Applied Physics Letters

A reference-free clustering method for the analysis of molecular break-junction measurements

Cite as: Appl. Phys. Lett. **114**, 143102 (2019); https://doi.org/10.1063/1.5089198
Submitted: 17 January 2019 • Accepted: 17 March 2019 • Published Online: 09 April 2019

Damien Cabosart, Maria El Abbassi, Davide Stefani, et al.

Applied Physics Letters ARTICLE scitation.org/journal/apl

A reference-free clustering method for the analysis of molecular break-junction measurements

Cite as: Appl. Phys. Lett. **114**, 143102 (2019); doi: 10.1063/1.5089198
Submitted: 17 January 2019 · Accepted: 17 March 2019 ·
Published Online: 9 April 2019

Damien Cabosart,[1,a)] Maria El Abbassi,[1] Davide Stefani,[1] Riccardo Frisenda,[2] Michel Calame,[3,4] Herre S. J. van der Zant,[1,b)] and Mickael L. Perrin[3,c)]

AFFILIATIONS

[1]Kavli Institute of Nanoscience, Delft University of Technology, 2600 GA Delft, The Netherlands
[2]2D Foundry, Instituto de Ciencia de Materiales de Madrid, Sor Juana Inès de la Cruz, 28049 Madrid, Spain
[3]Empa, Transport at Nanoscale Interfaces Laboratory, 8600 Dübendorf, Switzerland
[4]Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

[a)]damien.cabosart@gmail.com
[b)]h.s.j.vanderzant@tudelft.nl
[c)]mickael.perrin@empa.ch

ABSTRACT

Single-molecule break-junction measurements are intrinsically stochastic in nature, requiring the acquisition of large datasets of "breaking traces" to gain insight into the generic electronic properties of the molecule under study. For example, the most probable conductance value of the molecule is often extracted from the conductance histogram built from these traces. In this letter, we present an unsupervised and reference-free machine learning tool to improve the determination of the conductance of oligo(phenylene ethynylene)dithiol from mechanically controlled break-junction (MCBJ) measurements. Our method allows for the classification of individual breaking traces based on an image recognition technique. Moreover, applying this technique to multiple merged datasets makes it possible to identify common breaking behaviors present across different samples, and therefore to recognize global trends. In particular, we find that the variation in the extracted molecular conductance can be significantly reduced resulting in a more reliable estimation of molecular conductance values from MCBJ datasets. Finally, our approach can be more widely applied to different measurement types which can be converted to two-dimensional images.

The development of experimental tools for the collection of large datasets plays a key role in molecular electronics.[1–4] In particular, recent technological advancements in break-junction (BJ) techniques such as the scanning tunneling microscopy one (STM-BJ)[2,4] or the mechanically controlled one (MCBJ)[5] have made it possible to acquire statistically relevant datasets. Their measurement principle consists of repeatedly forming a quantum point contact in the presence of molecules while at the same time measuring the current flowing through it. Each breaking trace provides information about the conformation and configuration of the junction.[6] However, it is only through the statistical analysis of thousands of such traces, that an in-depth mapping of the breaking dynamics[7–9] and a meaningful interpretation of the molecular junction behavior can be obtained.[10–12]

In break-junction measurements, molecules are usually identified by the presence of plateau-like features after the breaking point of the last gold-gold atom connection, which has conductance equal to the conductance quantum ($G_0 = 2e^2/h$, where e is the elementary charge and h is Planck's constant). Since the behavior of the molecules in the junction can vary from one breaking trace to another (e.g., due to different injection points, number of molecules, electrode shapes, anchoring configuration, electronic coupling, level alignments, etc…), the recorded breaking traces may exhibit diverse features. A common way to process these traces and obtain statistical information about the most probable conductance value of the molecule (G_M) is by building a conductance histogram and fitting the prominent peak with a log-normal distribution.[13]

In the following, we illustrate why this approach may lead to inaccurate data interpretation and conclusions. For this purpose, we use a dataset recorded on an oligo(phenylene ethynylene)dithiol (OPE3) molecule consisting of more than 50 000 breaking curves.

These curves are obtained from six MCBJ samples and recorded at different breaking speeds and bias voltages [see details in Table S1

in Sec. II of the supplementary material), forming in total 16 datasets, all carried out at room temperature and under ambient conditions. Figure 1 shows two examples of conductance histograms built from these 16 datasets of breaking curves recorded at 100 mV. Although both datasets are recorded with the same experimental settings, they exhibit different molecular yields[14] (details about the calculation of the molecular yield are given in Sec. IV of the supplementary material). From this comparison plot, two main observations can be made: (i) the peak shapes and relative amplitudes are different for the two datasets, even though the same molecule is measured and (ii) the extracted values of G_M are not the same and differ by up to a factor of 4, when comparing the two most extreme values.

In total, 11 different datasets have been recorded with a bias voltage of 100 mV, each exhibiting different molecular yields, varying from 3% to 63%.[14] The inset of Fig. 1 presents the extracted value of G_M for all of them. The graph shows a considerable spread of about half an order of magnitude in conductance, and an apparent increase in G_M for the increasing molecular yield. The dependence of G_M on the molecular yield raises an important point about the correct interpretation of the extracted G_M values. This is particularly important for datasets exhibiting various types of breaking curves in which the most probable conductance value obtained from the raw histogram cannot be attributed to a unique molecular conformation, and more importantly, cannot be considered as a universal conductance value associated with that of a single and fully stretched molecule.

The problem of classification of breaking traces has recently been tackled using machine learning (ML) tools.[15,16] Generally speaking, ML algorithms can be subdivided into two main categories: supervised and unsupervised learning.[17,18] Supervised learning is used when the nature of the desired ML model output is known. For example, recently, supervised learning was used to train an artificial neural network for classifying experimental breaking curves of gold break-junctions based on labeled traces obtained by molecular dynamics simulations.[19] Moreover, a "deep" neural network has recently been applied to single-molecule measurements for DNA sequencing applications.[20] In contrast to that, the outcome of the unsupervised ML model is not predefined and the ML algorithms are used to detect the underlying structures of a given dataset. This approach allows, e.g., to classify the data according to specific characteristic features. Unsupervised learning has successfully been applied to the classification of breaking curves,[15,21,22] highlighting the importance of using more sophisticated tools to identify different types of breaking behavior in a given set of traces. However, the classification algorithm applied in those studies needed a reference vector, the choice of which may affect the clustering outcome, as shown in Fig. S2 of the supplementary material. Recently, an unsupervised clustering approach to identify the hierarchical data structure has been reported,[23] but in this case, the clustering required several parameters, and was performed on the 2D conductance-displacement histograms, and not on the individual breaking traces.

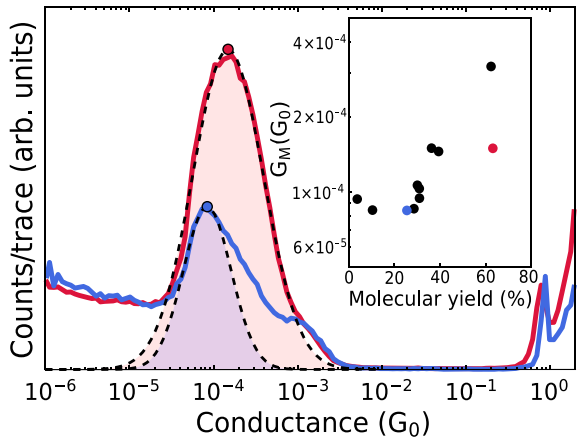

FIG. 1. Two unfiltered one-dimensional conductance histograms of the OPE3 molecule. The red and blue curves are built from breaking traces of the datasets related to samples 1b and 5a (see Table S1 in Sec. II of the supplementary material), respectively. The black dashed lines and the red/blue shaded regions correspond to log-normal distribution fits to the red/blue curves. The red/blue dots highlight the maximum of the log-normal distribution fits allowing to extract G_M. The inset shows G_M for all samples as a function of the molecular yield. The red/blue data points highlight the extracted G_M values using the red/blue histograms.

Our approach, schematically depicted in Fig. 2(a), aims to identify features in the experimental breaking traces and group the traces accordingly. The general workflow for the unsupervised ML classification can be summarized as follows: (i) construction of the *feature space* containing the relevant information about the shape of every breaking curve and (ii) applying a ML clustering algorithm in the constructed feature space that groups feature vectors into clusters. In the following, the workflow is explained in detail. The starting point of the method is the set of individual breaking traces, of which several examples are presented in Fig. 2(b). Each trace exhibits a specific shape and corresponds to a different breaking scenario of the junction. The blue trace, for example, shows a sharp conductance drop below 1 G_0 followed by an exponential conductance decrease as a function of the electrode displacement. This behavior is indicative of tunneling across a barrier when no molecule is bridging the electrodes. On the other hand, the green and red traces exhibit a step-like behavior below 1 G_0, which is associated with at least one OPE3 molecule trapped between the electrodes. However, the two curves do not have the same shape. The green trace has a conductance plateau around 1×10^{-3} G_0, while the red one exhibits a longer plateau at a lower conductance ($\approx 1 \times 10^{-4}$ G_0). These two traces exemplify the variability of molecular junctions during breaking and illustrate how MCBJ measurements allow to stochastically probe different molecular conformations/behaviors.

The creation of the applied feature space is partly inspired by the well-known MNIST dataset for handwritten digits, in which the images of the digits are reduced to 28×28 pixel images.[17] The MNIST dataset is often used to train various supervised ML models to identify handwritten digit images. During the learning process, the ML model identifies relevant features related to each digit in the training set using the discretized images. We employ the same principle to construct the feature space for our breaking trace classification approach, even though it is an unsupervised learning problem. As any image constitutes a two-dimensional (2D) representation of the shape of an object, one can naturally think of transforming every breaking curve into an individual 2D image. In our case, the created images are individual 2D histograms. Figure 2(c) illustrates how a breaking curve is transformed into an image by defining a region of interest (ROI), i.e., the boundaries of the final image. In the case of the OPE3 dataset, the ROI is defined in a conductance range between 1×10^{-6} and 1 G_0 and an electrode displacement range of 0–2 nm. The conductance range excludes the behavior of the metallic contact and of the measurement

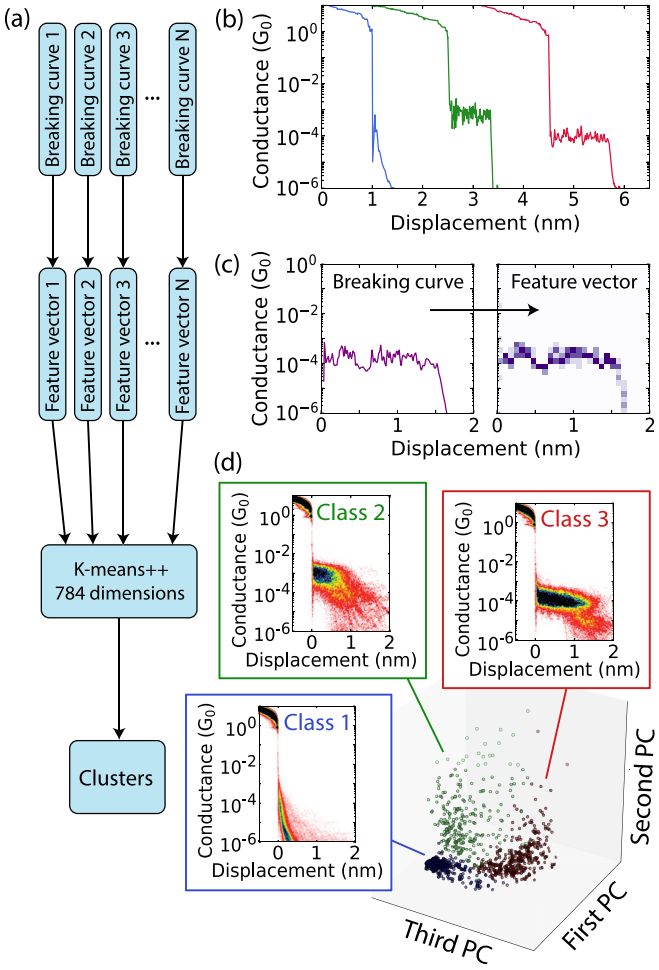

FIG. 2. (a) Schematic of the clustering algorithm. (b) Examples of breaking curves from the OPE3 dataset. For clarity, the traces are offset horizontally. (c) From left to right, transformation of a breaking trace into an individual 2D histogram. (d) Reduced feature space obtained using the principle component analysis in the case of the OPE3 dataset of sample 5a (see Table S1 of the supplementary material). The blue, green, and red points correspond to the reduced feature vectors related to the three different clusters/classes formed after using the K-means++ method. The 2D histograms built from the breaking curves of each class are displayed according to the cluster color.

noise-floor, while focusing purely on the behavior after the breaking of the junction. In addition to the ROI, the number of bins along the conductance and displacement axes is defined, i.e., the resolution of the individual 2D histograms. For instance, the breaking curve shown in Fig. 2(c) is transformed into an $M \times N$ pixel image. To construct the feature space, every 2D histogram is converted into a feature vector associated with a specific breaking curve. Each component of the obtained vectors represents one dimension in the feature space, meaning that one has to deal with a high-dimensional space. For example, in the case of 28×28 pixel images [Fig. 2(c)], the resulting feature space has 784 dimensions.

The last step of the classification task is to choose an appropriate clustering algorithm for the created feature space. For this work, we use the K-means++ method (from the Scikit-Learn Python library), which is one of the most popular clustering techniques (see details about the algorithm principle in Sec. VIII of the supplementary material). Its popularity is mainly due to its conceptual simplicity, low computational cost, and scalability, unlike more advanced clustering techniques. Another important advantage of the K-means++ algorithm is its ability to deal with high-dimensional spaces, which is a necessary requirement for our feature space. We note that other more advanced algorithms, e.g., taking into account nonisotropic cluster shapes, have been tested but fail in the case of such a high-dimensional feature space (see details in Sec. IX of the supplementary material).

To obtain a reduced representation of the high-dimensional space while preserving the maximum data variance, we employ a method called principal component analysis (PCA). This technique consists in projecting every feature vector onto the first three eigenvectors of the covariance matrix of the analyzed data. Figure 2(d) displays the reduced feature vector distribution, as well as the classification results obtained for one of the OPE3 datasets (sample 5a, see Table S1 in Sec. II of the supplementary material). We note that the PCA is not used in the clustering algorithm but only to reduce the 784-dimensional feature vectors to three dimensions for visualization purposes. The three clusters obtained from the high-dimensional clustering algorithm are plotted in different colors. It is important to realize that each point in the scatter plot corresponds to a single breaking trace. The clusters can subsequently be used to construct the 2D histograms belonging to the different classes, as shown in Fig. 2(d). The plot shows that the blue cluster (class 1) mainly contains breaking traces without any molecular signatures, while the green (class 2) and red (class 3) clusters are related to curves with plateau-like features. The green cluster contains breaking traces with conductance plateaus around 1×10^{-3} G_0, while for the red class, longer plateaus are observed at lower conductance values ($\approx 1 \times 10^{-4}$ G_0).

We now apply this approach to investigate the influence of the molecular yield on the most probable conductance value. For this purpose, we group all the datasets recorded at a fixed bias voltage (V) of 100 mV, i.e., in 11 sets with a cumulative amount of traces exceeding 40 000 curves (see Table S1 in the supplementary material). The conductance and 2D histograms for the full dataset are shown in Fig. 3(a), alongside the histograms of the three classes identified using our clustering method [see Figs. 3(b)–3(d)]. The obtained classes are similar to those previously obtained with a single OPE3 dataset [see Fig. 2(d)]. Class 1 contains curves without distinct molecular signatures [Fig. 3(b)], class 2 exhibits slanted plateaus starting around 1×10^{-3} G_0 [Fig. 3(c)], while the histograms of class 3 show flat and long plateau-like features around 1×10^{-4} G_0 [Fig. 3(d)]. In the following, we focus on class 2 and class 3.

As the clustering algorithm has been applied to the full dataset, and all the classes were found in all the datasets, the three identified clusters are universal, allowing for tracking of these classes and their respective occurrence across them (see details in Sec. V of the supplementary material). The extracted values of G_M for classes 2 and 3 as a function of the molecular yield are shown in Fig. 4(a). For both classes, G_M remains largely unaffected by the molecular yield, with extracted values of $5.0 \pm 1.1 \times 10^{-4}$ G_0 and $1.1 \pm 0.1 \times 10^{-4}$ G_0 for class 2 and class 3, respectively. The horizontal red dashed line indicates the most probable conductance value obtained by considering all curves belonging to class 3, while the shaded area corresponds to the standard

deviation of the red data points in Fig. 4(a). The plot also shows that the most probable conductance value of the unfiltered histogram is dominated by class 3, consisting of long traces up to 1.5 nm around 1×10^{-4} G_0. Such traces are commonly considered to originate from a single and fully stretched molecule bridging the two electrodes. However, the plot also demonstrates that it is only after the removal of class 1 and class 2 that the molecular conductance of these "ideal" molecular junctions is unraveled.

In contrast to the unfiltered data, the systematic dependence of G_M on the molecular yield is now absent for classes 2 and 3. In addition, the standard deviation in G_M for class 3 is about 5 times smaller than for the unfiltered dataset, allowing for a more accurate determination of the molecular conductance of the fully stretched molecule. Moreover, a large portion of the unfiltered conductance values lies outside the standard deviation of the conductance of class 3. These observations highlight the importance of data classification methods in break-junction measurements.

Finally, we also apply our method to a second dataset series with the aim to investigate the influence of the bias voltage on the determination of G_M. For this study, to avoid sample-to-sample variations, six OPE3 datasets successively recorded on the same sample for different bias voltages ($V = 50$, 100, 150, 200, 250, and 300 mV) are merged to obtain a dataset containing more than 10 000 curves [see the conductance and 2D histograms in Fig. S5(a) in Sec. VI of the supplementary material]. Three classes are formed using our clustering method, of which the conductance and 2D histograms are displayed in Figs. S5(b)–S5(d) in Sec. VI of the supplementary material. The resulting classes strongly resemble those obtained in Figs. 2(d), and 3(b)–3(d).

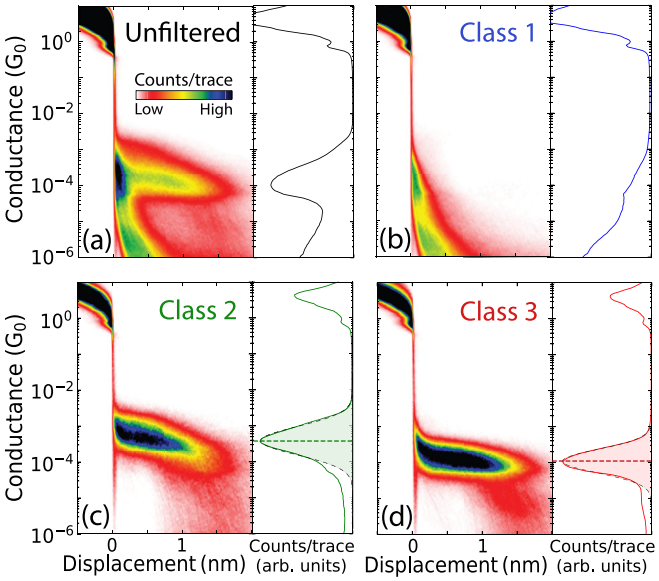

FIG. 3. (a) 2D (right panel) and 1D (left panel) conductance histograms built from all the breaking curves recorded at $V = 100$ mV, corresponding to 41 916 traces. (b)–(d) 2D (right panel) and 1D (left panel) conductance histograms built from the breaking curves of classes 1, 2, and 3, respectively, obtained with the clustering method. The black dashed lines and green/red shaded regions in (c) and (d), respectively, show log-normal distribution fits to the prominent peak in the histograms. The horizontal green/red dashed lines in (c) and (d), respectively, correspond to the mean of the log-normal distribution fits.

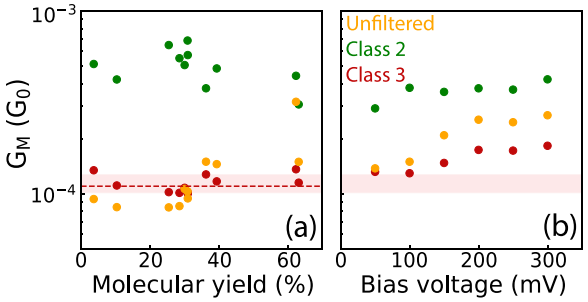

FIG. 4. (a) Most probable conductance G_M as a function of molecular yield. The orange, green, and red points are the extracted G_M in the case of the unfiltered data, class 2 and class 3, respectively. The horizontal red dashed line indicates the mean of the log-normal distribution fit obtained in Fig. 3(d) and the shaded area to the standard deviation of the red data points in (a). (b) G_M as a function of bias voltage. The orange, green, and red points are the extracted G_M in the case of the unfiltered data and classes 2 and 3, respectively. The red shaded region corresponds to the same conductance range of class 3 depicted in (a).

The extracted G_M values vs bias voltage are shown in Fig. 4(b). When considering the unfiltered data, a pronounced increase in G_M is observed for an increasing bias voltage. Classes 2 and 3, on the other hand, exhibit a smaller dependence. As a comparison, the graph highlights the extracted conductance range of class 3 for the yield dependence. While the most probable conductance of class 3 recorded at 50 mV and 100 mV lie at the edge of the conductance range, the extracted G_M in a bias voltage range of 150–300 mV lie outside. This observation suggests that the bias voltage indeed has the effect of increasing the conductance, as expected in the case of electron tunneling through a single, broadened level.

In our analysis, similar to previous studies,[15] the number of clusters is a free parameter and needs to be defined empirically. For the purpose of this letter, we have used 3 clusters. Nevertheless, this choice can be rationalized as follows: one can assume that each cluster corresponds to one particular molecular configuration with a distinct conductance. However, each cluster still includes local conformational and configurational changes which may lead to conductance fluctuations.

In our case, class 1 corresponds to the formation of junctions in which only single-barrier tunneling is observed, without any molecule bridging the gap. On the other hand, classes 2 and 3 show molecular signatures such as the formation of well-defined plateaus. Even though both classes have a molecular origin, clear differences between the two are observed. Class 2 is characterized by a slanted plateau with a higher conductance, whereas class 3 exhibits flat plateaus at lower conductance values. A possible explanation of this behavior may be related to variations in the anchoring of the molecule to the electrodes and the resulting changes in the injection point of the charges into the molecules. The long and flat plateaus present in class 3 are commonly believed to originate from a single-molecule bound to the two electrodes. In this scenario, charge transport occurs via the covalent bonds Au-S. For class 2, on the other hand, charges may also be injected through-space into the benzene rings via the overlap between its π-orbitals and the gold electrode wavefunctions. In this case, the sliding of the molecule on top of the electrodes and the resulting change in the orbital overlap may explain the gradual decay in conductance and

the shorter plateaus. To gain more insights into the junction formation, density functional theory and/or molecular dynamics calculations would be required. Such calculations, however, are beyond the scope of this letter.

To summarize, we demonstrated that unsupervised ML methods applied to MCBJ measurements allow for significant improvements in the extraction of the molecular conductance. Using our reference-free approach, we identify different molecular classes, which we track across several datasets. As such, we find that the two identified molecular classes are largely independent on the molecular yield, in contrast to the unfiltered data. Moreover, we find that the standard deviation of the extracted molecular conductance of the fully stretched molecule can be reduced by a factor of 5. Finally, by applying our approach on datasets with varying bias voltages, we observe a small dependence on the molecular conductance. The obtained results highlight the importance of using advanced and appropriate tools, such as ML algorithms, to efficiently analyze break-junction data and extract meaningful statistical molecular information.

See supplementary material for the influence of the reference vector in the method proposed by Lemmer *et al.*,[15] a description of the whole OPE3 datasets, the extraction of the most probable conductance, and the calculation of the molecular yield. A more detailed description of the clustering algorithm and results are also presented.

This work was partially funded by the FET open project QuIET (No. 767187). M.P. acknowledges the funding by the EMPAPOSTDOCS-II programme which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 754364.

REFERENCES

[1]R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, and J. M. van Ruitenbeek, "Measurement of the conductance of a hydrogen molecule," Nature **419**, 906–909 (2002).
[2]B. Q. Xu and N. J. Tao, "Measurement of single-molecule resistance by repeated formation of molecular junctions," Science **301**, 1221–1223 (2003).
[3]N. Agraït, A. L Yeyati, and J. M. van Ruitenbeek, "Quantum properties of atomic-sized conductors," Phys. Rep. **377**, 81–279 (2003).
[4]L. Venkataraman, J. E. Klare, I. W. Tam, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald, "Single-molecule circuits with well-defined molecular conductance," Nano Lett. **6**, 458–462 (2006).
[5]L. Wang, L. Wang, L. Zhang, and D. Xiang, "Advance of mechanically controllable break junction for molecular electronics," Top. Curr. Chem. **375**, 61 (2017).
[6]L. Venkataraman, J. E. Klare, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald, "Dependence of single-molecule junction conductance on molecular conformation," Nature **442**, 904 (2006).
[7]R. Frisenda and H. S. J. van der Zant, "Transition from strong to weak electronic coupling in a single-molecule junction," Phys. Rev. Lett. **117**, 126804 (2016).
[8]T. A. Su, H. Li, M. L. Steigerwald, L. Venkataraman, and C. Nuckolls, "Stereoelectronic switching in single-molecule junctions," Nat. Chem. **7**, 215 (2015).
[9]D. Stefani, K. J. Weiland, M. Skripnik, C. Hsu, M. L. Perrin, M. Mayor, F. Pauly, and H. S. J. van der Zant, "Large conductance variations in a mechanosensitive single-molecule junction," Nano Lett. **18**, 5981–5988 (2018).
[10]J. Ulrich, D. Esrail, W. Pontius, L. Venkataraman, D. Millar, and L. H. Doerrer, "Variability of conductance in molecular junctions," J. Phys. Chem. B **110**, 2462–2466 (2006).
[11]M. Perrin, C. Verzijl, C. Martin, A. Shaikh, R. Eelkema, J. van Esch, J. van Ruitenbeek, J. Thijssen, H. S. J. van der Zant, and D. Dulić, "Large tunable image-charge effects in single-molecule junctions," Nat. Nanotechnol. **8**, 282–287 (2013).
[12]M. L. Perrin, F. Prins, C. A. Martin, A. J. Shaikh, R. Eelkema, J. H. van Esch, T. Briza, R. Kaplanek, V. Kral, J. M. van Ruitenbeek *et al.*, "Influence of the chemical structure on the stability and conductance of porphyrin single-molecule junctions," Angew. Chem. Int. Ed. **50**, 11223–11226 (2011).
[13]M. T. González, S. Wu, R. Huber, S. J. Van Der Molen, C. Schönenberger, and M. Calame, "Electrical conductance of molecular junctions by a robust statistical analysis," Nano Lett. **6**, 2238–2242 (2006).
[14]R. Frisenda, D. Stefani, and H. S. J. van der Zant, "Quantum transport through a single conjugated rigid molecule, a mechanical break junction study," Acc. Chem. Res. **51**, 1359–1367 (2018).
[15]M. Lemmer, M. S. Inkpen, K. Kornysheva, N. J. Long, and T. Albrecht, "Unsupervised vector-based classification of single-molecule charge transport data," Nat. Commun. **7**, 12922 (2016).
[16]J. M. Hamill, X. Zhao, G. Mészáros, M. Bryce, and M. Arenz, "Fast data sorting with modified principal component analysis to distinguish unique single molecular break junction trajectories," Phys. Rev. Lett. **120**, 016601 (2018).
[17]A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* ("O'Reilly Media, Inc.," 2017).
[18]G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning* (Springer, 2013), Vol. 112.
[19]K. P. Lauritzen, A. Magyarkuti, Z. Balogh, A. Halbritter, and G. C. Solomon, "Classification of conductance traces with recurrent neural networks," J. Chem. Phys. **148**, 084111 (2018).
[20]T. Albrecht, G. Slabaugh, E. Alonso, and S. M. R. Al-Arif, "Deep learning for single-molecule science," Nanotechnology **28**, 423001 (2017).
[21]M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. C. Milan, R. J. Nichols, N. J. Long, and T. Albrecht, "New insights into single-molecule junctions using a robust, unsupervised approach to data collection and analysis," J. Am. Chem. Soc. **137**, 9971–9981 (2015).
[22]B. Li, M. Famili, E. Pensa, I. Grace, N. J. Long, C. Lambert, T. Albrecht, and L. F. Cohen, "Cross-plane conductance through a graphene/molecular monolayer/au sandwich," Nanoscale **10**, 19791–19798 (2018).
[23]B. H. Wu, J. A. Ivie, T. K. Johnson, and O. L. A. Monti, "Uncovering hierarchical data structure in single molecule transport," J. Chem. Phys. **146**, 092321 (2017).