

RESEARCH ARTICLE | JANUARY 11 2022

SERS-based ssDNA composition analysis with inhomogeneous peak broadening and reservoir computing

Phuong H. L. Nguyen ; Shimon Rubin; Pulak Sarangi; Piya Pal; Yeshaiahu Fainman 

 Check for updates

Appl. Phys. Lett. 120, 023701 (2022)

<https://doi.org/10.1063/5.0075528>

 CHORUS



View Online



Export Citation

CrossMark

Articles You May Be Interested In

An improved DNA force field for ssDNA interactions with gold nanoparticles

J. Chem. Phys. (June 2014)

Simulations of single grafted polyelectrolyte chains: ssDNA and dsDNA

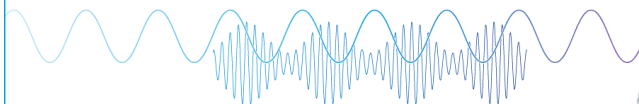
J. Chem. Phys. (February 2003)

In situ synthesis and direct immobilization of ssDNA on electron beam patterned hydrogen silsesquioxane

Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena (December 2009)

Webinar

Boost Your Signal-to-Noise Ratio with Lock-in Detection



Sep. 7th – Register now



Zurich Instruments

SERS-based ssDNA composition analysis with inhomogeneous peak broadening and reservoir computing

Cite as: Appl. Phys. Lett. **120**, 023701 (2022); doi: [10.1063/5.0075528](https://doi.org/10.1063/5.0075528)

Submitted: 16 October 2021 · Accepted: 30 December 2021 ·

Published Online: 11 January 2022



View Online



Export Citation



CrossMark

Phuong H. L. Nguyen,  Shimon Rubin, Pulak Sarangi, Piya Pal, and Yeshaiahu Fainman^{a)}

AFFILIATIONS

Electrical and Computer Engineering, University of California San Diego, La Jolla, California 92161, USA

^{a)} Author to whom correspondence should be addressed: fainman@ece.ucsd.edu

ABSTRACT

Surface-enhanced Raman spectroscopy employed in conjunction with post-processing machine learning methods is a promising technique for effective data analysis, allowing one to enhance the molecular and chemical composition analysis of information rich DNA molecules. In this work, we report on a room temperature inhomogeneous broadening as a function of the increased adenine concentration and employ this feature to develop one-dimensional and two dimensional chemical composition classification models of 200 long single stranded DNA sequences. Afterwards, we develop a reservoir computing chemical composition classification scheme of the same molecules and demonstrate enhanced performance that does not rely on manual feature identification.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0075528>

Surface-enhanced Raman spectroscopy (SERS) is a well-established nondestructive and label-free sensing technique, which admits high specificity inherent to Raman spectroscopy, as well as high sensitivity due to electromagnetic (EM)¹ and chemical enhancement (CE)^{2,3} mechanisms, and is, therefore, prospective to address various biological and medical needs.^{4–8} In particular, SERS-based classification of information rich DNA molecules⁹ based on their chemical composition holds promise for future applications in medical diagnosis^{10,11} and bio-analysis.^{12,13} However, the complexity of the signal and its low amplitude often limits applicability of SERS to meet modern clinical/medical¹⁴ and chemical applications,¹⁵ thus stimulating research directions uncovering additional features in SERS spectra such as power peak ratios of different dominant peaks^{16,17} and Raman peak broadening effects.^{18,19} In particular, the homogeneous (internal) broadening originates due to internal molecular properties, whereas the inhomogeneous (external) broadening stems from the perturbation imposed on the relevant molecule by the environment,²⁰ thus making the inhomogeneous broadening as a highly relevant feature when studying the interaction between the SERS active single stranded DNA (ssDNA) molecule and the adsorbing metal substrate. An independent and additional information processing based approach, which allows it to improve the accuracy of SERS, and Raman methods including various machine learning (ML)-based techniques of different

elaboration levels starting from linear regression models, which are based on manual extraction of key features in the spectra, through principle components analysis²¹ and neuromorphic-based schemes, to more advanced feedforward neural networks (FNNs) and deep learning techniques,^{22,23} have been successfully employed to achieve superior performance. Among different ML-based methods, reservoir computing (RC)^{24–26} that emerged as a subset of the recurrent-neural network (RNN) paradigm introduces a computationally plausible possibility to keep the internal weights fixed and tune only the output weights. It was demonstrated that this simplification allowed to bypass significant computational efforts needed to train the internal weights of RNNs, but at the same time still enhanced computational stability and in some cases even provided higher performance compared to RNNs,²⁷ leading to various applications including decision making,²⁸ dynamical systems control,^{27,29} classification,³⁰ etc. With those advantages in mind and the discussion above, it is reasonable to assume that RC is a promising method to be implemented on the feature-rich SERS spectra, and one can expect that future advancement of SERS and Raman sensing technology will rely on both identification of additional features in the corresponding signal spectra as well as development and implementation of dedicated ML-based methods.

In this Letter, we report an inhomogeneous peak broadening in the SERS signal at room temperature of the adenine specific mode p_A

(680–770 cm^{-1}) in ssDNA molecules. The width of the adenine peak is composed of homogeneous broadening due to temperature driven fluctuation, which is not expected to vary among different experiments, and an inhomogeneous broadening which increases with the number of adenine bases in ssDNA molecules. Therefore, effectively, only the inhomogeneous peak broadening effect is the dominant contributing factor to the observed broadening effect. While previous works reported the inhomogeneous Raman/SERS peak broadening as a function of temperature,¹⁸ the phase change transition from solid to liquid,^{31,32} or at cryogenic temperatures,^{18,19} our results indicate broadening at room temperature as a function of number of adenine bases in the molecule. We then employ this feature to build one- and two-dimensional (1D and 2D) regression models for ssDNA chemical composition analysis, performing the classification task depending on the number of adenine bases in the molecule. In particular, we employ the broadening feature in the 1D linear regression model (as illustrated in Fig. 1) and further combine it with the peak ratio feature¹⁷ in the 2D regression model. The two prominent peaks that are used to form the ratio feature are p_A and phosphate backbone-related mode p_{BB} (1050–1150 cm^{-1}) and described in Fig. S1 in Appendix A of the [supplementary material](#). We then implement a more general RC method

that does not depend on manual feature identification to demonstrate the enhanced performance compared to FNNs in terms of computation time and accuracy.

For our experiment, the SERS spectra are collected by using a Renishaw inVia Raman spectrometer. The settings include an excitation wavelength of 785 nm, a laser power of 50 mW, an acquisition time of 5 s, and one accumulation per spectrum. The objective magnification is $50\times$ with $\text{NA} = 0.75$. The grating type used is 1200/mm at 785 nm. The grating setting in the built-in spectrometer software is set to a static regime with acquired spectrum range extending between 600 and 1700 cm^{-1} . The resultant spectral resolution in our setup is approximately 1 cm^{-1} . In order to prevent the nanorod SERS substrate (see sample fabrication details in our previous work²²) from excessive oxidation, which would affect the signal to noise ratio (SNR), we made the substrates fresh the night before measurement, drop-cast the ssDNA solution on them to dry overnight, and measured all-samples within several minutes. In so doing, we also minimized potential differences due to oxidation-driven effects across all samples. Figure 1(a) presents average SERS spectra taken over 1000 clean measurements of 200-base long ssDNA molecules bonded to a silver nanorod surface capturing the adenine peak for different adenine molecular

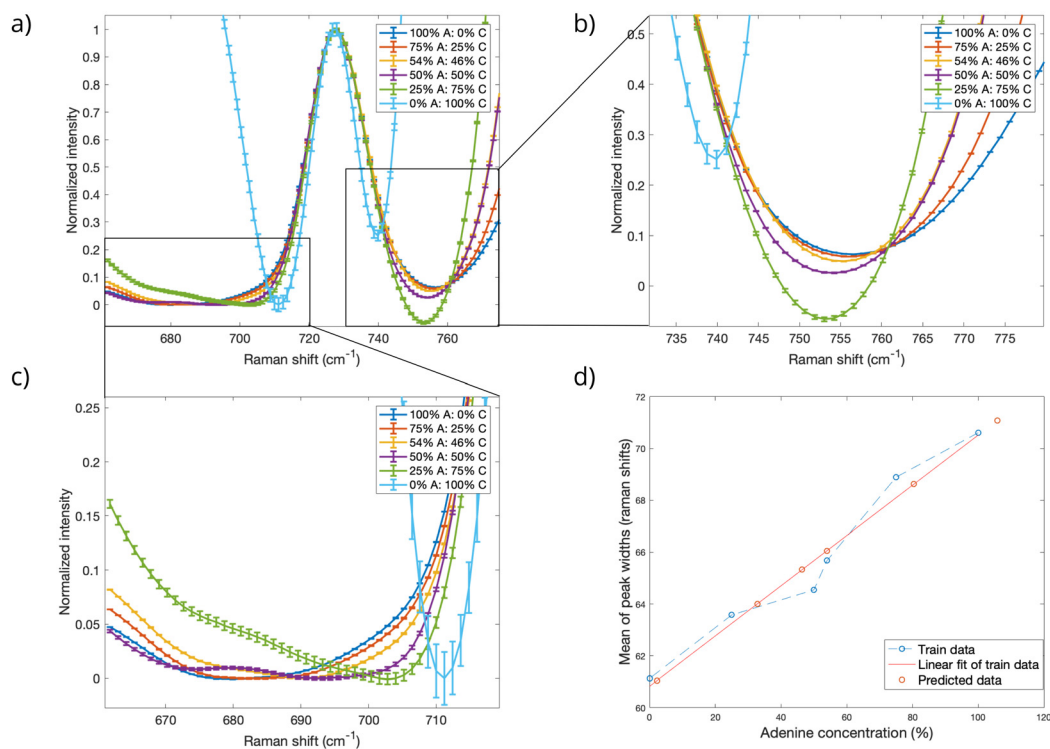


FIG. 1. (a) Experimental results of SERS spectra presenting the adenine related peak for different adenine concentrations in ssDNA molecules; we consider the following six cases: 0%, 25%, 50%, 54%, 75%, and 100%, where the concentration is defined as the number of adenine bases in ssDNA molecule divided by the total number of bases (200). All peaks are normalized and shifted so their Raman shifts are aligned to enhance the visibility of the peak broadening. The value of the adenine peak of the ssDNA molecule with 0% adenine concentration is only a fraction of ~ 25 compared to the value of the adenine peak that corresponds to 100% adenine concentration; its non-zero value at 0% signals that the signal stems from the presence of other molecular components in ssDNA molecules and is not related to the presence of adenine bases. Consequently, in terms of the linear regression model, the adenine peak value for the 0% case serves as a constant bias for all cases. (b) and (c) Magnified side section of the peak denoted by black rectangles in (a). (d) 1D linear regression model based on the peak broadening effect and the predicted test data. The horizontal and the vertical axes represent the adenine concentration (A) and average peak width (w), respectively, whereas the corresponding linear fit for the peak width is given by $w(A) = 0.097 \times A + 60.89$.

composition percentages, whereas Fig. 1(d) presents the corresponding inhomogeneous broadening as the adenine concentration, i.e., the number of adenine bases in the ssDNA molecule divided by the total number of bases, grows from 0% to 100%. While previous works employed increasing maximal power of the adenine peak as a feature for ssDNA classification,^{16,17} the normalized data presented in Figs. 1(a) and 1(b) clearly indicate that the width of the peak is an increasing function of the adenine concentration, albeit less pronounced than maximal power change. Note that the normalization, i.e., dividing the adenine peak in each spectrum by its maximal value, is performed for visualization purposes only in order to bring all peaks to the same unity height, thus eliminating peak amplitude modulation across different cases. In order to generate and train our models more effectively, first our input data are cleaned from possible outliers including those occurred due to unwanted systematic noise, cosmic ray, or impurities (dust) on the sample's surface using the Robust principal component analysis (PCA) technique. This technique works by finding the low-rank and sparse high-rank (noise) components of the collected raw data (see Appendix B of the [supplementary material](#)). SERS spectra with higher values of the sparse components are identified as outliers and eliminated from our data set before further processing. Appendix C of the [supplementary material](#) shows in detail how the data cleaning process happens in our work, where Fig. S3 illustrates how the inlier and outlier measurements look like. For the ease of interpretation, all the peaks here have been normalized to the same height and shifted to the same origin position on the Raman shift axis, whereas the peak width is quantified as a distance between the successive minima points around the corresponding peak [as shown in Figs. 1(b) and 1(c)]. While the different curves present minor changes near the peak, it shows more pronounced differences near the minima points, which appear in all five sets of measured data taken on different days, as illustrated in Fig. S4 in Appendix D of the [supplementary material](#). To employ this feature for the 1D linear regression classification model, we combine all data sets, then use two thirds as training data to derive the corresponding linear fit model [see Fig. 1(b)], and finally use it to predict the adenine percentage in the remaining third test data. We repeated the process of dividing the total data set into training and testing and performed model learning and classifying for five times. Table I second column presents the corresponding average root-mean-squared error (RMSE) of the predicted adenine composition for five different cases for this 1D model. For our previous work, we have utilized the 1D linear regression model for the peak ratio feature, stating an RMSE of 5.17% for gold and 5.46% for silver.²² In order to improve the detection efficiency in this work, we increase the number of features used in the regression model and construct a 2D model that employs the peak broadening feature together with the peak ratio feature. In particular, since the natural logarithm of the adenine-related peak p_A and the backbone-related peak p_{BB} ratio admits a linear dependence on the number of adenine bases in ssDNA molecules,^{17,22} we combine this quantity with adenine peak widths leading to the following quadratic fit:

$$A(w, r) = -558.22 + 33.09 \cdot w - 1848.78 \cdot r + 29.9 \cdot w \cdot r - 0.375 \cdot w^2 - 45.56 \cdot r^2. \quad (1)$$

Here, A is the adenine concentration in ssDNA molecules, whereas r and w are the logarithmic ratio of peak ratios and the peak width, respectively; the fit is illustrated in Fig. 2. The corresponding RMSE

TABLE I. RMSE values comparing the performance of 1D, 2D, and RC models of ssDNA molecule classification based on the number of adenine bases for five different data partitions into training and test.

Partition #	1D RMSE (%)	2D RMSE (%)	RC RMSE (%)
1	6.69	3.05	0.229
2	5.39	2.81	0.341
3	6.85	1.02	0.511
4	4.87	5.69	0.829
5	5.75	5.25	0.279
Average	5.91	3.56	0.438

for five different data partitions into training and test data is presented in Table I third column, indicating enhancement in the classification efficiency of the 2D model compared to the 1D model from 94% to 96.5%. It is worth mentioning that the quadratic model leads to superior performance when compared to the linear peak amplitude-based linear regression model (see Table II in Ref. 22).

In the second part of this work, we utilize the dedicated RC scheme, which allows us to employ the whole SERS spectra without manually identifying specific features (as opposed to the linear regression approach above) and furthermore allows us to demonstrate the superior performance compared to our previous work where we used the FNN.²² Figure 3 presents the schematic description of the proposed RC architecture, where the $N = 1021$ dimensional SERS spectra of a given measurement in the range between 608 and 1721 cm^{-1} , denoted as $u(n)$ where n is the positive integer, are concatenated together with the bias parameter, b , forming together a 1022 dimensional input vector. This input vector is then mapped by a fixed random matrix W_{in} into the reservoir space of dimension N , where each neuron $x(n)$ is subject to the following evolution equation:

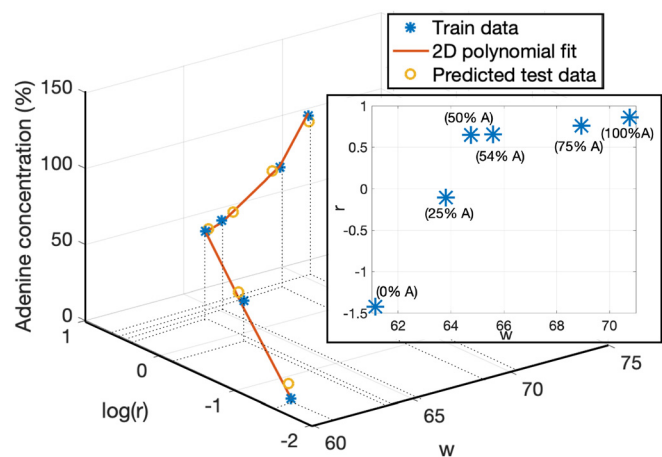


FIG. 2. 2D quadratic fit between the experimentally acquired peak width (w) and peak ratio ($r = p_A/p_{BB}$) to the corresponding adenine concentration (A). The quadratic fit is described in Eq. (1), and the inset presents projection on the w - $\log(r)$ plane indicating that w and $\log(r)$ are both a growing function of the adenine concentration.

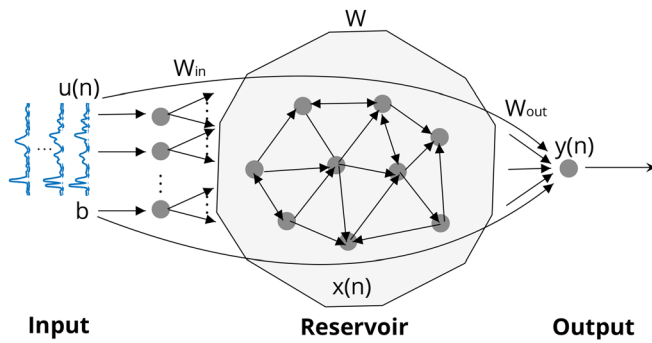


FIG. 3. Schematic description of the proposed RC architecture for SERS spectra chemical composition analysis/classification. The $N=1021$ dimensional SERS spectra vector serves as an input state and generates a reservoir state $x(n)$ of dimension $N \times 1$, subject to Eq. (2) with $a=1$, $b=1$ and random fixed matrices W_{in} and W of dimensions $N \times (N+1)$ and $N \times N$, respectively. Note that in the employed RC architecture both the input state and the reservoir state are mapped to the output layer, reflecting Eqs. (3) and (5).

$$x(n) = (1 - a)x(n - 1) + a \tanh \left(W_{in} \begin{pmatrix} b \\ u(n) \end{pmatrix} + Wx(n - 1) \right). \quad (2)$$

Here, W is a fixed $N \times N$ matrix from the reservoir space to itself, b is a constant (i.e., n -independent) bias value, and a is the so-called leaking rate. In our RC analysis, we employ a total of $K=6000$ SERS spectra corresponding to six different adenine concentrations in the ssDNA molecule (each concentration consists of 1000 SERS spectra) with each spectra serving as an input state $u(n)$, $n=1, \dots, K$. In particular, we employ two thirds of the spectra for training and one third for testing, corresponding to $2/3 \times 1000 \times 6 \simeq 4002$ and $1/3 \times 1000 \times 6 \simeq 1998$ spectra, respectively, and employ five arbitrary partitions of the data into training and testing to quantify heterogeneity of our dataset. To train the data, first we notice that the output of the RC model is given by

$$Y_{train} = W_{out} \cdot X_{train}, \quad (3a)$$

$$X_{train}^T = ((b, u(1), x(1)); \dots; (b, u(k), x(k))), \quad (3b)$$

where X_{train}^T is the $k \times (2N+1)$ dimensional matrix and $k=4002$ is the number of training measurements. The components of the $1 \times (2N+1)$ dimensional output matrix W_{out} are determined by employing ridge regression, leading to

$$W_{out} = Y_{train} X_{train}^T (X_{train} X_{train}^T + \beta I)^{-1}, \quad (4)$$

where β is the corresponding regularization parameter. After the training stage, the corresponding W_{out} then operates on the $(2N+1) \times l$ dimensional test matrix X_{test} where l is the number of input test spectra ($l=1998$), thus leading to the following expression for Y_{test} predicting the number of adenine bases in the ssDNA molecule

$$Y_{test} = W_{out} \cdot X_{test}, \quad (5a)$$

$$X_{test}^T = ((b, u(1), x(1)); \dots; (b, u(l), x(l))). \quad (5b)$$

Here, Y_{test} is a list of predicted adenine concentrations for the test spectra. Each predicted value in the list is depicted by $y(n)$, which

corresponds to each test input $u(n)$ (see Fig. 3). Our analysis indicates that the leaking rate value $a=1$ and reservoir dimensionality $N=1021$ are optimal in terms of the RMSE value (see Fig. S5 in the supplementary material).

In order to optimize our model efficiency, we use cross validation (CV) in our training process. CV is a technique where the entire data set can be divided into n subsets. Out of which $(n-1)$ subsets will be used in training, and one subset will be used for validation of the effectiveness of the model. The number of subsets is determined by the number of folds (number of times we repeat the process of division into subsets). We choose $n=5$ in our simulation, which means the division, training, and validating process will be repeated for five times before we pick out the process in which the model gives out the lowest RMSE. Furthermore, we rerun this CV process for five times, and each time we pick out a model with the lowest RMSE to be our best model. Table I fourth column lists the best RMSE values for five runs and the average RMSE of the whole process of this RC model. The classified (predicted) results are plotted against their ground truth values in Fig. 4 and almost completely overlap on the presented scale due to the small average error value presented in the last column in Table I. With a classification efficiency of about 99.5%, the RC model proves to be superior in accuracy to other models we employed so far such as linear regression, PCA linear regression, or FNNs.²²

To summarize, in this work, we reported three methods of different computational complexity to perform the SERS-based chemical composition analysis of 200 base long ssDNA molecules. The first linear model was implemented by employing the peak broadening effect in ssDNA molecules with the increased width of the adenine peak as a function of the increased molecular adenine concentration serving as only feature. In the second model, we employed the peak broadening effect with the previously reported peak power feature of the adenine peak to construct a quadratic model with enhanced performance. Finally, we constructed a dedicated RC computational platform capable of outperforming the first two models above and also our previously employed FNN model for the chemical composition model.²²

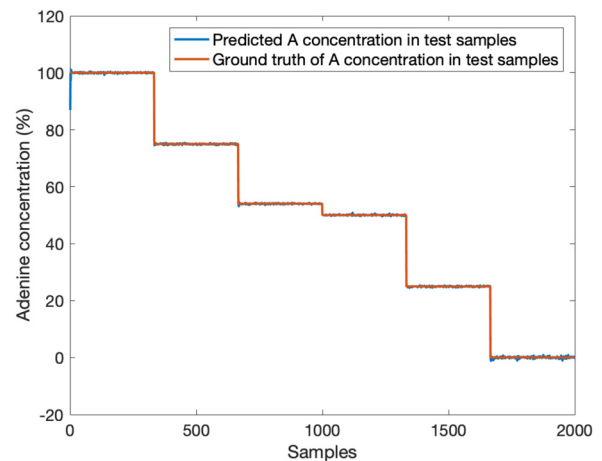


FIG. 4. Comparison between the RC-predicted concentration of adenine bases in ssDNA molecules and the actual value. Note that the small RMSE error 0.438% specified in Table I fourth column above leads to practical overlap of the two curves. For each type of the ssDNA molecule, we employed 2000/6 ~ 333 test samples.

both in training time and accuracy. In this work, we use adenine and cytosine DNA bases, because these admit non-intersecting peaks and, thus, are separable in the spectral domain. Nevertheless, other DNA bases may admit interfering peaks,³³ thus masking potential broadening effects unless the peaks are separated by modifying substrate properties (e.g., metal properties).

We believe that the observations and the methods reported in our work will stimulate future research of pure scientific directions aiming to enhance our understanding of both basic SERS effects and development of dedicated ML algorithms for SERS data analysis, as well as applicative directions, where employing additional features in ML methods will allow more efficient optical-based composition and chemical analysis of complex molecules. In particular, from a purely scientific perspective, the inhomogeneous broadening effect reported above provides an example where the CE effect in SERS enables to observe room-temperature broadening, opening another research direction aiming to map the factors that affect the intricate underlying interaction mechanisms between complex molecules and SERS substrates. While at this point, we can only hypothesize that the peak broadening for the adenine specific peak occurs due to increased number of binding configurations of adenine to the metal,^{33,34} each potentially admitting a different spatial orientation and the CE effect and leading, in turn, to degeneracy lifting of normal mode frequencies³⁵ and to higher spectra variance for sufficiently large number of adenine bases in the molecule; future research may lead to more detailed understanding. Another question our study raises from a materials science perspective: is it possible to control substrate properties in order to allow more binding configurations for a given base, thus selectively introducing an inhomogeneous broadening feature for some DNA bases and not for other bases for composition analysis scheme? Potentially, it may also allow the use of computationally inexpensive linear regression schemes to achieve composition analysis in cases where peak ratio feature is not well expressed. Another question is how the liquid environment, presenting ionic and molecular substances, affects the peak broadening of adsorbed molecules. From an application perspective, observation of the inhomogeneous peak broadening effect of DNA bases at room temperature may allow a variety of composition analysis technological applications without the need to employ energy expensive and bulky cooling systems. Furthermore, as we demonstrated within the framework of the quadratic model, implementing several features allows us to implement a computationally inexpensive regression model of enhanced accuracy compared to a single feature linear model. From the ML perspective, our results indicate that employing a dedicated RC method for molecular composition analysis opens a way to employ computationally feasible neural network schemes supporting faster data processing of information rich SERS spectra without compromising on the corresponding accuracy, thus serving as a stepping stone for future applications, where RC and other neuromorphic-based models allow reduced processing time and, hence, faster acquisition rate.

See the [supplementary material](#) for the full SERS spectra of the ssDNA adsorbed on silver nanorod substrate, further analysis of the data cleaning process using robust PCA, the repeatability of the peak broadening effect, choosing the optimum reservoir size, and the calculation of RMSE in this Letter.

This work was supported by the Defense Advanced Research Projects Agency (DARPA) DSO NAC Programs, the Office of Naval Research (ONR), the National Science Foundation (NSF) via Grant Nos. CBET-1704085, NSF ECCS-190184, and NSF ECCS-2023730, the San Diego Nanotechnology Infrastructure (SDNI) supported by the NSF National Nanotechnology Coordinated Infrastructure (Grant No. ECCS-2025752), the Quantum Materials for Energy Efficient Neuromorphic Computing—an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE) Office of Science, Basic Energy Sciences under Award No. DE-SC0019273, and the Cymer Corporation.

This work was also supported in part by the Office of Naval Research under Grant No. ONR N000141912256 and in part by the National Science Foundation under Grant No. NSF CAREER ECCS 1700506.

The authors cordially thank Prabhav Gaur for useful discussion and comments about Reservoir Computing.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹Surface-Enhanced Raman Scattering: Physics and Applications, Topics in Applied Physics, edited by K. Kneipp, M. Moskovits, and H. Kneipp (Springer, Berlin/New York, 2006), Vol. 103.
- ²M. Moskovits, "Surface-enhanced spectroscopy," *Rev. Mod. Phys.* **57**, 783–826 (1985).
- ³A. Campion, J. E. Ivaneky, C. M. Child, and M. Foster, "On the mechanism of chemical enhancement in surface-enhanced Raman scattering," *J. Am. Chem. Soc.* **117**, 11807–11808 (1995).
- ⁴J. R. Lombardi and R. L. Birke, "A unified approach to surface-enhanced Raman spectroscopy," *J. Phys. Chem. C* **112**, 5605–5617 (2008).
- ⁵B. Sharma, R. R. Frontiera, A.-I. Henry, E. Ringe, and R. P. Van Duyne, "SERS: Materials, applications, and the future," *Mater. Today* **15**, 16–25 (2012).
- ⁶S. Arsalani, T. Ghodselahi, T. Neishaboorynejad, and O. Baffa, "DNA detection based on localized surface plasmon resonance spectroscopy of Ag@Au biocomposite nanoparticles," *Plasmonics* **14**, 1419 (2019).
- ⁷W. H. Kim, J. U. Lee, S. Song, S. Kim, Y. J. Choi, and S. J. Sim, "A label-free, ultra-highly sensitive and multiplexed SERS nanoplasmonic biosensor for miRNA detection using a head-flocked gold nanopillar," *Analyst* **144**, 1768–1776 (2019).
- ⁸G. Braun, S. J. Lee, M. Dante, T.-Q. Nguyen, M. Moskovits, and N. Reich, "Surface-enhanced Raman spectroscopy for DNA detection by nanoparticle assembly onto smooth metal films," *J. Am. Chem. Soc.* **129**, 6378–6379 (2007).
- ⁹S. Khan, "DeepAcid: Classification of macromolecule type based on sequences of amino acids," arXiv preprint [arXiv:1907.03532](#) (2019).
- ¹⁰D. Capper, D. Jones, M. Sill *et al.*, "DNA methylation-based classification of central nervous system tumours," *Nature* **555**, 469–474 (2018).
- ¹¹P. S. Mischel, T. F. Cloughesy, and S. F. Nelson, "DNA-microarray analysis of brain cancer: Molecular classification for therapy," *Nat. Rev. Neurosci.* **5**, 782–792 (2004).
- ¹²Z. Wang, L. Gao, X. Guo, W. Lian, K. Deng, and B. Xing, "Development and validation of a novel DNA methylation-driven gene based molecular classification and predictive model for overall survival and immunotherapy response in patients with glioblastoma: A multiomic analysis," *Front. Cell Dev. Biol.* **8**, 576996 (2020).
- ¹³R. A. Young, "Biomedical discovery with DNA arrays," *Cell* **102**, 9–15 (2000).

- ¹⁴E. Pyrak, J. Krajczewski, A. Kowalik, A. Kudelski, and A. Jaworska, "Surface enhanced Raman spectroscopy for DNA biosensors—How far are we?," *Molecules* **24**, 4423 (2019).
- ¹⁵M. Fan, G. F. Andrade, and A. G. Brolo, "A review on recent advances in the applications of surface-enhanced Raman scattering in analytical chemistry," *Anal. Chim. Acta* **1097**, 1–29 (2020).
- ¹⁶L.-J. Xu, Z.-C. Lei, J. Li, C. Zong, C. J. Yang, and B. Ren, "Label-free surface-enhanced Raman spectroscopy detection of DNA with single-base sensitivity," *J. Am. Chem. Soc.* **137**, 5149–5154 (2015).
- ¹⁷L. M. Freeman, L. Pang, and Y. Fainman, "Self-reference and random sampling approach for label-free identification of DNA composition using plasmonic nanomaterials," *Sci. Rep.* **8**, 7398 (2018).
- ¹⁸C. Artur, E. C. Le Ru, and P. G. Etchegoin, "Temperature dependence of the homogeneous broadening of resonant Raman peaks measured by single-molecule surface-enhanced Raman spectroscopy," *J. Phys. Chem. Lett.* **2**, 3002–3005 (2011).
- ¹⁹P. G. Etchegoin and E. C. Le Ru, "Resolving single molecules in surface-enhanced Raman scattering within the inhomogeneous broadening of Raman peaks," *Anal. Chem.* **82**, 2888–2892 (2010).
- ²⁰W. Demtroder, *Laser Spectroscopy*, 4th ed. (Springer, 2008).
- ²¹D. Bharanidharan and N. Gautham, "Principal component analysis of DNA oligonucleotide structural data," *Biochem. Biophys. Res. Commun.* **340**, 1229–1237 (2006).
- ²²P. H. L. Nguyen, B. Hong, S. Rubin, and Y. Fainman, "Machine learning for composition analysis of ssDNA using chemical enhancement in SERS," *Biomed. Opt. Express* **11**, 5092 (2020).
- ²³C.-S. Ho, N. Jean, C. A. Hogan, L. Blackmon, S. S. Jeffrey, M. Holodniy, N. Banaei, A. A. E. Saleh, S. Ermon, and J. Dionne, "Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning," *Nat. Commun.* **10**, 4927 (2019).
- ²⁴H. Jaeger, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science* **304**, 78–80 (2004).
- ²⁵W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Comput.* **14**, 2531–2560 (2002).
- ²⁶D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Networks* **20**, 391–403 (2007).
- ²⁷M. Lukosevicius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.* **3**, 127–149 (2009).
- ²⁸G. Trierweiler Ribeiro, J. Guilherme Sauer, N. Fraccanabbia, V. Cocco Mariani, and L. dos Santos Coelho, "Bayesian optimized echo state network applied to short-term load forecasting," *Energies* **13**, 2390 (2020).
- ²⁹P. Tiño and A. Rodan, "Short term memory in input-driven linear dynamical systems," *Neurocomputing* **112**, 58–63 (2013).
- ³⁰M. Skowronski and J. Harris, "Minimum mean squared error time series classification using an echo state network prediction model," in *2006 IEEE International Symposium on Circuits and Systems* (IEEE, Island of Kos, Greece, 2006), pp. 3153–3156.
- ³¹A. K. Arora, M. Rajalakshmi, T. R. Ravindran, and V. Sivasubramanian, "Raman spectroscopy of optical phonon confinement in nanostructured materials," *J. Raman Spectrosc.* **38**, 604–617 (2007).
- ³²Y. Gao and P. Yin, "Origin of asymmetric broadening of Raman peak profiles in Si nanocrystals," *Sci. Rep.* **7**, 43602 (2017).
- ³³L. M. Freeman, L. Pang, and Y. Fainman, "Maximizing the electromagnetic and chemical resonances of surface-enhanced Raman scattering for nucleic acids," *ACS Nano* **8**, 8383–8391 (2014).
- ³⁴S. G. Harroun, "The controversial orientation of adenine on gold and silver," *ChemPhysChem* **19**, 1003–1015 (2018).
- ³⁵S. Rubin, P. H. L. Nguyen, and Y. Fainman, "The effect of DNA bases permutation on surface-enhanced Raman scattering spectrum," *Nanophotonics* **10**, 1581–1593 (2021).