

Machine-Learning Identification of the Sensing Descriptors Relevant in Molecular Interactions with Metal Nanoparticle-Decorated Nanotube Field-Effect Transistors

Long Bian,[†] Dan C. Sorescu,^{‡,§} Lucy Chen,[†] David L. White,[†] Seth C. Burkert,[†] Yassin Khalifa,^{||} Zhenwei Zhang,^{||} Ervin Sejdic,^{||} and Alexander Star^{*,†,¶}

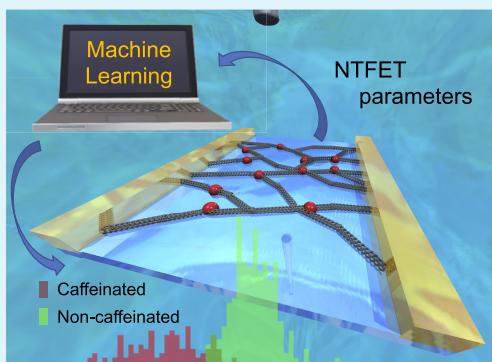
[†]Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

[‡]United States Department of Energy, National Energy Technology Laboratory, Pittsburgh, Pennsylvania 15236, United States

[§]Department of Chemical and Petroleum Engineering and ^{||}Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States

Supporting Information

ABSTRACT: Carbon nanotube-based field-effect transistors (NTFETs) are ideal sensor devices as they provide rich information regarding carbon nanotube interactions with target analytes and have potential for miniaturization in diverse applications in medical, safety, environmental, and energy sectors. Herein, we investigate chemical detection with cross-sensitive NTFETs sensor arrays comprised of metal nanoparticle-decorated single-walled carbon nanotubes (SWCNTs). By combining analysis of NTFET device characteristics with supervised machine-learning algorithms, we have successfully discriminated among five selected purine compounds, adenine, guanine, xanthine, uric acid, and caffeine. Interactions of purine compounds with metal nanoparticle-decorated SWCNTs were corroborated by density functional theory calculations. Furthermore, by testing a variety of prepared as well as commercial solutions with and without caffeine, our approach accurately discerns the presence of caffeine in 95% of the samples with 48 features using a linear discriminant analysis and in 93.4% of the samples with only 11 features when using a support vector machine analysis. We also performed recursive feature elimination and identified three NTFET parameters, transconductance, threshold voltage, and minimum conductance, as the most crucial features to analyte prediction accuracy.



KEYWORDS: carbon nanotubes, graphene, linear discriminant analysis, sensor arrays, support vector machine

INTRODUCTION

Sensor array approaches utilizing multivariate data analysis^{1–10} and machine-learning¹¹ techniques to classify different analytes have been emerging for biochemical and analytical applications. Materials such as metal–organic frameworks,^{12–16} imprinted polymers,^{17–19} peptides,⁴ carbon nanomaterials,⁸ and quantum dots⁹ have been widely investigated for incorporation into sensor arrays. For example, Anslyn and co-workers developed a cross-sensitive peptide-based ternary sensing array as a type of indicator displacement assay.⁴ Various types of wood used to age cachaça were distinguished and identified by processing via linear discriminant analysis (LDA) the changes in the UV spectrum from different tannin compositions interacting with the assay. In another example, Li and co-workers⁸ employed chemiresistive graphene sensors to discriminate and classify the structurally and chemically similar volatile organic compounds toluene, *o*-xylene, *p*-xylene, and mesitylene. Leveraging feature selection process and principal component analysis, their approach improved the prediction accuracy to 92%.⁸ In a recent study, Parkin and co-workers⁹

utilized a surface functionalized with a fluorescent quantum dots array to detect and differentiate among five different explosives. The fingerprints from the differential quenching of quantum dot–analyte interaction were analyzed by LDA and support vector machine (SVM), achieving up to 100% classification accuracy.

Carbon nanotubes are highly sensitive to changes in the local environment and serve as an ideal platform for the fabrication of label-free nanoelectronic sensors. To date, various analytes have been detected with carbon nanotube-based chemiresistor or field-effect transistor (NTFET) devices used as electrical transducers when suitable molecular recognition layers were employed.^{20–26} The characteristic NTFET curves, i.e., the dependence of the source–drain conductance on the applied gate voltage ($G - V_G$), contain a wealth of information as compared to resistive measurements

Received: September 11, 2018

Accepted: December 14, 2018

Published: December 14, 2018



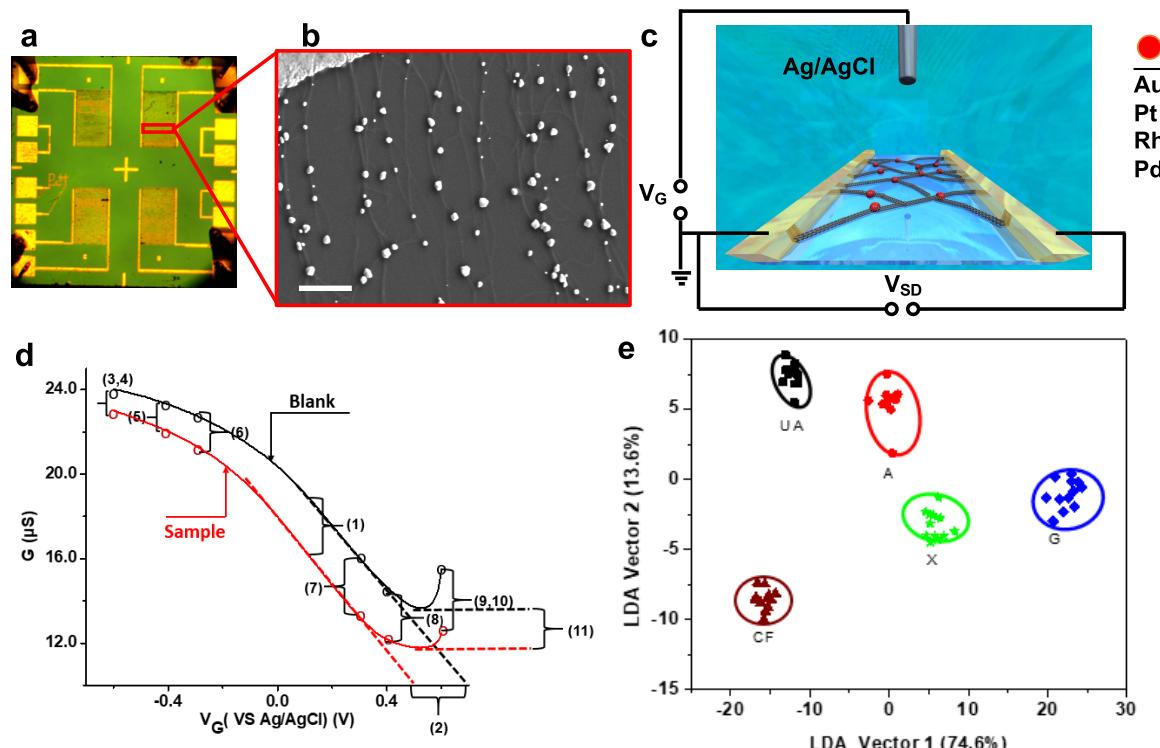


Figure 1. Device schematics and operational principle. (a) Optical microscope image of the Si chip with four gold interdigitated electrodes (IDEs). (b) Scanning electron microscope (SEM) image of a network of SWCNTs spanning the fingers of IDE after electrodeposition of Au NPs. Scale bar is 1 μm . (c) Schematic diagram of the liquid-gated carbon nanotube field-effect transistor (NTFET). (d) Transfer characteristics of the platinum nanoparticle-decorated NTFET in blank and in caffeine (1 mM) solutions. Selected features (11) were calculated from the NTFET curves in blank and for the sample solutions. (e) LDA plots for purine compounds discrimination.

and have proved useful for analyzing the mechanism of interaction with target analytes.²⁷ Integrating the NTFET data from the interaction of chemically distinct sensor surfaces to each analyte provides a novel avenue for deeper analysis, which in turn enables the detection of analytes without unique designs of molecular recognition layers. For example, we recently employed a series of metal nanoparticle-functionalized NTFET devices as cross-sensitive sensor arrays, demonstrating electronic “tongue” functions and ability to analyze different cell types.²⁸ We selected 11 features from each $G - V_G$ curve based on the NTFET biosensing mechanisms to form a data matrix, which was further analyzed using LDA. On the basis of this approach, we were able to successfully discriminate between tissues and malignant and nonmalignant cells.²⁸ Instead of examining individual parameters of NTFET and trying to correlate their changes to the chemical nature of the analytes, the methodology developed in our previous work²⁸ takes a more holistic approach in analyzing multiple responses (i.e., multiple parameters of NTFET curves) by constructing the pattern recognition of the analytes. We expect that further analysis based on a machine-learning approach will allow us to learn which parameters or sensor elements enhance discrimination of specific analytes leading to better sensor arrays designs and improve feature extraction toward fully automated NTFET measurements.

In this work, we selected five representative purine derivatives, i.e., adenine (A), guanine (G), xanthine (X), uric acid (UA), and caffeine (CF), which were measured using a sensor array comprised of metal nanoparticle-decorated NTFET devices²⁹ to discriminate each purine compound. To ensure that the sensors work differently while remaining cross-

sensitive to the analytes in the sensor array, four NTFET devices were decorated with metal nanoparticles (Au, Pt, Pd, and Rh), whereas one NTFET device was left bare, due to the work function-dependent potential barriers at the interface of metal and single-walled carbon nanotubes (SWCNTs).³⁰ The interactions of purine derivatives with extended Au metal surfaces and Au nanoparticle SWCNT hybrid systems were further investigated by density functional theory (DFT) calculations to determine which compounds have the largest adsorption energies and sensitivity in sensor measurements. The sensor array was further tested for the detection of caffeine, a popular stimulant of the human central nervous system, whose consumption has been associated with cardiovascular health and calcium balance.^{31–33} We developed and evaluated the robustness and accuracy of the caffeine-prediction models built from a total of 55 features (i.e., 11 features extracted from five different NTFET devices) using a stratified 10-fold cross-validation. The highest accuracy achieved was 95% using LDA with 48 features, whereas a 93.4% accuracy was obtained by using the support vector machine (SVM) method with only 11 features. Furthermore, we determined the accuracy and sensitivity of the algorithm to different types of features and identified, based on a recursive feature elimination with cross-validation (RFEVC) approach, those features which are essential in determining the accuracy of the model.

RESULTS AND DISCUSSION

Figure 1a depicts the Si chip used for sensing experiments with each of the four devices having interdigitated electrodes (IDEs). Oxidized SWCNTs were deposited from solution on

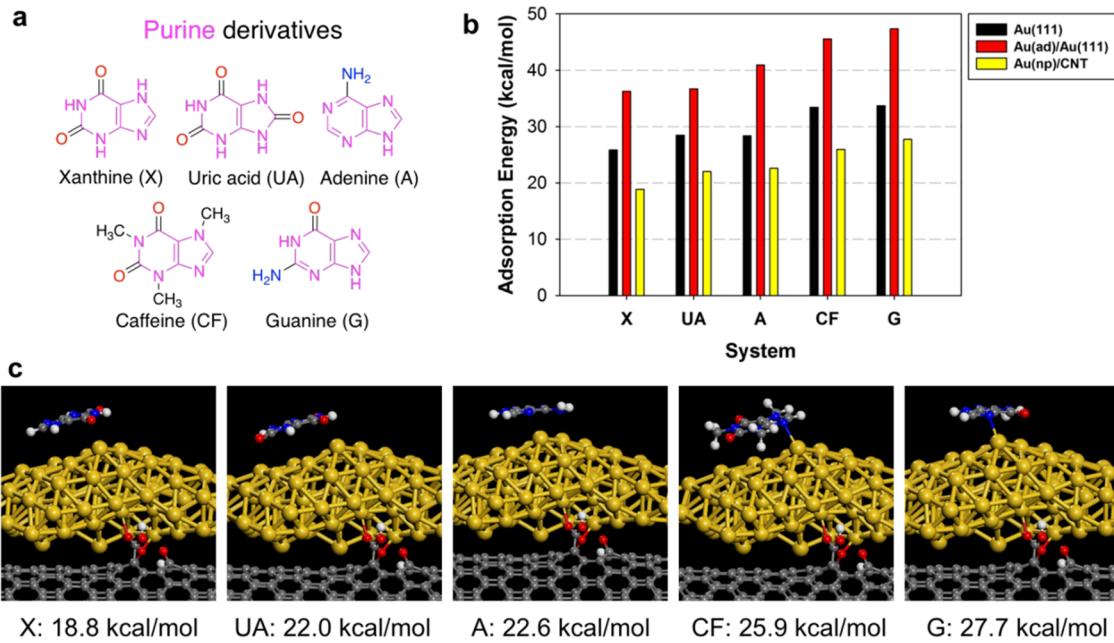


Figure 2. (a) List of purine derivatives considered in this work. (b) The DFT-calculated adsorption energies of purine compounds on a bare Au(111) surface (black filled bars), on Au(111) surface with an extra Au adatom (Au(ad)/Au(111)) (red), and on an Au₈₀ nanoparticle decorating a (14,0) SWCNT (Au(np)/CNT) (yellow). Graphical representation of the adsorption configurations of purine derivatives on Au(111) and Au(ad)/Au(111) surfaces are indicated in Supporting Information Figures S4 and S5. (c) Pictorial views of the DFT-calculated structures of purine derivatives on the Au(np) adsorbed at an oxygenated defect on a (14,0) SWCNT together with the corresponding adsorption energies indicated at the bottom of respective panels. The C atoms are indicated in gray, N atoms in blue, O atoms in red, H atoms in white, and Au atoms in orange. The acronyms indicated in panels (b) and (c) are defined in panel (a).

the chip surface via dielectrophoresis (DEP),³⁴ and metal nanoparticles (NPs) were subsequently nucleated on the oxidized defects on SWCNTs by electrodeposition.^{29,35} A scanning electron microscopy (SEM) image of Au NPs on SWCNTs is shown in Figure 1b. Each chip has four carbon nanotube devices decorated with the same type of metal nanoparticles. We repeated the process for three other metals (Pt, Rh, and Pd) and for one SWCNTs sample without any metal electrodeposition. Consequently, the sensor array is composed of 20 total sensors (4 sensors on each chip × 5 different functionalizations). Each sensor was tested in a liquid-gated field-effect transistor configuration (Figure 1c). We applied a bias voltage (+0.05 V) between the source and the drain and swept the gate voltage from +0.6 to −0.6 V to get the NTFET characteristics for both blank (H₂O) and purine solution (1 mM). G-V_G curves demonstrate p-type semiconducting characteristics with a modest on/off ratio due to the presence of metallic nanotubes in the SWCNT sample (Figure 1d). NTFET transfer characteristics were collected for each device type in H₂O as a baseline followed by the test solutions. Solutions were introduced to the chip surface and allowed to incubate for 5 min. For every solution, a minimum of 12 measurements were taken per surface type. For example, upon exposure to caffeine solution, the NTFET transfer characteristics of Pt-SWCNT device change by shifting to more negative gate voltages and lower source-drain conductances (Figure 1d). SEM images and the changes of NTFET characteristics for each device type in response to incubation to all five purine solutions are presented in Figure S1. All of the measured devices exhibited changes in NTFET characteristics that can be viewed as a combination of shifting and tilting of the G-V_G curves. The response of each device was different to each purine compound, and different types of

metal NP-decorated sensors responded differently toward the same analyte due to the differences in the work function of the metal and in the associated Schottky barriers formed across metal NP–nanotube interfaces.^{27,30,36,37} For example, Au and Pt have work functions of 5.0 and 5.65 eV, respectively, therefore the gate voltage would shift more on a Au surface than on a Pt surface for the same compound (Figure S1).

A supervised learning algorithm, LDA, was subsequently applied to discriminate between all five purine derivatives based on the combined response of the sensor array. Following our previously published approach,²⁸ 11 features that capture the changes in NTFET transfer curves (particularly related to curves “shifting” or “tilting”) were extracted from the NTFET transfer curves, as defined in Figures S2 and S3. To optimize the collection of the large NTFET sensor dataset (over 40 000 NTFET curves), we have designed a feature extractor (with 11 parameters extracted from each pair of curves, Figure 1d). The corresponding data extraction script was written in Python language and is available for download from our GitHub address.³⁸ A data matrix containing the responses and the replicates of Au-SWCNTs, Pt-SWCNTs, Rh-SWCNTs, Pd-SWCNTs, and bare oxidized SWCNTs to adenine, caffeine, guanine, uric acid, and xanthine were input into the PAST software,³⁹ where LDA was subsequently performed. The objective of LDA was to maximize the differences between different classes and to minimize the differences within each class. In the plot comprised of the first two LDA vectors (Figure 1e), the data points for different purines are seen to be well differentiated into separate regions, indicating an effective discrimination of purine derivatives despite their structural similarities.

To clarify the underlying mechanism of the purine derivatives discrimination and to understand the interaction

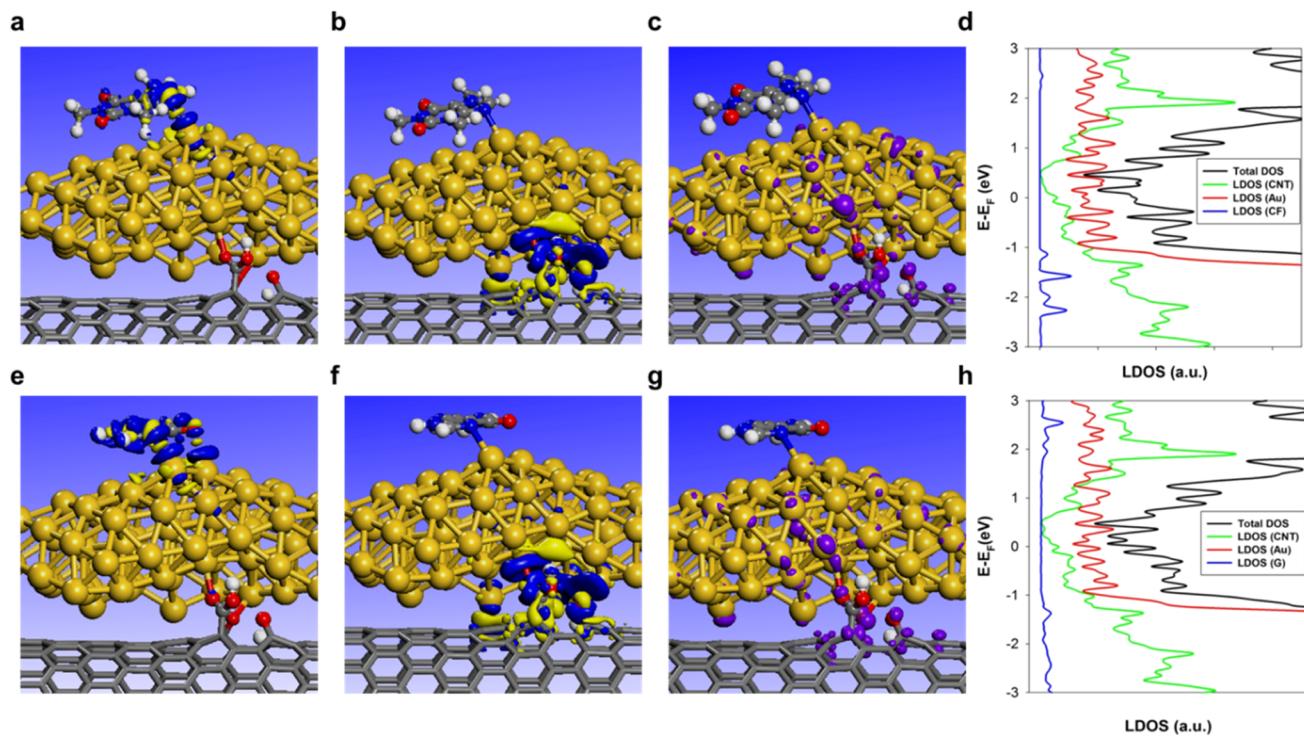


Figure 3. Charge difference maps for the two most stable purine derivatives identified in this study: caffeine (panels a, b) and guanine (panels e, f) adsorbed on the hybrid Au(np)–SWCNT (14,0) system. The indicated isosurfaces are $+0.01 \text{ e}^-/\text{\AA}^3$ (yellow) and $-0.01 \text{ e}^-/\text{\AA}^3$ (blue). The results are shown for the interface between respective molecules and Au nanoparticles (panels a, e) and between Au nanoparticles and SWCNT (panels b, f). The partial charge density isosurface ($0.01 \text{ e}^-/\text{\AA}^3$) corresponding to an energy range of 0.2 eV above the Fermi level is shown in panels c) for caffeine and (g) for guanine. Total density of states (DOS) and the corresponding local density of states (LDOS) projected on individual SWCNT, Au nanoparticle, and adsorbed molecules are illustrated in panels (d) caffeine and (h) guanine. The color scheme for different atoms is the same, as described in Figure 2.

between purine compounds and metal nanoparticle-decorated SWCNTs,⁴⁰ we conducted DFT calculations. The first objective of these theoretical investigations was to identify the strength of interaction between purine derivatives and Au surface, and for this purpose three different cases were considered. These correspond to molecular adsorption directly on a bare Au(111) surface (see Figure S4), adsorption mediated by an Au adatom on Au(111) surface (see Figure S5) and adsorption on an Au(80) nanoparticle decorating a (14,0) SWCNT (Figure 2c). In the last case, the coupling of the Au nanoparticle to the SWCNT takes place at an O-functionalized defect. For all molecules investigated (see Figure 2a), several different lying down or vertical orientations were considered as well as different binding locations on the corresponding surfaces. The illustrated results indicate only the most stable configurations identified. As seen in Figure S4, all purine compounds adsorb in a lying-down configuration to maximize the interaction between the respective rings and the metallic surface; in particular the interaction between the N and O sites with Au atoms. Among the purine set investigated, the most stable are CF and G compounds for which adsorption energies of 33.4 and 33.7 kcal/mol were determined. In the case when molecular adsorption is mediated by an Au adatom on the Au(111) surface (see Figure S5), this facilitates formation of covalent bonds primarily with N (for X, A, CF, and G molecules) or with O (in the case of UA) sites and helps to decrease the repulsive interactions between H atoms and Au surface. As a result, the calculated adsorption energies (see Figure 2b) increase relative to the case when adsorption takes place on the bare surface. Interestingly, the relative order of the

adsorption energies for various molecules does not change for the surface with or without an Au adatom. The same relative energetic trend is maintained for the case when molecules adsorb on the Au(np)–SWCNT hybrid system (see Figure 2c). The smaller size of the Au nanoparticle relative to the previous infinite Au slab models leads to a decrease in the corresponding long-range van der Waals interactions and a corresponding drop of the adsorption energies to values in the 18.8–27.7 kcal/mol range. Similar to the previous cases when adsorption takes place on extended metallic surfaces, the relative order in adsorption energies of various purine compounds is unchanged and overall, both CF and G are found to be the most stable bonded molecules. For both these two molecules, direct bonding of N atoms with a protruding Au atom located on the nanoparticle surface is seen to take place (see Figure 2c). Overall, these results indicate that binding of purine derivatives can take place on either Au(111), Au(ad)/Au(111), or Au(np)/CNT and in each case the largest adsorption energies are found for CF and G compounds. In particular, adsorptions of CF and G on Au(np)/CNT are enhanced by Au–N bond formation. We note that in the absence of Au from Au–SWCNT hybrid system, the adsorption of purine derivatives on bare SWCNT is governed only by weak van der Waals interactions leading to adsorption energy values significantly smaller than those observed when Au nanoparticles are present on SWCNT (see Figure S6).

The interaction between purine derivatives and Au surface also leads to important charge-transfer effects. These can be seen from the charge difference maps represented in Figure

3a,b for CF and 3e,f for G when adsorbed on Au(np)–CNT hybrid system. The first set of panels (a) and (e) indicate the charge transfer at molecule–Au interface, whereas the second set of panels provide similar type of information at the Au–CNT interface. On the basis of Bader charge analysis, it is found that charges as high as 0.34e for CF–Au–CNT and 0.36e for G–Au–CNT are transferred from Au nanoparticle to the SWCNT, respectively. In NTFET characteristics, the charge transfer can be correlated to the change in transconductance (slope), threshold voltage shifting, and minimum conductance change,^{27,30} and the results from DFT calculations support the feature selection results in SVM model (vide infra). The electronic properties of purine Au–SWCNT hybrid system for CF and G are analyzed in Figure 3d,h, respectively. In each case, we illustrate the total density of states (DOS) and the localized density of states (LDOS) on the respective components, i.e., the adsorbed molecules, the Au nanoparticles, and the SWCNT. For both CF and G systems, there are nonzero contributions just above the Fermi level, indicating a semimetallic or metallic character. The contributing states in the range of 0.2 eV above the Fermi level are represented in the partial charge distribution plots shown in Figure 3c for CF and 3g for G. In both cases, the indicated charge densities are delocalized on both Au nanoparticle and on SWCNT, in support of the sensing functionality of the entire hybrid system.

On the basis of the DFT results, Au NP-decorated SWCNTs demonstrate the potential to discriminate purine molecules. However, due to the close adsorption energies among some of these compounds, for example, UA and A, it is hard to discriminate them exclusively using energetic criteria on Au NP-functionalized SWCNTs surfaces. In fact, NTFET characteristics that can be correlated to adsorption energies such as the change in transconductance (slope), threshold voltage shifting, and minimum conductance for Au NP-decorated devices (Figure S7) were similar for all five purine derivatives, indicating that individual NTFET characteristics cannot adequately distinguish between the compounds. Therefore, we extended the analysis to the case of different metallic surfaces, and we demonstrate that by utilizing the subtle interactions between each compound with different metallic surfaces, combined with an LDA analysis, discrimination of various purine compounds becomes possible.

In order to evaluate possible applications of the purine discrimination model, we examined its robustness. We marked one caffeine sample as unknown to see if the model can correctly classify unknowns. As shown in Figure S8, the brown empty triangle is the marked unknown sample and this was classified outside the caffeine region. Upon addition of an unknown sample the scatter of the data points in Figure S8 is practically a mirror image to that in Figure 1e, which is indicative of overfitting.

The overfitting of data shows that the trained model has poor generalization ability and is highly dependent on the initial training set. To address this problem and more reliably predict the presence of purine compounds, we chose to examine the effect of algorithm selection on predictive capability and robustness. We considered the case of caffeine which, as was shown by DFT calculations, provides a strong charge-transfer effect to metallic nanoparticle. We expanded the collected dataset by including additional training data. The samples used for the training dataset are summarized in Table 1. A total of 398 measurements, separated into caffeinated and

Table 1. Caffeinated and Noncaffeinated Samples for Building Prediction Models and Their Concentrations

class	name	concentration ^a
caffeinated	Coca-Cola (caffeine)	0.4 mM
	Diet Coke (caffeine)	0.4 mM
	Coca-Cola zero sugar (caffeine)	0.4 mM
	Pepsi (caffeine)	0.4 mM
	Dr. Pepper (caffeine)	0.4 mM
	aqueous solution (glucose and caffeine)	0.3 M and 0.5 mM
	aqueous solution (caffeine)	0.5 mM 1 mM 2 mM 3 mM
	non	non
	non	non
	non	non
noncaffeinated	non	non
	non	non

^aConcentrations refer to the components in the parenthesis.

noncaffeinated classes, were taken from 32 different solutions with various concentrations. Caffeine prediction models were constructed for the new dataset and validated by a 10-fold stratified cross-validation approach. Both LDA and SVM models discussed below were tested using this procedure. For this purpose, the dataset was split into 10 parts, each part having the ratio of caffeine to noncaffeine 1:2.54 in accordance to the ratio of the 2 classes in the dataset, 9 parts were used to build the model and 1 part was used as an unknown sample to test the robustness of the model.

We tested the LDA model and used the recursive feature elimination with cross-validation (RFECV)⁴¹ to determine the relationship between the number of features and the model prediction accuracy (Figure S9). In the RFECV procedure, the feature selection is performed recursively by reducing the set of features based on a weighting scheme with the weights being initially assigned to the features from an estimator. With the corresponding number of features at each iteration, the algorithm outputs an accuracy based on stratified 10-fold cross-validation. The estimator was trained on the initial set of features with the importance of a feature ranked by its importance in the linear model (i.e., the features that contribute to a higher classifier accuracy typically have larger weights and consequently are more important). The

Table 2. Selected FET Features and Their Prediction Accuracy in SVM Analysis after RFECV

FET features ^a	Au	Pt	Au + Pt	Rh	Pd	bare
1 relative change in transconductance (slope)	✓	✓	✓✓		✓	
2 threshold voltage (V_{th}) shift	✓	✓	✓✓	✓	✓	✓
3 $\Delta G/G_0$ at $-0.6V_G$	✓				✓	
4 $\Delta G/G_0@V_{th}$ at $-0.6V_G$	✓			✓	✓	
5 $\Delta G/G_0@V_{th}$ at $-0.4V_G$	✓		✓✓	✓	✓	✓
6 $\Delta G/G_0@V_{th}$ at $-0.3V_G$			✓✓	✓	✓	
7 $\Delta G/G_0@V_{th}$ at $+0.3V_G$	✓		✓✓		✓	
8 $\Delta G/G_0@V_{th}$ at $+0.4V_G$					✓	✓
9 $\Delta G/G_0@V_{th}$ at $+0.6V_G$		✓	✓✓	✓	✓	
10 $\Delta G/G_0$ at $+0.6V_G$					✓	
11 $\Delta G/G_0$ at minimum G	✓	✓	✓✓	✓	✓	✓
accuracy	85.4%	86.4%	90.9%	71.8%	77.6%	75.5%

^aRefer to Figures 1d and S3 for definitions of the FET features.

importance/ranking was calculated from the weight square (w^2) of each feature in the model.⁴¹ In an LDA model (Figure S10), the best accuracy approaches 95% for a model with 48 features out of a maximum of 55 available (Figure S9a). The model was further tested against a titration regime to confirm specific response to caffeine. We prepared three caffeine-spiked samples (Coca-Cola) as the test data with total molarities of caffeine of 1, 2, and 3 mM. The corresponding data were analyzed by the LDA model generating an accuracy of 85%.

To explore the effect of algorithm choice on the number of features as well as to identify the most important features across different models for caffeine classification, we performed RFECV on a SVM model. SVM⁴² is a powerful machine-learning algorithm widely used for classification^{9,43} and gene selection^{41,44} in chemistry and biology. The classifier itself is defined by a separating hyperplane defined in the dimension of the overall feature space. For example, given the labeled data (training data), SVM outputs an optimal hyperplane which classifies new data. For the caffeine-prediction dataset, the best accuracy of the SVM model was found to be around 93.4% with only 11 features across all 5 sensor surfaces (Figure S9b). The features of interest identified in this case are the following: 1-Au, 2-Au, 11-Au, 2-Pt, 5-Pt, 7-Pt, 9-Pt, 2-Rh, 6-Pd, 5-Bare, and 8-Bare, where the features numbers for each corresponding surface are detailed in Figures 1d and S3. The Python code to perform feature selection can be found in Supporting Information. Selection of too many features in the model not only increases the model running time and noise but also can potentially lead to overfitting. As a result, based on the comparable accuracy of LDA and SVM and the significantly smaller number of features required by SVM, we selected SVM as our classification model algorithm for the detection of caffeine.

The SVM model for caffeine prediction shows great potential as a classifier leading to high accuracy with comparatively few input features. Analysis of the selected features in the SVM model reveals that Au and Pt surfaces have the largest contributions to the classification, as 7 out of 11 features belong to these 2 surfaces. Features from the other surfaces Rh, Pd, and bare only marginally improve the accuracy, as shown below. To examine the underlying features for each metal surface, we performed RFECV on each single surface and summarized the selected features (black checks) and the corresponding accuracy in Table 2. The corresponding confusion matrices for different surfaces before and after RFECV are presented in Table S1. The model accuracies from

Au and Pt surfaces are 85.4 and 86.4%, respectively, the highest among the five surfaces investigated. In fact, when combined, these two surfaces provide an accuracy of 90.9% with the selected features in Table 2. This accuracy is somewhat smaller than the 93.4% accuracy observed when considering all five surfaces. Additionally, we found that the three features selected on Au surface, which are 1-Au, 2-Au, and 11-Au, can give 85.4% prediction accuracy. These three selected FET features correspond to changes in transconductance, threshold voltage, and minimum conductance, parameters associated with the charge transfer between analyte molecules and nanotube sensors. This observation is consistent with our DFT results, which demonstrate the existence of an important charge transfer between caffeine and Au NP-decorated SWCNTs hybrid system (vide supra). However, individual NTFET characteristics such as transconductance, threshold voltage, and minimum conductance (Figure S7) cannot adequately distinguish between the related compounds.

The above presented analysis indicates that the supervised learning techniques can be successfully applied to NTFET sensor data and allows identification of analytes with high accuracy, complementing traditional DFT calculations.

CONCLUSIONS

We were able to use metal nanoparticle-decorated SWCNTs as a FET sensor array combined with supervised machine-learning techniques to discriminate purine compounds. Initial viability of purine compound sensing was established by DFT calculations demonstrating both sufficient interaction energy as well as charge transfer that would impact FET performances. Upon applying LDA to a set of five metal nanoparticle-decorated carbon nanotube NTFET devices, we encountered overfitting when we tried to predict the presence of different purine compounds. We made our model more robust by expanding the database of caffeinated and noncaffeinated samples combined with a stratified 10-fold cross-validation approach and evaluating the effect of algorithm choice. For the large amount of feature extraction tasks from our extended experimental database, we created a feature extractor written in Python language with the automated results showing little difference from the manually processed ones. The primary advantage of using LDA for caffeine prediction was to achieve a 95% accuracy in successful classification. In this approach, a set of 48 out of a total of 55 potential features were needed to achieve 95% accuracy. The SVM model was found to be advantageous for greatly reducing the number of features to 11

with a slight decrease in classification accuracy to 93.4%. For SVM models, Au and Pt surfaces provide the critical measurements, and the corresponding NTFET characteristics, i.e., the transconductance, the threshold voltages, and the minimum conductance, were identified as the main sensing descriptors necessary for the overall classification. These main features corroborate our initial DFT results related to purine binding and to the associated charge transfer to metallic surfaces. Overall, this work demonstrates the successful combination of sensor array experiments, supervised learning, and DFT calculations to discriminate among five different purine compounds, an approach with potential for discriminating structurally similar analytes.

EXPERIMENTAL SECTION

Materials. Commercially available SWCNTs (P3-SWNT) were purchased from Carbon Solutions, Inc. Poly(dimethylsiloxane) (PDMS) was procured from Ellsworth Adhesives. Metal salts HAuCl₄ (99.9%), H₂PtCl₆ (99.9%), Na₃RhCl₆ (98%), Pd(Ac)₂ (98%), and anhydrous caffeine (C₈H₁₀N₄O₂), L-glucose (C₆H₁₂O₂), citric acid (C₆H₈O₇) were acquired from Sigma-Aldrich. Sodium chloride (NaCl) was purchased from Fisher Scientific. All reagents were analytical grade and were used as received. Soft drinks (Coca-Cola, Diet Coke, Coca-Cola zero sugar, caffeine free Coca-Cola Classic, caffeine free Diet Coke, Pepsi, and Dr. Pepper) were purchased directly from Giant Eagle and Walmart and were degassed for experimental use. Nanopure water from Thermo Scientific Barnstead Nanopure System with resistivity >18.2 MΩ cm was used to prepare all solutions.

NTFET Fabrication and Decoration with Metal Nanoparticles. Si wafers (100 nm) with 1000 chips comprising four gold interdigitated electrodes (IDEs) were fabricated via a standard photolithography process, as previously described.⁴⁵ Individual chips ($2 \times 2 \text{ mm}^2$) were then wire bonded into standard 40-pin ceramic dual in-line packages (CerDIP) and secured with PDMS to protect and passivate the electrical contacts. Liquid fluid chambers (1 mL) were fixed to the packages using PDMS to contain blank and sample solutions. SWCNTs were incorporated between electrodes using AC dielectrophoretic (DEP) deposition (0.01 mg/mL SWCNT in water, 10 V_{pp} at 10 MHz for 60 s). Metal nanoparticles (Au, Pt, Pd, and Rh) were then electrodeposited through bulk electrolysis using a CH Instruments electrochemical analyzer in a three-electrode setup (1 M Ag/AgCl reference electrode, Pt counter electrode, and IDEs as working electrodes) from a metal salt solution (1 mM in 0.1 M HCl). Size of metal nanoparticles was optimized for these experiments through control of the concentration of the metal salt solution (1 mM), deposition voltage (-0.2 V), and time (30 s) based on the previous studies.^{28,40} This electrodeposition process yields discrete metal nanoparticles in sizes ranging from 10 to 50 nm anchored to defect sites of the SWCNTs, as characterized by scanning electron microscopy (SEM).

Scanning Electron Microscopy Imaging. Scanning electron microscopy (SEM) was performed with Zeiss Sigma VP scanning electron microscope.

NTFET Measurements. NTFET transfer characteristics ($G-V_G$) were recorded by introducing the chosen solution into the liquid chamber and addressing the device with a Ag/AgCl reference electrode (CHI111, CH Instruments, Inc), which was calibrated against the standard new Ag/AgCl electrode to ensure the stability. Nanopure water was used. Water is electrochemically active due to autoprotolysis^{46,47} and has an ionic conductivity of 10^{-6} S/cm. Gate voltage (V_G) was swept from +0.6 to -0.6 V with a constant 50 mV source-drain voltage (V_{SD}) and source-drain current (I_{SD}) recorded as a function of gate voltage using two Keithley 2400 SourceMeters. Source-drain conductance (G) was then calculated (I_{SD}/V_{SD}) and plotted as a function of the applied V_G . NTFET transfer characteristics ($G-V_G$) were collected from each device type in nanopure water as a baseline. NTFET transfer characteristics were stable after 5 min

incubation (Figure S11). Subsequently, sample solutions (listed in Table 1) were introduced to the liquid chamber and allowed to incubate with the chip for 5 min before the second NTFET transfer characteristic was collected.

Computational Methods. Investigation of the adsorption and electronic properties of the set of purine derivatives indicated in Figure 2a on Au and on hybrid Au–SWCNT surfaces has been done using plane-wave DFT calculations, as implemented in Vienna ab initio simulation package.^{48,49} Perdew–Burke–Ernzerhof⁵⁰ exchange correlation functional corrected to include Grimme-D3⁵¹ long-range dispersion interactions was used in these calculations together with a cutoff energy of 400 eV for structure optimizations and 500 eV for electronic properties calculations. Three surface models were investigated. The first one corresponds to a bare (5×5) Au(111) slab model with four layers among which the top two layers were relaxed, whereas the bottom two were kept frozen at bulk optimized values. In the second model, adsorption of purine derivatives takes place on a similar Au(111) surface but with an extra Au adatom, initially located on the surface at a hollow site. For both of these two models, a vacuum width of 16 Å was used together with a $3 \times 3 \times 1$ Monkhorst–Pack⁵² k -point grid. In the case of Au–SWCNT hybrid system, a (14,0) SWCNT with six repeating units (taken along Oy axis) has been considered with the Au80 nanoparticle bonded to the CNT surface at a C–C defect functionalized with a mixture of COOH, O, and H species. The k -point grid considered for optimizations of this hybrid systems contained a single Γ point which was increased to a $1 \times 7 \times 1$ Monkhorst–Pack⁴⁸ grid for DOS calculations. The adsorption energies of purine molecules on either Au or hybrid Au–SWCNT surfaces have been determined as $E_{\text{ads}} = (nE_{\text{molec}} + E_{\text{slab}} - E_{(\text{molec}+\text{slab})})/n$, where E_{molec} is the energy of isolated purine molecule at its optimized gas-phase geometry, n represents the number of adsorbate molecules in the simulation cell, E_{slab} is the total energy of the slab (Au or hybrid Au–SWCNT), and $E_{(\text{molec}+\text{slab})}$ is the total energy of the adsorbate/slab system. Positive adsorption energies correspond in this sign convention to stable binding configurations. The amount of charge exchanged between adsorbates and Au or Au–SWCNT systems was evaluated based on the Bader charges⁵³ using Henkelman et al.⁵⁴ algorithm. The effect of solvation upon the adsorption energies of purines on Au(111) surface (see Figure S12) has been investigated using the Poisson–Boltzmann implicit solvent model as developed by Hennig et al.^{55,56} and implemented in VASP. The calculated adsorption energies in the presence of solvation follow closely the results obtained in gas phase.

ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsami.8b15785.

SEM and NTFET characteristics of different devices; feature extraction methods; Python code; LDA and SVM plots; confusion matrices; and DFT computational studies (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: astar@pitt.edu.

ORCID

Alexander Star: 0000-0001-7863-5987

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Characterization Facility (NFCF) at the University of Pittsburgh for access to the SEM instrumentation.

■ REFERENCES

- (1) Diehl, K. L.; Anslyn, E. V. Array Sensing Using Optical Methods for Detection of Chemical and Biological Hazards. *Chem. Soc. Rev.* **2013**, *42*, 8596–8611.
- (2) Askim, J. R.; Mahmoudi, M.; Suslick, K. S. Optical Sensor Arrays for Chemical Sensing: The Optoelectronic Nose. *Chem. Soc. Rev.* **2013**, *42*, 8649–8682.
- (3) Albert, K. J.; Lewis, N. S.; Schauer, C. L.; Sotzing, G. A.; Stitzel, S. E.; Vaid, T. P.; Walt, D. R. Cross-reactive Chemical Sensor Arrays. *Chem. Rev.* **2000**, *100*, 2595–2626.
- (4) Ghanem, E.; Afsah, S.; Fallah, P. N.; Lawrence, A.; LeBovidge, E.; Raghunathan, S.; Rago, D.; Ramirez, M. A.; Telles, M.; Winkler, M.; Schumm, B.; Makhneja, K.; Portillo, D.; Vidal, R. C.; Hall, A.; Yeh, D.; Judkins, H.; da Silva, A. A.; Franco, D. W.; Anslyn, E. V. Differentiation and Identification of Cachaça Wood Extracts Using Peptide-Based Receptors and Multivariate Data Analysis. *ACS Sens.* **2017**, *2*, 641–647.
- (5) Bright, C. J.; Nallon, E. C.; Polcha, M. P.; Schnee, V. P. Quantum Dot and Polymer Composite Cross-Reactive Array for Chemical Vapor Detection. *Anal. Chem.* **2015**, *87*, 12270–12275.
- (6) Nallon, E. C.; Schnee, V. P.; Bright, C. J.; Polcha, M. P.; Li, Q. Discrimination Enhancement with Transient Feature Analysis of a Graphene Chemical Sensor. *Anal. Chem.* **2016**, *88*, 1401–1406.
- (7) Mahmoudi, M.; Lohse, S. E.; Murphy, C. J.; Suslick, K. S. Identification of Nanoparticles with a Colorimetric Sensor Array. *ACS Sens.* **2016**, *1*, 17–21.
- (8) Nallon, E. C.; Schnee, V. P.; Bright, C.; Polcha, M. P.; Li, Q. Chemical Discrimination with an Unmodified Graphene Chemical Sensor. *ACS Sens.* **2016**, *1*, 26–31.
- (9) Peveler, W. J.; Roldan, A.; Hollingsworth, N.; Porter, M. J.; Parkin, I. P. Multichannel Detection and Differentiation of Explosives with a Quantum Dot Array. *ACS Nano* **2016**, *10*, 1139–1146.
- (10) Bender, M.; Bojanowski, N. M.; Seehafer, K.; Bunz, U. H. F. Immobilized Poly(aryleneethynylene) pH Strips Discriminate Different Brands of Cola. *Chem. - Eur. J.* **2018**, *24*, 1–6.
- (11) Cao, B.; Adutwum, L. A.; Oliynyk, A. O.; Luber, E. J.; Olsen, B. C.; Mar, A.; Buriak, J. M. How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics. *ACS Nano* **2018**, *12*, 7434–7444.
- (12) Chidambaram, A.; Stylianou, K. C. Electronic Metal–Organic Framework Sensors. *Inorg. Chem. Front.* **2018**, *5*, 979–998.
- (13) Germain, M. E.; Knapp, M. J. Discrimination of Nitroaromatics and Explosives Mimics by a Fluorescent Zn(salicylaldimine) Sensor Array. *J. Am. Chem. Soc.* **2008**, *130*, 5422–5423.
- (14) Gustafson, J. A.; Wilmer, C. E. Optimizing Information Content in MOF Sensor arrays for Analyzing Methane-air Mixtures. *Sens. Actuators, B* **2018**, *267*, 483–493.
- (15) Qin, L.; Wang, X.; Liu, Y.; Wei, H. 2D-Metal–Organic-Framework-Nanozyme Sensor Arrays for Probing Phosphates and Their Enzymatic Hydrolysis. *Anal. Chem.* **2018**, *90*, 9983–9989.
- (16) Yao, M.-S.; Tang, W.-X.; Wang, G.-E.; Nath, B.; Xu, G. MOF Thin Film-Coated Metal Oxide Nanowire Array: Significantly Improved Chemiresistor Sensor Performance. *Adv. Mater.* **2016**, *28*, 5229–5234.
- (17) Haupt, K.; Mosbach, K. Molecularly Imprinted Polymers and Their Use in Biomimetic Sensors. *Chem. Rev.* **2000**, *100*, 2495–2504.
- (18) Greene, N. T.; Morgan, S. L.; Shimizu, K. D. Molecularly imprinted polymer sensor arrays. *Chem. Commun.* **2004**, 1172–1173.
- (19) Greene, N. T.; Shimizu, K. D. Colorimetric Molecularly Imprinted Polymer Sensor Array using Dye Displacement. *J. Am. Chem. Soc.* **2005**, *127*, 5695–5700.
- (20) Star, A.; Joshi, V.; Skarupo, S.; Thomas, D.; Gabriel, J.-C. P. Gas Sensor Array Based on Metal-Decorated Carbon Nanotubes. *J. Phys. Chem. B* **2006**, *110*, 21014–21020.
- (21) Feigel, I. M.; Vedala, H.; Star, A. Biosensors Based on One-dimensional Nanostructures. *J. Mater. Chem.* **2011**, *21*, 8940–8954.
- (22) Meyyappan, M. Carbon Nanotube-Based Chemical Sensors. *Small* **2016**, *12*, 2118–2129.
- (23) Fennell, J. F.; Liu, S. F.; Azzarelli, J. M.; Weis, J. G.; Rochat, S.; Mirica, K. A.; Ravnsbæk, J. B.; Swager, T. M. Nanowire Chemical/Biological Sensors: Status and a Roadmap for the Future. *Angew. Chem., Int. Ed.* **2016**, *55*, 1266–1281.
- (24) Schnorr, J. M.; Swager, T. M. Emerging Applications of Carbon Nanotubes. *Chem. Mater.* **2011**, *23*, 646–657.
- (25) Swager, T. M. Sensor Technologies Empowered by Materials and Molecular Innovations. *Angew. Chem., Int. Ed.* **2018**, *57*, 4248–4257.
- (26) Ellis, J. E.; Star, A. Carbon Nanotube Based Gas Sensors toward Breath Analysis. *ChemPlusChem* **2016**, *81*, 1248–1265.
- (27) Heller, I.; Janssens, A. M.; Mannik, J.; Minot, E. D.; Lemay, S. G.; Dekker, C. Identifying The mechanism of Biosensing With Carbon Nanotube Transistors. *Nano Lett.* **2008**, *8*, 591–595.
- (28) Silva, G. O.; Michael, Z. P.; Bian, L.; Shurin, G. V.; Mulato, M.; Shurin, M. R.; Star, A. Nanoelectronic Discrimination of Non-malignant and Malignant Cells Using Nanotube Field-Effect Transistors. *ACS Sens.* **2017**, *2*, 1128–1132.
- (29) Kauffman, D. R.; Sorescu, D. C.; Schofield, D. P.; Allen, B. L.; Jordan, K. D.; Star, A. Understanding the Sensor Response of Metal-Decorated Carbon Nanotubes. *Nano Lett.* **2010**, *10*, 958–963.
- (30) Kauffman, D. R.; Star, A. Chemically Induced Potential Barriers at the Carbon Nanotube–Metal Nanoparticle Interface. *Nano Lett.* **2007**, *7*, 1863–1868.
- (31) Nawrot, P.; Jordan, S.; Eastwood, J.; Rotstein, J.; Hugenholz, A.; Feeley, M. Effects of Caffeine on Human Health. *Food Addit. Contam.* **2003**, *20*, 1–30.
- (32) Xie, C.; Cui, L.; Zhu, J.; Wang, K.; Sun, N.; Sun, C. Coffee Consumption and Risk of Hypertension: A Systematic Review and Dose–response Meta-analysis of Cohort Studies. *J. Hum. Hypertens.* **2018**, *32*, 83–93.
- (33) Heckman, M. A.; Weil, J.; Gonzalez de Mejia, E. Caffeine (1, 3, 7-trimethylxanthine) in Foods: A Comprehensive Review on Consumption, Functionality, Safety, and Regulatory Matters. *J. Food Sci.* **2010**, *75*, R77–R87.
- (34) Li, J. Q.; Zhang, Q.; Yang, D. J.; Tian, J. Z. Fabrication of Carbon Nanotube Field Effect Transistors by AC Dielectrophoresis Method. *Carbon* **2004**, *42*, 2263–2267.
- (35) Quinn, B. M.; Dekker, C.; Lemay, S. G. Electrodeposition of Noble Metal Nanoparticles on Carbon Nanotubes. *J. Am. Chem. Soc.* **2005**, *127*, 6146–6147.
- (36) Ding, M.; Tang, Y.; Star, A. Understanding Interfaces in Metal–Graphitic Hybrid Nanostructures. *J. Phys. Chem. Lett.* **2013**, *4*, 147–160.
- (37) Ding, M.; Sorescu, D. C.; Kotchey, G. P.; Star, A. Welding of Gold Nanoparticles on Graphitic Templates for Chemical Sensing. *J. Am. Chem. Soc.* **2012**, *134*, 3472–3479.
- (38) <https://github.com/Pitt-Star-Group/IVg-Analysis>.
- (39) <https://folk.uio.no/ohammer/past/>.
- (40) Michael, Z. P.; Shao, W.; Sorescu, D. C.; Euler, R. W.; Burkert, S. C.; Star, A. Probing Biomolecular Interactions with Gold Nanoparticle-Decorated Single-Walled Carbon Nanotubes. *J. Phys. Chem. C* **2017**, *121*, 20813–20820.
- (41) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422.
- (42) Cortes, C.; Vapnik, V. Support-vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (43) Ivanciu, O. Applications of Support Vector Machines in Chemistry. In *Journal of Hypertension*; John Wiley & Sons, Inc., 2007; Vol. 23.
- (44) Zhang, X.; Lu, X.; Shi, Q.; Xu, X.-q.; Leung, H.-c. E.; Harris, L. N.; Iglesias, J. D.; Miron, A.; Liu, J. S.; Wong, W. H. Recursive SVM Feature Selection and Sample Classification for Mass-spectrometry and Microarray Data. *BMC Bioinf.* **2006**, *7*, No. 197.
- (45) Gou, P.; Kraut, N. D.; Feigel, I. M.; Bai, H.; Morgan, G. J.; Chen, Y.; Tang, Y.; Bocan, K.; Stachel, J.; Berger, L.; Mickle, M.; Sejdić, E.; Star, A. Carbon Nanotube Chemiresistor for Wireless pH Sensing. *Sci. Rep.* **2014**, *4*, No. 4468.

- (46) Sonmez, B. G.; Ertop, O.; Mutlu, S. Modelling and Realization of a Water-Gated Field Effect Transistor (WG-FET) Using 16-nm-Thick Mono-Si Film. *Sci. Rep.* **2017**, *7*, No. 12190.
- (47) Cramer, T.; Campana, A.; Leonardi, F.; Casalini, S.; Kyndiah, A.; Murgia, M.; Biscarini, F. Water-gated Organic Field Effect Transistors – Opportunities for Biochemical Sensing and Extracellular Signal Transduction. *J. Mater. Chem. B* **2013**, *1*, 3728–3741.
- (48) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976**, *13*, 5188–5192.
- (49) Kresse, G.; Furthmüller, J. Efficiency of Ab-initio Total Energy Calculations for Metals and Semiconductors Using a Plane-wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (50) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab initio Total-energy Calculations Using a Plane-wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (51) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (52) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to The Projector Augmented-wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775.
- (53) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: New York, 1994.
- (54) Sanville, E.; Kenny, S. D.; Smith, R.; Henkelman, G. Improved Grid-based Algorithm for Bader Charge Allocation. *J. Comput. Chem.* **2007**, *28*, 899–908.
- (55) Fishman, M.; Zhuang, H. L.; Mathew, K.; Dirschka, W.; Hennig, R. G. Accuracy of Exchange-correlation Functionals and Effect of Solvation On The Surface Energy of Copper. *Phys. Rev. B* **2013**, *87*, No. 245402.
- (56) Mathew, K.; Sundararaman, R.; Letchworth-Weaver, K.; Arias, T. A.; Hennig, R. G. Implicit Solvation Model for Density-functional Study of Nanocrystal Surfaces and Reaction Pathways. *J. Chem. Phys.* **2014**, *140*, No. 084106.