The NTD Nanoscope Kit

Stephen Winters-Hilt[1,2,*], Eric Morales[1], and Evenie Horton-Chao[1]

[1] Dept. of Computer Science, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148, USA
[2] Meta Logos Incorporated, 6218 Waldo Dr., New Orleans, LA 70122, USA

*email: winters@cs.uno.edu; winters@meta-logos.com

Abstract

We present the first nanopore transduction detector (NTD) kit. When nanopore transduction events are examined in conjunction with machine learning based signal processing software, a 'nanoscope' is obtained. To expedite set-up and calibration of the NTD Nanoscope kit components and signal processing software, the kit is being designed to include a set of test buffers and control molecules. In more sophisticated pattern recognition tasks data can be streamed off the detector to data-processing servers for NTD analysis, data sharing, and device trouble-shooting. Details on the kit's capabilities, cost, and minimal peripheral lab equipment needs are described. The NTD Kit could provide the means for a powerful biosensing platform for highly specific DNA (or RNA) detection, DNA aptamer-based detection, or detection via DNA-linked binding moieties. The NTD Kit is the first deployment of this technology to academia (biochemistry labs, etc.) and industry (medical labs, etc.). The deployment is being made so manageable, however, and the cost and peripheral lab expense so small (the device is inexpensive, easy to operate, and only needs a small portable hood for prep work), that current 'dry-labs', e.g. computationally intensive bioinformatics labs, can enable sufficient wet-lab capability for NTD experiments. This is a remarkably cross-disciplinary approach to biomolecular science and engineering and opens the door for non-traditional wet-lab research efforts to 'get into the game' by directly designing their own molecular probes and biosensing tests, according to the commoditized construction guidelines to be described.

1 Introduction

Preparations are underway to make a nanopore transduction detector (NTD) kit. When nanopore transduction events are examined in conjunction with machine learning based signal processing software, a 'nanoscope' is obtained [1,2].

To expedite set-up and calibration, the NTD Nanoscope kit is being designed to include a set of test buffers and control molecules. The NTD Nanoscope is designed to be as self-calibrating as possible to provide enhanced, autonomous reliability: signals are normalized computationally with respect to physical parameters (e.g. temperature, ph, salt concentration, etc.), eliminating the need for complex (and expensive) physical feedback systems to stabilize the device. In addition, specially engineered calibration probes have been designed to enable real-time self-calibration by generating a standard "carrier signal." These probes can also be added to samples being analyzed to provide a run-by-run self-calibration. These redundant self-calibration capabilities result in a stable, user-friendly, device that can be operated by an entry level lab technician.

The more sophisticated pattern recognition tasks data can be streamed off the detector to the NTD kit provider's data-processing servers for NTD analysis, and device trouble-shooting. Advanced signal processing tools will be accessible via web-service when the servers are set up, both for real-time (on-line), or for off-line (batch) signal processing/analysis. On-line calibration services and troubleshooting analysis for kits will be made possible via use of the control molecules (some described in what follows) as test references.

The NTD Kit could provide the means for a powerful biosensing platform for DNA (or RNA) detection, DNA aptamer-based detection, or detection via DNA-linked binding moieties (where the DNA part is designed to be drawn into the detector and establish a modulatory channel blockade signal). The NTD Kit also enables a variety of assaying protocols to be performed, as will be shown.

2 Background

2.1 Nanopore Detection

A nanometer-scale channel is used to associate ionic current measurements with single-molecule channel blockades. The α-hemolysin toxin self-assembles forming a trans membrane channel, leading to an inexpensive and reproducible nanopore detector. (Fig. 1, Left) The a-hemolysin based nanopore detector is well-suited for the observation of biomolecular kinetics through the modulation of ionic current through its limiting aperture. A DNA hairpin molecule that is a channel modulator is shown superimposed onto the cross sectional view of the α-hemolysin channel. In Fig. 1, Right.

By periodically reversing the applied potential, hundreds of molecules can be sampled every second (Fig. 2, Left). Standard Data acquisition hardware is then used to take the analog signal from the patch clamp amplifier and deliver it to a computer (Fig. 2, Right).

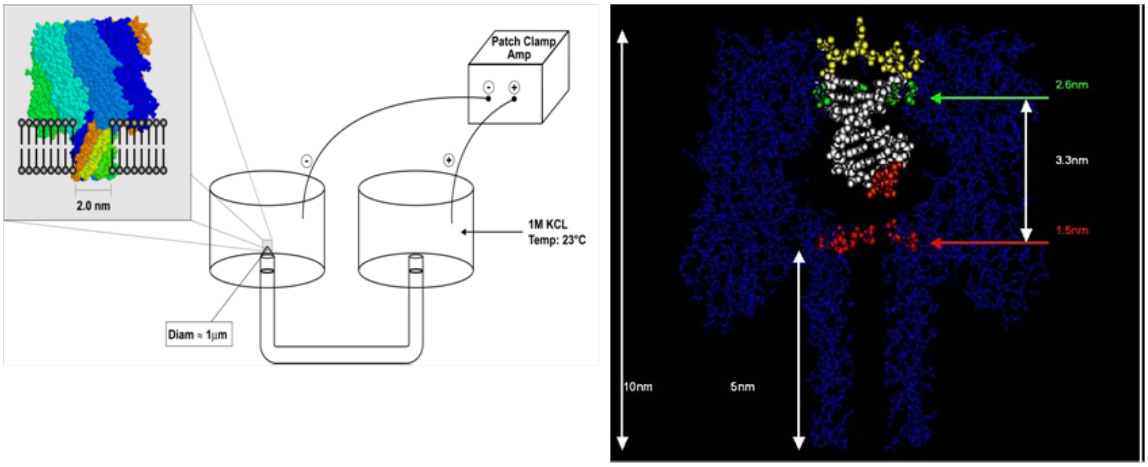

Fig. 1 Nanopore Transduction Detector (left) and α-hemolysin channel (right).

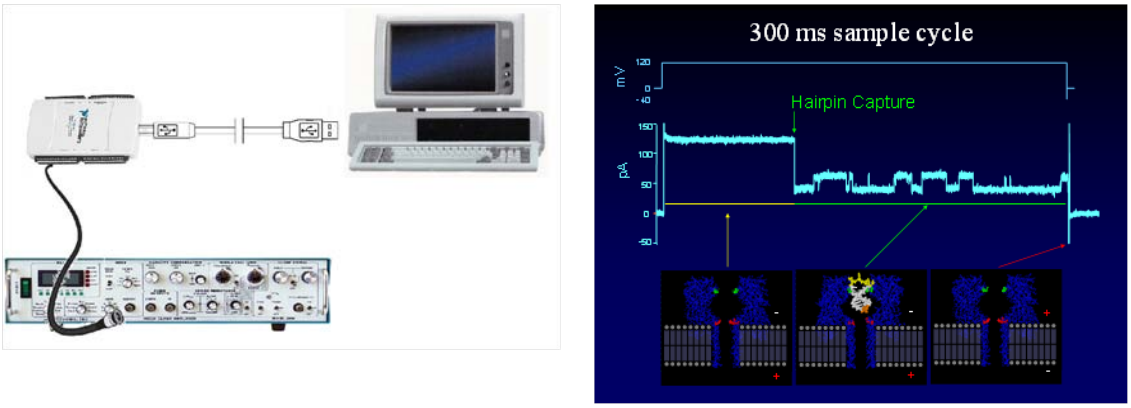

Fig. 2 Sampling Protocol (left) and DAQ hardware (right).

2.2 Nanopore Transduction Detection

The NTD Nanoscope offers the means to critically complete the SNP and RNA-seq data processing pipeline. Current methods used by industrial for DNA seqeucing have error rates of approximately one-in-a-thousand. This is a limitation due to the enzymes used in the methods themselves having error rates of approximately one-in-a-thousand. DNA sequencing is fast becoming incredibly inexpensive, but at this one-in-a-thousand error rate. The problem is that disease conditions more rare than one-in-a thousand can be overlooked or mis-diagnosed, and most genetic diseases are more rare than this. What is needed is targeted re-sequencing and SNP discovery in the 'important' parts of the genome to obtain reduced error rates, and to do so in an inexpensive way.

In the NTD functionalization the transducer molecule is drawn into the channel by an applied potential but is too big to translocate, instead becoming stuck in a bistable capture such that it modulates the channel's ion-flow with stationary statistics in a distinctive way. If the channel modulator is bifunctional in that one end is meant to be captured and modulate while the other end is linked to an aptamer or antibody or ssDNA for specific binding (or annealing), then we have the basis for a remarkably sensitive and specific event transducer.

The NTD Nanoscope offers a means to do sensing with better than one-in-a-thousand accuracy on any SNP variant or antigen. The NTD Nanoscope description in what follows also directly establishes a general-target (non-DNA) biosensor with sensitivity and specificity limited by that of the aptamer or antibody involved.

The nanopore transduction detection platform (Fig. 3) involves functionalizing a nanopore detector platform in a new way that is cognizant of signal processing and machine learning capabilities and advantages (see [3] for recent results on NTD 'Model' Systems, such that a highly sensitive biosensing capability is achieved [1,2]. The core idea in the NTD functionalization of the nanopore detector is to design a molecule that can be drawn into the channel (by an applied potential) but will be too big to translocate, instead becoming stuck in a bistable 'capture' such that it modulates the ion-flow in a distinctive way. An approximately two-state 'telegraph signal' has been engineered for a number of NTD modulators.

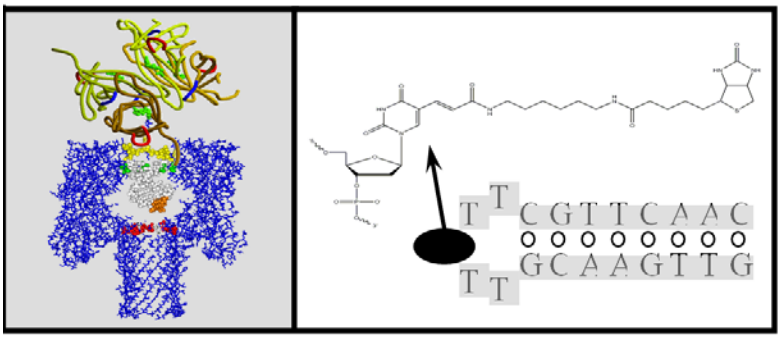

Fig. 3. A nanopore transducer molecule. The nanopore detector consists of a single pore in a lipid bilayer that is created by the oligomerization of the staphylococcal alpha-hemolysin toxin, and a patch clamp amplifier capable of measuring pico Ampere channel currents. Transduction is then done my introduction of a transdcuer molecule, typically one that produces a signal with stationary stattics that is not at a 'fixed level', e.g., its typically a telegraph like signal dominated by two levels.

3 Methods

3.1 SSA Protocol Server Users

The SSA Protocol entails methods for signal acquisition, feature extraction, classification, and clustering (see [1,4,5] for details). For Nanoscope Device or Nanoscope Kit users, a further refinement in the standard web-interfaces to SSA Web-Server methods [4] is to provide a streaming-data SSA-Server access. The latter setting offers Nanoscope calibration and troubleshooting capabilities in operational settings, as well as access to distributed real-time speedup on core computational methods. Thus, the entire pipeline of signal processing needed, to go from raw streaming data to concise results, can be passed to the SSA-Server. Usually it is more efficient, however, to separate the acquisition module (the tFSA) and bundle that with the device (or kit) to enable a bare-bones operational capability 'locally', for signal acquisition, then enable feature extraction and classification/clustering enhancements via centralized maintenance of an SSA-Server algorithm toolset and data library. Many of the latter analyses can be done in real-time, for the experiments shown in the Methods for example, where a networked device is already set up as described. This same deployment is, thus, also sought in the Nanoscope Device or Nanoscope Kit development and is thereby already validated with the results on the existing system over the past several years. What this means for the signal processing needs is that the latency of a large raw data transfer is largely eliminated by the local pre-processing to perform rudimentary signal acquisition, which only passes acquired signal onto the network. For gene

finding in mammals, for example, the gene regions amount to about 3% of the genome, so crude recognition of such would reduce acquisition and transmission of data by a factor of 33. The drop in bandwidth needs in NTD channel current signal processing can be even more pronounced.

3.2 NTD KIT Components and Software

Kit Package:
(1) NTD Kit Device Components and chemicals
(2) NTD Set-up Manual
(3) configured NTD Kit Computer with local SSA software and meta-logos SSA link
(4) buffer controls
(5) model systems/controls
(6) Data analysis and data repository services
(7) Setup, calibration, and troubleshooting services

3.3 Channel Current Cheminformatics (CCC) Architecture

A protocol has been developed for the discovery, characterization, and classification of localizable, approximately-stationary, statistical signal structures in stochastic sequential data, such as channel current data. Some of the CCC methods involved are shown in Fig. 4.

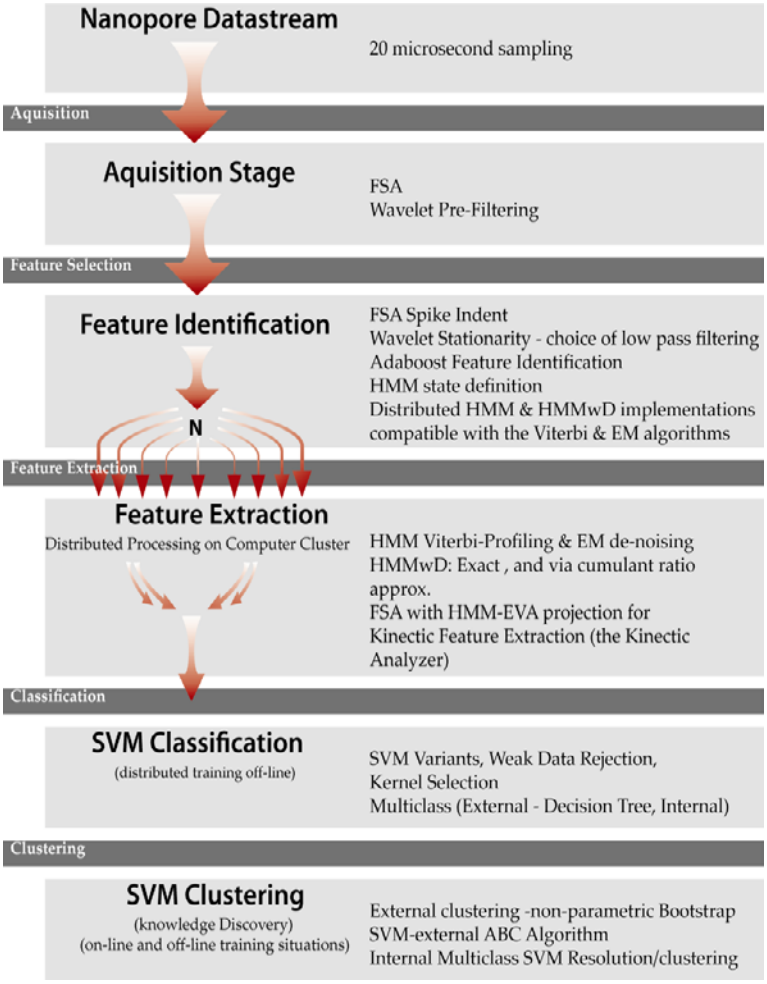

Fig. 4. Nanopore cheminformatics & data-flow control architecture. Aside from the modular design with the different machine learning methods shown (HMMs, SVMs, etc.), recent augmentations to this architecture for real-time processing include use of a networked server to link to the patch-clamp amplifier, and the 'real-time' pattern recognition informed signal processing architecture (the latter shown in Fig. 6).

4 Results

4.1 Kit Components

In Fig. 5 is shown the NTD kit core Assembly with Peltier-based temperature controller model 350B manufactured by Newport Temperature. The A-M systems patch clamp amplifier model 2400 is used to impose a constant voltage across electrodes while monitoring picoampere fluctuations. A micro manipulator is used to deliver profusion tubing to *cis* chamber during channel acquisition. A TMC bench top vibration isolation table is used to support the core assembly and reduce vibration shock. With the NTD Kit we, thus, achieve a nanoscale single-molecule lab construction. This is obviously a low-cost method to perform nanoscale experiments. There is also a remarkable versatility in the incorporation of transducer, as will become more apparent in what follows.

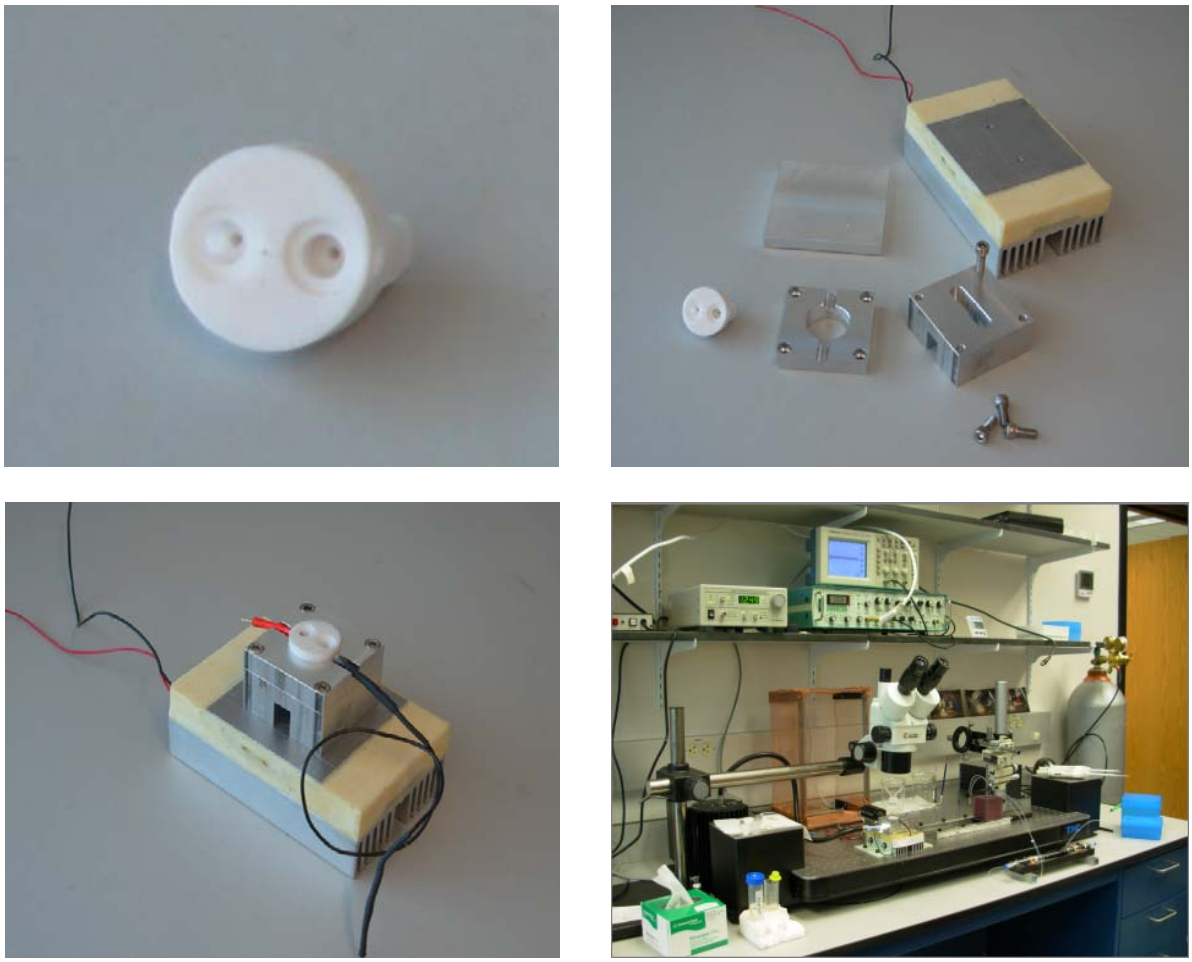

Fig. 5. The NTD Kit Components. (Upper Left) One inch diameter Teflon core machined to contain two chambers that are connected by a segment of Teflon tubing. This tubing has one end open while the other is capped to restrict it's opening to a tiny pin hole (Aperture) approximately 25 microns in diameter (not shown). (Upper Right) Aluminum square block is machined to support the core and provide access for illumination and electrode ports, and have easy assembly and mounting on the Peltier device. (Lower Left) The Teflon and aluminum components are shown seated on the Peltier device for temperature control of the wells. Silver chloride (Ag/AgCl) pellet electrodes are shown inserted into the cis and trans chambers port wells. (Lower Right) A NTD Nanoscope kit arrangement that uses a table-top vibration isolation table, has perfusion apparatus, Faraday cage, and linkage to a computer (to left of workstation, not shown).

4.2 Stochastic Phase Modulation (SPM)

in NTD applications of the SSA Protocol, due to the molecular dynamics of the captured transducer molecule, a unique reference signal with stationary (or approximately stationary) statistics is engineered to be generated during transducer blockade, analogous to a carrier signal in standard electrical engineering signal analysis. The adaptive SSA machine learning algorithms for real-time analysis of the stochastic signal generated by the transducer molecule offer a "lock and key" level of signal discrimination. The heart of the signal processing algorithm is an adaptive Hidden Markov Model (AHMM) based feature extraction method, implemented on a distributed processing platform for real-time operation. For real-time processing, the AHMM is used for feature extraction on channel blockade current data, while classification and clustering analysis are implemented using a Support Vector Machine. In addition, the design of the machine learning based algorithms allow for scaling to large datasets, real-time distributed processing, and are adaptable to analysis on any channel-based dataset, including resolving signals for different nanopore substrates (e.g. solid state configurations) or for systems based on translocation technology. The machine learning software has also been integrated into the nanopore detector for "real-time" pattern-recognition informed (PRI) feedback [6,7] (see Fig. 6). The methods used to implement the PRI feedback include *distributed* HMM and SVM implementations, which enable the processing speedup that is needed.

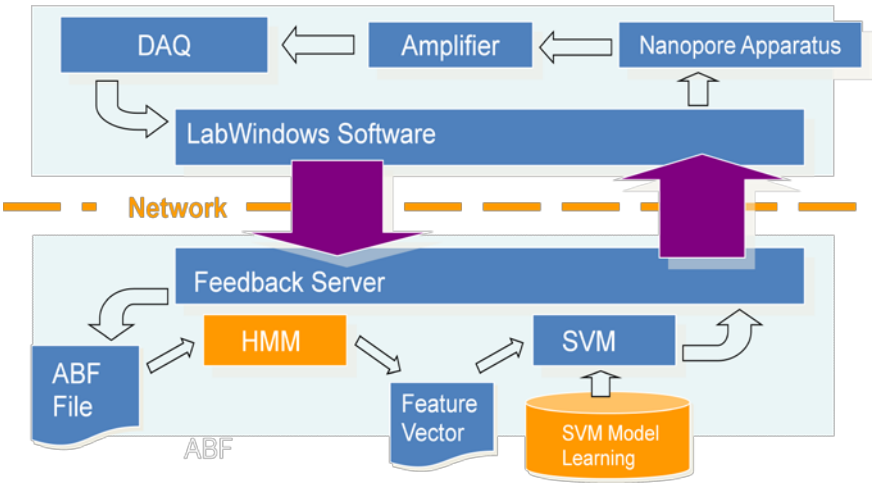

Fig. 6. PRI Sampling Control (see [7] for specific details). Labwindows Feedback Server Architecture with Distributed CCC processing. The HMM learning (on-line) and SVM learning (off-line), denoted in orange, are network distributed for N-fold speed-up, where N is the number of computational threads in the cluster network.

A mixture of two DNA hairpin species (denoted {9TA, 9GC} in [8]) is examined in an experimental test of the PRI system [7]. In separate experiments, data is gathered for the 9TA and 9GC blockades in order to have known examples to train the SVM pattern recognition software. A nanopore experiment is then run with a 1:70 mix of 9GC:9TA, with the goal to eject 9TA signals as soon as they are identified, while keeping the 9GC's for a full 5 seconds (when possible, sometimes a channel-dissociation or melting event can occur in less than that time). The results showing the successful operation of the PRI system is shown in Fig. 7 as a 4D plot, where the radius of the event 'points' corresponds to the duration of the signal blockade (the 4th dimension). The result in Fig. 7 demonstrates an approximately 50-fold speedup on data acquisition of the desired minority species.

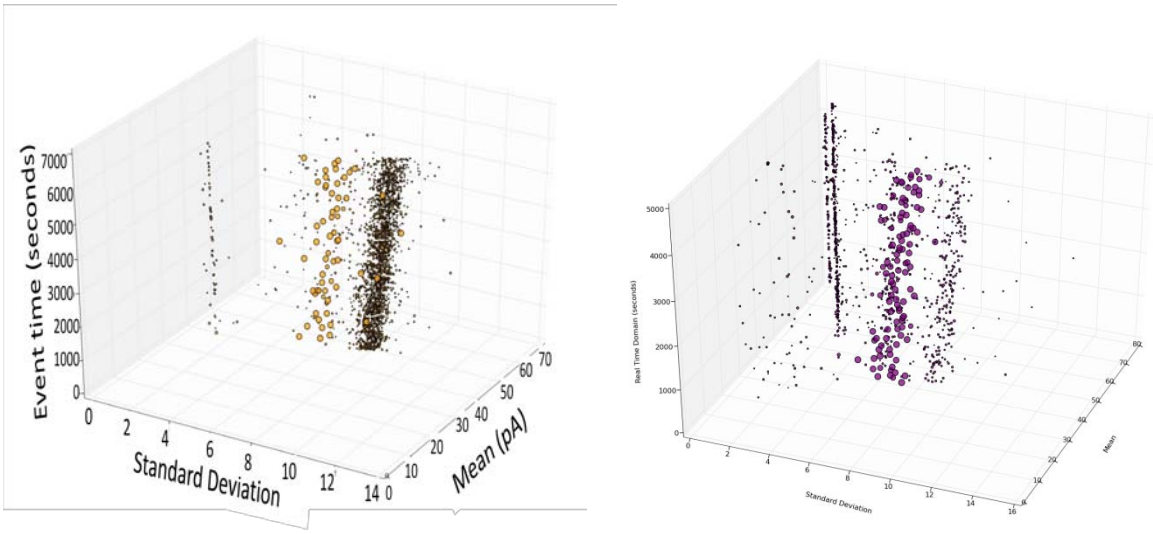

Fig. 7. PRI Mixture Clustering Test with 4D plot [7]. **Left.** The vertical axis is the event observation time, and the plotted points correspond to the standard deviation and mean values for the event observed at the indicated event time. The radius of the points correspond to the duration of the corresponding signal blockade (the 4th dimension). Three blockade clusters appear as the three vertical trajectories. The abundant 9TA events appear as the thick band of small-diameter (short duration, ~100ms) blockade events. The 1:70 rarer 9GC events appear as the band of large-diameter (long duration, ~ 5s) blockade events. The third, very small, blockade class corresponds to blockades that partially thread and almost entirely blockade the channel. **Right. 1:280 rarer events.** Five-species pattern recognition informed sampling: greater than 99.9% accuracy inherited from off-line version

4.3 The SSA Toolset

Although the nanopore transduction detector can be a self-contained 'device' in a lab, external information can be used, for example, to update and broaden the operational information on control molecules ('carrier references'). For the general NTD 'kit' user, carrier reference signals and other systemically-engineered constructs can be used, for example, for a wide range of thin-client arrangements (where they typically have minimal local computational resource). The paradigm for both nanopore device and NTD kit implementations involve system-oriented interactions, where the kit implementation may operate on more of a data service/data repository level and thus need 'real-time' (high bandwidth) system processing of data-service requests or data-analysis requests. Although not as system-dependent on database-server linkages, the more self-contained 'device' implementation will still typically have, for example, local networked (parallelized) data-warehousing, and fast-access, for distributed processing speedup on real-time experimental operations.

For the general-user and the NTD-Kit user, the server interfaces for the SSA Server facilities are being designed to provide services ranging from web-interface based services to an active, streaming, experimental calibration/troubleshooting effort. Nanopore Device and Nanopore Kit users will have the signal acquisition module, the tFSA, bundled with their hardware (on a preconfigured computer, see Table 1). This allows for bare-bones functionality with the bundling as a stand-alone device. When there is need for more refined signal processing, such as when using the weakness recovery protocol to acquire signal when the FSA fails, the SSA Server will have the full panoply of methods available to 'get a lock' on the signal, and examine it further. The transfer of data at this juncture will also be efficiently pre-processed (e.g., the FSA

acquisition will already done) and will thereby allow for much lower bandwidth and latency in the networked signal processing. Biology has invented such structure repeatedly, such as with the retinal pre-processing in the human eye.

KIT-user on-site Software Toolset	**General-user Web-Portal Toolset**
FSA	FSA, FSA-tuning
HMM*	HMM, HMMD, etc.
SVM*	generalized SVM
	Tuning Metaheuristics
	NTD Calibration Services
	NTD Troubleshooting Servies
	CCC Shared Reference Signal Library
(small data for HMM and SVM, no distributed processing)	(large datasets for HMM and SVM training, with use of distributed processing and speciality hardware such as GPUs)

Table 1. Software Toolsets. *specialized enhancement packages to be made available from Meta Logos in future refinements to their Nanoscope effort.

5 Discussion

The nanopore transduction detector is a powerful new tool for single molecule detection. Automation of the detector, with pattern recognition informed feedback, promises to greatly enhance the potential of this device technology. A kit deployment of this technology is being developed by Meta Logos Inc. In prototype experiments a kit appears to be possible, with certain components bought separately (vibration isolation table; patch clamp amplifier, etc) for as little as $10,000 in total. Two elements are missing in a practical deployment of this very new technology: (1) model test systems – in response to this Meta Logos is developing test buffers and test molecules with its kit along the lines of the successful experiments described in [1,3,5] and listed below; and (2) software with both local acquisition tools (see [5]) and off-site server or web-interfaced tool-sets (see [4]), the latter with more elaborate data analysis methods and services being made available via Meta Logos server-linked signal processing with its kit or consulting arrangements (see capability demonstrated in PRI Results shown).

5.1 Model Systems In-Preparation (along lines of experiments described in [3,9])

(1) The streptavidin biosensor model system.
(2) The DNA annealing model system.
(3) The targeted pathogen model system
(4) The SNP minority/majority population detection system
(5) The aptamer-transduced detection of thrombin system
(6) The antibody-transduced detection system using (commoditized) immuno-PCR style tagging with DNA, where complement is engineered to form a Y-transducer, or related, detection system.

6 Conclusion

A promising use for the Nanopore Transduction Detector (NTD) is highly specific SNP detection. Due to the inherent single-molecule nature of this device, and the ability to observe the stationary statistics of single molecule interactions for very long periods, NTD allows for stochastic carrier wave (SCW) signal discrimination boosting, offering very low error-rate SNP detection, using a NTD Nanoscope, in next-generation medical diagnostics and drug discovery.

7 Acknowledgements

The author would like to thank lab technicians Amanda Alba and Eric Morales for help performing the nanopore experiments, and University of New Orleans students Jorge Chao, and Joshua Morrison for help with the nanopore experiments and the channel current cheminformatics analysis. The author would like to thank the University of New Orleans, NIH, NSF, NASA, and the Louisiana Board of Regents for research support. The author would also like to thank META LOGOS Inc., for research support and a research license. (META LOGOS was co-founded by the author in 2009, when it obtained exclusive license to the NTD and machine-learning based signal processing intellectual property.) The author would also like to thank Robert Adelman (CEO META LOGOS, Inc.), Andrew Peck (CEO PxBioSciences), and Mike Lewis (Professor, University of Missouri-Columbia), for insights into the potential impact of the NTD approach.

8 References

[1] Winters-Hilt, S. and R. Adelman. Method and System for Characterizing or Identifying Molecules and Molecular Mixtures. USPTO Filing. Meta Logos Inc. 2010.

[2] Winters-Hilt, S., and Pincus, S. Nanopore-based biosensing. PATENT, UNO filing, 2004.

[3] Winters-Hilt, S. Single nucleotide polymorphism detection via cheminformatics analysis of nanopore-blockades engineered for event transduction. 3rd International Conference on Bioinformatics and Computational Biology (BICoB). March 23 - 25, 2011**.** New Orleans, Louisiana USA.

[4] Winters-Hilt, S., J. Chao, and J. Morrison. Web Access Portals to the Stochastic Sequential Analysis Protocol and Methods. MCBIOS 2011 Proceedings.

[5] Winters-Hilt, S., J. Morrison, and J. Chao. *Ad Hoc* Signal Acquisition and Stochastic Carrier-Wave Signal Processing via FSAs, generalized HMMs, and generalized SVMs. MCBIOS 2011 Proceedings.

[6] Baribault, C and Winters-Hilt S. A novel, fast, HMM-with-Duration implementation -- for application with a new, pattern recognition informed, nanopore detector. BMC Bioinformatics 2007, 8 S7: S19.

[7] Eren AM, Amin I, Alba A, Morales E, Stoyanov A, and Winters-Hilt S. Pattern Recognition Informed Feedback for Nanopore Detector Cheminformatics. Accepted paper in book “Advances in Computational Biology”, Springer: Advances in Experimental Medicine and Biology, June 2010

[8] Winters-Hilt, S., W. Vercoutere, V. S. DeGuzman, D. Deamer, M. Akeson, and D. Haussler, "Highly Accurate Classification of Watson-Crick Base-Pairs on Termini of Single DNA Molecules," *Biophys. J.* Vol. 84, pg 967, 2003.

[9] Winters-Hilt, S., E. Horton-Chao, and E. Morales. The NTD Nanoscope. MCBIOS 2011 Proceedings.