyounsi, A. Larbot, M. Persin, J. Sarrazin, L. Cot, *J. Membrane Sci.* 1995, 102, 123.

90. E. J. F. Dickinson, H. Ekstrom, E. Fontes, *Electrochem. Commun.* 2014, 40, 71.

91. R. Kumar, M. Muthukumar, *J. Chem. Phys.* 2009, 131, 194903.
92. M. Muthukumar, *J. Chem. Phys.* **2010**, *132*, 195101.

93. L. Shen, A. Guo, X. Y. Zhu, *Surf. Sci.* **2011**, *605*, 494.

94. O. Joshi, J. McGuire, *Appl. Biochem. Biotech.* **2009**, *152*, 235.

95. B. Y. Hua, K. Y. Han, R. B. Zhou, H. J. Kim, X. H. Shi, S. C. Abeysirigunawardena, A. Jain, D. Singh, V. Aggarwal, S. A. Woodson, T. Ha, *Nat. Methods* **2014**, *11*, 1233.

96. J. K. Towns, F. E. Regnier, *Anal. Chem.* **1991**, *63*, 1126.

97. C. J. Van Rijsbergen, “*Information Retrieval*”, 1979, Butterworth-Heinemann, Massachusetts, USA.

98. J. K. Rosenstein, M. Wanunu, C. A. Merchant, M. Drndic, K. L. Shepard, *Nat. Methods* **2012**, *9*, 487.

99. R. M. M. Smeets, U. F. Keyser, N. H. Dekker, C. Dekker, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 417.

100. V. Tabard-Cossa, D. Trivedi, M. Wiggin, N. N. Jetha, A. Marziali, *Nanotechnology* **2007**, *18*, 305505.

101. J. D. Uram, K. Ke, M. Mayer, *ACS Nano* **2008**, *2*, 857.

102. G. Holmes, A. Donkin, I. E. Witten, in Proceedings of ANZIIS ‘94Australian New Zealand Intelligent Information Systems Conference, Brisbane, Queensland, Australia, 1994.

103. S. Geisser, “*Predictive Inference*”, 1993, Chapman and Hall, New York, USA.

104. J. J. Rodriguez, L. I. Kuncheva, *IEEE T. Pattern. Anal.* **2006**, *28*, 1619.

105. W. Tonomura, M. Tsutsui, A. Arima, K. Yokota, M. Taniguchi, T. Washio, T. Kawai, *Lab Chip* **2019**, *19*, 1352.

106. M. Taniguchi, T. Ohshiro, Y. Komoto, T. Takaai, T. Yoshida, T. Washio, *J. Phys. Chem. C* **2019**, *123*, 15867.

107. K. Noto, M. H. Saier, C. Elkan, *Lect. Notes Artif. Int.* **2008**, *5360*, 202.
Figure Captions

Fig. 1 High aspect ratio and low aspect ratio nanopores. (a) The principle of a solid-state nanopore system. The solid-state nanoporous material is composed of a membrane (light blue) and a support substrate (green). When the analyte (blue) passes through the nanopore, the ion current is hindered by the analyte, such that the ion current decreases. (b) Classification of high aspect ratio and low aspect ratio nanopores. The membrane thickness and pore diameter are $L$ and $D$, respectively.

Fig. 2 Typical analysis of an ion current–time waveform. (a) Schematic diagram of the ion current–time waveform when the analyte passes through the nanopore continuously. The peak ion current, the duration of the ion current, and the time between the ion current peaks are defined as $I_p$, $t_d$, and $t_f$, respectively. (b) Schematic diagram of the ion current–time waveform when the analyte collides with the nanopore. The collision produces a smaller $I_p$ and $t_d$ than does an analyte passing through the nanopore. (c) $I_p$ histogram and (d) a two-dimensional map of $I_p$ and $t_d$ extracted from the ion current–time waveform obtained when viruses A and B are measured.

Fig. 3 Measurement of viruses and bacteria using solid-state nanopores. (a) Filament virus measurement using nanopores with $L/D = 0.77$. Atomic force microscopy (AFM) image and ion current–time waveform on mica. (b) Measurement of tobacco mosaic virus using nanopores with $L/D < 1.2$. TEM image and ion current–time waveform of tobacco mosaic virus. (c) Measurement of influenza virus using solid-state nanopores with $L/D = 0.17$. Influenza structure and ion current–time waveform. (d) Measurement of liposome using solid-state nanopores with $L/D = 0.8$. TEM image of liposome and ion current–time waveform. (e) Measurement of Escherichia coli and Bacillus subtilis using solid-state nanopores with $L/D = 0.013$. SEM images of E. coli and B. subtilis. Ion current–time waveform obtained by superimposing 200 waveforms
of *E. coli* and *B. subtilis*. Scale bars denote 2 µm.

Fig. 4 Aspect ratio dependence of electric field strength obtained by Finite element analysis calculations. Electric field strengths of (a) a low aspect ratio nanopore, (b) an *L/D* ~ 1 nanopore, and (c) a high aspect ratio nanopore. (d) Aspect ratio dependence of electric field strength in the radial direction. (e) and (f) Aspect ratio dependence of the electric field strength in the z direction. *D* and *ρ* indicate the diameter and radial direction of the nanopore, respectively.

Fig. 5 Viral and bacterial flow dynamics. Viruses and bacteria are negatively charged. (a) A case in which the surface charge of the nanopore is negative. When electrophoresis > electroosmotic flow + fluid resistance, viruses and bacteria pass through the nanopore. (b) Another case wherein the surface charge of the nanopore is negative. When electrophoresis < electroosmotic flow + fluid resistance, viruses and bacteria cannot pass through the nanopore. (c) The case in which the surface charge of the nanopore is positive. Viruses and bacteria pass through the nanopore at a fast rate.

Fig. 6 Analysis of the ion current–time waveform obtained from bacteria. (a) Electron microscopic image of streptococci. Green is a false color. Scale bar denotes 0.5 µm. (b) Typical ion current–time waveform obtained from streptococci. Measured with low aspect ratio nanopores with *L/D* = 0.042. (c) Ion current–time waveform dependence of streptococci. *N* streptococci showed *N* peak current values. (d) Ion current–time waveform of *Streptococcus* by multiphysics simulation.

Fig. 7 Chemical modification of the solid-state nanopore surface. (a) Conceptual diagram of an
unmodified nanopore and the ion current–time waveform. Aβ peptides were measured with SiN nanopores with a diameter of 33 nm and a thickness of 22 nm. (b) Conceptual diagram and ion current–time waveform of a solid-state nanopore modified with a lipid bilayer.

Fig. 8 Analysis of ion current–time waveforms using machine learning. (a) In the learning process for two types of viruses, the ion current–time waveform of each virus is acquired. A classifier suitable for classifying two types of viruses is learned using the features of each waveform. For example, when the features are $I_p$ and $t_{eb}$, the two types of viruses are not successfully classified. If two types of features are used, the two types of viruses are classified with high accuracy. In the reasoning process, the ion current–time waveform of the virus does not indicate which virus was acquired. The waveform is classified using the feature that has the highest accuracy in the learning process. The classification result is evaluated by the $F$-measure value. This technique can classify which virus has produced a given waveform. (b) Learning process and reasoning process. The reasoning model uses the feature and classifier that achieved the highest accuracy in the learning process. Data combination refers to the mixing of virus A and virus B data. See Figure 9 for an example of undersampling and evaluation methods.

Fig. 9 Undersampling and evaluation methods. Undersampling is used when the number of data is not uniform; the number of data is set to the smaller number. The five-fold cross-validation method is one of the evaluation methods. The data groups obtained by undersampling are randomly divided into five groups. Four out of five groups are used as training data to build the reasoning model. A group of data is inputted to the determined reasoning model, and the $F$-measure is calculated. The $F$-measure obtained by repeating the same operation five times is averaged to evaluate the identification accuracy.
Fig. 10 Analysis of virus A and virus B using machine learning. (a) Confusion matrix. The numerical value of the matrix corresponds to the number of data. (b) Precision, recall, and $F$-measure calculated from the confusion matrix.

Fig. 11 Analysis of viral and bacterial ion current–time waveforms using machine learning. (a)–(e) Identification of influenza virus. (a) Ten defined features. (b) Distribution of $F$-measure for two types of influenza viruses. $F$-measure values obtained from all combinations of 60 feature vectors and 71 classifiers. (d) Pulse number dependency of the identification accuracy of two types of viruses. (e) Relationship between $P_{rec}$ and features. Physical interpretation of features. (f)–(h) Discrimination between *Escherichia coli* and *Bacillus subtilis*. (f) The defined feature. (g) A histogram of the aspect ratio dependence of the $F$-measure. (h) Dependence of $F$-measure on aspect ratio.
Fig. 1 High aspect ratio and low aspect ratio nanopores. (a) The principle of a solid-state nanopore system. The solid-state nanoporous material is composed of a membrane (light blue) and a support substrate (green). When the analyte (blue) passes through the nanopore, the ion current is hindered by the analyte, such that the ion current decreases. (b) Classification of high aspect ratio and low aspect ratio nanopores. The membrane thickness and pore diameter are $L$ and $D$, respectively.
Fig. 2 Typical analysis of an ion current–time waveform. (a) Schematic diagram of the ion current–time waveform when the analyte passes through the nanopore continuously. The peak ion current, the duration of the ion current, and the time between the ion current peaks are defined as $I_p$, $t_d$, and $t_f$, respectively. (b) Schematic diagram of the ion current–time waveform when the analyte collides with the nanopore. The collision produces a smaller $I_p$ and $t_d$ than does an analyte passing through the nanopore. (c) $I_p$ histogram and (d) a two-dimensional map of $I_p$ and $t_d$ extracted from the ion current–time waveform obtained when viruses A and B are measured.
Fig. 3 Measurement of viruses and bacteria using solid-state nanopores. (a) Filament virus measurement using nanopores with $L/D = 0.77$. Atomic force microscopy (AFM) image and ion current–time waveform on mica. (b) Measurement of tobacco mosaic virus using nanopores with $L/D < 1.2$. TEM image and ion current–time waveform of tobacco mosaic virus. (c) Measurement of influenza virus using solid-state nanopores with $L/D = 0.17$. Influenza structure and ion current–time waveform. (d) Measurement of liposome using solid-state nanopores with $L/D = 0.8$. 
TEM image of liposome and ion current–time waveform. (e) Measurement of *Escherichia coli* and *Bacillus subtilis* using solid-state nanopores with $L/D = 0.013$. SEM images of *E. coli* and *B. subtilis*. Ion current–time waveform obtained by superimposing 200 waveforms of *E. coli* and *B. subtilis*. Scale bar denote 2μm.
Fig. 4 Aspect ratio dependence of electric field strength obtained by Finite element analysis calculations. Electric field strengths of (a) a low aspect ratio nanopore, (b) an $L/D \sim 1$ nanopore, and (c) a high aspect ratio nanopore. (d) Aspect ratio dependence of electric field strength in the radial direction. (e) and (f) Aspect ratio dependence of the electric field strength in the $z$ direction. $D$ and $\rho$ indicate the diameter and radial direction of the nanopore, respectively.
Fig. 5 Viral and bacterial flow dynamics. Viruses and bacteria are negatively charged. (a) A case in which the surface charge of the nanopore is negative. When electrophoresis > electroosmotic flow + fluid resistance, viruses and bacteria pass through the nanopore. (b) Another case wherein the surface charge of the nanopore is negative. When electrophoresis < electroosmotic flow + fluid resistance, viruses and bacteria cannot pass through the nanopore. (c) The case in which the surface charge of the nanopore is positive. Viruses and bacteria pass through the nanopore at a fast rate.
Fig. 6 Analysis of the ion current–time waveform obtained from bacteria. (a) Electron microscopic image of streptococci. Green is a false color. Scale bar denotes 0.5 μm. (b) Typical ion current–time waveform obtained from streptococci. Measured with low aspect ratio nanopores with $L/D = 0.042$. (c) Ion current–time waveform dependence of streptococci. $N$ streptococci showed $N$ peak current values. (d) Ion current–time waveform of $Streptococcus$ by multiphysics simulation.
Fig. 7 Chemical modification of the solid-state nanopore surface. (a) Conceptual diagram of an unmodified nanopore and the ion current–time waveform. Aβ peptides were measured with SiN nanopores with a diameter of 33 nm and a thickness of 22 nm. (b) Conceptual diagram and ion current–time waveform of a solid-state nanopore modified with a lipid bilayer.
Fig. 8 Analysis of ion current–time waveforms using machine learning. (a) In the learning process for two types of viruses, the ion current–time waveform of each virus is acquired. A classifier suitable for classifying two types of viruses is learned using the features of each waveform. For example, when the features are $I_p$ and $t_d$, the two types of viruses are not successfully classified. If two types of features are used, the two types of viruses are classified with high accuracy. In the reasoning process, the ion current–time waveform of the virus does not indicate which virus was acquired. The waveform is classified using the feature that has the highest accuracy in the learning process. The classification result is evaluated by the $F$-measure value. This technique can classify which virus has produced a given waveform. (b) Learning process and reasoning process. The
reasoning model uses the feature and classifier that achieved the highest accuracy in the learning process. Data combination refers to the mixing of virus A and virus B data. See Figure 9 for an example of undersampling and evaluation methods.
Fig. 9 Undersampling and evaluation methods. Undersampling is used when the number of data is not uniform; the number of data is set to the smaller number. The five-fold cross-validation method is one of the evaluation methods. The data groups obtained by undersampling are randomly divided into five groups. Four out of five groups are used as training data to build the reasoning model. A group of data is inputted to the determined reasoning model, and the $F$-measure is calculated. The $F$-measure obtained by repeating the same operation five times is averaged to evaluate the identification accuracy.
Fig. 10 Analysis of virus A and virus B using machine learning. (a) Confusion matrix. The numerical value of the matrix corresponds to the number of data. (b) Precision, recall, and $F$-measure calculated from the confusion matrix.
Fig. 11 Analysis of viral and bacterial ion current–time waveforms using machine learning. (a)–(e) Identification of influenza virus. (a) Ten defined features. (b) Distribution of $F$-measure for two types of influenza viruses. $F$-measure values obtained from all combinations of 60 feature vectors and 71 classifiers. (d) Pulse number dependency of the identification accuracy of two
types of viruses. (e) Relationship between $P_{\text{rec}}$ and features. Physical interpretation of features. (f)–(h) Discrimination between \textit{Escherichia coli} and \textit{Bacillus subtilis}. (f) The defined feature. (g) A histogram of the aspect ratio dependence of the $F$-measure. (h) Dependence of $F$-measure on aspect ratio.
