THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC

Article

Grid-Based Correlation Analysis to Identify Rare Quantum Transport Behaviors

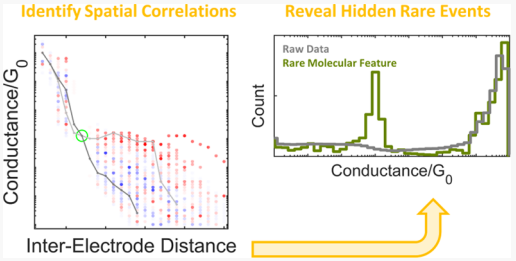
Nathan D. Bamberger, Dylan Dyer, Keshaba N. Parida, Dominic V. McGrath, and Oliver L. A. Monti*

Cite This: J. Phys. Chem. C 2021, 125, 18297−18307

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Most single-molecule transport experiments produce large and stochastic data sets containing a wide range of behaviors, presenting both a challenge to their analysis and an opportunity for discovering new physical insights. Recently, several unsupervised clustering algorithms have been developed to help extract and separate distinct features from single-molecule transport data. However, these clustering approaches have been primarily designed and used to extract major data set components and are consequently likely to struggle with identifying very rare features and behaviors that may nonetheless contain physically meaningful information. In this work, we thus introduce a completely new analysis framework specifically designed for rare event detection in single-molecule break junction data to help unlock such information and provide a new perspective with different implicit assumptions than clustering. Our approach leverages the concept of correlations of breaking traces with their own history to robustly identify paths through distance−conductance space that correspond to reproducible rare behaviors. As both a demonstrative and important example, we focus on rare conductance plateaus for short molecules, which can be essentially invisible when examining raw data. We show that our grid-based correlation tools successfully and reproducibly locate these rare plateaus in real experimental data sets, including in situations that traditional clustering approaches find challenging. This result enables a broader variety of molecules to be considered in the future and suggests that our new approach is a useful tool for detecting rare-yet-meaningful behaviors in single-molecule transport data more generally.



1. INTRODUCTION

Single-molecule electronics have the potential to enable cheap and efficient circuit fabrication at the ultimate size limit[1] and also provide an appealing test-bed for exploring intriguing physical phenomena at the nanoscale such as quantum interference,[2−4] spin filtering,[5,6] and interfacial coupling.[7−9] A significant and ongoing challenge in the investigation of transport through single-molecule systems, however, is extracting meaning from the large and stochastic data sets typically produced by experimental techniques such as the scanning tunneling microscope break junction (STM-BJ)[10−16] and mechanically controlled break junction (MCBJ).[17−24] Both of these methods involve forming and then breaking a thin metal constriction to create a single-molecule junction in the nanogap between two metal electrodes. The primary data collected is the conductance ($G = I/V$) through the junction during the breaking process as a function of how much the two sides have been pulled apart, known as a "breaking trace". Because of the inherently stochastic nature of both the breaking process and how/whether a molecule diffuses into and binds in the nanogap, a wide range of breaking trace behaviors are observed for each single-molecule system.[25] The process is thus repeated to collect thousands of breaking traces, with the results commonly summarized in one-dimensional (1D)[10] and two-dimensional (2D)[26,27] histograms. Conductance in these histograms is near-universally displayed on a logarithmic scale due to the large dynamic range of conductance values exhibited by most molecules.

These 1D and 2D breaking trace histograms are a powerful tool to reveal the average and/or most common behaviors in a data set, such as an exponentially decaying conductance in the absence of molecules ("tunneling behavior") or a relatively constant conductance over the length of a bound molecule ("molecular plateaus"). However, creating histograms inherently excludes all "trace history" information, i.e., the specific paths through distance/log(conductance) space followed by different traces. This makes it difficult to distinguish qualitatively different behaviors that may be present in the same data set since histograms will effectively average these behaviors together. A particular challenge is presented by behaviors that occur in only a minority of traces, which can become effectively invisible due to histogram averaging. However, previous work has demonstrated that such rare events may nonetheless correspond to physically important

Received: May 31, 2021
Revised: July 27, 2021
Published: August 17, 2021

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

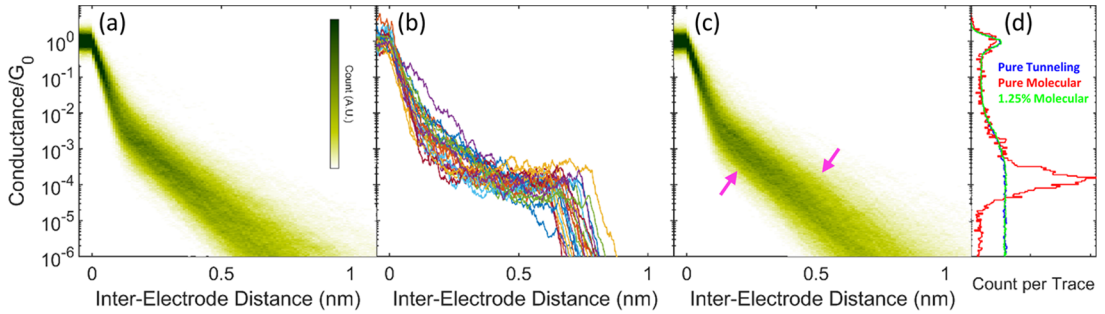


Figure 1. Demonstration, using simulated data, of how a short and rare molecular plateau feature can be effectively invisible when looking at histograms of raw data. (a) 2D histogram of 2000 simulated tunneling traces. (b) Twenty-five overlaid simulated molecular traces. (c) 2D histogram for data set containing 1975 of the simulated tunneling traces from panel (a), combined with the 25 molecular traces from panel (b). Despite looking identical to the 2D histogram in panel (a), the presence of the 25 molecular traces will produce a positive correlation between different locations along the molecular plateau region (e.g., pink arrows); our new approach thus quantifies such spatial correlations to identify rare-yet-meaningful trace behaviors. (d) Overlaid 1D conductance histograms for the data sets from panels (a–c), again showing that the plateau feature in the mixed data set is invisible in the raw data.

behaviors, such as switching between different spin states[28] or different sequences of binding behaviors.[29] Critically, the molecular signature itself—i.e., molecular plateaus—can in some circumstances become an easily lost rare event. Due to the stochastic nature of molecular binding, such plateaus are typically only observed in a fraction of all breaking traces, with the remaining traces displaying tunneling behavior. The magnitude of this so-called "molecular yield" varies depending on the binding group strength,[30−32] molecular concentration,[33] and other unknown or uncontrolled variables.[34,35] Figure 1 uses simulated traces to illustrate that, for short molecules whose molecular plateaus mostly overlap with the tunneling background, low molecular yield can make the molecular signature functionally impossible to identify in both the 1D and 2D histograms. Partially for this reason, most break junction experiments focus on systems with molecular yields >10%,[8,11,27,32,35,36] and often approaching 100%,[11,30,31] because this produces histograms with clear molecular features. However, high molecular yields increase the risk of measuring multi-molecule rather than the desired single-molecule features.[35] Moreover, requiring high molecular yields restricts the binding modalities under consideration, which may end up excluding optimal molecular structures for specific applications. For example, a large fraction of all molecules studied in break junctions employ thiol linker groups,[37] in part, because the strong sulfur−gold bond typically produces high molecular yields. However, the gold−sulfur bond also has a tendency for strong Fermi-level pinning,[38−42] which can negatively impact tunability of the quantum transport properties. A major benefit of rare event detection in break junction data, therefore, is that identifying infrequent molecular plateaus in the case of low molecular yields could allow a broader variety of molecular structures and metal−organic interfaces to be studied.

One strategy for separating qualitatively different behaviors that may overlap in 1D and 2D histograms is to employ clustering. Indeed, over the past 5 years, several clustering approaches have been designed specifically for breaking traces[34,43−53] and related data,[54,55] and these approaches have had varied success in extracting known and potential features, including "hidden" features, from real and simulated data sets. However, during their design and demonstration, these clustering algorithms have been primarily used to investigate features occurring fairly frequently in their respective data sets. Therefore, while such approaches can detect rare behaviors in certain cases, they will in general struggle with this task that falls on the edge of what they were designed for. A few types of filtering or plateau detection algorithms are able to extract weak molecular features from background behavior,[14,35,56−58] but these tend to be specifically tuned to a single system and/or rely on arbitrary cutoffs. There is therefore a need for a new type of robust approach, which is specifically targeted to the challenge of rare event detection, unlike current clustering approaches, and which unlike filtering requires only minimal a priori knowledge of the type of feature to be identified.

Here, we introduce such a new approach by focusing exactly on the very information that is lost when making 1D and 2D histograms: trace history and, in particular, any correlation between the future trajectory of a trace and its past behavior. While previous analysis tools for break junction data have considered correlations between the number of points at each conductance[59] or between conductance cuts at different distances,[50] these approaches reduce the natural space of breaking traces from two dimensions to one. In contrast, in our approach, we calculate correlations between different locations in the full two-dimensional distance/log(conductance) space, which allows us to identify rare paths followed by a subset of traces. For example, due to the 25 molecular traces contained in the data set shown in Figure 1c, there will be a positive correlation between the two areas indicated with pink arrows because more traces pass through both areas than would be expected if traces had no history and simply progressed like random walks through the 2D histogram. This positive correlation will then map out the shape of the rare event, e.g., the molecular plateau. In this work, we therefore define a new framework for rigorously defining two-dimensional spatial correlations and also present tools based on this framework that can identify rare events like the one contained in Figure 1c. We stress that our approach is designed to be applicable for identifying many types of rare events in break junction data, but here we focus on the particular challenge of recognizing rare molecular plateaus both as a concrete example and because of its importance.

In the remainder of this paper, we first present details of both our collection of experimental breaking traces and our generation of simulated breaking traces. We then use the simulated data to introduce our new approach, which starts by using coarse-gridding to define pairwise correlations between

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

discrete locations in distance/log(conductance) space. Markov-Chain Monte Carlo simulations are then used to extract particular types of rare features and identify interesting regions for further analysis. We use simulated data for this purpose to demonstrate how our approach works on data sets containing known rare plateau features. Finally, in the last section, we apply our new framework to experimental data sets, thereby validating that we can successfully and reproducibly identify rare plateau features in practice. This application also reveals that molecular signatures may be more common than previously thought, enabling research on systems with low molecular yield and with potential implications for our understanding of the nanoscopic environment of the junction.

2. METHODS

2.1. Generation of Simulated Breaking Traces. Due to the atomic-scale complexity of single-molecule junctions, creating physical models that faithfully reproduce all features and properties of experimental breaking traces remains a significant challenge. For this work, we therefore instead used an empirical model to generate simulated breaking traces that capture at least the most obvious properties of observed breaking traces. Such traces are sufficient for our purposes, despite obvious shortcomings in terms of capturing the full richness of experimental breaking traces, because we only use them to demonstrate *how* our grid-based correlation tools operate, not as a means of training or validating these tools.

Two types of simulated breaking traces were created: tunneling and molecular. All traces were generated on a logarithmic conductance scale with 500 data points per nm of inter-electrode distance. The tunneling traces consist of three sections: a prerupture plateau near 1 G_0, followed by a sharp drop-off to represent snap-back,[31,60] and finally a shallower linear drop-off to represent tunneling. The parameters defining each of these sections (e.g., the slope of the tunneling drop-off) were fixed for each individual trace and normally distributed across the set of all tunneling traces. Low-amplitude random noise was then added on top of each trace to create a more realistic shape. The molecular traces were generated in the same way as the tunneling traces, except that they contained two additional sections—a very gradually sloped plateau and a fairly sharp drop-off from the end of this plateau—which occur after the tunneling drop-off has decreased to the chosen conductance value for the plateau. See Section S.2 for full details of simulated trace generation.

Two thousand tunneling traces were generated to produce the 2D histogram in Figure 1b, and then 25 simulated molecular traces were combined with the first 1975 of those tunneling traces to produce the data set shown in Figure 1c and used throughout Section 3.

2.2. MCBJ Fabrication and Experimental Setup. MCBJ samples were fabricated and run following the methods of Bamberger et al.[50] Each sample was fabricated on a substrate of 0.5 mm thick phosphor bronze coated with a few-micron insulating layer of polyimide. The pattern of a thin metal wire with an ∼100 nm constriction in the center was defined using electron beam lithography. The wire itself was created by thermally evaporating a 4 nm titanium adhesion layer followed by 80 nm of gold. Finally, reactive ion etching with an O_2/CHF_3 plasma was used to turn the central constriction into an ∼1 μm free-standing gold bridge by underetching the polyimide.

MCBJ experiments were performed in air at room temperature. Each sample was clamped into a custom-built three-point bending apparatus in which a push rod was used to bend the sample and thereby thin and break the gold bridge. To collect each breaking trace, a stepper motor (ThorLabs DRV50) was first used to adjust the push rod until the conductance through the gold bridge was between 5 and 7 G_0 (where G_0 is the quantum of conductance, equal to 77.48 μS).[61] When this set point was reached, the collection of a single breaking trace was triggered by raising the push rod 40 μm at 60 μm/s using a linear piezo actuator (ThorLabs PAS009) while simultaneously recording the conductance through the bridge at 20 kHz using a custom high-bandwidth Wheatstone bridge amplifier.[62] The piezo was then retracted, after which the process can be repeated to collect another breaking trace. Custom LabVIEW software was used to automatically collect thousands of consecutive breaking traces. During trace collection, the bending apparatus was placed on a vibrationally isolated table to reduce mechanical noise and inside a copper Faraday cage to reduce electrical and acoustic noise.

2.3. Collection of Experimental Data Sets. The molecules studied in this work (OPV2-2SMe, OPV2-2BT, and OPV2-2SAc; see below) were each synthesized on-site (Section S.1.1) and characterized using NMR spectroscopy (Section S.1.2) and mass spectrometry (Section S.1.3). Each molecule was dissolved in HPLC-grade (>99.7%, Alfa Aesar) dichloromethane (DCM) to form ∼1 μM and/or ∼10 μM solutions. Note that the acetyl-protected binding groups in OPV2-2SAc are known to spontaneously deprotect on the gold surface for experiments performed in air, forming free thiol binding groups.[63]

Data set collection also followed the method of Bamberger et al.[50] Each MCBJ sample was cleaned with O_3/UV and rinsed with ethanol shortly before use. For each sample, we initially deposited ∼10 μL of pure DCM, using a clean glass syringe, on the center of the junction with the aid of a Kalrez gasket. We then collect an "empty" or "tunneling" data set of a few thousand breaking traces, both as a negative control and to calculate an attenuation ratio that was then applied to all subsequent data sets collected on that same MCBJ sample. Next, the LabVIEW program was paused with the gold bridge fully broken, and 10−20 μL of molecular solution was deposited on the center of the junction using a clean glass syringe. The LabVIEW program was then allowed to continue collecting breaking traces.

The experimental data in this work were collected using four different MCBJ samples. Due to events such as multiple depositions of molecular solution and/or a full relaxation of the push rod, the set of traces from each sample was broken into multiple chunks of 2000+ sequentially collected traces, and each of these data sets was analyzed independently (see Section S.3 for details).

3. RESULTS AND DISCUSSION

In this section, we begin by developing our new framework for considering two-dimensional spatial correlations, which statistically identifies deviations from random-walk behavior in order to define a quantity we call connection strength. Next, we use connection strength as the basis for a "feature-finder" tool that can extract, with minimal input parameters, different types of rare features from a breaking trace data set. Finally, we use our new framework and feature-finder tool to extract weak

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

molecular plateau features from experimental data sets for several short molecules and validate the results, demonstrating the power and utility of this new approach.

3.1. Calculating Correlations and Defining Connection Strength. Before we can calculate correlations between different locations in distance/log(conductance) space, we first need to formally define such locations. To this end, we superimpose each trace onto a coarse grid and then represent it as a series of lattice points from the grid, which we call "nodes". As shown in Figure 2a, each successive node in a coarsened trace increases by exactly one grid unit in x but can increase or decrease by any amount in y. The advantage of using coarsened traces is that nodes represent finite and discrete locations in distance/log(conductance) space that each trace unambiguously either does or does not pass through, making it straightforward to consider spatial correlations (see Sections S.4.1 and S.4.2 for details). It is important to note that the size of the coarse grid can impact the correlations we are subsequently able to find: if the grid is very coarse, then subtle correlations can be drowned out by the great number of uncorrelated traces passing through each node; whereas if the grid is very fine, an ever-larger total number of traces is needed to provide the statistical power to identify correlations in the first place. In practice, however, there is a broad range of gridding sizes over which our approach works well and yields reasonable results. Throughout this work, we thus use a grid with 25 nodes per nm and 10 nodes per conductance decade, but our main conclusions are not overly sensitive to modest changes in this grid size (Section S.7.1).

The main idea behind our strategy for identifying when trace behavior is correlated with past trace history is the realization that if no such correlations existed, then traces would simply proceed as (weighted) random walks through distance/log(conductance) space (i.e., whether each trace visits a given node would depend only on the single node visited immediately prior). Therefore, we can find such correlations by looking for the "least-random-walk-like" trace behaviors. This idea is illustrated on an extremely simplified case in Figure 2b,c: suppose we have a set of 502 traces (Figure 2b) with 100 traces each sloping downwards through five parallel nodes (to represent tunneling traces) and two traces that proceed horizontally (to represent molecular plateaus). Based on all of the traces that pass through a given node, we can calculate "exit probabilities" (Figure 2c) for each node to its neighboring nodes (Section S.4.3). Using the exit probabilities in Figure 2c, it is clear that *if* traces behaved like random walks, then, on average, only $(2\%)^4 = 0.000016\%$ of the traces passing through node X would also pass through node Y. However, for the actual traces (Figure 2b), 2/102 = 2% of the traces from X also pass through Y. Because $2\% \gg 0.000016\%$—that is, *more* traces go from X to Y than expected under random-walk behavior—we conclude that nodes X and Y are *positively* correlated.

To quantify this measure of correlation while also rigorously accounting for the contribution of random chance, we use a pair of one-sided binomial hypothesis tests. As explained in Section S.4.4, these tests allow us to calculate the probability, under the null hypothesis of random-walk behavior, of more or fewer traces than observed passing between two nodes. For example, in Figure 2b,c, these tests compute the probability of seeing two or more traces going from X to Y under random-walk conditions as only $5.5 \times 10^{-10}\%$ and the probability of seeing two or fewer such traces under random-walk conditions

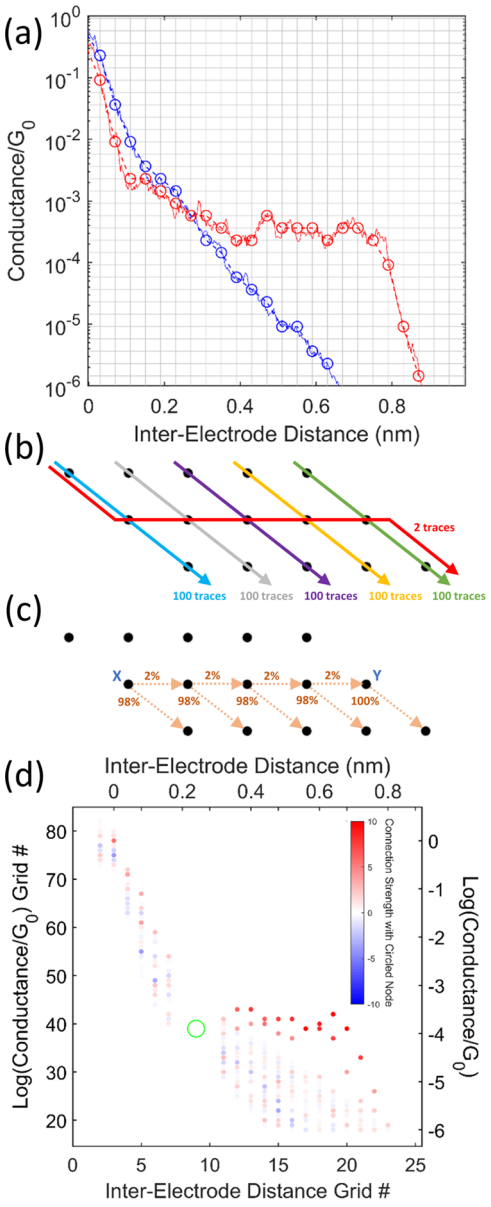


Figure 2. (a) Example of coarse-gridding two traces by representing each as a series of lattice points from the same grid. The original traces are represented with solid lines, the coarsened traces are represented with circles connected by dotted lines, and the grid is represented with gray lines. To make the grid easily visible, a coarser gridding is used for this plot than that used elsewhere in this work. (b) Hypothetical example of a small set of nodes with 502 traces passing through them. (c) Exit probabilities calculated based on the traces in panel (b). Based solely on these exit probabilities, the probability of going from node X to node Y is vanishingly small, but in fact two traces followed this path, revealing their correlation with their own history. (d) Connection strength distribution for all nodes vs the circled green node for the data sets from Figure 1c. The most positive connection strength nodes (red) clearly pick out the molecular plateau feature that was invisible in the raw data.

as ~100%. We refer to these two types of probabilities as $p_{\text{above}}(X,Y)$ and $p_{\text{below}}(X,Y)$, respectively.

Positively correlated node pairs will have small p_{above} and large p_{below} values, while negatively correlated node pairs—i.e., cases in which fewer traces pass between both nodes than would be expected under random-walk conditions—will have large p_{above} and small p_{below} values, and node pairs with little-to-

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

no correlation will have relatively large p_{above} and p_{below} values. We therefore combine both values into a single measure of pairwise node correlation that we call "connection strength", or CS, using the definition (see Section S.4.5 for details)

$$\mathrm{CS}(X, Y) = -\ln(p_{above}(X, Y)) + \ln(p_{below}(X, Y))$$

Taking logarithms allows us to easily differentiate very small probabilities from one another, and the sign choices ensure that CS will be positive (negative) for positively (negatively) correlated node pairs. Moreover, in information theory, log-probabilities represent the self-information or "surprisal" of an event,[64] making this a well-motivated definition for our goal of identifying correlations that are surprising relative to the expectation of random-walk behavior. To make CS symmetric with respect to nodes X and Y, X is always chosen as the node farther left and Y as the node farther right, since traces are only able to proceed from left to right. For nodes in the same column, CS is set to zero, indicating no correlation (since it is impossible by design for a single trace to pass through both such nodes).

This concept of pairwise connection strength between nodes is central to our ability to use spatial correlation to identify rare events in breaking traces. To demonstrate this, Figure 2d shows the connection strength distribution for the simulated data set in Figure 1c for every node compared to the green-circled node. The positive correlations clearly form a plateau shape, indicating that more traces passing through the circled node follow a plateau-like path than would be expected from pure random-walk behavior. This shows that there must exist a subset of traces that do *not* behave like random walks, but rather follow this plateau path instead of the average behavior of the data set; in other words, pairwise connection strength has exactly picked out the rare event that was purposefully built into this simulated data set! While "higher-order" correlations—e.g., if a trace passes through nodes X and Y, is it more or less likely than a random-walk to also pass through node Z?—can be calculated in principle, the number of observations for such events in a given data set, and hence our statistical power, would drop exponentially. For this reason, we focus exclusively on pairwise connection strength in this work.

3.2. Identifying Significant Features. Figure 2d demonstrates that connection strength distributions are a powerful way of visualizing spatial correlations and can reveal rare events. However, for each data set, there will be as many different distributions like the one in Figure 2d as there are nodes, making it unrealistic to examine all of them. The connection strength distribution with a specific node is thus most useful as a tool for investigating locations in distance/log(conductance) space that a researcher is already interested in. To perform a less directed exploration, we developed a new tool to identify such "interesting nodes" in the first place, with what counts as an "interesting node" naturally depending on what type of feature/rare event is under consideration.

The basis of this new tool is determining where in distance/log(conductance) space the "least-random-walk-like" node sequences meeting certain criteria can be found. To formalize the concept of "least-random-walk-like", we define the "significance" of a sequence of nodes as the average connection strength between all possible node pairs chosen from that sequence (Section S.4.6). High-significance node sequences intuitively represent paths through distance/log(conductance) space that real traces followed significantly more often than random-walk traces would have, and which may thus correspond to physically meaningful rare events. To solve the problem of identifying high-significance node sequences, we make use of Markov-Chain Monte Carlo (MCMC) simulation, which is a way to estimate the distribution of multidimensional objects according to a predefined weighting. In our case, the multidimensional objects are node sequences meeting specifiable criteria and the weighting of each sequence is exp(significance/T), where T is an "effective temperature" controlling how flattened vs peaked the distribution will be. This tool, which we refer to as the "MCMC feature-finder", thus produces a distribution of node sequences heavily weighted toward high-significance paths, allowing us to find those very paths and locate the "interesting nodes" introduced above (see Section S.5 for all MCMC details).

As a first illustration of the MCMC feature-finder, we will consider eight-node sequences for the simulated data set from Figure 1c, with no additional restrictions placed on the sequences' shape, slope, etc. Because a distribution of node sequences is difficult to represent visually, we instead examine the feature-finder results by plotting the frequency with which each individual node was included in the sequences produced by the MCMC (Figure 3a). Instead of following the rare

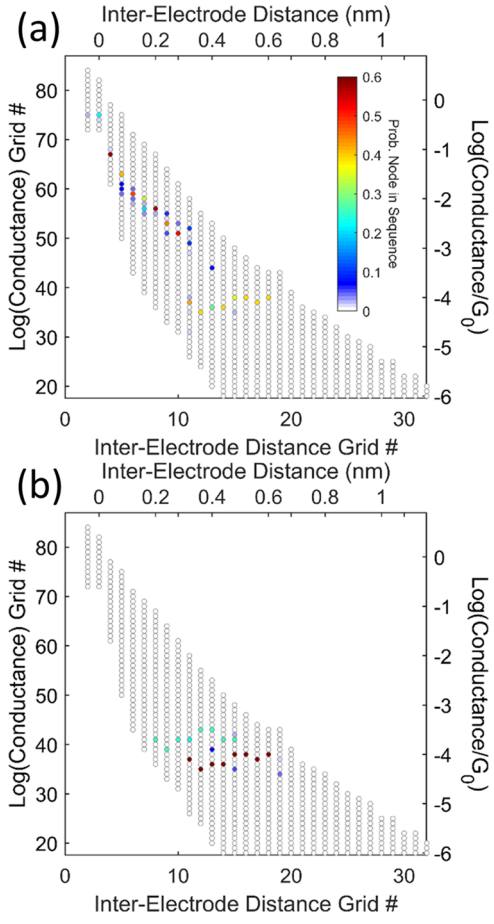


Figure 3. (a) Final distribution of nodes produced by running the MCMC feature-finder with eight-node sequences and no additional restrictions on the simulated data set from Figure 1c. The main feature discovered is a tunneling-like feature at high conductance because tunneling traces are also correlated with their own history. (b) To focus our search on plateau-like features in particular, we add slope criteria to the MCMC feature-finder, leading to the successful recovery of the rare plateau feature hidden in the data set in Figure 1c.

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

molecular plateau feature, these node sequences are concentrated in the high-conductance tunneling region. This is because tunneling is also not a random-walk behavior; as explained in Section 2, each simulated tunneling trace has a fixed average slope, so its future trajectory is correlated with its past behavior (and this is likely true of experimental tunneling traces as well). This demonstrates a critically important point: traces in the same data set can possess correlation with their histories in multiple different ways. To identify a specific *type* of correlation in a given data set using the MCMC feature-finder, therefore, we include the option within this tool to impose extra criteria on the node sequences being generated.

For example, in this paper, we are focused on the specific case of plateau-like features. In our second illustration of the MCMC feature-finder (Figure 3b), we thus require the eight-node sequences to have slopes of no more than 2.5 decades/nm (for details on slope calculation, see Section S.5.2). The inclusion of this relatively weak restriction results in the MCMC feature-finder highlighting exactly the rare molecular plateau feature that was built into this simulated data set. The high-probability nodes in Figure 3b thus represent "interesting nodes" in the context of plateau-like behavior, which could be further explored using connection strength distributions as in Figure 2d. We note that these nodes and the rare plateau feature they constitute were identified without knowing anything a priori about its location in distance/log-(conductance) space. The only input needed was what type of rare event we were looking for, i.e., relatively flat sequences that are eight nodes (~0.32 nm) in length.

While specifying the type of feature ahead of time does provide an avenue of influence for the user's biases, we note that user choice and biases are necessarily involved in any type of analysis, for example, through the selection of which unsupervised machine-learning algorithm to use. We therefore believe that it is advantageous to have the user make certain decisions explicitly and consciously, rather than making them implicitly and perhaps unknowingly via algorithm design. This is especially true for the case of choosing a feature type because constructing a single universal and automated algorithm for identifying *every* type of rare event in *every* possible data set is likely unfeasible. Our MCMC feature-finder should therefore be thought of as a guided exploration tool used to locate rare events of a type loosely defined by the user based on their physical intuition and/or the context of their particular application. Under this view, the ability of our MCMC feature-finder to be targeted at different types of rare behavior in the same data set, depending on the user's focus, is an advantage.

3.3. Using Grid-Based Correlation Tools to Find Rare Molecular Features in Experimental Data. In the previous section, we used a simulated data set containing a known rare molecular plateau feature to demonstrate that our new grid-based correlation tools can detect such rare behaviors in principle, and to explain how and why this is the case. Simulated breaking traces, however, can differ in critical ways from experimental breaking traces. Validating that these new tools achieve their goals in practice thus requires turning to such experimental data.

Figure 4a shows overlaid 1D conductance histograms for three data sets collected on the same MCBJ sample with the short molecule OPV2-2SMe (see Section S.3 for 2D histograms). These data sets collectively form an ideal test case for our ability to detect rare molecular plateau features.

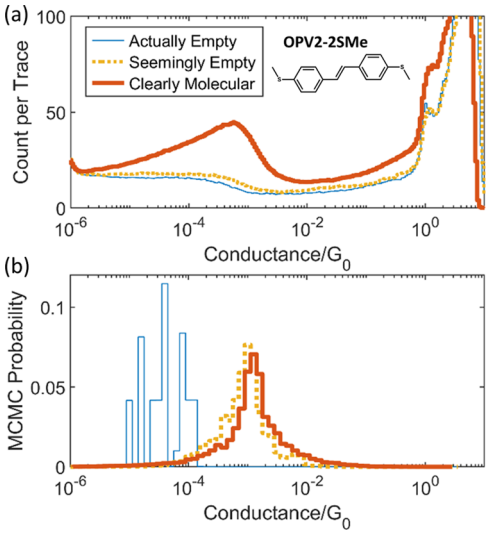


Figure 4. (a) Overlaid 1D conductance histograms for three experimental MCBJ data sets collected on the same sample with the molecule OPV2-2SMe (inset). The "actually empty" data set was collected before any molecules were deposited, the "seemingly empty" data set was collected after molecular deposition, but its raw data looks indistinguishable from the actually empty data set, and the "clearly molecular" data set was collected after molecular deposition and shows a clear molecular peak. (b) Node probabilities, projected onto the conductance axis, from the MCMC feature-finder when it was applied to the three data sets from panel (a) to search for plateau-like features. The same plateau feature is discovered in both the "seemingly empty" and "clearly molecular" data sets and not discovered in the "actually empty" data set.

The "actually empty" data set (blue in Figure 4a) is composed of breaking traces collected before any molecule was introduced to the system and so serves as a control in which no molecular plateaus should exist. The "clearly molecular" data set (red in Figure 4a) contains breaking traces collected after a solution of the molecule OPV2-2SMe was deposited on the sample. As the name suggests, this data set contains a clear molecular feature and thus serves as a positive control for where the plateaus for this particular molecule are expected to appear. Crucially, the "seemingly empty" data set (yellow in Figure 4a) contains breaking traces that were also collected after molecular deposition and yet appears nearly identical to the "actually empty" data set when examining histograms. This "seemingly empty" data set thus serves as our test case because it could plausibly contain OPV2-2SMe plateaus—since OPV2-2SMe was physically present on the MCBJ sample during data collection—but if so they must be rare, which would inform our understanding of the junction environment (see below).

We applied our MCMC feature-finder to all three data sets from Figure 4a, with the MCMC simulation set to generate 12-node sequences restricted to have a slope of no more than 2.5 decades/nm (see Section S.5.2). As shown in Figure 4b and Section S.6, the MCMC discovered a plateau feature in the "seemingly empty" data set that is nearly identical to the plateau feature discovered in the "clearly molecular" data set and distinct from the feature found in the "actually empty" data set. Due to the presence of both positive and negative controls, this result provides strong evidence that we have indeed successfully identified rare molecular plateaus for this OPV2-2SMe molecule. To demonstrate the advantages of our approach for such rare plateau detection, in Section S.10, we

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

have analyzed the "seemingly empty" data set from Figure 4a using a few representative clustering strategies. The results show that these strategies, while having some success, are generally challenged by this task, especially as the "rareness" of the feature increases. Our approach performs better at this task because it was specifically designed with rare events in mind.

To validate the results in Figure 4b, two potential concerns must be addressed. First, there is a concern that the plateau shapes discovered by the MCMC feature-finder are not representative of the original experimental data. This concern arises because the node sequences in the MCMC simulation are not restricted to node sequences that occurred in the actual experimental traces. Moreover, as described above, in order to focus on rare plateaus in particular, we applied a slope restriction to the MCMC simulations. There is thus a potential risk that the final MCMC results were "forced" into plateau-like shapes for the "seemingly empty" data set.

To address this concern and connect back to the original breaking traces, we start by considering the connection strength distribution with respect to one of the "interesting nodes" identified by the MCMC feature-finder (Figure 5a). Next, we introduce a new grid-based correlation tool by scoring each trace passing through the selected node by the average connection strength vs the selected node of all of the other nodes the trace passes through (Section S.8). Intuitively, higher scores identify the least-random-walk-like experimental traces passing through the selected node. Finally, we plot 2D and 1D histograms of just the top-10% scoring of these traces (Figure 5b,c). This reveals that plateau-like features are indeed present in the actual experimentally collected traces in the "seemingly empty" data set. No slope criteria or restrictions are used at any point in making Figure 5; we just selected the least-random-walk-like traces through a particular node, and those traces turn out to have plateau features. This scoring tool is thus a useful way to validate that a feature discovered with the MCMC feature-finder really exists in the experimental data. It is also a great way to extract for further analysis the traces in a data set that actually correspond to a particular type of rare behavior. The results in Figure 5 are robust to other choices of high-probability nodes from the MCMC output (Section S.9).

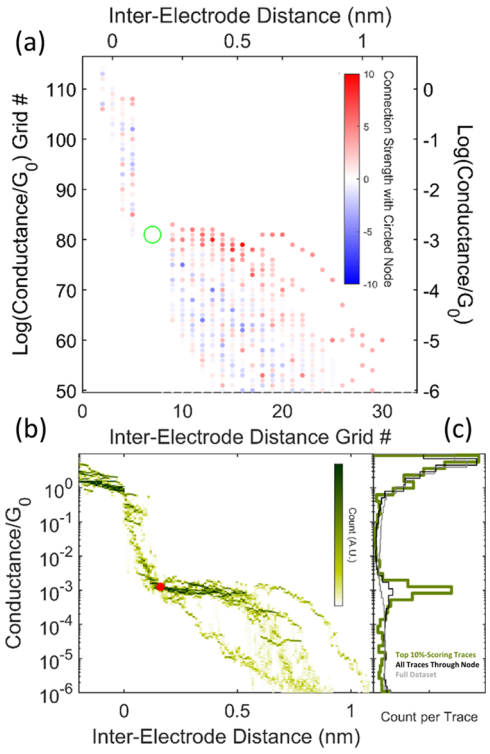


Figure 5. (a) Connection strength distribution for the "seemingly empty" data set from Figure 4 vs the node circled in green, which was identified as a high-probability node by the MCMC feature-finder. To connect these results back to the original traces, we use this distribution to score each of the traces passing through the selected node by the average connection strength of the other nodes they visit. (b) 2D histogram for the top-10% scoring traces through the selected node (now shown in red) demonstrates that plateaus *are* present in the original experimental traces, not just the MCMC results. (c) These top-scoring traces show a very clear molecular peak (green) that is completely invisible in the raw data (gray) and only weakly visible in all of the traces passing through the selected node (black).

The second potential concern about the MCMC feature-finder results in Figure 4b is that the plateau feature discovered in the "seemingly empty" data set could be located at the same conductance as the plateau feature in the "clearly molecular" data set simply by chance. After all, the MCMC feature-finder is designed to always find *some* feature matching its input criteria because some path through distance/log(conductance) space will always be *least*-random-walk-like. For example, a "plateau-like" feature (albeit at a different conductance) was still discovered in the "actually empty" data set despite our expectation that plateaus will not occur when no molecules are present. This is, in fact, an inherent and unavoidable challenge for any attempt at rare event detection in a highly stochastic and/or noisy system: in any single data set, it can be difficult to distinguish whether an identified rare event is "real" or just due to random noise-like behavior. The solution is to consider multiple data sets since "real" rare events should be detected consistently in the same place.

We thus applied our MCMC feature-finder, with the same settings used in Figure 4b (12-node sequences, slope $\leq$2.5 decades/nm) to additional examples of "actually empty", "seemingly empty", and "clearly molecular" data sets collected in the presence of three different short molecules (see Section S.3 for all data set details and histograms). As shown in Figure 6, for each molecule, the MCMC feature-finder consistently identifies essentially the *same* plateau feature in both the "seemingly empty" and the "clearly molecular" data sets, and this result is robust to changes in the MCMC feature-finder parameters (Section S.7). This constitutes strong evidence that the rare plateaus detected in the "seemingly empty" data sets are, in fact, molecular signatures and not just random/noisy behavior. That the identified plateau features appear at different conductances for the different molecules provides yet more evidence that these plateaus originate from the experimental data and are not simply an artifact of the MCMC simulation. In contrast, there is significant variation in the conductance of the features found in the "actually empty" data sets, and these features are mostly not robust to changes in the MCMC feature-finder parameters (Section S.7). This is in agreement with our hypothesis that these data sets do not contain consistent plateau features but rather just random behaviors that sometimes approximate plateau-like shapes by chance.

Besides helping to validate our ability to detect rare events in practice, the data in Figure 6 also provide potential insights into the nanoscopic environment of single-molecule junctions. We note that most of the "seemingly empty" data sets were

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

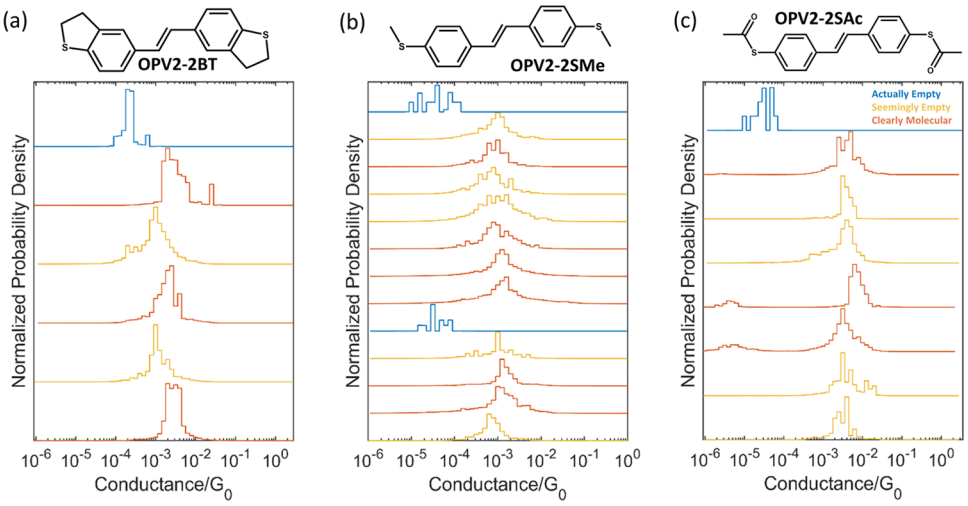


Figure 6. (a–c) Conductance distributions for the plateau-like features identified by the MCMC feature-finder in 27 different experimental data sets for three different short molecules. The distributions for "actually empty", "seemingly empty", and "clearly molecular" data sets are shown in blue, yellow, and red, respectively (see Section S.3 for details). For each molecule, the location of the identified plateau is quite consistent across all "seemingly empty" and "clearly molecular" data sets, providing strong evidence that the rare plateaus discovered by the MCMC feature-finder are, in fact, signatures of the molecules in question.

collected under essentially identical conditions to their "clearly molecular" counterparts (Section S.3). While the reasons for such variation in event frequency are not well studied in the field, one hypothesis is that the concentration of molecules in the immediate vicinity of the junction (and hence the frequency of molecular plateaus) is controlled, in large part, by random chance, and is only loosely dependent on the concentration of the deposited molecular solution. Therefore, our identification of molecular behavior in so many "seemingly empty" data sets may suggest that, at room temperature, these molecules are mobile enough that they can still find their way into the junction even when their concentration in the immediate vicinity is quite low. Additionally, the fact that there is little-to-no systematic difference between the conductances of the plateaus identified in the "seemingly empty" and "clearly molecular" data sets is consistent with the widely held assumption that MCBJ experiments are truly measuring single-molecule junctions, with little-to-no contribution from multi-molecule junctions.

4. CONCLUSIONS

Extracting as much meaningful information as possible from the large, stochastic data sets created by single-molecule break junction experiments continues to be a challenging multi-dimensional problem that will consequently require multiple different approaches to solve. Much effort has recently gone into developing different types of filtering and, especially, clustering algorithms for break junction data. While these approaches have many advantages and documented successes, to date they have mostly been designed for and applied to the problem of extracting prominent data set features, rather than the related but distinct challenge of identifying rare behaviors in huge samples of breaking traces. In this work, we specifically addressed this latter challenge by introducing a novel approach that uses the trace history information that is ignored by traditional histogram-based analysis to calculate pairwise correlations between discretized locations in distance/log-(conductance) space. This framework is quite distinct from other published analysis tools for single-molecule transport data, which we believe has value; all analysis tools make implicit assumptions about data set structure through their design, and increasing the variety of these designs and assumptions is important for gaining new perspectives and generating new hypotheses. Using simulated breaking traces, we demonstrated how a suite of tools based on our framework can be used to detect different types of rare features in the same data set.

To evaluate the utility of these new rare event identification tools, we chose to focus on the specific challenge of detecting rare molecular plateaus in the case of short molecules for which those plateaus overlap with a strong tunneling background signal. Using experimental MCBJ data sets collected for three separate short molecules, we demonstrated a consistent ability to detect molecular plateau features corresponding to the molecular species that was known to be physically present but whose signature was invisible in the raw data. These rare and hidden molecular plateaus were identified without any a priori knowledge or assumptions about their conductance values, and multiple controls and validation tests supported our inference that they are, in fact, signatures of the molecules in question.

The successful detection of rare molecular plateaus in several experimental data sets is important for multiple reasons. First, it provides us with important insights into the junction environment, suggesting, for example, that molecules can find their way to the very middle of the junction even when the overall frequency of molecular bridging events is quite low. In future studies, our tools may help yield insights into the causes of variable bridging frequency, such as varying local concentrations. Second, we believe that addressing the challenge of identifying rare molecular plateaus is of particular importance because if very weak molecular signatures can be reliably detected, then single-molecule researchers are empowered to explore a greater variety of molecular binding groups, concentration regimes, or perhaps single-molecule chemical reactions. Third, despite the importance of rare plateaus in particular, we stress that our grid-based correlation approach is not limited by design to this one type of rare event. Therefore, rigorously validating that the new tools presented

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307

here can successfully detect one form of rare behavior is also important because it suggests that they can be used for rare event detection more generally. This could potentially allow new types of physically meaningful break junction behaviors to be found and understood. We also note that our approach was able to identify molecular plateaus in the "clearly molecular" data sets as well, suggesting that while these tools were designed for rare event detection in particular, they may, in fact, have applicability for extracting more common events as well.

In contemplating the possible extension of our grid-based correlation approach to other types of rare events, however, it is also important to consider its limitations. For example, the MCMC feature-finder is designed to focus on the *single* most-correlated feature of a given type. If multiple features match the user-specified criteria—e.g., a data set with two rare plateau features—then the MCMC feature-finder is likely to be heavily weighted toward the one with even just slightly higher internal correlation (though the second could then be found by modifying the MCMC criteria). An example of this situation is shown in Section S.11.1 using simulated data. Another limitation of our approach is that it is designed to identify rare events that are localized in distance/log(conductance) space. This is relevant, for example, for the consideration of rare switching events between multiple molecular conductance states. As illustrated in Section S.11.2 using simulated data, our approach is capable of identifying such switching events if they occur at a preferred distance (e.g., if the switching is caused by stretching-induced conformational change) but is not suitable for detecting switching that stochastically occurs across a broad range of distances (e.g., if the switching mechanism is light-induced).

To enable these and other extensions of our work, we have made all of the MATLAB code needed to implement our new approach freely and publicly available as part of our SMAUG Toolbox at github.com/LabMonti/SMAUG-Toolbox.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.1c04794.

> Molecular synthesis; details of simulated trace generation; experimental data sets used in this work; further details on grid-based correlation framework; details of MCMC feature-finder; additional data from MCMC results; robustness of MCMC results; details on trace scoring; and robustness of conclusions drawn from trace scoring. The MATLAB code used for this work is available free of charge at github.com/LabMonti/SMAUG-Toolbox. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Oliver L. A. Monti − *Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona 85721, United States; Department of Physics, University of Arizona, Tucson, Arizona 85721, United States;* orcid.org/0000-0002-0974-7253; Email: monti@u.arizona.edu

Authors

Nathan D. Bamberger − *Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona 85721, United States;* orcid.org/0000-0001-5348-5695

Dylan Dyer − *Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona 85721, United States*

Keshaba N. Parida − *Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona 85721, United States*

Dominic V. McGrath − *Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona 85721, United States;* orcid.org/0000-0001-9605-2224

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.1c04794

Author Contributions

N.D.B. and O.L.A.M. conceived the research ideas. K.N.P. synthesized the molecules, directed by D.V.M. N.D.B. and D.D. fabricated the MCBJ samples and collected the break junction data. N.D.B. designed, developed, and implemented the grid-based correlation framework, with advice and input from O.L.A.M. N.D.B. wrote the manuscript with advice and input from all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support from the National Science Foundation, Award No. DMR-1708443, is gratefully acknowledged. Plasma etching of MCBJ samples was performed using a Plasmatherm reactive ion etcher acquired through an NSF MRI grant, Award No. ECCS-1725571. MCMC simulations were performed using High Performance Computing (HPC) resources supported by the University of Arizona TRIF, UITS, and RDI and maintained by the UA Research Technologies department. Quality control was performed using a scanning electron microscope in the W. M. Keck Center for Nano-Scale Imaging in the Department of Chemistry and Biochemistry at the University of Arizona with funding from the W. M. Keck Foundation Grant.

■ REFERENCES

(1) Forrest, S. R. The Path to Ubiquitous and Low-Cost Organic Electronic Appliances on Plastic. *Nature* **2004**, *428*, 911−918.

(2) Liu, J.; Huang, X.; Wang, F.; Hong, W. Quantum Interference Effects in Charge Transport through Single-Molecule Junctions: Detection, Manipulation, and Application. *Acc. Chem. Res.* **2019**, *52*, 151−160.

(3) Chen, Z.; Chen, L.; Li, G.; Chen, Y.; Tang, C.; Zhang, L.; Liu, J.; Chen, L.; Yang, Y.; Shi, J.; Liu, J.; Xia, H.; Hong, W. Control of Quantum Interference in Single-Molecule Junctions via Jahn-Teller Distortion. *Cell Rep. Phys. Sci.* **2021**, *2*, No. 100329.

(4) Camarasa-Gómez, M.; Hernangómez-Pérez, D.; Inkpen, M. S.; Lovat, G.; Fung, E.-D.; Roy, X.; Venkataraman, L.; Evers, F. Mechanically-Tunable Quantum Interference in Ferrocene-Based Single-Molecule Junctions. *Nano Lett.* **2020**, *20*, 6381−6386.

(5) Pal, A. N.; Li, D.; Sarkar, S.; Chakrabarti, S.; Vilan, A.; Kronik, L.; Smogunov, A.; Tal, O. Nonmagnetic Single-Molecule Spin-Filter Based on Quantum Interference. *Nat. Commun.* **2019**, *10*, No. 5565.

(6) Qiu, S.; Miao, Y.-Y.; Zhang, G.-P.; Ren, J.; Wang, C.; Hu, G.-C. Manipulating Current Spin Polarization in Magnetic Single-Molecule Junctions via Destructive Quantum Interference. *J. Phys. Chem. C* **2020**, *124*, 12144−12152.

(7) Leary, E.; Rosa, A. L.; González, M. T.; Rubio-Bollinger, G.; Agraït, N.; Martín, N. Incorporating Single Molecules into Electrical Circuits. The Role of the Chemical Anchoring Group. *Chem. Soc. Rev.* **2015**, *44*, 920−942.
(8) Zeng, B.-F.; Wang, G.; Qian, Q.-Z.; Chen, Z.-X.; Zhang, X.-G.; Lu, Z.-X.; Zhao, S.-Q.; Feng, A.-N.; Shi, J.; Yang, Y.; Hong, W. Selective Fabrication of Single-Molecule Junctions by Interface Engineering. *Small* **2020**, *16*, No. 2004720.
(9) Isshiki, Y.; Fujii, S.; Nishino, T.; Kiguchi, M. Fluctuation in Interface and Electronic Structure of Single-Molecule Junctions Investigated by Current versus Bias Voltage Characteristics. *J. Am. Chem. Soc.* **2018**, *140*, 3760−3767.
(10) Xu, B.; Tao, N. J. Measurement of Single-Molecule Resistance by Repeated Formation of Molecular Junctions. *Science* **2003**, *301*, 1221−1223.
(11) McNeely, J.; Miller, N.; Pan, X.; Lawson, B.; Kamenetska, M. Angstrom-Scale Ruler Using Single Molecule Conductance Signatures. *J. Phys. Chem. C* **2020**, *124*, 13427−13433.
(12) Medina, S.; García-Arroyo, P.; Li, L.; Gunasekaran, S.; Stuyver, T.; José Mancheño, M.; Alonso, M.; Venkataraman, L.; L Segura, J.; Cordón, J. C. Single-Molecule Conductance in a Unique Cross-Conjugated Tetra(Aminoaryl)-Ethene. *Chem. Commun.* **2021**, *57*, 591−594.
(13) Schmidt, M.; Wassy, D.; Hermann, M.; Teresa Gonzalez, M.; Agrait, N.; Angela Zotti, L.; Esser, B.; Leary, E. Single-Molecule Conductance of Dibenzopentalenes: Antiaromaticity and Quantum Interference. *Chem. Commun.* **2021**, *57*, 745−748.
(14) Inkpen, M. S.; Lemmer, M.; Fitzpatrick, N.; Milan, D. C.; Nichols, R. J.; Long, N. J.; Albrecht, T. New Insights into Single-Molecule Junctions Using a Robust, Unsupervised Approach to Data Collection and Analysis. *J. Am. Chem. Soc.* **2015**, *137*, 9971−9981.
(15) Li, H. B.; Xi, Y.-F.; Hong, Z.-W.; Yu, J.; Li, X.-X.; Liu, W.-X.; Domulevicz, L.; Jin, S.; Zhou, X.-S.; Hihath, J. Temperature-Dependent Tunneling in Furan Oligomer Single-Molecule Junctions. *ACS Sens.* **2021**, *6*, 565−572.
(16) Wang, Y.-H.; Yan, F.; Li, D.-F.; Xi, Y.-F.; Cao, R.; Zheng, J.-F.; Shao, Y.; Jin, S.; Chen, J.-Z.; Zhou, X.-S. Enhanced Gating Performance of Single-Molecule Conductance by Heterocyclic Molecules. *J. Phys. Chem. Lett.* **2021**, *12*, 758−763.
(17) Martin, C. A.; Ding, D.; van der Zant, H. S. J.; van Ruitenbeek, J. M. Lithographic Mechanical Break Junctions for Single-Molecule Measurements in Vacuum: Possibilities and Limitations. *New J. Phys.* **2008**, *10*, No. 065008.
(18) Huisman, E. H.; Trouwborst, M. L.; Bakker, F. L.; van Wees, B. J.; van der Molen, S. J. The Mechanical Response of Lithographically Defined Break Junctions. *J. Appl. Phys.* **2011**, *109*, No. 104305.
(19) Vrouwe, S. A. G.; van der Giessen, E.; van der Molen, S. J.; Dulic, D.; Trouwborst, M. L.; van Wees, B. J. Mechanics of Lithographically Defined Break Junctions. *Phys. Rev. B* **2005**, *71*, No. 035313.
(20) Muller, C. J.; van Ruitenbeek, J. M.; de Jongh, L. J. Conductance and Supercurrent Discontinuities in Atomic-Scale Metallic Constrictions of Variable Width. *Phys. Rev. Lett.* **1992**, *69*, 140−143.
(21) Liu, Y.; Ornago, L.; Carlotti, M.; Ai, Y.; El Abbassi, M.; Soni, S.; Asyuda, A.; Zharnikov, M.; van der Zant, H. S. J.; Chiechi, R. C. Intermolecular Effects on Tunneling through Acenes in Large-Area and Single-Molecule Junctions. *J. Phys. Chem. C* **2020**, *124*, 22776−22783.
(22) Kobayashi, S.; Kaneko, S.; Kiguchi, M.; Tsukagoshi, K.; Nishino, T. Tolerance to Stretching in Thiol-Terminated Single-Molecule Junctions Characterized by Surface-Enhanced Raman Scattering. *J. Phys. Chem. Lett.* **2020**, *11*, 6712−6717.
(23) Huang, C.; Jevric, M.; Borges, A.; Olsen, S. T.; Hamill, J. M.; Zheng, J.-T.; Yang, Y.; Rudnev, A.; Baghernejad, M.; Broekmann, P.; Petersen, A. U.; Wandlowski, T.; Mikkelsen, K. V.; Solomon, G. C.; Brøndsted Nielsen, M.; Hong, W. Single-Molecule Detection of Dihydroazulene Photo-Thermal Reaction Using Break Junction Technique. *Nat. Commun.* **2017**, *8*, No. 15436.
(24) Chen, L.-C.; Zheng, J.; Liu, J.; Gong, X.-T.; Chen, Z.-Z.; Guo, R.-X.; Huang, X.; Zhang, Y.-P.; Zhang, L.; Li, R.; Shao, X.; Hong, W.; Zhang, H.-L. Nonadditive Transport in Multi-Channel Single-Molecule Circuits. *Small* **2020**, *16*, No. 2002808.
(25) Li, Z.; Mejía, L.; Marrs, J.; Jeong, H.; Hihath, J.; Franco, I. Understanding the Conductance Dispersion of Single-Molecule Junctions. *J. Phys. Chem. C* **2021**, *125*, 3406−3414.
(26) Martin, C. A.; Ding, D.; Sørensen, J. K.; Bjørnholm, T.; van Ruitenbeek, J. M.; van der Zant, H. S. J. Fullerene-Based Anchoring Groups for Molecular Electronics. *J. Am. Chem. Soc.* **2008**, *130*, 13198−13199.
(27) Kamenetska, M.; Koentopp, M.; Whalley, A. C.; Park, Y. S.; Steigerwald, M. L.; Nuckolls, C.; Hybertsen, M. S.; Venkataraman, L. Formation and Evolution of Single-Molecule Junctions. *Phys. Rev. Lett.* **2009**, *102*, No. 126803.
(28) Harzmann, G. D.; Frisenda, R.; van der Zant, H. S. J.; Mayor, M. Single-Molecule Spin Switch Based on Voltage-Triggered Distortion of the Coordination Sphere. *Angew. Chem., Int. Ed.* **2015**, *54*, 13425−13430.
(29) Li, Z.; Smeu, M.; Park, T.-H.; Rawson, J.; Xing, Y.; Therien, M. J.; Ratner, M. A.; Borguet, E. Hapticity-Dependent Charge Transport through Carbodithioate-Terminated [5,15-Bis(Phenylethynyl)-Porphinato]Zinc(II) Complexes in Metal−Molecule−Metal Junctions. *Nano Lett.* **2014**, *14*, 5493−5499.
(30) Moreno-García, P.; Gulcur, M.; Manrique, D. Z.; Pope, T.; Hong, W.; Kaliginedi, V.; Huang, C.; Batsanov, A. S.; Bryce, M. R.; Lambert, C.; Wandlowski, T. Single-Molecule Conductance of Functionalized Oligoynes: Length Dependence and Junction Evolution. *J. Am. Chem. Soc.* **2013**, *135*, 12228−12240.
(31) Hong, W.; Manrique, D. Z.; Moreno-García, P.; Gulcur, M.; Mishchenko, A.; Lambert, C. J.; Bryce, M. R.; Wandlowski, T. Single Molecular Conductance of Tolanes: Experimental and Theoretical Study on the Junction Evolution Dependent on the Anchoring Group. *J. Am. Chem. Soc.* **2012**, *134*, 2292−2304.
(32) Yoo, P. S.; Kim, T. Linker-Dependent Junction Formation Probability in Single-Molecule Junctions. *Bull. Korean Chem. Soc.* **2015**, *36*, 265−268.
(33) Vladyka, A.; Perrin, M. L.; Overbeck, J.; Ferradás, R. R.; García-Suárez, V.; Gantenbein, M.; Brunner, J.; Mayor, M.; Ferrer, J.; Calame, M. In-Situ Formation of One-Dimensional Coordination Polymers in Molecular Junctions. *Nat. Commun.* **2019**, *10*, No. 262.
(34) Cabosart, D.; El Abbassi, M.; Stefani, D.; Frisenda, R.; Calame, M.; van der Zant, H. S. J.; Perrin, M. L. A Reference-Free Clustering Method for the Analysis of Molecular Break-Junction Measurements. *Appl. Phys. Lett.* **2019**, *114*, No. 143102.
(35) Frisenda, R.; Stefani, D.; van der Zant, H. S. J. Quantum Transport through a Single Conjugated Rigid Molecule, a Mechanical Break Junction Study. *Acc. Chem. Res.* **2018**, *51*, 1359−1367.
(36) Lachmanová, Š. N.; Kolivoška, V.; Šebera, J.; Gasior, J.; Mészáros, G.; Dupeyre, G.; Lainé, P. P.; Hromadová, M. Environmental Control of Single-Molecule Junction Evolution and Conductance: A Case Study of Expanded Pyridinium Wiring. *Angew. Chem., Int. Ed.* **2021**, *60*, 4732−4739.
(37) Metzger, R. M. Unimolecular Electronics. *Chem. Rev.* **2015**, *115*, 5056−5115.
(38) Kim, B.; Choi, S. H.; Zhu, X.-Y.; Frisbie, C. D. Molecular Tunnel Junctions Based on π-Conjugated Oligoacene Thiols and Dithiols between Ag, Au, and Pt Contacts: Effect of Surface Linking Group and Metal Work Function. *J. Am. Chem. Soc.* **2011**, *133*, 19864−19877.
(39) Rodriguez-Gonzalez, S.; Xie, Z.; Galangau, O.; Selvanathan, P.; Norel, L.; Van Dyck, C.; Costuas, K.; Frisbie, C. D.; Rigaut, S.; Cornil, J. HOMO Level Pinning in Molecular Junctions: Joint Theoretical and Experimental Evidence. *J. Phys. Chem. Lett.* **2018**, *9*, 2394−2403.
(40) Van Dyck, C.; Geskin, V.; Cornil, J. Fermi Level Pinning and Orbital Polarization Effects in Molecular Junctions: The Role of Metal Induced Gap States. *Adv. Funct. Mater.* **2014**, *24*, 6154−6165.
(41) Xie, Z.; Bâldea, I.; Frisbie, C. D. Determination of Energy-Level Alignment in Molecular Tunnel Junctions by Transport and

Spectroscopy: Self-Consistency for the Case of Oligophenylene Thiols and Dithiols on Ag, Au, and Pt Electrodes. *J. Am. Chem. Soc.* **2019**, *141*, 3670−3681.
(42) Heimel, G.; Romaner, L.; Brédas, J.-L.; Zojer, E. Interface Energetics and Level Alignment at Covalent Metal-Molecule Junctions: π-Conjugated Thiols on Gold. *Phys. Rev. Lett.* **2006**, *96*, No. 196806.
(43) Wu, B. H.; Ivie, J. A.; Johnson, T. K.; Monti, O. L. A. Uncovering Hierarchical Data Structure in Single Molecule Transport. *J. Chem. Phys.* **2017**, *146*, No. 092321.
(44) Korshoj, L. E.; Afsari, S.; Chatterjee, A.; Nagpal, P. Conformational Smear Characterization and Binning of Single-Molecule Conductance Measurements for Enhanced Molecular Recognition. *J. Am. Chem. Soc.* **2017**, *139*, 15420−15428.
(45) Hamill, J. M.; Zhao, X. T.; Mészáros, G.; Bryce, M. R.; Arenz, M. Fast Data Sorting with Modified Principal Component Analysis to Distinguish Unique Single Molecular Break Junction Trajectories. *Phys. Rev. Lett.* **2018**, *120*, No. 016601.
(46) Lauritzen, K. P.; Magyarkuti, A.; Balogh, Z.; Halbritter, A.; Solomon, G. C. Classification of Conductance Traces with Recurrent Neural Networks. *J. Chem. Phys.* **2018**, *148*, No. 084111.
(47) Huang, F.; Li, R.; Wang, G.; Zheng, J.; Tang, Y.; Liu, J.; Yang, Y.; Yao, Y.; Shi, J.; Hong, W. Automatic Classification of Single-Molecule Charge Transport Data with an Unsupervised Machine-Learning Algorithm. *Phys. Chem. Chem. Phys.* **2020**, *22*, 1674−1681.
(48) Magyarkuti, A.; Balogh, N.; Balogh, Z.; Venkataraman, L.; Halbritter, A. Unsupervised Feature Recognition in Single Molecule Break Junction Data. *Nanoscale* **2020**, *12*, 8355−8363.
(49) Vladyka, A.; Albrecht, T. Unsupervised Classification of Single-Molecule Data with Autoencoders and Transfer Learning. *Mach. Learn.: Sci. Technol.* **2020**, *1*, No. 035013.
(50) Bamberger, N. D.; Ivie, J. A.; Parida, K.; McGrath, D. V.; Monti, O. L. A. Unsupervised Segmentation-Based Machine Learning as an Advanced Analysis Tool for Single Molecule Break Junction Data. *J. Phys. Chem. C* **2020**, *124*, 18302−18315.
(51) Liu, B.; Murayama, S.; Komoto, Y.; Tsutsui, M.; Taniguchi, M. Dissecting Time-Evolved Conductance Behavior of Single Molecule Junctions by Non-Parametric Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 6567−6572.
(52) Lin, L.; Tang, C.; Dong, G.; Chen, Z.; Pan, Z.; Liu, J.; Yang, Y.; Shi, J.; Ji, R.; Hong, W. Spectral Clustering to Analyze the Hidden Events in Single-Molecule Break Junctions. *J. Phys. Chem. C* **2021**, *125*, 3623−3630.
(53) El Abbassi, M.; Overbeck, J.; Braun, O.; Calame, M.; van der Zant, H. S. J.; Perrin, M. L. Benchmark and Application of Unsupervised Classification Approaches for Univariate Data. *Commun. Phys.* **2021**, *4*, No. 50.
(54) Lemmer, M.; Inkpen, M. S.; Kornysheva, K.; Long, N. J.; Albrecht, T. Unsupervised Vector-Based Classification of Single-Molecule Charge Transport Data. *Nat. Commun.* **2016**, *7*, No. 12922.
(55) Albrecht, T.; Slabaugh, G.; Alonso, E.; Al-Arif, S. M. R. Deep Learning for Single-Molecule Science. *Nanotechnology* **2017**, *28*, No. 423001.
(56) Quek, S. Y.; Venkataraman, L.; Choi, H. J.; Louie, S. G.; Hybertsen, M. S.; Neaton, J. B. Amine−Gold Linked Single-Molecule Circuits: Experiment and Theory. *Nano Lett.* **2007**, *7*, 3477−3482.
(57) Jang, S.-Y.; Reddy, P.; Majumdar, A.; Segalman, R. A. Interpretation of Stochastic Events in Single Molecule Conductance Measurements. *Nano Lett.* **2006**, *6*, 2362−2367.
(58) Brooke, R. J.; Szumski, D. S.; Vezzoli, A.; Higgins, S. J.; Nichols, R. J.; Schwarzacher, W. Dual Control of Molecular Conductance through PH and Potential in Single-Molecule Devices. *Nano Lett.* **2018**, *18*, 1317−1322.
(59) Halbritter, A.; Makk, P.; Mackowiak, Sz.; Csonka, Sz.; Wawrzyniak, M.; Martinek, J. Regular Atomic Narrowing of Ni, Fe, and V Nanowires Resolved by Two-Dimensional Correlation Analysis. *Phys. Rev. Lett.* **2010**, *105*, No. 266805.
(60) Yanson, A. I.; Bollinger, G. R.; van den Brom, H. E.; Agraït, N.; van Ruitenbeek, J. M. Formation and Manipulation of a Metallic Wire of Single Gold Atoms. *Nature* **1998**, *395*, 783−785.
(61) Ohnishi, H.; Kondo, Y.; Takayanagi, K. Quantized Conductance through Individual Rows of Suspended Gold Atoms. *Nature* **1998**, *395*, 780−783.
(62) Johnson, T. K.; Ivie, J. A.; Jaruvang, J.; Monti, O. L. A. Fast Sensitive Amplifier for Two-Probe Conductance Measurements in Single Molecule Break Junctions. *Rev. Sci. Instrum.* **2017**, *88*, No. 033904.
(63) González, M. T.; Leary, E.; García, R.; Verma, P.; Herranz, M. Á.; Rubio-Bollinger, G.; Martín, N.; Agraït, N. Break-Junction Experiments on Acetyl-Protected Conjugated Dithiols under Different Environmental Conditions. *J. Phys. Chem. C* **2011**, *115*, 17973−17978.
(64) Jones, D. S. *Elementary Information Theory*; Oxford University Press, 1979.

https://doi.org/10.1021/acs.jpcc.1c04794
J. Phys. Chem. C 2021, 125, 18297−18307