Imperial College London

Machine-learning-enhanced charge transport studies of organometallic molecules

By

Mario Lemmer

A dissertation submitted to

The Department of Chemistry

The Imperial College of Science, Technology, and Medicine

In partial fulfilment of the requirements for the degree of

Doctor of Philosophy

May 2017

London, United Kingdom

ABSTRACT

Molecular electronics is an intriguing area of research and it has received a considerable amount of attention in the recent years. This attention is partially fuelled by possible application of single molecules as microelectronic devices and by fundamental research about charge transport processes. In many areas of molecular electronics, large data sets are recorded to account for the large fluctuations on the nanoscale.

Over the course of this research project, a new, machine-learning-based approach to data analysis was developed, namely MPVC. It is demonstrated herein, how this fundamentally new approach to data analysis can provide greater insight into complex single molecule conductance data. At the same time, this approach does not make assumptions about the data.

The new analysis approach was then applied to single molecule conductance data of organometallic molecules. These types of molecules have received considerable attention lately because of their potential applications as molecular wires or single molecular switches. It is demonstrated how MPVC analysis can provide additional insights into these data.

Finally, results on single molecular conductance data for four ring-shaped, organometallic compounds will be presented. This is the first time that single molecule conductance data has been reported on for ring-shaped compounds. Ring-shaped molecules are of particular interest for the molecular electronics community because of their potential application as single molecular switches. Their spatially separated current pathways could provide a testbed for charge transport studies with a special focus on quantum interference.

Overall, the work presented within this thesis demonstrates the use of machine learning techniques in molecular electronics and provides novel insights into the charge transport on the nanoscale with a special focus on organometallics and is thus a significant contribution to the field.

DECLARATION

I, Mario Lemmer, hereby state that the work described in this thesis is original and was carried out by myself. The contributions of others are cited or acknowledged. The work in this dissertation has not been submitted before to fulfil any degree requirements at this or any other university.

Copyright Notice

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my research supervisor Dr Tim Albrecht. I am thankful for the privilege of joining your research group and I have immensely profited from the experience. Over the course of the PhD I have valued your constructive advice, patience, support, scientific expertise and trust. Thank you also for the, certainly memorable, group trip to Scotland!

During the PhD I had the privilege of working with two of Imperials finest postdocs. My deepest appreciation goes to Michael Inkpen. I had the good fortune of working with Mike for almost two years and benefited very much from the experience. I learned much about molecular electronics from you and appreciate the time you took to teach me. Thank you also for the double-burger-nights, which are some of my fondest memories of London!

I am also very grateful to Eva Pensa. I have learned a lot from your experience and it was an absolute pleasure working in the lab with you! Thank you for all your help, the good times and for welcoming me into your flat! Thank you also to all my colleagues in the Albrecht and Long Groups. You made the time at Imperial special and enjoyable! Special thanks go to Lucy for the analytes, doughnuts, and good times! I had the good luck of working with a number of fantastic co-workers/friends. Thank you to Fahmida, Mariana, Max Ili, Harvey, Leo, Anil, Silvia, Raquel, Dan, Billy and Pippa for making the time at Imperial special.

I am also thankful to The Leverhulme Trust for funding my research project and my time at Imperial. One of my most valuable non-scientific experiences at Imperial I owe to Karuna from the Imperial College Chaplaincy. Thank you for helping me to grow as a human being.

I also would like to express my deepest appreciation for my parents and sisters. You have been very supportive my whole life and especially during the PhD. When times get tough, I know I can rely on you to be there with empathetic ears and kind words. Special thanks go to my dad for scientific discussion and continuing support.

Last but certainly not least, I am deeply indebted to my wife Steph. I sincerely appreciate you taking the bold step into the unknown and moving to a new country with me. You lifted my spirit when the PhD seemed overwhelming and you believed in me when I struggled. You are a constant source of happiness in my life and without you this endeavour would not have been possible. I dedicate this thesis to you.

LIST OF PUBLICATIONS

- "New insights into single-molecule junctions using a robust, unsupervised approach to data collection and analysis", Michael S. Inkpen, Mario Lemmer, Nathan Fitzpatrick, David Costa-Milan, Richard J. Nichols, Nicholas J. Long and Tim Albrecht, in *Journal of the American Chemical Society*, 2015
- "Unsupervised vector-based classification of single-molecule charge transport data", Mario Lemmer, Michael S. Inkpen, Katja Kornysheva, Nicholas J. Long, and Tim Albrecht in *Nature Communications,* 2016
- "Single-Molecule Conductance Studies of Organometallic Complexes Bearing 3-Thienyl Contacting Groups", Sören Bock, Oday Al-Owaedi, Samantha Eaves, David Costa-Milan, Mario Lemmer, Brian Skelton, Henrry Osorio, Richard J. Nichols, Simon J. Higgins, Pilar Cea, Nicholas J. Long, Tim Albrecht, Santiago Martín, Colin J. Lambert, and Paul J. Low in *Chemistry - A European Journal*, 2017

TABLE OF CONTENTS

Chapter 1 Introduction 21

1.1 Motivation Molecular Electronics 22

1.2 Scanning Tunnelling Microscope 23

1.2.1 Quantum Mechanical Tunnelling 24

1.2.2 STM Concept 26

1.3 Single Molecule Conductance Measurements 27

1.3.1 STM-I(s) Technique 28

1.3.2 STM Break-Junction 29

1.3.3 Data Analysis 29

1.4 Charge Transport 30

1.4.1 Conductance Quantisation 30

1.4.2 Charge Transport Mechanisms 31

1.4.3 Quantum Interference 34

1.4.4 Organometallics in Molecular Electronics 35

1.5 Electrochemistry 36

1.5.1 The Three-Electrode Electrochemical Cell 36

1.5.2 Electrode Processes 37

1.5.3 Cyclic Voltammetry 37

1.5.4 Electrochemical STM 39

1.6 Machine Learning in Single Molecular Electronics 40

1.6.1 Dimensionality Reduction 40

1.6.2 Clustering and Data Selection 41

1.7 References 45

Chapter 2 Materials and Methods 50

2.1 Molecules under Study 51

2.1.1 Reference Systems 51

2.1.2 Ferrocene Systems 52

2.1.3 Ruthenium Systems 54

2.2 Scanning Tunnelling Microscope 55

2.2.1 Scanning Tunnelling Microscope General 55

2.2.2 Substrate and Cleaning Procedure 56

2.2.3 STM Tip Preparation 57

2.2.4 STM Data Collection 58

2.2.5 Simulated I(s) Data 60

2.2.6 Data Analysis Considerations 63

2.3 Electrochemistry 63

2.3.1 General....63
2.3.2 Procedure for Solution CV....64
2.3.3 Procedure for Surface Immobilized CV....64
2.3.4 Method for Single Crystal Voltammetry....65
2.4 References....66

Chapter 3 Multivariate Pattern Analysis....67
3.1 Motivation for Vector-Based Analysis....68
3.2 Analysis with MPVC....68
3.3 MPVC Analysis of Simulated Data....71
3.4 MPVC Analysis of ODT....74
3.5 MPVC Analysis of OPE....79
3.6 Summary and Conclusion....82
3.7 References....83

Chapter 4 Ferrocene Compounds....85
4.1 Introduction Ferrocene Compounds....86
4.2 Linear Ferrocene Compounds....87
4.2.1 Conductance of the Au-(s-FDTM)-Au Junction....87
4.2.2 Discussion and Summary s-FDTM....93
4.2.3 Conductance of the Au-(s-FDT)-Au Junction....94
4.2.4 Discussion and Summary s-FDT....98
4.2.5 Conductance of the Au-(s-FPT)-Au Junction....99
4.2.6 Discussion and Summary s-FPT....103
4.3 Double-Branched Ferrocene....104
4.4 Biferrocene Compounds....105
4.4.1 BFTM Introduction....105
4.4.2 Electrochemistry of BFTM....106
4.4.3 Conductance of the Au-(BFTM)-Au Junction....107
4.4.4 Discussion and Summary BFTM....115
4.4.5 Introduction BFP....116
4.4.6 Electrochemistry of BFP....116
4.4.7 Conductance of the Au-(BFP)-Au Junction....117
4.4.8 Discussion and Summary BFP....123
4.5 Summary and Conclusion....125
4.6 References....127

Chapter 5 Ruthenium Macrocycles....129
5.1 Introduction....130
5.2 Conductance of the Au-(T-Ru_2M)-Au Junction....131

5.2.1 T-Ru_2M General....131
5.2.2 T-Ru_2M at 0.5 V and 8 nA....131
5.2.3 Cross Validation with Plateau Finder....135
5.2.4 Control Experiment....135
5.2.5 Discussion and Summary....136
5.3 Conductance of the Au-(sPy-Ru_4M)-Au Junction....137
5.3.1 Single Molecule Conductance at 0.5 V....138
5.3.2 Plateau Finder Analysis....142
5.3.3 Control Experiment....143
5.3.4 Conductance Series....144
5.3.5 Discussion and Summary....145
5.4 Conductance of the Au-(lPy-Ru_4M)-Au Junction....146
5.4.1 lPy-Ru_4M General....146
5.4.2 I(s) Spectroscopy....146
5.4.3 Control Experiment....152
5.4.4 lPy-Ru_4M Discussion and Summary....154
5.5 T-Ru_4M....155
5.5.1 T-Ru_4M General....155
5.5.2 T-Ru_4M Electrochemistry....155
5.5.3 T-Ru_4M Spectroscopy under Argon....157
5.5.4 T-Ru_4M in Air....164
5.5.5 T-Ru_4M - Discussion and Summary....167
5.6 Ruthenium Macrocycles: Summary and Discussion....168
5.7 References....170
Chapter 6 Conclusion and Outlook....172
Appendices....176
Appendix I: Additional Figures....177
Appendix II: MATLAB Scripts....187
Appendix III: Permission Documents....194

LIST OF FIGURES

Figure 1-1: Citations of reference 16 (Aviram and Ratner, "Molecular Rectifiers", 1974) as function of year. First publication of STM was in 1982. In 2003, Tao and Xu reported on the conductance of single molecules (dashed line). Citation record from ISI Web of Knowledge. *24*

Figure 1-2: Band diagram of STM tip and substrate at four different conditions: ***A****) Large distance between tip and substrate.* ***B****) Tunnelling distance, no bias.* ***C****) Forward bias between tip and substrate.* ***D****) Reverse bias between tip and substrate.* *25*

Figure 1-3: Schematic of the STM tip rastering over a substrate surface. Adatoms on the surface are matched by z-piezo adjustment of the tip. *26*

Figure 1-4: Frame of "A Boy and his Atom". The short film was created by IBM in 2013, by moving single atoms. Courtesy of International Business Machines Corporation, © (2013) International Business Machines Corporation.[24] *27*

Figure 1-5: Schematic of the STM-I(s) technique with 1,8-Octanedithiol (ODT) in the junction. Three stages of junction formation are shown: 1) Tip at initial distance 2) Molecular junction formed. 3) Rupture of the junction. *29*

Figure 1-6: Band diagram of a molecular junction. VL is the vacuum level, T the transmission probability, E_F *the Fermi level and* Φ *the barrier height.* *32*

Figure 1-7: Mach-Zehnder resonance in a para- (left) and meta- (right) substituted benzene ring. Constructive (left) and destructive (right) interference is observed for the two cases. Arrows indicated spatially separated current paths. *35*

Figure 1-8: Band diagrams explaining two different charge transport models for electron transport across a redox active molecule. ***A****) resonant tunnelling and* ***B****) a two-step electron transport. Reprinted with permission from Guo, 2013.*[77] *36*

Figure 1-9: Schematic of three-electrode electrochemical cell.[78] *36*

Figure 1-10: ***A****) Electrical circuit of the cell.* ***B****) Applied voltage as function of time for cyclic voltammetry.* ***C****) Current response of an ideal system without Faradaic current.* ***D****) Double layer at the electrode-electrolyte interface.*[78,81] *37*

Figure 1-11: Cyclic voltammogram of ferricyanide. Data from Echem Analyst of Gamry Instruments inc, ... *38*

Figure 1-12: Tunnelling current as function of substrate potential. Bringing the redox level in resonance, between the electrode Fermi levels leads to an increase in tunnelling current. Figuure was reprinted with permission from Albrecht et al., 2005.[85] *39*

Figure 1-13: Positively correlated two-dimensional data is projected onto a new axis, u_1*, while minimizing the projection error. Projection error is indicated by dashed grey lines.* *41*

Figure 1-14: Example of an unlabelled data with two features, x and y. *41*

Figure 1-15: Different iterations of a k-means cluster process of the data from the previous figure. "x" indicate cluster centroids. Drawn by adapting an algorithm of the "Machine Learning" online course by Andrew Ng.[88] *43*

Figure 1-16: Experimental I(s) traces of OPE ***A****) with and* ***C****) without plateau.*[31] ***B****) and* ***D****) are the corresponding single-trace histograms as well as the plateau-determining bin count (dashed line, PDBC).* *44*

Figure 2-1: Structure of ODT *51*

Figure 2-2: Structure of OPE. *51*

Figure 2-3: Structures of s-FDTM, s-FDT and s-FPT. *52*

Figure 2-4: Structure of the double-branched ferrocene complex, b-FPT. *53*

Figure 2-5: Structure of BFTM *53*

Figure 2-6: Structure of BFP *53*

Figure 2-7: Structure of the ruthenium macrocycle, T-Ru_2*M. The structure was provided by Lucy E. Wilson.* *54*

Figure 2-8: Ruthenium Macrocycle (sPy-Ru_4M) with short pyridine linkers. Structure was used with permission from Stefan Scheerer. 54
Figure 2-9: Ruthenium macrocycle, lPy-Ru_4M, with long pyridine linkers. The structure was modified with permission from Stefan Scheerer. 55
Figure 2-10: Ruthenium Macrocycle with thioacetyl linkers. The structure was modified with permission from Stefan Scheerer. 55
Figure 2-11: Image of the 3D printed tip etching setup (left). Au wire is fed from a stock roll through a needle into the etching solution. The needle is connected to the working electrode of the bipotentiostat. Rack and pinion are used to lower the wire into solution and to remove the wire from solution after the etching process. The ring-shaped gold electrode in solution is connected to the counter electrode of the bipotentiostat. A schematic of the etching process is shown on the right. 57
Figure 2-12: Spectroscopy window of PicoView 1.14.4 during data collection. Left side of the image shows a typical "clean" I(s) trace. Subsection of window is not found because it is obscured by I(s) trace. This is indicated by the red inset. In the case of a "dirty" trace (right hand side), the subsection is found by the macro scheduler, as indicated by green inset. 59
Figure 2-13: Simulated I(s) trace with different noise levels of 0.1, 0.3 and 0.5. Dark to light green, traces were offset for clarity. 61
Figure 2-14: Simulated data, without plateau (top left in red) and plateaus at various plateau currents (top left in green). Different break-off distances (top right), I_0 (bottom left) and decay factors (bottom right). 62
Figure 2-15: Noise-free simulated data with plain plateau, telegraphic noise, and sine-shaped plateau, from left to right. Traces were offset for clarity. 63
Figure 2-16: Image of the STM cell setup for surface immobilized cyclic voltammetry on Au(111) single crystals (left). Right hand side of the figure is a schematic of the cross section of the STM cell along the dashed line. 65
Figure 3-1: Simulated I(s) traces with different plateaus are shown in different colours. Black traces are experimental traces with ODT in the junction. The dashed, magenta trace is a typical reference trace as described below. Traces were offset for clarity The figure was redrawn with permission from Lemmer et al., 2016.[4] 69
*Figure 3-2: Principle of MPVC - **A**) For illustration purposes I(s) traces contain only two current values, I_n and I_{n+1}. Every I(s) trace can thus be represented by a vector in two dimensions, $\mathbf{I}_m$. Similar traces have a similar distance, $|\mathbf{Y}_m|$ and angle, θ_m, to a common reference vector, **R**. Four different types of I(s) traces are shown in the graph in different colours. **B**) Illustrative polar plot showing that different types of I(s) traces form different clusters when ΔX and θ of many traces are plotted. The figure was redrawn with permission from Lemmer et al., 2016.*[4] 70
*Figure 3-3: **A**) Simulated I(s) traces with plain, telegraphic, sine-shaped and without plateau. Inset shows a magnification of the telegraphic plateau, emphasizing the binary switching. Traces are offset for clarity. **B**) Logarithmic 2D current histogram of all traces in the simulated data set. The figures were reproduced with permission from reference 4.* 72
Figure 3-4: Polar plot of the whole data set. Dimensionality was reduced with the MPVC approach. Data forms four distinct clusters corresponding to the four different types of I(s) traces. Colour assignment was done based on parameter distribution (vide infra). The figure was redrawn with permission from Lemmer et al., 2016.[4] 72
*Figure 3-5: Distribution-based clustering of the simulated data set. **A**) ΔX was fitted with a Gaussian distribution. Dashed lines are the 95 % confidence intervals. **B**) Second fit of ΔX. Shown in magenta are all traces assigned to the telegraph cluster. Shown in black are traces not yet assigned to any cluster. **C**) In a last clustering step, traces in the black cluster are assigned to two different clusters based on a fit of the θ with two Gaussian distributions. **D**) Histogram of h_r is shown for completeness. The figure was redrawn with permission from Lemmer et al., 2016.*[4] 73
*Figure 3-6: Logarithmic 2D current histograms of the four individual clusters. **A**) The plain exponential cluster. **B**) The cluster with traces with telegraphic signal in the plateau. **C**) Traces with plain plateau. **D**) Sine-shaped plateau. The figure was redrawn with permission from Lemmer et al., 2016.*[4] 74

Figure 3-7: Structure of ODT. *74*
*Figure 3-8: 1D **A**) and **B**) 2D current histograms without any data selection. **C**) shows the whole data set, reduced to three parameters per trace with MPVC. Figure was reprinted from Lemmer et al., 2016.*[4] *75*
*Figure 3-9: 1D current histogram of the red **A**) and blue cluster **C**). **B**) and **D**) show the logarithmic 2D histograms of the red and blue cluster, respectively. **E**) is the cylinder plot after cluster assignment with FCM using the GK algorithm. The figure was redrawn with permission from Lemmer et al., 2016.*[4] . *76*
Figure 3-10: Cylinder plot of the whole data set viewed along the h_r-axis. Colour coding was done according to cluster membership. Cluster assignment was performed with FCM using the GK algorithm. Figure was reprinted from Lemmer et al., 2016.[4] *77*
*Figure 3-11: Left column - 1D current histogram of the high **A**) and low **C**) conductance fraction of traces. Right column - 2D current histogram of the high **B**) and low **D**) conductance fraction of traces. **E**) Cylinder plot where the traces with 5 % highest (lowest) ΔX are shown in magenta (green). Figure was reprinted from Lemmer et al., 2016.*[4] *78*
Figure 3-12: Structure of OPE. *79*
*Figure 3-13: **A**) 1D and **B**) 2D current histogram of the whole data set without any data selection. **C**) Cylinder plot of the whole data set in the MPVC representation. Figure was reprinted from Lemmer et al., 2016.*[4] *79*
Figure 3-14: Cylinder plot of the whole data set where cluster membership is indicated by colour coding. Cluster assignment was determined with FCM using the GK algorithm. Figure was reprinted from Lemmer et al., 2016.[4] *80*
*Figure 3-15: 1D and logarithmic 2D current histograms of the blue cluster (**A** and **B**), red cluster (**C** and **D**) and green cluster (**E** and **F**). The blue cluster contains only plain exponential traces, red contains traces with high conductance plateaus. The green cluster contains I(s) traces with low conductance plateaus. Figure was reprinted from Lemmer et al., 2016.*[4] *81*
Figure 3-16: Flow chart of analysis processes. Typically, data sets are sorted by hand or automated by some criterion and the plateau-containing traces are analysed in histograms (left two branches). MPVC analysis is a conceptual departure, where traces are considered to be vectors. Dimensionality reduction is applied and then clustering algorithm split up the data into K meaningful subgroup. Individual subgroups are then analysed individually. Adapted with permission from reference 4. ..... *82*
Figure 4-1: Structures of linear ferrocenes s-FDTM, s-FDT and s-FPT. *86*
Figure 4-2: Structure of the double-branched ferrocene, b-FPT. *86*
Figure 4-3: Structure of biferrocene compounds, BFP and BFTM. *87*
Figure 4-4: Structure of s-FDTM. *87*
*Figure 4-5: **A**) Log 2D conductance histogram. **B**) Linear 1D conductance histogram of the whole data set without any selection. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) $0.13 \cdot 10^{-4}$ G_0.* *88*
Figure 4-6: Logarithmic point density of the data set. The point cloud from MPVC analysis was projected onto the ΔX/θ-plane and the point density was calculated with binning increments of 3.6° and 0.02 in angular and radial direction, respectively. The colour scheme corresponds to the exponent of the point density. *88*
Figure 4-7: Cylinder plot of the whole data set viewed along the h_r-axis. Colour assignment was done per cluster membership. Clustering was performed with FCM using the GK algorithm. *89*
*Figure 4-8: **A**) Logarithmic 2D conductance histogram of the red cluster. **B**) Linear 1D conductance histogram of the traces that contain conductance features. Plateau traces were selected per MPVC analysis using GK FCM. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) $0.13 \cdot 10^{-4}$ G_0* *89*
Figure 4-9: Representative selection of conductance traces in the red cluster. Traces are offset in the left panel and overlaid in the right panel with their median shown in red. *90*
*Figure 4-10: **A**) Event distribution of the plateau traces over the course of the experiment. **B**) Absolute (tip-substrate) break-off distance histogram of the red cluster. Shown in black is the initial tip-substrate separation. Bin width 1 Å.* *90*

*Figure 4-11: **A**) Logarithmic 2D conductance histogram of the predominantly plain exponential cluster. **B**) Linear 1D conductance histogram of the traces in the blue cluster. The members of the blue cluster contain predominantly plain exponential traces. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) $0.13 \cdot 10^{-4}$ G_0..... 91*

*Figure 4-12: **A**) Logarithmic 2D conductance histogram. **B**) Linear 1D conductance histogram of the traces that contained conductance features. Plateau traces were selected by plateau finder. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) $0.13 \cdot 10^{-4}$ G_0..... 92*

*Figure 4-13: Logarithmic 2D conductance histogram of the plateau traces identified by **A**) MPVC and **B**) plateau finder. Data in both histograms was filtered with a low-pass Bessel filter to improve visibility of the described effect. Binning: **A**) and **B**) 0.04 Å and 50 bins per decade. 92*

*Figure 4-14: Decay factor as a function of distance of the plateau traces. Plateau traces were found by **A**) MPVC and **B**) plateau finder. Bin widths: 0.2 Å and 0.03 $Å^{-1}$. 93*

Figure 4-15: Structure of s-FDT. 94

*Figure 4-16: Logarithmic **A**) 2D and **B**), **C**) logarithmic 1D conductance histogram of the data set recorded at -0.3 V. Data selection was made with the PF algorithm. Bin widths: **A**) 0.16 Å and 50 bins per decade in **A**) and **B**). Colours in **C**) correspond to estimated peak conductance with error intervals. 95*

Figure 4-17: Representative sample traces with estimated conductance (dashed lines) and error margins (coloured intervals). Traces were offset for increased clarity. 96

Figure 4-18: Event distribution of the plateau traces over the course of the experiment. 96

Figure 4-19: Conductance of s-FDT as function of applied bias voltage. Different colours represent different conductance classes. 97

Figure 4-20: Break-off distances as a function of bias voltage for s-FDT. 97

Figure 4-21: Structure of s-FPT. 99

*Figure 4-22: Logarithmic **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.15 V. Data selection was made with linear plateau finder algorithm. Bin widths: **A**) 0.04 Å and 50 bins per decade in **A**) and **B**). Data in **B**) was filtered with a low-pass Bessel filter. 100*

*Figure 4-23: **A**) Filtered logarithmic, **B**) filtered linear and **C**) unfiltered linear conductance histograms of plateau-containing traces. Histogram colour was set to black to improve contrast. Peak conductance is indicated by dashed lines, error margins by shaded intervals. Binning: 50 bins per decade for logarithmic histogram. Bin width for linear histogram: $0.01 \cdot 10^{-4}$ G_0. ..... 101*

Figure 4-24: Representative sample traces. Dashed lines correspond to the most likely conductance values as estimated from the logarithmic histogram in the previous figure. Coloured intervals correspond to the error margins associated with the conductance values. Traces were offset for clarity. 101

Figure 4-25: Summary of peak conductance values of s-FPT as determined from their log 1D conductance histograms. Different colours denote different conductance classes. 102

Figure 4-26: Break-off distances of s-FPT as a function of bias voltage. 102

*Figure 4-27: Example data set with long plateaus. Logarithmic **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.6 V. Data selection was made with linear plateau finder algorithm. Bin widths **A**) 0.16 Å and 50 bins per decade **B**) $6.5 \cdot 10^{-7}$ G_0. Data in **B**) was filtered with a low-pass Bessel filter. 103*

Figure 4-28: Comparison of peak conductance of s-FDT and s-FPT. s-FDT is shown as "x" and s-FPT as "o". Error bars were omitted for improved clarity. 104

Figure 4-29: Structure of the double-branched ferrocene molecule, b-FPT. 104

*Figure 4-30: **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.2 V and 50 nA. Data selection was made with linear plateau finder algorithm. Bin widths: **A**) 0.08 Å and 50 bins per decade in **A**) and **B**). Data in **A**) and **B**) was filtered with a low-pass Bessel filter. 105*

Figure 4-31: Structure of BFTM, a biferrocene molecule with thiomethyl linkers. 105

*Figure 4-32: **A**) Cyclic voltammograms with increasing scan rate (Inner to outer traces 25 mV/s to 600 mV/s). **B**) Peak current of the oxidation/reduction wave. **C**) Peak separation (squares) and equilibrium*

potentials (triangles) of the oxidation/reduction step. Graphs were corrected for solution resistance. Colour coding in all panels according to scan rate. Measurements were performed in degassed 0.1 M $TBAPF_6$/DCM under argon atmosphere. The same substrate was used for CV and for spectroscopy experiments..... 107

Figure 4-33: Cyclic voltammogram recorded at a scan rate of 25 mV/s..... 107

Figure 4-34: Cylinder plot of the whole data set. Colour assignment was done per cluster membership. Clustering was performed with GK FCM. Shown in red are plateau traces and data points in blue represent plain exponential traces. 108

*Figure 4-35: Logarithmic **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.1 V and 6.5 nA. Data selection was made with MPVC and GK FCM. Bin widths: **A**) 0.16 Å and 50 bins per decade in **A**) and **B**). Data was filtered with a low-pass Bessel filter..... 109*

Figure 4-36: Cylinder plot of the whole data set in the MPVC representation. Colour coding was done after cluster assignment with FCM clustering. The cylinder plot is viewed diagonally, along two orthogonal ΔX directions and along Hamming axis. 110

*Figure 4-37: **A**), **B**) 2D and **C**), **D**) 1D conductance histogram of the orange and purple cluster, respectively. Data selection was made with MPVC and GK FCM. Histograms in **C**) and **D**) were fitted with a Gaussian distribution. Bin widths: 0.16 Å and 50 bins per decade. Data was filtered with a low-pass Bessel filter..... 111*

Figure 4-38: Representative sample traces of both conductance groups. Traces were filtered with a low-pass Bessel filter and offset for clarity. Dashed lines are most probable conductance of both conductance classes as determined from Gaussian fit of 1D log histogram..... 112

*Figure 4-39: Event distribution over the course of the experiment of the **A**) orange low-conductance class and **B**) purple high-conductance class..... 112*

Figure 4-40: Break-off distance histograms of the orange and purple cluster. Shown in grey is the estimated initial tip-substrate distance. Bin width: 2 Å. 113

*Figure 4-41: Histograms of plateau traces recorded at various voltages. The first column contains the log 1D histograms and the second column the log 2D histograms. Bias voltages: **A**) and **B**) -0.2 V, **C**) and **D**) 0.4 V, **E**) and **F**) 0.5 V and **G**) and **H**) 0.6 V. Data selection was performed with MPVC for **E**) and **F**) and logarithmic PF for the rest. Note that data was filtered with a low-pass Bessel filter. Binning: 0.16 Å and 50 bins per decade. 114*

Figure 4-42: Structure of biferrocene with meta-substituted pyridine anchors, BFP. 116

Figure 4-43: Surface immobilised cyclic voltammogram of BFP in 0.1 M TBAHFP in DCM. Peak intensity decreased over time. Note that potential was not measured against a reference since the analyte disappeared before a reference could be added..... 117

*Figure 4-44: **A**) All data log 2D and **B**) log 1D conductance histogram recorded before incubation of the substrate. 2000 traces were recorded at 0.3 V and 20 nA setpoint current. 118*

*Figure 4-45: **A**) 2D and **B**) 1D conductance histogram of the whole data set recorded at 0.5 V and 32.5 nA. Bin widths: **A**) 0.16 Å and 50 bins per decade **B**) $2.6 \cdot 10^{-6}$ G_0. ..... 118*

Figure 4-46: Cylinder plot of the whole data set in the vector representation. The vector representation was produced with the MPVC approach. Shown in blue are plain exponential traces, green and purple "x" are members of high and low conductance cluster, respectively. Shown in black are noisy traces that are connected to the analyte, but do not provide further insight into the data set..... 119

*Figure 4-47: **A**), **B**) 2D and **C**), **D**) 1D conductance histogram of the low conductance purple, and high conductance green cluster, respectively. Data selection was made with MPVC and GC FCM. Bin widths: **A**) and **B**) 0.04 Å and 50 bins per decade.**C**) and **D**) $2.6 \cdot 10^{-6}$ G_0. Trace count: **A**) and **C**) 34. **B**) and **D**) 49..... 120*

Figure 4-48: Sample traces of the low conductance (purple) and high conductance (green) cluster. Dashed lines represent most likely conductance values. Traces were offset for clarity. 121

*Figure 4-49: Event distribution of the low conductance (**A**, purple) and high conductance (**B**, green) cluster over the course of the experiment..... 121*

*Figure 4-50: Absolute break-off distance histograms of the **A**) low conductance and **B**) high conductance cluster. Grey bars are the initial tip-substrate distances. Bin width: 2 Å 122*

*Figure 4-51: **A**), **B**) 2D and **C**), **D**) 1D conductance histogram of data sets recorded at 0.3 V and 0.9 V, respectively. Data selection was made with linear plateau finder. Bin widths: 0.08 Å and 50 bins per decade 123*

Figure 5-1: Structure of the ruthenium macrocycle with thiomethyl linkers. (T-Ru_2M) 130

Figure 5-2: Overview of tetraruthenium macrocycle and the variations in anchor groups. 130

Figure 5-3: Structure of the diruthenium macrocycle with thiomethyl linkers 131

*Figure 5-4: All data log 2D (**A**) and log 1D (**B**) conductance histogram, consisting of all 1950 recorded I(s) traces. Note that data in the log 1D histogram was filtered with a low-pass Bessel filter. Bin width: 0.16 Å and 50 bins per decade. 132*

Figure 5-5: Logarithmic point density of the cylinder plot viewed along the h_r-axis. ΔX was normalised for this plot. The point cloud from MPVC analysis was projected onto the $\Delta X/\theta$-plane and the point density was calculated with binning increments of 3.6° and 0.02 in angular and radial direction. ... 132

*Figure 5-6: Cylinder plot of the whole data set in the MPVC representation. Viewed diagonally, along 0°-axis, along 90°-axis and along h_r-axis. (**A** to **D**) Members of the plateau-containing cluster are coloured in red, members of the plain exponential cluster are shown in blue. 133*

*Figure 5-7: Log 2D (**A**) and log 1D (**B**) conductance histogram of the plateau cluster after MPVC analysis. Note that the data in the log 1D current histogram was filtered with a Bessel low-pass filter to account for the low plateau currents. Bin widths: 0.16 Å and 50 bins per decade 134*

*Figure 5-8: **A**) Event distribution and **B**) break-off distance histogram of the plateau-containing cluster. Bin width: 1 Å. 135*

*Figure 5-10: **A**) All data log 2D and **B**) log 1D conductance histogram recorded before incubation of the substrate. 1600 traces were recorded at 0.3 V and 20 nA setpoint current. Note that the data in the log 1D current histogram was filtered with a Bessel low-pass filter for comparison with the previous histograms. Bin widths: 0.16 Å and 50 bins per decade 136*

Figure 5-11: Structure of the short tetraruthenium macrocycle (sPy-Ru_4M), which was kindly provided by the group of Rainer Winter in Konstanz. 137

*Figure 5-12: **A**) Log 2D conductance and **B**) log 1D conductance histograms of the whole data set recorded at 0.5 V bias after incubation of the substrate with the analyte. The data in the log 1D histogram was filtered with a low-pass Bessel filter. Bin widths: 0.08 Å and 50 bins per decade 138*

Figure 5-13: Logarithmic point density of the whole data set in the MPVC representation, projected onto the $\theta/\Delta X$-plane. Triangle is an enlarged view of the white triangle in the polar plot. 139

Figure 5-14: Cylinder plot of the data set in the MPVC representation containing the whole data after cluster assignment with FCM GK. The plot is viewed along the h_r-axis 139

*Figure 5-15: Logarithmic **A**) 2D and **B**) 1D conductance histogram of the red cluster. This cluster contained predominantly traces with plateaus as determined by MPVC FCM with the GK algorithm. Data for both histograms were filtered with a low-pass Bessel filter. Bin widths: 0.08 Å and 50 bins per decade 140*

Figure 5-16: Representative sample traces of the red, plateau-containing cluster. Traces are offset for clarity. The dashed line represents the conductance calculated from the log 1D conductance histogram in the previous figure. 140

Figure 5-17: Cumulative count of the plateau cluster as a function of trace number during the experiment. 141

Figure 5-18: Histogram of the break-off distance on an absolute scale with zero being the substrate surface. The light red histogram is the initial distance at the setpoint current. The schematic on top of the histogram illustrates the relative and absolute scale. Bin width: 0.5 Å. 141

Figure 5-19: Cylinder plot of the whole data set. Clustering was performed by plateau finder algorithm. The polar plot is viewed along the h_r-axis. Red traces contain traces with plateaus, blue traces plain exponentially decaying I(s) traces. 142

*Figure 5-20: **A**) log 2D and **B**) log 1D conductance histogram of the control experiment. Both histograms have been filtered with a low-pass Bessel filter. Bin widths: 0.08 Å and 50 bins per decade 143*

Figure 5-21: Cylinder plot of the MPVC representation of the whole data set recorded in the control experiment. The plot is viewed along the h_r-axis..... 144

*Figure 5-22: **A**) Logarithmic 1D conductance histograms for data recorded at bias voltages from 0.5 V to -0.4 V. Histograms are offset for clarity and filtered with a low-pass Bessel filter. **B**) Peak conductance values of log 1D conductance histograms. Bin width: 50 bins per decade..... 145*

Figure 5-23:Structure of the tetraruthenium macrocycle with extended pyridine linkers. The molecule was kindly provided by the group of Rainer Winter in Konstanz, Germany. 146

*Figure 5-24: **A**) Log 2D and **B**) log 1D conductance histogram of the whole data set without any data selection. The log 2D conductance histogram is unfiltered to visualise the quantised nature of the current on this scale, the log 1D conductance histogram is filtered with a low-pass Bessel filter. Bin widths: 0.16 Å and 50 bins per decade 147*

Figure 5-25: Cylinder plot of the whole data set in the MPVC representation. Colour assignment was performed per cluster membership. Cluster membership was determined by FCM with the GK algorithm. All 5000 traces of the data set are displayed in the graph. 148

*Figure 5-26: **A**) 2D and **B**) 1D conductance histograms of the red cluster. This cluster contains predominantly traces with plateaus, as determined by MPVC. Data in **B**) was filtered with a low-pass Bessel filter. Bin widths: 0.16 Å and 50 bins per decade 148*

*Figure 5-27: **A**) Cumulative counts of the red cluster as a function of the trace number of the experiment. **B**) Break-off distance histogram of the red cluster. Bin width: 1 Å. 149*

Figure 5-28: Representative examples of high (left four traces in red) and low (green traces) conductance traces. The dashed line marks the conductance of the high conductance class ($2 \cdot 10^{-5}$ G_0). Traces were off-set for clarity. 150

Figure 5-29: Cylinder plot of the whole data set in the MPVC representation. Coloured in blue are the plain exponential traces, in semi-transparent red are plateau traces. Red "x" are the high conductance traces and green "x" are the low conductance traces selected in the previous figure. 151

*Figure 5-30: Log 2D histogram of the presumed high (**A**) and low (**B**) conductance group. Log 1D conductance histogram of the high (**C**) and low (**D**) conductance group as well as their break-off distance histograms (**E** and **F**). Bin width: 1 Å. Data was filtered with a low-pass Bessel filter. 152*

*Figure 5-31: Log 2D (**A**) and log 1D (**B**) conductance histograms of the control experiment. The Au(111) substrate was incubated in pure DCM without analyte before the experiment. Note that data in **B** was filtered with a low-pass Bessel filter to be comparable to the previous graphs..... 153*

*Figure 5-32:**A**) Event distribution of the control experiment. Plateaus were selected by the PF algorithm. **B**) Break-off histogram of plateaus in the control experiment. Bin width: 1 Å. 154*

Figure 5-33:Structure of T-Ru_4M (thioacetyl tetraruthenium macrocycle). A sample of the molecule was kindly provided by the group of Rainer Winter in Konstanz, Germany. Structure was used with permission from P. Anders. 155

*Figure 5-34: **A**) Cyclic voltammograms with increasing scan rate (Inner to outer traces). **B**) Peak current of the 0/+2 oxidation/reduction step. **C**) Peak separation (squares) and half-way potentials (triangles) of the 0/+2 oxidation/reduction step. Corrected for solution resistance. Colour coding in all panels according to scan rate. Measurements were performed in degassed 0.1 M TBAPF_6/DCM under argon atmosphere..... 157*

*Figure 5-35: Log 2D **A**) and log 1D **B**) conductance histograms of the whole data set without any data selection. Both histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm. Data in **B**) was filtered with a low-ass Bessel filter. 158*

Figure 5-36: Cylinder plot of the whole data set. Coloured in blue are predominantly plain exponential decays, coloured in red are mainly plateau-containing traces as determined by MPVC with FCM GK. 159

*Figure 5-37: **A**) Log 2D and **B**) log 1D conductance histogram of the red cluster. Note that the log 1D histogram has been filtered with a low-pass Bessel filter and the first 10 % of each trace were excluded to limit the magnitude of the setpoint peak. Binning: 50 bins per decade for conductance and 500 distance bins per 8 nm..... 160*

Figure 5-38: Sample G(s) traces of the red cluster on a semi-logarithmic scale. Traces are offset for clarity. Dashed lines indicate the peak conductance values from the log 1D conductance histogram. 160

*Figure 5-39: **A**) Event distribution of the red cluster over the course of the experiment. **B**) Break-off distance histogram of the red cluster. Bin width: 1 Å. 161*

*Figure 5-40: Log 2D (**A**) and log 1D (**B**) conductance histograms of data found by plateau finder. Note that the log 1D histogram has been filtered with a low-pass Bessel filter and the first 10 % of each trace were excluded to limit the magnitude of the "setpoint peak". Both histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm..... 162*

Figure 5-41: Cylinder plot of all traces in the data set. Colours are per cluster membership as determined by the plateau finder algorithm. 163

*Figure 5-42: Log 2D (**A**) and log 1D (**B**) conductance histograms of the control experiment after application of the plateau finder algorithm. 4 % of all traces were selected by the algorithm. Recorded at 0.3 V and 20 nA. Both histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm..... 163*

*Figure 5-43: Log 2D and log 1D conductance histograms of two different experiments where the low conductance group was observed. 2000 traces each were recorded at 0.2 V (**A** and **B**) and -0.4 V (**C** and **D**). Plateau traces were selected by the PF algorithm. Data in all histograms was filtered with a low-pass Bessel filter. All histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm. 165*

*Figure 5-44: Log 2D and log 1D conductance histograms of two different experiments where the high-conductance group was observed. 2000 traces each were recorded at 0.04 V (**A** and **B**) and 0.05 V (**C** and **D**). Plateau traces were selected by the PF algorithm. Data in all histograms was filtered with a low-pass Bessel filter. All histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm..... 166*

Figure 5-45: Break-off distance histogram of the high conductance class shown in the previous figure (0.04 V). Black bars correspond to the initial distance. Red bars to the absolute break-off distance. 167

Appendix

Figure App. 1: Cumulative counts of the traces found by the plateau finder algorithm. 177

Figure App. 2: Break-off distance histogram of the traces found by the plateau finder algorithm. ... 177

Figure App. 3: Break-off distance histograms with lognormal fit at various bias voltages for s-FDT. 178

Figure App. 4: Conductance histograms of s-FDT at different bias voltages. Traces were filtered with a low-pass Bessel filter. 179

Figure App. 5: Event distribution of s-FPT. 180

Figure App. 6: Conductance histograms of s-FPT at different bias voltages. Traces were filtered with a low-pass Bessel filter. 181

Figure App. 7: Break-off distances of experiments at positive bias with s-FPT. 182

Figure App. 8: Break-off distances of experiments at negative bias with s-FPT. 183

*Figure App. 9: **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.1 V and 6.5 nA. Data selection was made with MPVC and GC FCM. Bin widths: 0.16 Å and 50 bins per decade. Data was filtered with a low-pass Bessel filter. 183*

*Figure App. 10: log 2D (**A**) and log 1D (**B**) conductance histogram, consisting of all 1950 recorded I(s) traces. Note that data in the log 1D histogram was filtered with a low-pass Bessel filter. Histograms are made up of 43 of total 2000 traces. 184*

Figure App. 11: log 2D (A) and log 1D (B) conductance histogram, consisting of all 1950 recorded I(s) traces. Note that data in the log 1D histogram was filtered with a low-pass Bessel filter. Histograms are made up of 95 of total 2000 traces. Data was selected with the ln PF algorithm..... 184

Figure App. 12: Logarithmic point density of the cylinder plot projected onto the ΔX/θ-plane. The view is along the h_r-axis. The semi-circular cluster pointing towards the origin contains plain exponentials. The right diffuse end of this cluster overlaps with the plateau cluster. (50 radial and 50 angular bins) 185

Figure App. 13: Log 1D conductance histogram with Gaussian fit of the data set presented in the main text. Data was selected with FCM GK. 185

Figure App. 14: 1D log conductance histogram of the high (left) and low (right) conductance group. Data was filtered with a low-pass Bessel filter. 186

Figure App. 15: Event distribution of the control experiment. Plateaus were selected by the plateau finder algorithm..... 186

LIST OF TABLES

Table 1-1: Summary of temperature, voltage, and distance dependencies of different charge transport regimes.[41] *33*

Table 4-1: Summary of peak conductance and error margins of the experiment conducted at -0.3 V. *95*

Table 4-2: Approximate conductance values at different bias voltages. G_{low} *and* G_{high} *representing the limits of the error interval given by the full width half maximum of the Gaussian fit. Note that conductance values given are only a rough estimate as no well-defined conductance peaks were observed. The corresponding conductance histograms are shown in Figure 4-41.* *113*

Table 5-1: Summary of experimental conditions, conductance results and their junction formation probabilities for T-Ru_4*M.* **Experiment was performed under argon atmosphere.****JFP does not distinguish here between low and high conductance group found in the same data set as they were not clearly separable.* *155*

Appendix

Table App. 1: Summary of permission status of third party content. *194*

TABLE OF ABBREVIATIONS

λ	*wave length*	*ODT*	*1,8-Octanedithiol*
μ	*chemical potential*	*OPE*	*α,ω-dithiol terminated oligo phenylene(ethynylene)*
μ_r	*cluster centroid r*	*p*	*momentum*
A	*area*	*PCA*	*principle component analysis*
BJ	*break-junction*	*PDBC*	*plateau-determining bin count*
C	*capacitance*	*PF*	*plateau finder*
c-AFM	*conductive probe atomic force microscopy*	*q*	*elementary charge*
CE	*counter electrode*	*Q*	*charge*
CV	*cyclic voltammetry*	R_s	*resistance*
D_0	*diffusion coefficient*	*R*	*gas constant*
DCM	*dichloromethane*	$\boldsymbol{R}$	*reference vector*
DOS	*density of states*	*RE*	*reference electrode*
$E_{1/2}$	*half wave potential*	*roi*	*region of interest*
EC	*electrochemical*	*s*	*distance*
E_F	*Fermi energy*	s_0	*initial distance*
E_{vacuum}	*vacuum level*	*SAM*	*self-assembled monolayer*
F	*Faraday constant*	s_b	*break-off distance*
FCM	*fuzzy c-means*	s_{mol}	*molecular length*
FWHM	*full width at half maximum*	*STDEV*	*standard deviation*
G	*conductance*	*STM*	*scanning tunnelling microscope*
G_0	*conductance quantum*	$T(\varepsilon)$	*transmission probability*
GK	*Gustafson-Kessel*	*t*	*time*
h	*Planck constant*	*T*	*temperature*
HCl	*hydrochloric acid*	*THF*	*tetrahydrofuran*
HOMO	*highest occupied molecular orbital*	*V*	*bias voltage*

h_r	*reduced Hamming distance*	*WE*	*working electrode*
I	*current*	$\boldsymbol{Y}$	*distance vector*
I(s)	*current-distance*	β	*exponential decay factor*
I(t)	*current-time*	Γ	*surface concentration*
I_0	*setpoint current*	ΔE	*peak separation*
i_p	*peak current*	Δn	*excess charge*
I_p	*plateau current*	ΔX	*norm*
JFP	*junction formation probability*	ε	*electron energy*
LUMO	*Lowest unoccupied molecular orbital*	θ	*angle*
m	*electron mass*	*v*	*scan rate*
MCBJ	*mechanically controlled break junction*	v_e	*drift velocity*
MPVC	*multi-parameter vector-based classification*	σ	*sigma*
n	*number of observations*	Φ	*energy barrier*

Chapter 1

Introduction

1.1 Motivation Molecular Electronics 22
1.2 Scanning Tunnelling Microscope 23
1.3 Single Molecule Conductance Measurements 27
1.4 Charge Transport 30
1.5 Electrochemistry 36
1.6 Machine Learning in Single Molecular Electronics 40
1.7 References 45

Synopsis: *In this chapter, a brief introduction into molecular electronics will be given, followed by the theoretical foundations of the Scanning Tunnelling Microscope (STM), single molecular conductance measurements as well as the charge transport on the nanoscale. Finally, at the end of the chapter, an overview of relevant electrochemistry and machine learning theory will be given.*

1.1 Motivation Molecular Electronics

Molecular electronics is a field of research, which is concerned with molecules that either singly, or in parallel monolayer arrays, act as active or passive electronic components.[1] Molecular electronics is a highly interdisciplinary field and touches many areas of research from biology, where the charge transport across DNA is studied, to engineering, where enhanced understanding of thermal and electric transport is converted into increased efficiency of thermoelectric devices.[2,3]

Initially, one major driving force behind molecular electronics was the application as single-molecular transistors.[4] It was thought that single molecules would take over when the limits of miniaturizations would be reached for silicon technology. According to Moore's law, the number of transistors per integrated circuit chip doubles roughly every two years.[5] For a MOSFET, this translates into a reduction of channel length of factor $\sqrt{2}$. It was thought that silicon-based semiconductor elements eventually would not be able not keep up with the requirements for shrinking transistors[5] and single molecular electronics would take its place.

Even though there is some indication that the day is coming where single molecular electronics replaces silicon technology,[6] silicon technology keeps complying with Moore's law and can now produce feature sizes down to 7 nm.[7,8] Conventional semiconductor technology is thus already approaching the size range of molecular electronics and has the advantage of decades of process experience.

An advantage that molecular electronics does have over silicon technology, however, is that molecular electronics typically is a "bottom-up" process compared to the "top-down" process of silicon technology. Top-down means that conventional silicon technology starts up with macroscopic, ultra-pure silicon wafers and then patterns those wafers with considerable technological and financial effort to produce nanoscale transistors. The opposite direction would be a bottom-up process as coined by Richard Feynman.[9] In bottom-up molecular electronics we can use chemical synthesis to fabricate atomically precise molecules with relative ease. Molecular electronics is thus not affected by lithography diffraction limits and increasing prices with decreasing size.[10] A possible way in for molecular electronics into semiconductor devices could be hybrid devices, which could combine the best of the two worlds.

Possibly more exciting than fixing limitations of silicon technology, are the new possibilities for technology and science that molecular electronics might enable. Lessons

learned in molecular electronics could be applied to construct a quantum computer.[11] Single molecular electronics has been proposed to have applications as chemical sensors with sensitivity down to single analyte molecules.[12,13] Fundamentally new circuit elements have been proposed, like the single molecular memristor.[14]

Apart from immediate applications, the study of charge carrier transport is of particular interest to the molecular electronics community. On the one hand, the details of the charge transport across a molecule determine the properties of any electronic device derived from it. On the other hand, electrons define the nature of the molecule. They are responsible for forming molecular orbitals, which define the molecule and bonds connecting it to the electrodes. Refined knowledge about charge transport might lead to a deeper understanding of the thermoelectric effect in single molecules. New charge transport phenomena like the Kondo effect have already been discovered in single molecules.[15] Other well-known phenomena are being applied to the single molecular world. The optical double-slit experiment is being expanded to double-branched single molecules. Electrical interference can be probed with this system to investigate quantum mechanical processes.

In the following sections of this chapter, the relevant theoretical foundations as well as experimental techniques used will be presented. It will be explained how machine learning can help understanding the charge transport processes on the nanoscale. In Chapter 2, the experimental conditions will be described in detail. In Chapter 3, a novel analysis approach using machine learning will be presented and applied to single-molecule conductance data of two well-studied reference systems. In Chapter 4, single molecule conductance results of several ferrocene-based molecules will be presented. Chapter 5 contains results for macrocyclic ruthenium molecules. Key results will be summarized at the end of the dissertation in Chapter 6.

1.2 Scanning Tunnelling Microscope

Molecular electronics is a relatively old field of research. The theoretical foundations were laid as early as the 1970s with the seminal paper by Aviram and Ratner.[16] But at the time, without experimental methods to test the proposed theory, the field did not gain momentum, Figure 1-1.

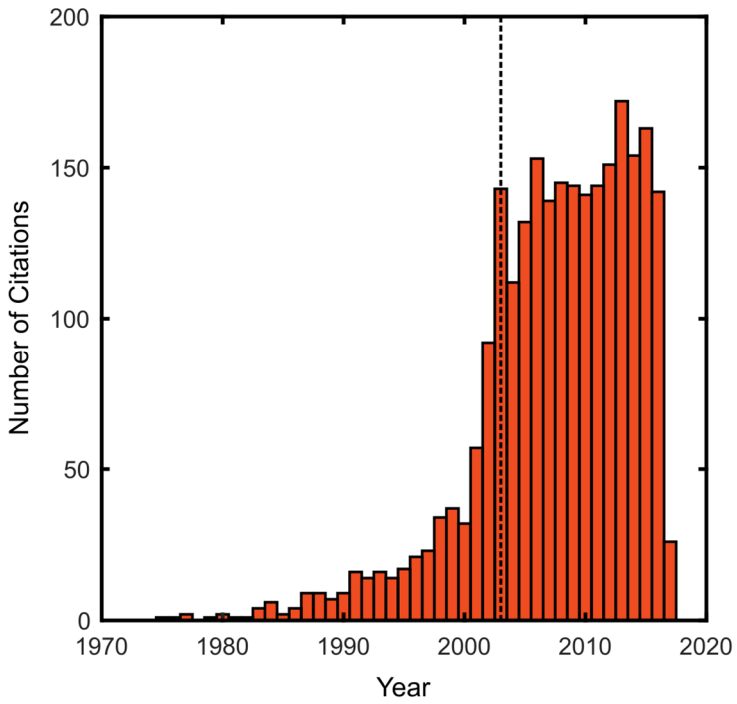


Figure 1-1: Citations of reference 16 (Aviram and Ratner, "Molecular Rectifiers", 1974) as function of year. First publication of STM was in 1982. In 2003, Tao and Xu reported on the conductance of single molecules (dashed line). Citation record from ISI Web of Knowledge.

This changed fundamentally, when in the 1980s and 1990s advances in technology sparked the renaissance of molecular electronics. For the first time, it was possible to experimentally measure the conductance of single molecules and to validate the proposed theories. One invention of particular interest, for the molecular electronics community in general and this dissertation in particular, is the scanning tunnelling microscope (STM). The first STM was developed in 1981 by a team with Binnig and Rohrer at IBM in Switzerland.[17] It had a profound effect on nanotechnology in general and its inventors were awarded with a Nobel prize in 1986, only five years after their original paper was first published.[18] In 2003 Tao and Xu used the STM to measure the conductance of single molecules (see dashed line in histogram, Figure 1-1).[19] The principles and theory behind the STM will be presented in the following sections with a special focus on how the STM can be used to measure single molecule conductance.

1.2.1 Quantum Mechanical Tunnelling

The basic concept of the STM relies on a quantum mechanical tunnelling current between a sharp tip and a conductive substrate. In the STM, tip and substrate are typically metals as

indicated in the band diagram in Figure 1-2. The electronic states in both metals are filled up with electrons to the Fermi Energy, E_F.[20] In panel **A**, there is a large vacuum gap. The electrodes are far apart. The vacuum gap does not contain any states and is thus an energetic barrier to the electrons. The wave functions of the electrons decay exponentially with distance in the vacuum gap between the electrodes. Wave functions of tip and substrate do not overlap and thus charge carriers cannot travel between tip and substrate. Φ_T and Φ_S are the work function of tip and substrate, describing the energy required to remove an electron from the solid to the vacuum level, E_{vacuum}.

If both metals are brought into close proximity of each other, *i.e.* in the nanometer range, their wave functions begin to overlap (Panel **B**). Despite the electrodes not touching, the potential barrier is sufficiently narrow for electrons to tunnel through it. Without applied voltage, electrons flow between the electrodes, leading to interfacial charging. This electron flow continues until the Fermi levels of both metals are aligned. Note that this tunnelling current is not an electrical discharge, but rather an effect of quantum mechanics which cannot be explained with classical physics.

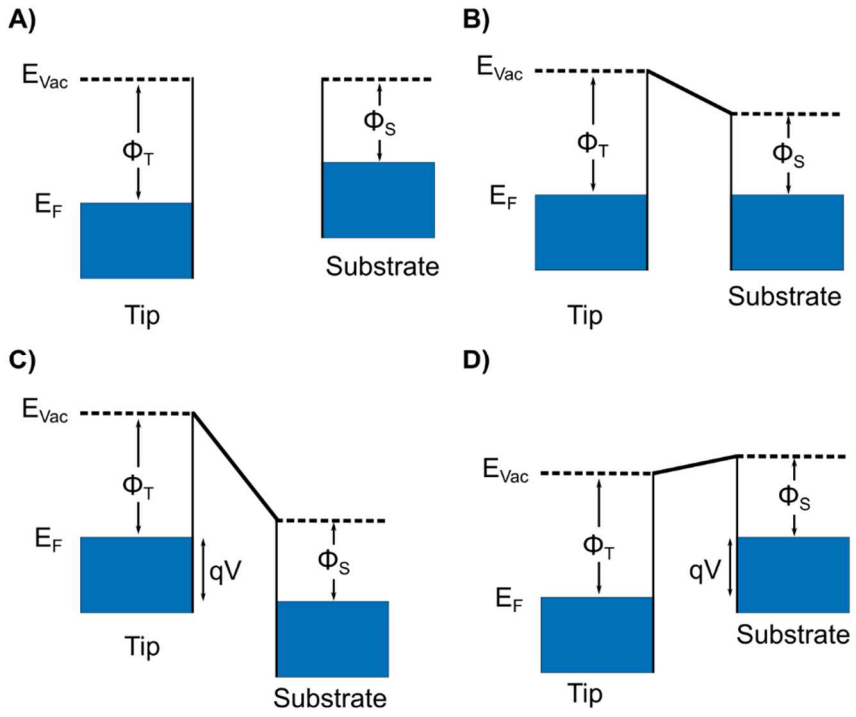


*Figure 1-2: Band diagram of STM tip and substrate at four different conditions: **A**) Large distance between tip and substrate. **B**) Tunnelling distance, no bias. **C**) Forward bias between tip and substrate. **D**) Reverse bias between tip and substrate.*

When a bias voltage is applied to the junction, the energy levels of substrate and tip are offset by qV. Where q is the elementary charge and V the applied bias voltage. The direction

of the offset is given by the polarity of the applied bias. This is visualised in panel **C** and **D** for different polarity.

In 1963 John Simmons published a model, describing the tunnelling current density between two conductors separated by a thin insulator of thickness s.[22,23]

$$J = Be^{-\beta s} \tag{1.1}$$

The preexponential factor, B, is proportional to the applied bias voltage, V, and the square root of the tunnelling barrier height, Φ_B.

$$B = \frac{q^2 V}{h^2 d}\sqrt{2m\Phi_B} \tag{1.2}$$

Where q is the electron charge, h is the Planck constant and m the electron mass. The exponential factor β is proportional to the square root of the barrier height:

$$\beta = \frac{4\pi}{h}\sqrt{2m\Phi_B} \tag{1.3}$$

The exponential dependency of the tunnelling current on the distance is key for the remarkable resolution of the STM. For example, an increase of tip-substrate separation from 0.5 nm to 0.6 nm will lead to a decrease of tunnelling current by almost an order of magnitude.[17]

1.2.2 STM Concept

The basic concept of the STM is that an ideally atomically sharp metal tip and a conductive substrate at different potentials are brought into close proximity of each other. As described in the previous section, a distance dependent tunnelling current will flow. The tip is mounted on a 3D piezo system which can move the tip in all directions. The tip is then rastered across the surface of the substrate as shown in Figure 1-3.

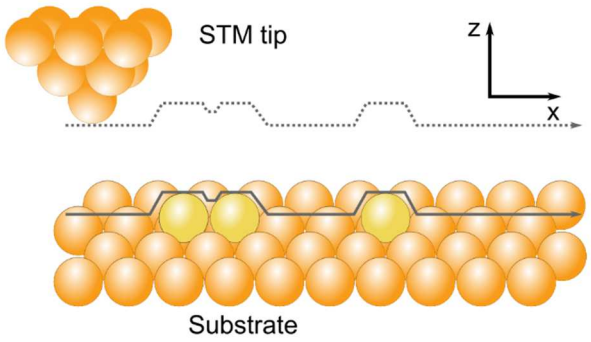


Figure 1-3: Schematic of the STM tip rastering over a substrate surface. Adatoms on the surface are matched by z-piezo adjustment of the tip.

In constant current mode, the tunnelling current is kept constant with help of a feedback loop which adjusts the z-piezo movement normal to the surface. From the z-piezo movement

the topography of the surface can then be reconstructed. The STM provides a relatively simple way to image systems in real-space with molecular and even atomic resolution. The STM can be thought of as the swiss army knife of microscopes. Apart from imaging experiments, the STM can be used as a tool to manipulate single atoms. The stick figure in Figure 1-4 was constructed by picking up single atoms with a STM tip and positioning them on a substrate. By changing atom by atom, a short movie was created by IBM.

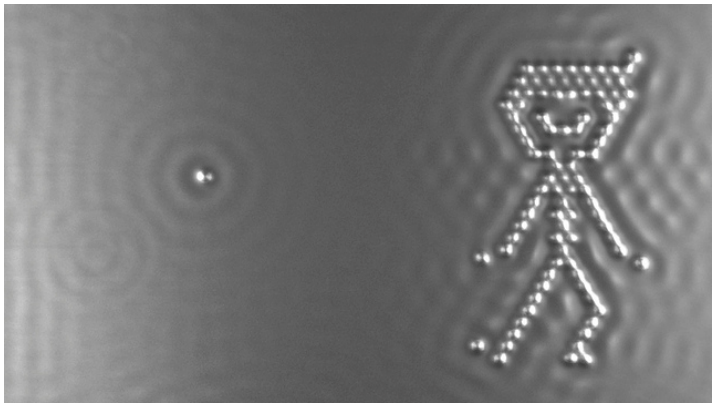


Figure 1-4: Frame of "A Boy and his Atom". The short film was created by IBM in 2013, by moving single atoms. Courtesy of International Business Machines Corporation, © (2013) International Business Machines Corporation.[24]

A different way of using the STM is spectroscopy mode. In the next section, it will be discussed how the STM can be used to measure the conductance of single molecules.

1.3 Single Molecule Conductance Measurements

Today, there are many different techniques in use to measure the conductance and to probe charge transport across single molecules. Among the well-established techniques are mechanically controlled break-junctions (MCBJ),[25] conductive probe atomic force microscopy (c-AFM),[26] as well as STM-based techniques like STM break-junction (STM-BJ),[19] the current-distance (STM-$I(s)$)[27] technique and the current-time (STM-$I(t)$)[28] technique. Since the STM was used throughout this dissertation, the following sections are dedicated to two STM-based techniques to measure single molecule conductance. Single molecule conductance data in this dissertation was collected using the *I(s)* technique. The *I(s)* technique was chosen because it is a gentler method of junction formation compared to the break-junction. For weak molecule-electrode coupling, there is typically no material transfer between tip and substrate (*vide infra*) and it is presumably less invasive with respect to the analyte.

1.3.1 STM-*I(s)* Technique

In the STM-*I(s)* technique, a conductive substrate is covered with a sub-monolayer of an analyte. The analyte is equipped with two anchor groups. One anchor group binds to the substrate and is used to assemble the analyte (sub)-monolayer. The other anchor group can bind to the STM tip. See Figure 1-5 for schematic. A bias voltage is applied between the substrate and the STM tip. When the STM tip is brought into proximity of the substrate, a tunnelling current can be measured. The initial distance, s_0, of tip and substrate is set, so that a predefined tunnelling current, I_0, is measured. The STM tip is then pulled away from the substrate and the tunnelling current, I, is then measured as function of distance, s.[27]

In the absence of a molecule in the gap between tip and substrate, the tunnelling current will decay exponentially with distance until it drops below the noise limit of the setup.[22] If there is a molecule present in the gap, a molecular junction can form. If the tip is pulled away, initially, the tunnelling current decays exponentially as direct tip-substrate tunnelling dominates the current. With increasing distance, direct tunnelling is less efficient and the current is dominated by the transport through the molecule. In a first approximation, the measured tunnelling current is independent of the distance in this case. The current plateaus. Eventually, when the tip-substrate distance becomes too large, the molecular junction ruptures. The current goes back to exponential decay. The tunnelling current can be so small at this point that it is below noise level current and no longer measurable. Examples of *I(s)* traces with and without molecular junction formation are shown in Figure 1-16, **A** and **C**. For the case of weak coupling between the molecule and the electrodes, the junction likely ruptures at the anchor group. For thiols, it has been established that the breaking force required to break the gold-thiol bond is on the order of that of a gold-gold bond.[29] It can be observed that strings of gold are pulled out of the electrodes in this case.[30]

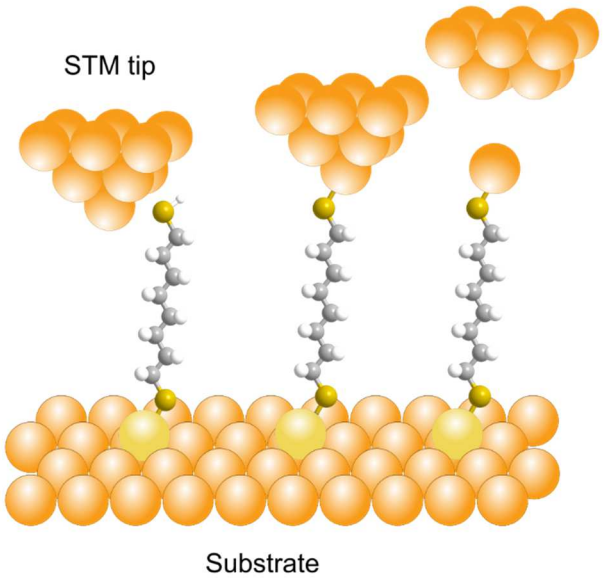


Figure 1-5: Schematic of the STM-I(s) technique with 1,8-Octanedithiol (ODT) in the junction. Three stages of junction formation are shown: 1) Tip at initial distance 2) Molecular junction formed. 3) Rupture of the junction.

1.3.2 STM Break-Junction

As one of the pioneers of experimental molecular electronics, Tao *et al.* were among the first to introduce single molecule conductance measurements with the STM break-junction technique.[19] The STM-BJ works similarly to the *I(s)* technique. The difference is that instead of positioning the tip at a close distance to the substrate, the tip is driven into the surface of the substrate. After the tip makes contact with the substrate, the tip is withdrawn from the surface and the tunnelling current, again, is recorded as a function of tip-substrate distance. As with the *I(s)* technique, the current is thought to plateau while current is flowing through the molecule. Compared to the *I(s)* technique, the BJ technique has a higher junction formation probability. Tip and substrate need to be the same material since material transfer between the two is likely. In molecular electronics, the experimental conductance can depend on the experimental method and the electrode configuration that it generates.[32] It has been reported that for the same molecule, high conductance groups are more easily identified with the STM BJ method.[33,34]

1.3.3 Data Analysis

In an experiment, fluctuations of the measurement scale with $1/\sqrt{n}$, where n is the number of observations, meaning states of the system.[35] For macroscopic conductors n is large and fluctuations are thus small compared to the measurement. On a single molecule scale these

fluctuations can be on the order of the signal. Variations in the measured signal are significant. To overcome the influence of fluctuations, it has been established to record many *I(s)* traces and compile these traces into current or conductance histograms.[19] Histograms can be constructed with linear or logarithmic binning. Peaks in those histograms are then interpreted as most likely conductance of the analyte.

Aside from statistical fluctuations, variations in the contact geometry at the electrodes and conformational changes of the molecule can have a strong influence on the molecular conductance.[36,37] To overcome the low signal-to-noise ratio and the limited junction formation probability of single molecule conductance measurements, data selection and filtering can be applied.[38,34] Data selection will be discussed in more detail below.

1.4 Charge Transport

On the macroscopic scale the current through a conductor is proportional to the applied voltage. Ohm first described this relationship and it is hence known as Ohm's law. If the dimensions of the conducting material are reduced to the nanoscale, this law does not always hold true. In the following section, a short overview over some of the peculiarities of charge transport processes on the nanoscale will be given.

1.4.1 Conductance Quantisation

Let us consider a one-dimensional wire that is connecting two reservoirs of electrons. The wire perfectly conducts electricity, meaning the transmission probability, of an electron from one reservoir to the other is 1. The reservoirs are at two different potentials. One reservoir will be populated up to a chemical potential μ_1, the other one up to μ_2.[20] The energy difference between them is $q \cdot V$, where q is the elementary charge and V is the applied potential. The current between the two reservoirs can then be described by:

$$I = \Delta n q v_e = \frac{D(\varepsilon) q V}{L} q v_e \tag{1.4}$$

Where Δn is the excess charge density flowing along the potential gradient, v_e is the drift velocity, $D(\varepsilon)$ is the density of states (DOS), ε is the electron energy and L is the length of the wire. Together with an expression for the density of states, this simplifies to:[20]

$$I = \frac{2}{h v_e} v_e q^2 V = \frac{2q^2}{h} V \tag{1.5}$$

Dividing the current by the potential thus gives a conductance of the wire that is dependent on two fundamental constants. This conductance is called the conductance quantum,

G_0.[39] This quantized nature of conductance is typically observed in break-junction experiments, where integer multiples of G_0 are observed. It is noted, however, that this quantisation is dependent on the electrodes and is not observed in every case.[40]

$$G_0 = \frac{2q^2}{h} \approx 7.748 \cdot 10^{-5}\ S \tag{1.6}$$

If we consider a wire that is not perfectly conductive, an electron can either be transmitted through the wire or it can be reflected. In this more general case the conductance is described by the Landauer equation:[20]

$$G(\varepsilon_F) = G_0 \mathcal{T}(\varepsilon_F) \tag{1.7}$$

Where ε_F is the Fermi energy and $\mathcal{T}(\varepsilon_F)$ is the transmission probability of a charge carrier at the Fermi energy.

1.4.2 Charge Transport Mechanisms

A molecular junction consists of a molecule that is coupled to two electrodes, which connect the molecule to a macroscopic circuit. A band diagram of a molecular junction is shown in Figure 1-6. The mechanism of charge transport in the junction depends on several factors. Important features, which influence charge transport, include the magnitude and sign of the applied potential, temperature, the electrodes, the geometry of the junction, the interfaces between electrodes and molecule and the molecule itself (*vide infra*).

The electrodes are typically high-work function metals. They have a continuous DOS without a band gap and the states are filled up to the Fermi level. Characteristic to the electrodes is their material, which determines their density of states and work function.[20,41]

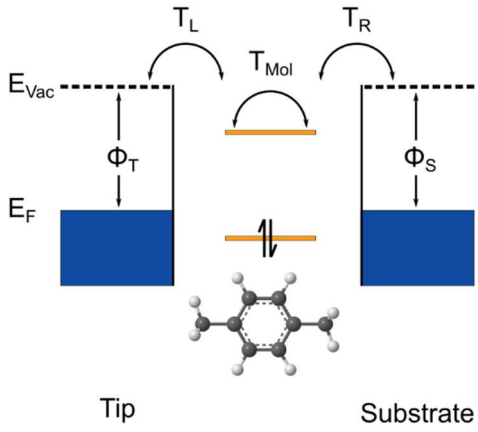


Figure 1-6: Band diagram of a molecular junction. VL is the vacuum level, T the transmission probability, E_F the Fermi level and Φ the barrier height.

The molecule itself is characterized by discrete states. Of those states, the "frontier orbitals" usually govern the charge transport. The "frontier orbitals" are the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO).[41]

The interface between electrodes and molecule is largely influenced by the anchor groups of the molecule. Weak coupling into a physisorbed molecule for example leads to a high barrier. Strong coupling on the other hand, for example through a strong covalent bond, results in effective coupling with a small barrier or, in the ideal case, no barrier (Ohmic contact).[41]

Coherent Tunnelling

In the simplest case of a molecular junction, a low potential is applied, there are no molecular states accessible and the distance between electrodes is short. In this case, the transport mechanism is coherent tunnelling. In the first instance the barrier can be assumed to be rectangular and the tunnelling current can be estimated with the Simmons tunnelling model, equation 1.1. Low potential means here that the energy barrier is much larger than $q \cdot V << \Phi_B$.

Under high applied bias ($eV >> \Phi_B$), Fowler-Nordheim (also called field emission) tunnelling becomes dominant. Here the tunnelling barrier is deformed under the applied bias that the effective barrier is smaller and tunnelling becomes more effective.

Both tunnelling mechanisms are independent of temperature in the first instance. The electrodes are effectively coupled to the molecule. Both processes are fast and elastic, meaning that the charge carrier does not reside in junction for a significant amount of time.

Superexchange is a charge transfer process that is used to explain coherent tunnelling over distances that would be too long for coherent tunnelling alone.[23] The idea is that the charge

interacts with molecular orbitals of the molecule in the junction for an increased electron transfer rate.[23] Even with superexchange, tunnelling is not typically effective at lengths greater than 2.5 nm.[42] Although, Lu *et al*. reported a transition length of 2.75 nm.[43]

Incoherent tunnelling

Incoherent tunnelling is a charge transport process where the charge carrier tunnels through multiple barriers and resides in a well between barriers long enough to interact with the atoms of the molecule. The charge carrier loses its coherence. The process is not temperature dependent and is observed for long molecules.[41]

Hopping

Hopping is a thermally activated process where a charge carrier passes a potential barrier by hopping over it, rather than tunnelling through it. The temperature dependence follows the Arrhenius relationship.[41] Hopping has a lower length-dependence than tunnelling and is predominant over long distances.[44,45]

Thermionic Emission

Thermionic emission is a charge transport process at the interface between electrode and molecule. The barrier between molecule and electrode is overcome here by hopping over the potential that is deformed by the applied bias voltage.

Generally, tunnelling is favoured through thin, but high barriers. Hopping is the dominant process for shallow barriers and high temperatures.[41] In some cases, the above mentioned charge transport processes across molecules can be significantly influenced by quantum interference. The temperature, voltage and distance dependencies of different charge transport mechanisms are summarized in Table 1-1.

Table 1-1: Summary of temperature, voltage, and distance dependencies of different charge transport regimes.[41]

Conduction Mechanism	Temperature Dependence	Voltage Dependence	Distance Dependence
Coherent Tunnelling	None	$I \sim V$	$\ln(I) \sim d$
Incoherent Tunnelling	None	$I \sim V$	$\ln(I) \sim d$

Fowler-Nordheim Tunnelling	None	$\ln(\mathrm{I}/V^2) \sim 1/V$	$\ln(I) \sim d$
Hopping	$\ln(\mathrm{I}) \sim 1/T$	$I \sim V$	$I \sim 1/d$
Thermionic Emission	$\ln(\mathrm{I}/T^2) \sim 1/T$	$\ln(I) \sim \sqrt{V}$	$\ln(I) \sim 1/\sqrt{d}$

1.4.3 Quantum Interference

Interference is a well-studied phenomenon in physics. Especially in optics, interference is ubiquitous. Experiments with optical double slits and the resulting interference patterns convinced physicists of the wave nature of light. The concept of interference was interpreted as the superposition of two or more wave functions where the amplitudes of those wave functions would be amplified or cancelled out depending on the phase of the wave functions. The concept of interference was later extended to include classical particles. De Broglie postulated that a particle with a momentum p would have a wavelength of $\lambda = h/p$, where h is the Planck constant.[20]

For molecular electronics quantum interference of electron waves is of utmost importance. Quantum interference takes many forms in molecular electronics. In a molecule which is not connected to any electrodes, interference is responsible for the formation of discrete energetic molecular orbitals.[46] In the case of a molecule connected to electrodes, an external electron with energy of a continuous energy distribution can enter the molecule and experience constructive or destructive interference. Transmission through the molecule depends on the energy and thus de Broglie wavelength of the electron.[46] Experimentally, quantum interference has only been observed a few years ago.[47,48,49,50,51,52,53]

Three manifestations of quantum interference in single molecules are typically distinguished in molecular electronics, namely Breit-Wigner resonances and Fano resonances and Mach-Zehnder interferometry.[46] In Breit-Wigner resonances, an electron passing through a molecule interferes with a state of the molecule. A Fano resonance describes the case of a charge carrier, traversing the molecule, interfering with the state of a side group along the backbone of a molecule. Mach-Zehnder resonances are arguably the most prominent manifestation of quantum interference in molecular electronics.[54] They occur when there are two or more distinct electric pathways present through a molecule. Upon reconnecting, the waves can interfere constructively or destructively, Figure 1-7.

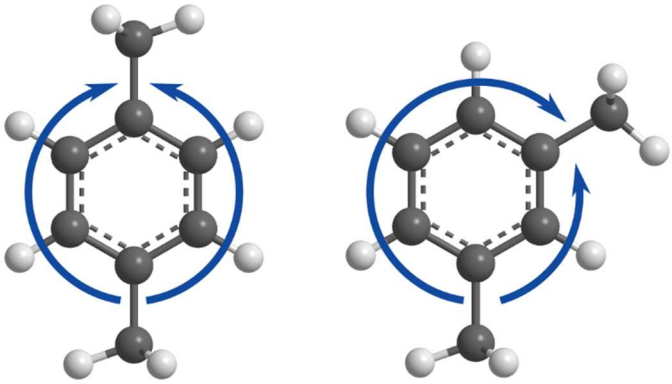


Figure 1-7: Mach-Zehnder resonance in a para- (left) and meta- (right) substituted benzene ring. Constructive (left) and destructive (right) interference is observed for the two cases. Arrows indicated spatially separated current paths.

Potential applications for engineered quantum interference are plentiful. Quantum interference has been shown to be useful to create giant thermopower in thermoelectric materials,[56] or can be used to tune conductance through a molecule by manipulating a side group.[57] Transistors operating with quantum interference have also been proposed.[58,59,55]

1.4.4 Organometallics in Molecular Electronics

Organometallic molecules are molecules that contain carbon-metal bonds.[60] They have received considerable attention in recent years in the molecular electronics community, partially because they show potential for application as thermoelectric materials.[61] Organometallic compounds have also been proposed as molecular wires due to their low attenuation and effective charge transport while keeping wave coherence or as single-molecular transistors.[62,63,64,65,66,67] They have been shown to give rise to rectification, which makes them interesting as possible candidates for molecular diodes.[68] They exhibit increased conductance compared to their purely organic equivalents. The increased conductance is due to the excellent alignment of molecular frontier orbitals with the electrode Fermi levels.[69,70] Beyond that, metal centres give organometallics redox and photochemical addressability.[71,72]

The charge transport process across organometallics is aided by the redox state of the molecule. Schmickler *et al.* suggested a one-step coherent resonant tunnelling process assisted by the redox level.[73,74] Kuznetsov and Ulstrup created a model for a sequential two-step process where the redox molecule is briefly oxidized/reduced.[75,76,77] Vibrational relaxation takes place here.[70] Schematic for both models are displayed in Figure 1-8.

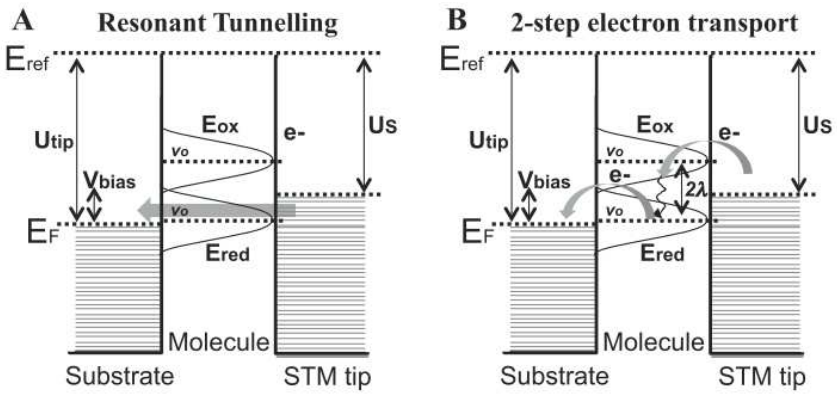


*Figure 1-8: Band diagrams explaining two different charge transport models for electron transport across a redox active molecule. **A**) resonant tunnelling and **B**) a two-step electron transport. Reprinted with permission from Guo, 2013.*[77]

1.5 Electrochemistry

Electrochemistry is a subsection of physical chemistry. It is at the intersection between electrical and chemical phenomena, where changes in electrical potential lead to chemical changes and *vice versa*.[78] This section of the Introduction Chapter is dedicated to electrochemical methods and principles that are relevant for this dissertation. It is not exhaustive and interested readers are referred to the large body of literature available on this subject.[55,78,79]

1.5.1 The Three-Electrode Electrochemical Cell

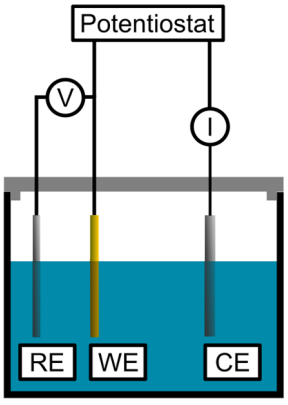


Figure 1-9: Schematic of three-electrode electrochemical cell.[78]

In a typical three-electrode electrochemical cell, three electrodes are submerged into a polar solvent. See Figure 1-9 for schematic. The solvent contains a supporting electrolyte that is responsible for current flow in the cell. The three electrodes in the cell are the working electrode (WE), counter electrode (CE), and reference electrode (RE). The reference electrode itself does not carry a significant amount of current. The working electrode is in contact with the analyte, which is either in solution or on the electrode itself. The counter electrode balances the charge of the electrodes. A potential function can be applied to the system and the current response of the system, including an analyte in solution or immobilized in the working electrode surface, is then measured.[78,80]

1.5.2 Electrode Processes

When a potential is applied to the cell, ions in solution will drift to the working electrode, attracted by the opposite charge on the electrode surface. Charged species and polarized solvent molecules at the electrode-solution interface form the electrical double layer, Figure 1-10 **D**. The electrical response of this cell can be modelled as an electrical circuit with a capacitor and resistor in series, **A**. Where the working electrode and double layer form a capacitor of capacitance C_d and the solution acts as a resistor with resistance R_s.[78,81] When an electroactive analyte is added to the cell and a suitable potential is applied, a Faradaic current, I_F, is observed in addition to the capacitive current, I_C.[80]

$$I = I_C + I_F = C_d \frac{dE}{dt} + I_F = \nu C_d + I_F \tag{1.8}$$

Where I is the total observed current, E the applied potential, t the time and ν the scan rate.

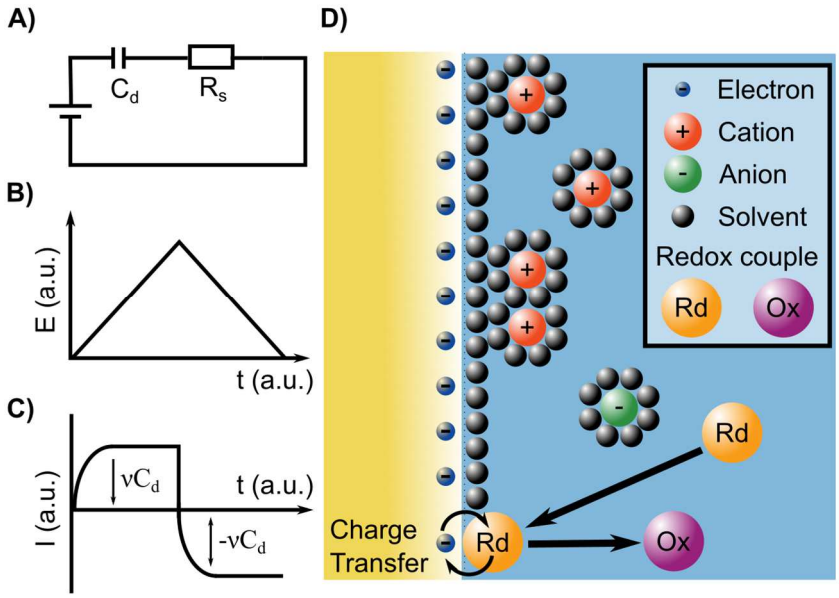


*Figure 1-10:**A**) Electrical circuit of the cell. **B**) Applied voltage as function of time for cyclic voltammetry. **C**) Current response of an ideal system without Faradaic current. **D**) Double layer at the electrode-electrolyte interface.*[78,81]

1.5.3 Cyclic Voltammetry

Cyclic voltammetry (CV) is a potential sweep technique. It is a powerful and very sensitive technique to investigate small quantities of analyte molecules. During cyclic voltammetry, the potential in a three-electrode cell is swept, back and forth, linearly between two potential limits, Figure 1-10 **B**. The current as a function of potential is recorded. (See panel

C for idealized current response without analyte present). If an electroactive analyte is present in solution, the Faradaic part of the current is governed by the diffusion kinetics of the analyte to the working electrode. At the working electrode, the analyte undergoes a redox process and diffuses away from the surface, Figure 1-10 **D**. It can be derived that at 25°C the Faradaic peak current, I_p, is proportional to the square root of the scan rate.[78]

$$I_p = 2.69 \cdot 10^5 {n_e}^{3/2} A D_0^{1/2} C_0^* \nu^{1/2} \quad (1.9)$$

Where n_e is the number of electrons involved in the redox process, A is the area of the electrode in cm^2, $D_0^{1/2}$ is the diffusion coefficient in cm^2/s, C_0^* is the concentration in mol/cm^3 and ν is the scan rate in V/s. A typical voltammogram with the key parameters is shown in Figure 1-11. E_p is the peak potential, i_p is the peak current of the anodic and cathodic process. $E_{1/2}$ is the half-wave potential and ΔE is the peak separation.

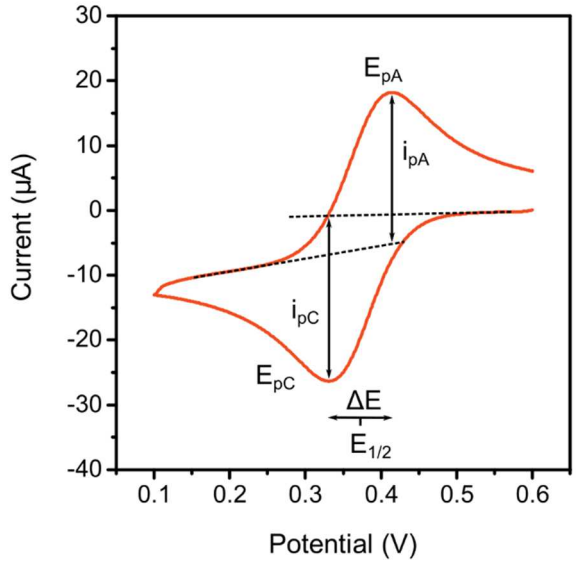


Figure 1-11: Cyclic voltammogram of ferricyanide. Data from Echem Analyst of Gamry Instruments inc,

Instead of providing the analyte in solution, it is possible to functionalise the working electrode with the analyte before the experiment. Self-assembled monolayers (SAM) are an elegant and convenient way to assemble monolayers of molecules on a substrate surface. They are typically created by immersing a substrate in a solution of an analyte. The analyte is equipped with a moiety that physisorbs or chemisorbs to the substrate. After an incubation period in solution, an ideally single-molecular layer of molecules assembles on the substrate.[82,83,84]

When the analyte is immobilized on the electrode surface and no analyte is present in solution, the current is no longer governed by diffusion effects. The peak current is now proportional to the scan rate.[78]

$$I_p = \frac{n^2 F^2 A \Gamma}{4RT} \nu \tag{1.10}$$

Where F is the Faraday constant, Γ the surface concentration. R is the gas constant and T is the temperature. Surface immobilized voltammograms were recorded during this dissertation to confirm that the analyte assembled on the surface and stayed intact over the course of the experiment. They were also used to measure the surface concentration of the analyte

1.5.4 Electrochemical STM

The STM can also be operated as under electrochemical control (EC-STM). The same setup as a conventional STM can be used, but two additional electrodes are introduced into the STM cell to function as reference and counter electrode. The substrate acts as the working electrode. A bipotentiostat is used to change potential levels of the electrodes. The electrochemical STM cell is filled with an electrolyte solution. The tip needs to be coated with an insulator layer, leaving only the apex exposed. This is necessary to reduce the Faradaic current and thus enable a stable tunnelling junction.

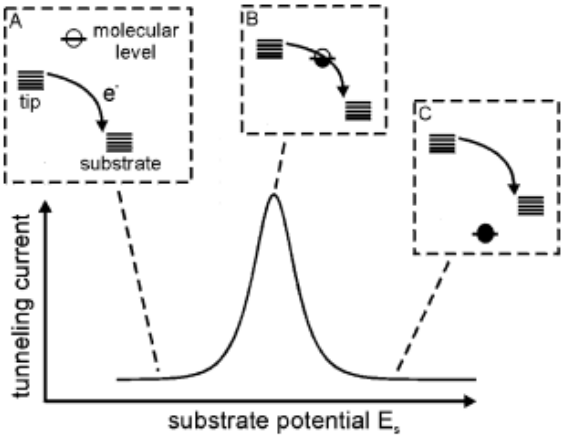


Figure 1-12: Tunnelling current as function of substrate potential. Bringing the redox level in resonance, between the electrode Fermi levels leads to an increase in tunnelling current. Figuure was reprinted with permission from Albrecht et al., 2005.[85]

The EC-STM has the advantage compared to the STM that there is precise electrochemical control over the potential. Analyte species on the substrate surface can be oxidized or reduced during spectroscopy or imaging experiments. In an EC-STM, the reference electrode can act as the gate electrode in a three-terminal device. This gate electrode or the applied potential can be used to bring molecular energy levels in and out of resonance (Figure 1-12).[85] This means that the redox level can be brought between the Fermi levels of the electrodes. Alignment of the redox state with the electrodes leads to an increase in tunnelling current. Albrecht *et al*. observed a 50-fold

current amplification between on and off resonance current.[85] This three-terminal current-voltage relationship resembles those of traditional transistors.[86]

1.6 Machine Learning in Single Molecular Electronics

In 1959, Artur Samuel reported on a machine learning procedure for the game of checkers, which enabled the computer to beat Samuel himself, who had programmed it. He defined machine learning as programming computers to learn from experience without being explicitly programmed.[87] Machine learning has come a long way since 1959 and is now affecting our lives profoundly. Prominent examples range from spam filters in email programs and facial recognition software to recommender systems for online shopping and anomaly detection for credit card fraud detection.[88] Also, in the sciences, machine learning has made an entry. In astronomy for example machine learning is used to deal with a "data avalanche" from increasingly powerful telescopes.[89,90] In neuroscience, machine learning is used to find out how the brain stores and processes information.[91] Typically, machine learning is applied to analyse large data sets. It is therefore logical to apply machine learning techniques to the field of molecular electronics and the vast amounts of data collected in this field. In the following section, a brief overview of relevant machine learning techniques will be given.

1.6.1 Dimensionality Reduction

A concept which is frequently used in machine learning is dimensionality reduction. It is applied to reduce the dimensionality of large data sets while preserving as much of the information as possible. Dimensionality reduction can be useful for single molecular conductance data. A spectroscopy data set consists of M *I(s)* traces, where each trace has N data points. Typical values for M are 1000 traces per data set with 2000 current values each. Disregarding the distance component, which is identical for each trace, this results in a 2000 by 1000 entry matrix which makes up the data set. Or phrased differently, the data set consists of 1000 vectors in 2000-dimensional space. Each dimension is one feature of the data set. In this high-dimensionality representation, visualisation of the data set is difficult. Dimensionality reduction can then be applied to transform high dimensionality data into lower dimensions. In lower dimensionality, coherent patterns are easier to detect.[92]

Arguably the most prominent example of dimensionality reduction is principle component analysis (PCA). PCA finds a surface onto which to project the data with minimal projection error. For our example PCA would find a K-dimensional subspace onto which every data point can be projected with minimal projection error. Where $K \leq N$.[88]

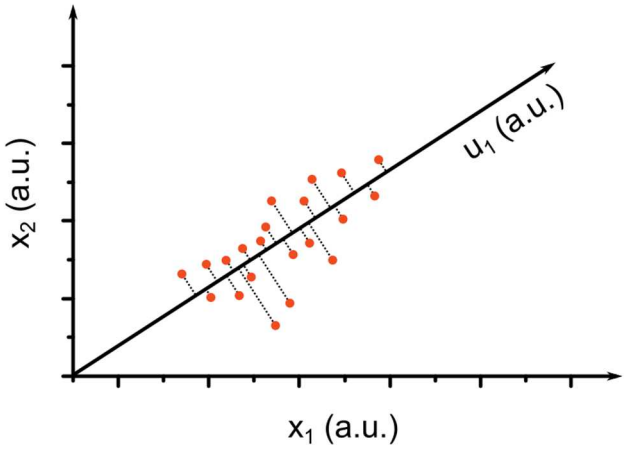


Figure 1-13: Positively correlated two-dimensional data is projected onto a new axis, u_1, while minimizing the projection error. Projection error is indicated by dashed grey lines.

For two dimensions, this concept is visualized in Figure 1-13. The red data points are positively correlated. The data points are projected onto a new axis, u_1, maximizing the variance of the data on this new axis. The dimensionality of the data set was thus reduced from two to one dimension while minimizing the loss in information. Dashed grey lines indicate the projection error associated with the dimensionality reduction.

Dimensionality reduction can be applied to aid with single molecular data analysis. In Chapter 3, a customized dimensionality reduction algorithm will be presented.

1.6.2 Clustering and Data Selection

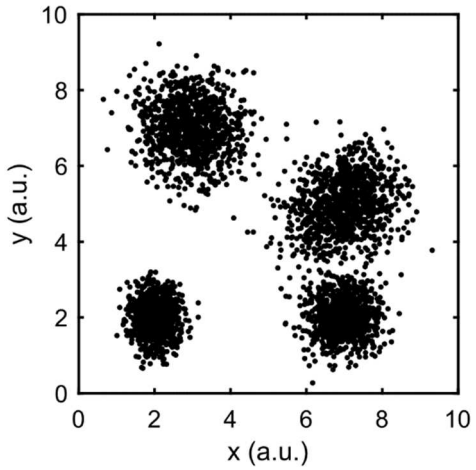


Figure 1-14: Example of an unlabelled data with two features, x and y.

After applying dimensionality reduction to a data set, clustering algorithms are frequently used to group the data into meaningful sub-groups.[92] Within machine learning two of the most common types of algorithms are supervised and unsupervised algorithms.[88] Clustering algorithms are examples of unsupervised learning algorithms. This means that the data is not labelled and an algorithm groups the data into sub-clusters by optimizing some criterion. Cluster assignment is optimized so that members of the same cluster are similar and members of different clusters are dissimilar. An example of an unlabelled data set with two features is shown in Figure 1-14.

Distribution-based Clustering

The most straightforward way to assign data points to clusters might be distribution based clustering. Here, histograms of the features are fitted with distributions. Data points with features within a certain confidence interval of the distribution mean are considered to be part of the same cluster. In the case of the data shown in Figure 1-14, distributions fitted to histograms of x and y could be used to determine cluster membership.

Density Clustering

Laio *et al.* proposed a clustering algorithm which assigns data points to clusters based on point density in the vicinity of the data point.[93] No distributions need to be fitted for this approach and the algorithm determines the number of clusters in the data set. The drawback for this method is that in a decision graph, a manual decision about cut-off point density needs to be made.

k-means clustering

By far the most common clustering algorithm is k-means clustering.[88] It is an iterative algorithm which is initialised by choosing R random cluster centroids μ_r, where R is the integer number of clusters. The iterative k-means process consists of two steps: (i) First all data points are assigned to their closest cluster centroid. (ii) Then the cluster centroid is moved to the new cluster mean. These steps are repeated until a predefined criterion is reached. The optimization objective of the algorithm is to minimize the squared distance between all data points and their assigned cluster centroids.[88] If the centroids do not move between iterations the algorithm has converged. To avoid local optima, the k-means clustering algorithm is typically restarted many times with different random starting points. Figure 1-15 shows different iterations of the k-means algorithm applied on the data set of the previous figure. The number of clusters was selected to be four. Cluster centroids are indicated by "x". After 12 iterations, the algorithm converged in this example and the cluster assignment matches the optical intuition.

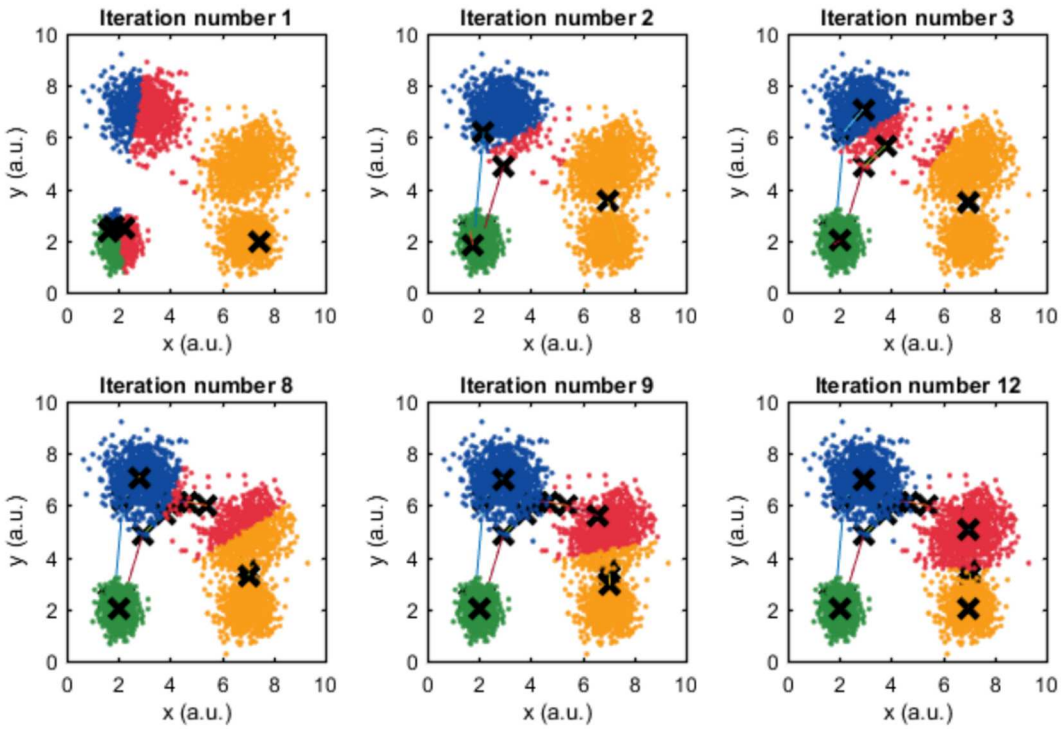


Figure 1-15: Different iterations of a k-means cluster process of the data from the previous figure. "x" indicate cluster centroids. Drawn by adapting an algorithm of the "Machine Learning" online course by Andrew Ng.[88]

Selecting a sensible number of clusters is a common problem for k-means clustering. There is usually not an exact solution to this issue, but there are rules of thumb that can be helpful in choosing a number. One example is the "elbow method", where the squared residual error after clustering is plotted against the number of clusters. An edge in the graph is interpreted as the optimal number of clusters.[88]

Fuzzy c-means (FCM) clustering - Gustafson-Kessel (GK) Algorithm

Fuzzy clustering is an adaptation of k-means clustering. For k-means clustering, data points are always assigned to exactly one cluster centroid. Fuzzy clustering on the other hand allows for weighted membership in several clusters.[94] The weighted membership describes to which degree data points belong to a certain cluster. The Gustafson-Kessel algorithm is an extension of FCM, applying an adaptive distance norm in order to detect clusters of different geometrical shapes within one data set.[95,96,97]

Plateau Finder Algorithm

A different approach to data selection was proposed by Inkpen *et al.*[98] This method was tailored to single molecule conductance data. Rather than grouping traces according to

similarity, the plateau finder splits a data set into three groups by predefined criteria. The three groups corresponded to traces with plateau, traces without plateau and noisy traces. To distinguish noisy traces from non-noisy traces, a noise factor was defined for every *I(s)* trace. A noise-factor cut-off was then selected and traces below the cut-off were considered non-noisy, traces above were considered noisy. The non-noisy traces were then separated into plateau-traces and plain exponential traces. The criterion used here was the presence of peaks of height larger than a plateau-determining bin count (PDBC) in the single-trace histograms of certain binning parameters. Figure 1-16 demonstrates this concept with two *I(s)* traces. The red trace in **A** contains a plateau and the corresponding single-trace histogram in **B** has a peak above the PDBC threshold. The blue *I(s)* trace in **C** does not have a plateau and no peak above the PDBC threshold in **D**.

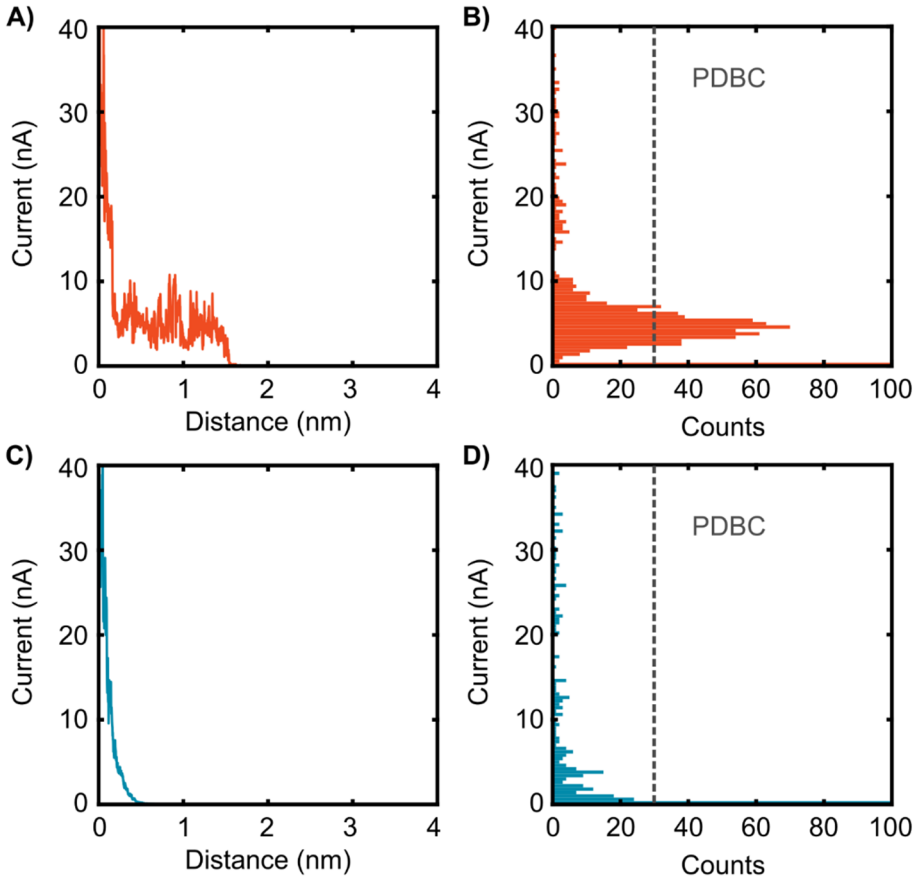


*Figure 1-16: Experimental I(s) traces of OPE **A**) with and **C**) without plateau.*[31] *__B__) and __D__) are the corresponding single-trace histograms as well as the plateau-determining bin count (dashed line, PDBC).*

1.7 References

1 R. M. Metzger, in *Intelligent Materials*, 2007, pp. 205–230.

2 L. Cui, R. Miao, C. Jiang, E. Meyhofer and P. Reddy, *J. Chem. Phys.*, 2017, **146**, 92201–26.

3 S. L. and J. H. and O. S. and P. H. and P. J. and P. Z. and S. C. and S. Huang, *Nanotechnology*, 2010, **21**, 262001.

4 M. L. Perrin, E. Burzurí and H. S. J. van der Zant, *Chem. Soc. Rev.*, 2015, **44**, 902–919.

5 G. Moore, *Science*, 2003, **299**, 210–211.

6 Q. Cao, S.-J. Han, J. Tersoff, A. D. Franklin, Y. Zhu, Z. Zhang, G. S. Tulevski, J. Tang and W. Haensch, *Science*, 2015, **350**, 68–72.

7 C. Vu, IBM Research Alliance Produces Industry's First 7nm Node Test Chips, http://www-03.ibm.com/press/us/en/pressrelease/47301.wss, (accessed 4 April 2017).

8 Ultra-fast, energy-sipping devices powered by Intel, http://www.intel.com/content/www/us/en/silicon-innovations/intel-14nm-technology.html, (accessed 4 April 2017).

9 S. L. Garrett and M. E. Poese, *Eng. Sci.*, 1960, 22–36.

10 W. Lu and C. M. Lieber, *Nat. Mater.*, 2007, **6**, 841–50.

11 J. L. O. T.D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, *Nature*, 2010, **464**, 45–53.

12 J.-O. L. Koen Besteman Frank G. M. Wiertz, Hendrik A. Heering, and and C. Dekker, *Nano Lett*, 2003, **3**, 727–730.

13 T. Albrecht, *Nat. Commun.*, 2012, **3**, 829.

14 Y. Li, G. W. Holloway, S. C. Benjamin, A. D. Briggs, J. Baugh and J. A. Mol, 2016, **1**, 1–20.

15 W. Liang, M. P. Shores, M. Bockrath, J. R. Long and H. Park, *Nature*, 2002, **417**, 725–729.

16 A. . R. Aviram M. A., *Chem. Phys. Lett.*, 1974, **29**, 277–283.

17 M. R. Castell, in *Nanocharacterisation*, ed. A. I. K. and S. J. Haigh, Royal Society of Chemistry, Oxford, 2nd edn., 2015, pp. 80–107.

18 G. Binnig, H. Rohrer, C. Gerber and E. Weibel, *Phys. Rev. Lett.*, 1982, **49**, 57–61.

19 B. Xu and N. Tao, *Science*, 2003, **301**, 1221–1223.

20 C. Kittel, *Introduction to solid state physics*, John Wiley & Sons, Inc., Berkeley, 8th edn., 2005, vol. 3.

21 R. J. Hamers, *Annu. Rev. Phys. Chem.*, 1989, **40**, 531–559.

22 J. G. Simmons, *J. Appl. Phys.*, 1963, **34**, 1793–1803.

23 R. L. McCreery, *Chem. Mater.*, 2004, **16**, 4477–4496.

24 IBM, A Boy and his Atom, http://www.research.ibm.com/articles/madewithatoms.shtml#fbid=IF6Ms94PXMb, (accessed 5 April 2017).

25 M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin and J. M. Tour, *Science*, 1997, **278**, 252–254.

26 X. D. Cui, a Primak, X. Zarate, J. Tomfohr, O. F. Sankey, a L. Moore, T. a Moore, D. Gust, G. Harris and S. M. Lindsay, *Science*, 2001, **294**, 571–574.

27 W. Haiss, H. van Zalinge, S. J. Higgins, D. Bethell, H. Hoebenreich, D. J. Schiffrin and R. J. Nichols, *J. Am. Chem. Soc.*, 2003, **125**, 15294–15295.

28 W. Haiss, R. J. Nichols, H. Van Zalinge, S. J. Higgins, D. Bethell, D. J. Schiffrin and others, *Phys Chem Chem Phys*, 2004, **6**, 4330–4337.

29 M. Frei, S. V. Aradhya, M. S. Hybertsen and L. Venkataraman, *J. Am. Chem. Soc.*, 2012, **134**, 4003–4006.

30 D. Krüger, H. Fuchs, R. Rousseau, D. Marx and M. Parrinello, *Phys. Rev. Lett.*, 2002, **89**, 186402.

31 M. Lemmer, M. S. Inkpen, K. Kornysheva, N. J. Long and T. Albrecht, *Nat. Commun.*, 2016, **7**, 17–19.

32 W. Haiss, S. Martín, E. Leary, H. Van Zalinge, S. J. Higgins, L. Bouffier and R. J. Nichols, *J. Phys. Chem. C*, 2009, **113**, 5823–5833.

33 O. Al-owaedi, S. Eaves, M. Lemmer, B. W. Skelton, H. M. Osorio, S. J. Higgins, P. Cea, N. J. Long, C. J. Lambert, P. J. Low, S. Bock, O. Al-owaedi, S. Eaves, D. C. Milan, M. Lemmer, B. W. Skelton, H. M. Osorio, R. J. Nichols, S. J. Higgins, P. Cea, N. J. Long, T. Albrecht, S. Martin, C. J. Lambert and P. J. Low, *Chem. - A Eur. J.*, 2017, **23**, 2133–2143.

34 A. Vezzoli, I. M. Grace, C. Brooke, R. J. Nichols, C. J. Lambert and S. J. Higgins, *J. Chem. Phys.*, 2017, **146**, 92307-1–7.

35 M. Ratner, *Nat. Nanotechnol.*, 2013, **8**, 378–381.

36 J. Ulrich, D. Esrail, W. Pontius, L. Venkataraman, D. Millar and L. H. Doerrer, *J. Phys. Chem. B*, 2006, **110**, 2462–2466.

37 T. A. Su, H. Li, M. L. Steigerwald, L. Venkataraman and C. Nuckolls, *Nat. Chem.*,

2015, **7**, 215–220.

38 T. Hines, I. Diez-Perez, J. Hihath, H. Liu, Z. S. Wang, J. Zhao, G. Zhou, K. Mullen and N. Tao, *J. Am. Chem. Soc.*, 2010, **132**, 11658–11664.

39 B. J. Van Wees, H. Van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. Van Der Marel and C. T. Foxon, *Phys. Rev. Lett.*, 1988, **60**, 848–850.

40 F. Pauly, M. Dreher, J. K. Viljas, M. Häfner, J. C. Cuevas and P. Nielaba, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2006, **74**, 1–21.

41 S. Karthäuser, *J. Phys. Condens. Matter*, 2011, **23**, 13001.

42 S. H. Choi, B. Kim and C. D. Frisbie, *Science*, 2008, **320**, 1482–1486.

43 Q. Lu, K. Liu, H. Zhang, Z. Du, X. Wang and F. Wang, *ACS Nano*, 2009, **3**, 3861–3868.

44 N. J. Tao, *Nat. Nanotechnol.*, 2006, **1**, 173–181.

45 H. Kim and D. Segal, *ArXiv e-prints*, 2017, 1–11.

46 C. J. Lambert, *Chem. Soc. Rev.*, 2015, **44**, 875–888.

47 R. S. H. Vazquez S. Schneebeli, M. Kamenetska, R. Breslow, L. Venkataraman, M. S. Hybertsen, *Nat. Nanotechnol.*, 2012, **7**, 663–667.

48 S. Ballmann, R. Härtle, P. B. Coto, M. Elbing, M. Mayor, M. R. Bryce, M. Thoss and H. B. Weber, *Phys. Rev. Lett.*, 2012, **109**, 1–5.

49 S. V. Aradhya, J. S. Meisner, M. Krikorian, S. Ahn, R. Parameswaran, M. L. Steigerwald, C. Nuckolls and L. Venkataraman, *Nano Lett.*, 2012, **12**, 1643–1647.

50 V. Kaliginedi, P. Moreno-Garcia, H. Valkenier, W. Hong, V. M. García-Suárez, P. Buiter, J. L. H. Otten, J. C. Hummelen, C. J. Lambert and T. Wandlowski, *J. Am. Chem. Soc.*, 2012, **134**, 5262–5275.

51 S. V Aradhya and L. Venkataraman, *Nat. Nanotechnol.*, 2013, **8**, 399.

52 C. R. Arroyo, S. Tarkuc, R. Frisenda, J. S. Seldenthuis, C. H. M. Woerde, R. Eelkema, F. C. Grozema, H. S. J. van der Zant and H. S. J. Van Der Zant, *Angew. Chemie - Int. Ed.*, 2013, **52**, 3152–3155.

53 C. M. Guédon, H. Valkenier, T. Markussen, K. S. Thygesen, J. C. Hummelen and S. J. van der Molen, *Nat. Nanotechnol.*, 2012, **7**, 305–309.

54 D. Walter, D. Neuhauser and R. Baer, *Chem. Phys.*, 2004, **299**, 139–145.

55 S.-H. Ke, W. Yang and H. U. Branger, *Nano Lett.*, 2008, **8**, 43.

56 C. Finch, V. García-Suárez and C. Lambert, *Phys. Rev. B*, 2009, **79**, 33405.

57 T. Papadopoulos, I. Grace and C. Lambert, *Phys. Rev. B*, 2006, **74**, 193306.

58 Y. Li, J. Mol, S. Benjamin and A. Briggs, 2016, 1–7.

59 A. A. Kocherzhenko, L. D. A. Siebbeles and F. C. Grozema, *J. Phys. Chem. Lett.*, 2011, **2**, 1753–1756.

60 H. Hart, L. E. Craine, D. J. Hart and C. M. Hadad, *Organische Chemie*, 2007.

61 H. Nakamura, T. Ohto, T. Ishida and Y. Asai, *J. Am. Chem. Soc.*, 2013, **135**, 16545–16552.

62 G. Sedghi, V. M. García-Suárez, L. J. Esdaile, H. L. Anderson, C. J. Lambert, S. Martín, D. Bethell, S. J. Higgins, M. Elliott, N. Bennett, J. E. Macdonald and R. J. Nichols, *Nat. Nanotechnol.*, 2011, **6**, 517–523.

63 J. Park, A. N. Pasupathy and J. I. Goldsmith, 2002, **417**, 722–725.

64 V. M. García-Suárez, J. Ferrer and C. J. Lambert, *Proc. 2006 Int. Conf. Nanosci. Nanotechnology, ICONN*, 2006, 43–45.

65 B. Kim, J. M. Beebe, C. Olivier, S. Rigaut, D. Touchard, J. G. Kushmerick, X.-Y. Zhu and C. D. Frisbie, *J. Phys. Chem. C*, 2007, **111**, 7521–7526.

66 C. Olivier, B. Kim, D. Touchard and S. Rigaut, *Organometallics*, 2008, **27**, 509–518.

67 G. Sedghi, K. Sawada, L. J. Esdaile, M. Hoffmann, H. L. Anderson, D. Bethell, W. Haiss, S. J. Higgins and R. J. Nichols, *J. Am. Chem. Soc.*, 2008, **130**, 8582–8583.

68 R. Liu, S. H. Ke, W. Yang and H. U. Baranger, *J. Chem. Phys.*, 2006, **124**, 1–6.

69 K. Sugimoto, Y. Tanaka, S. Fujii, T. Tada, M. Kiguchi and M. Akita, *Chem. Commun.*, 2016, **52**, 5796–5799.

70 G. Kastlunger and R. Stadler, *Monatshefte für Chemie - Chem. Mon.*, 2016, **147**, 1675–1686.

71 P. J. Low, *Dalt. Trans.*, 2005, 2821–2824.

72 X. Xiao, D. Brune, J. He, S. Lindsay, C. B. Gorman and N. Tao, *Chem. Phys.*, 2006, **326**, 138–143.

73 W. Schmickler and C. Widrig, *J. Electroanal. Chem.*, 1992, **336**, 213–221.

74 W. Schmickler and N. Tao, *Electrochim. Acta*, 1997, **42**, 2809–2815.

75 J. Zhang, A. M. Kuznetsov, I. G. Medvedev, Q. Chi, T. Albrecht, P. S. Jensen and J. Ulstrup, *Chem. Rev.*, 2008, **108**, 2737–2791.

76 J. Zhang, Q. Chi, A. M. Kuznetsov, A. G. Hansen, H. Wackerbarth, H. E. M. Christensen, J. E. T. Andersen and J. Ulstrup, *J. Phys. Chem. B*, 2002, **106**, 1131–1152.

77 S. Guo, J. M. Artés and I. Díez-Pérez, *Electrochim. Acta*, 2013, **110**, 741–753.

78 A. J. Bard and L. R. Faulkner, *Electrochemical Methods Fundamentals and Applications*, John Wiley & Sons, inc., New York, second edi., 2012.

79 G. A. Mabbott, *J. Chem. Educ.*, 1983, **60**, 697.

80 C. Brett and A. Brett, *Springer*, 1993, 427.

81 J. Izquierdo and C. Kranz, *Eur. Polym. J.*, 2016, **83**, 428–449.

82 A. Ulman, *Chem. Rev.*, 1996, **96**, 1533–1554.

83 C. Vericat, M. E. Vela, G. Benitez, P. Carro and R. C. Salvarezza, *Chem. Soc. Rev.*, 2010, **39**, 1805.

84 K. Walzer, E. Marx, N. C. Greenham, R. J. Less, P. R. Raithby and K. Stokbro, *J. Am. Chem. Soc.*, 2004, **126**, 1229–34.

85 T. Albrecht, A. Guckian, J. Ulstrup and J. G. Vos, *Nano Lett.*, 2005, **5**, 1451–1455.

86 A. M. Kuznetsov and J. Ulstrup, *J. Electroanal. Chem.*, 2004, **564**, 209–222.

87 A. L. Samuel, *IBM J. Res. Dev.*, 1959, **3**, 210–229.

88 A. Ng, Coursera - Machine Learning, www.coursera.org/learn/machine-learning, (accessed 1 April 2017).

89 A. Szalay and J. Gray, *Science*, 2001, **293**, 2037–2041.

90 Y. Zhang and Y. Zhao, *Astron. Astrophys.*, 2004, **422**, 1113–1121.

91 F. Pereira, T. Mitchell and M. Botvinick, *Neuroimage*, 2009, **45**, S199–S209.

92 C. Ding and X. He, *Proc. twenty-first Int. Conf. Mach. Learn.*, 2004, **Cl**, 29.

93 A. Rodriguez and A. Laio, *Science*, 2014, **344**, 1492–1496.

94 Wolfram Alpha - Fuzzy Clustering, http://reference.wolfram.com/legacy/applications/fuzzylogic/Manual/12.html, (accessed 4 April 2017).

95 D. E. Gustafson and W. C. Kessel, *IEEE CDC, San Diego, Calif.*, 1978, 761–766.

96 M. S. Yang, *Math. Comput. Model.*, 1993, **18**, 1–16.

97 B. Balasko, J. Abonyi and B. Feil, *Clustering and Data Analysis Toolbox*, 2005.

98 M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. Costa-Milan, R. J. Nichols, N. J. Long and T. Albrecht, *J. Am. Chem. Soc.*, 2015, **137**, 9971–9981.

Chapter 2

Materials and Methods

2.1 Molecules under Study 51
2.2 Scanning Tunnelling Microscope 55
2.3 Electrochemistry 63
2.4 References 66

Synopsis: *This chapter describes the experimental details used throughout the thesis. It is divided into three sections. At first the analyte molecules and their assembly conditions are introduced, then the data acquisition for single molecule conductance measurements is explained. Finally, electrochemical techniques to characterize the analyte molecules are described.*

2.1 Molecules under Study

2.1.1 Reference Systems

1,8-Octanedithiol (ODT)

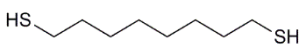


Figure 2-1: Structure of ODT

ODT (see structure in Figure 2-1, ≥ 97 %, Sigma Aldrich, St. Louis, USA) was used as purchased without additional purification. Solutions of 0.01 mM ODT in ethanol absolute (VWR International, Radnor, USA) were used to assemble (sub)-monolayers of ODT on the substrate. The substrates were incubated for 40 s and subsequently rinsed with ethanol to remove any unbound species and dried in a stream of nitrogen (zero grade, BOC, Guildford, UK).

α,ω-dithiol terminated oligo phenylene(ethynylene) (OPE)

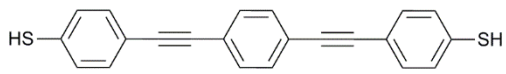


Figure 2-2: Structure of OPE

Thioacetyl terminated OPE (Figure 2-2, after removal of the acetyl protecting group) was synthesized by M. S. Inkpen, as described in reference 1. The analyte was assembled on the substrate by submersion in a 0.01 mM solution of the analyte in tetrahydrofuran (THF, Chromasolv plus for HPLC ≥ 99.9 %, Sigma Aldrich, St. Louis, USA). Prior to use, the analyte in solution was deprotected, meaning the acetyl group was removed, by addition of 1 µl NH_4OH per 1 ml of solution and subsequent incubation for 10 min to 15 min at room temperature. The substrate was then immersed for 45 s followed by rinsing with ethanol to remove unbound species and deprotection agent. After rinsing, the substrate was dried in a stream of nitrogen.

2.1.2 Ferrocene Systems

single-branched ferrocene compounds

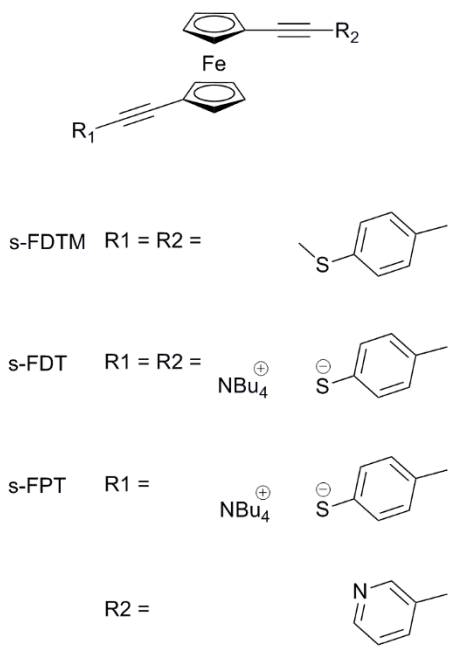


Figure 2-3: Structures of s-FDTM, s-FDT and s-FPT.

OPE type molecules incorporating ferrocene centres with thiometyl linkers (s-FDTM, Figure 2-3) were synthesised by L. E. Wilson and M. S. Inkpen, following a procedure described by M. S. Inkpen *et al.*[2] Monolayers were assembled by substrate incubation in 0.01 mM to 6 mM solutions in dichloromethane (DCM, for HPLC ≥ 99.8 %, Acros Organics, Geel, Belgium) or THF. Incubation times were varied between 40 s and 50 h. After incubation, substrates were rinsed with DCM or ethanol and dried in a stream of nitrogen.

Linear ferrocene molecules with thiol anchors (s-FDT, Figure 2-3) were synthesized by M. S. Inkpen[2] and used in concentrations between 1 mM and 0.01 mM in THF. After incubation for 40 s to 60 s, substrates were rinsed with ethanol and dried in a stream of nitrogen.

Single-branched ferrocene molecules with pyridine and thiol anchors (s-FPT, Figure 2-3) were synthesized by M. S. Inkpen, as described in reference 3. 1 mM solutions of the analyte in THF were used to incubate the substrate for 20 s to 60 s. After incubation, the substrate was rinsed with ethanol and dried in a stream of nitrogen.

b-FPT (branched Ferrocene Pyridine Thiol)

Branched ferrocene complexes with pyridine and two thiol linkers (Figure 2-4) were synthesized by M. S. Inkpen, as described in reference 3. Analogous to s-FPT, 1 mM solutions of the analyte were used to incubate the substrate for 40 s, followed by ethanol rinsing and drying in a stream of nitrogen.

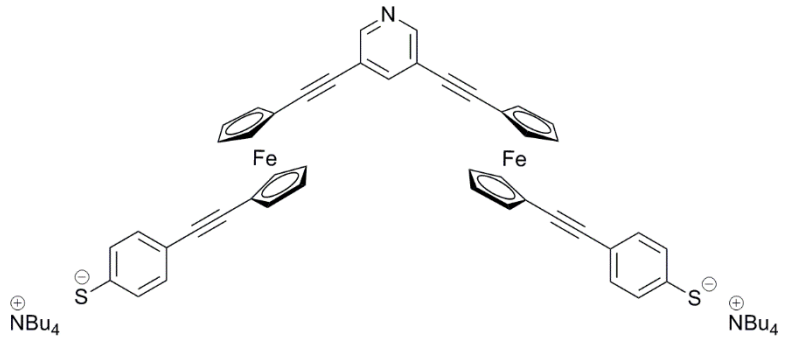


Figure 2-4: Structure of the double-branched ferrocene complex, b-FPT.

BFTM (Bi-Ferrocene Thiomethyl)

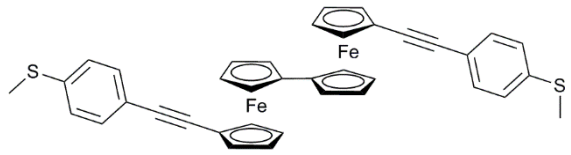


Figure 2-5: Structure of BFTM

OPE-type molecules containing two ferrocene centres and thiomethyl anchors in the para configuration (Figure 2-5) were synthesized by L. E. Wilson as described in reference 4. Incubation concentrations were varied between 1 mM and 6 mM, incubation times between 1 h and 70 h. After incubation, the substrates were rinsed with DCM and dried in a stream of nitrogen.

BFP (meta Bi-Ferrocene Pyridine)

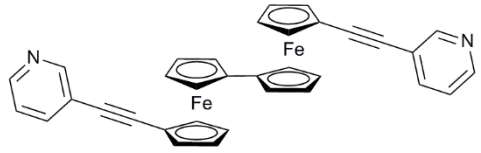


Figure 2-6: Structure of BFP

Linear bi-ferrocenes with pyridine anchors in the meta configuration (Figure 2-6) were synthesized by L. E. Wilson as described in reference 4. Incubation concentrations were varied between 2 mM and 4 mM, incubation times between 1 h and 27 h. After incubation, the substrates were rinsed with DCM and dried in a stream of nitrogen.

2.1.3 Ruthenium Systems

T-Ru$_2$M (ThioMethyl Ruthenium Macrocycle)

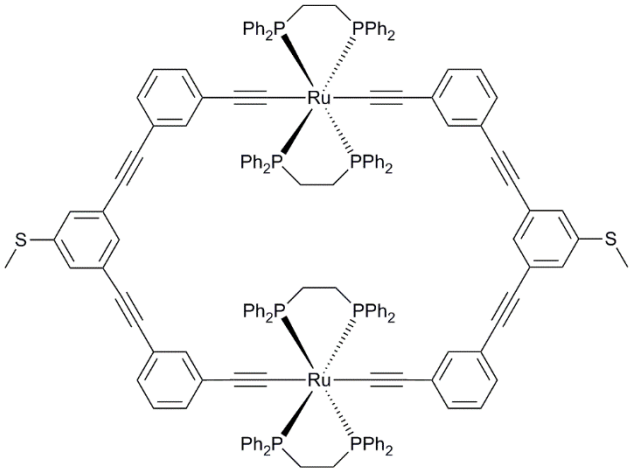


Figure 2-7: Structure of the ruthenium macrocycle, T-Ru$_2$M. The structure was provided by Lucy E. Wilson.

Macrocyclic ruthenium with thiomethyl linkers (Figure 2-7) was synthesized in a combined effort by L. E. Wilson and M. S. Inkpen. A publication describing the synthesis is currently in preparation by Wilson *et. al.* Substrates were incubated in 2 mM T-Ru$_2$M in Chloroform ($CHCl_3$, ≥ 99 %, anhydrous, Sigma Aldrich, St. Louis, USA) for 1 h. After incubation, the substrates were rinsed with $CHCl_3$ and dried in a stream of nitrogen.

sPy-Ru$_4$M (short Pyridine Ruthenium Macrocycle)

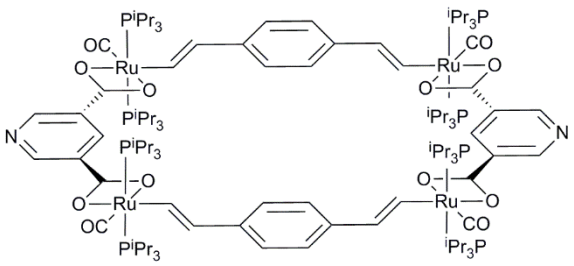


Figure 2-8: Ruthenium Macrocycle (sPy-Ru$_4$M) with short pyridine linkers. Structure was used with permission from Stefan Scheerer.

Ruthenium macrocycles with short pyridine anchors (Figure 2-8) were synthesized by P. Anders *et al.*[5] Incubation concentration was varied between 0.2 mM and 0.9 mM, incubation times between 40 s and 3 d. After incubation, the substrates were rinsed with DCM and dried in a stream of nitrogen.

lPy-Ru$_4$M (long Pyridine Ruthenium Macrocycle)

Ruthenium macrocycles with long pyridine anchors (Figure 2-9) were synthesized by P. Anders *et al.* at the group of Rainer Winter in Konstanz, Germany. Incubation concentration was varied between 0.04 mM and 1 mM. After incubation for 10 min, the substrates were rinsed with DCM and dried in a stream of nitrogen.

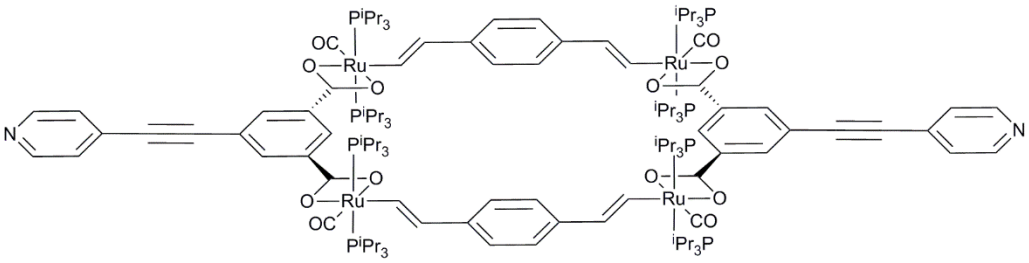


Figure 2-9: Ruthenium macrocycle, lPy-Ru$_4$M, with long pyridine linkers. The structure was modified with permission from Stefan Scheerer.

T-Ru$_4$M (Thioacetyl Ruthenium Macrocycle)

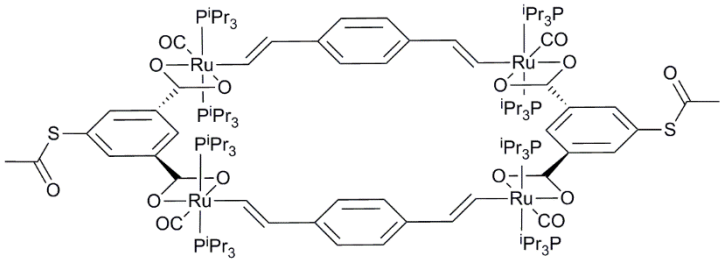


Figure 2-10: Ruthenium Macrocycle with thioacetyl linkers. The structure was modified with permission from Stefan Scheerer.

A ruthenium macrocycle with thioacetyl anchors (Figure 2-10) was synthesized by P. Anders *et al.* at the group of Rainer Winter in Konstanz, Germany. Incubation concentration was varied between 0.004 mM and 0.6 mM, incubation times between 10 min and 25 h. After incubation, the substrates were rinsed with DCM and dried in a stream of nitrogen.

2.2 Scanning Tunnelling Microscope

2.2.1 Scanning Tunnelling Microscope General

The scanning tunnelling microscope (STM) used throughout the project was an Agilent Technologies AFM/STM 5100 (Keysight, Santa Rosa, USA). A linear scanner with 10 µm range was used with a preamplifier sensitivity of 10 nA/V and a maximum sampling frequency of 20 kHz. To minimize mechanical noise, the setup was located on an optical bread board, levitating on an optical table isolator (Model PBI52515 and SDA90120 respectively, Thorlabs, Newton, USA). The system was encased in a sound proofing box, wrapped in copper mesh, to provide additional shielding from noise. Within the encasing, the microscope was rested on tennis balls to reduce vibrational coupling between table and instrument. Imaging and spectroscopy were performed in air and under ambient conditions, unless otherwise stated.

2.2.2 Substrate and Cleaning Procedure

Substrates used for imaging and spectroscopy were highly orientated Au(111) single crystals (diameter 11 mm, thickness 1 mm, 99.999 % purity, polished with roughness < 0.01 µm and orientation accuracy < 0.1 °, MaTeck, Jülich, Germany).

Prior to use, single crystals were electrochemically polished, meaning the surface was electrochemically oxidized and the oxide was subsequently removed to restore a pristine Au(111) surface for each experiment. The method applied here was adapted from Tang *et al.*[6] For the polishing process, the single crystal was rested on a gold coil (purity > 99.9 %, diameter 1 mm, Goodfellow, Huntingdon, UK) and submerged into a 0.1 M H_2SO_4 solution (95 %, VWR International, Radnor, USA). A ring-shaped Pt counter electrode with a ring diameter of approximately 10 mm was placed in solution around the submerged disc (Wire diameter 0.5 mm, purity > 99.9 %, Goodfellow, Huntingdon, UK). A potential of 6 V was applied between gold disc (anode) and ring (cathode), using a DC Power Supply unit (Model HY3003D, Rapid Electronics, Colchester, UK). Upon application of the potential, a red gold oxide layer started forming on the disc. After approximately 5 minutes the oxidation process was stopped and the disc was rinsed with deionized water (Purite Select Fusion, resistivity > 18.2 MΩ·cm). Subsequently, the red gold oxide was removed by submersion in 1 M hydrochloric acid (HCl: 37 %, VWR International, Radnor, USA). After the oxide removal, the disc was thoroughly rinsed with water. The polishing process was repeated a minimum of three times.

To smoothen out an increased number of surface defects possibly introduced during the etching process, the disc was annealed.[7] For this purpose, the gold disc was rested on an alumina boat (Z561754, Sigma Aldrich, St. Louis, USA) and placed in a furnace (Nabertherm LE 4/11/R6, Lilienthal, Germany). The substrate was heated to 850 °C and annealed for a total time of 36 h.

Immediately prior to use, the disc was cleaned to remove any surface contaminants. The substrate was placed in an alumina boat and heated with a hydrogen flame (Hydrogen zero grade, BOC, Guildford, UK) until bright red glow. The substrate was kept at this colour for a period of 10 s and then allowed to cool down for another 10 s. This process was repeated three times. The substrate was then left to cool to ambient temperature before use in the STM or incubation in an analyte solution.

2.2.3 STM Tip Preparation

Gold tips for both imaging and spectroscopy were prepared by electrochemical etching, adapting a method published by Wang *et al.*[8] In the process, a quasi-continuous gold wire (diameter 0.25 mm, 99.99 % as drawn, Goodfellow, Huntingdon, UK) was submersed about 1 mm deep into a mixture of (1:1 v/v) ethanol absolute and hydrochloric acid. A ring-shaped gold counter electrode (diameter 1 mm, 99.99 % as drawn, Goodfellow, Huntingdon, UK) with a ring diameter of about 10 mm was positioned so that the gold wire was in the centre of the counter electrode (Figure 2-11).

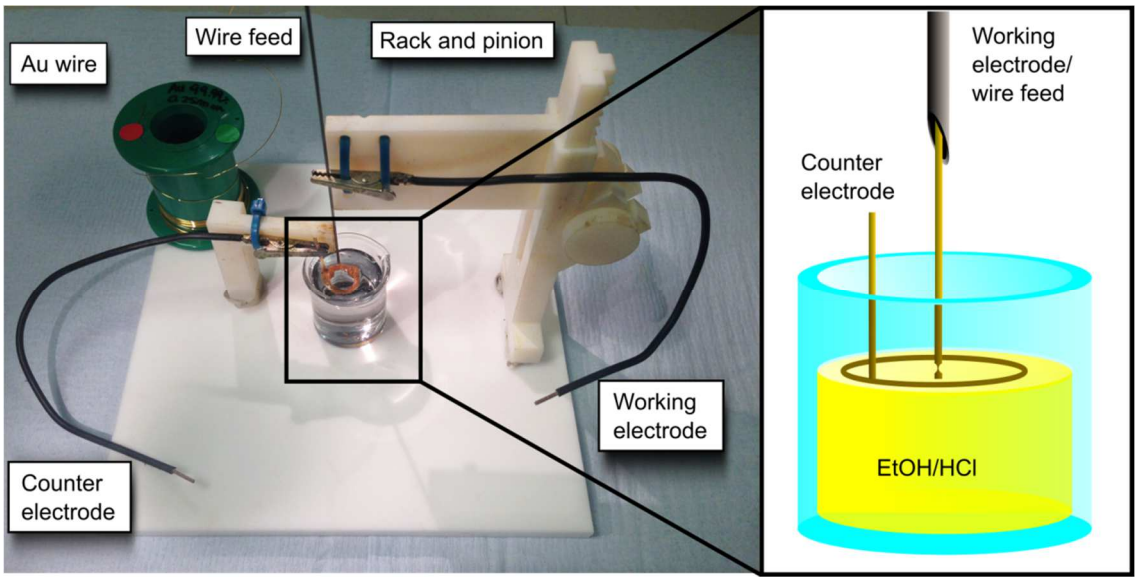


Figure 2-11: Image of the 3D printed tip etching setup (left). Au wire is fed from a stock roll through a needle into the etching solution. The needle is connected to the working electrode of the bipotentiostat. Rack and pinion are used to lower the wire into solution and to remove the wire from solution after the etching process. The ring-shaped gold electrode in solution is connected to the counter electrode of the bipotentiostat. A schematic of the etching process is shown on the right.

The counter electrode was also submerged into solution to about 1 mm depth. A CHI760C potentiostat (CH Instruments, Austin, USA) was used in the potentiostatic *I(t)* method to apply a potential of 2.4 V between the tip (anode) and the counter electrode (cathode), starting the etching process. At the meniscus, which the solution formed with the gold wire, the etching rate is increased compared to the rest of the submersed wire. The increased etching-rate leads to a thinning of the wire at the meniscus. After an etching period of typically 40 s to 80 s the lower part of the wire, which was submerged in solution, fell down. The potential was turned off immediately after the etching current dropped to zero. The un-submersed bit of the gold wire formed a sharp tip. The wire with the tip was then cut to a length of approximately 15 mm and dipped in H_2O and ethanol to remove any hydrochloric acid.

Subsequently, the tips were lightly bent into an s-shape to be positioned in the STM scanner tube, leaving about 5 mm of the tip outside the scanner. About 50 % of tips could maintain a stable tunnelling junction, meaning trace and retrace during imaging in the STM would align closely.

To provide a robust and reproducible setup for tip etching, a tip-etching station was designed with FreeCAD and printed in 3D. Printable CAD files can be obtained from the author.

2.2.4 STM Data Collection

Data collection during *I(s)* spectroscopy experiments on the STM had been automated previously by M. S. Inkpen[9] using Macro Scheduler (MJTNet Ltd., London, UK). For the automated experiment, all experimental parameters were input into an excel spreadsheet. The macro scheduler would then read the spreadsheet and run the experiment accordingly.

I(s) Spectroscopy

For *I(s)* spectroscopy experiments, the following parameters were predefined: Integral and proportional (I and P) gain 10 % and 15 %, respectively. (Note that during the actual withdrawal of the tip, the scanner feedback loop was deactivated. I and P gain were only active during the periods between scans.) Current-distance traces were typically recorded from 0 nm to 8 nm, relative to the setpoint distance s_0, i.e. tip-substrate distance at setpoint tunnelling current I_0. Current-distance traces were recorded over a period of 1 s, sampling the current 2000 times, thus resulting in a withdrawal speed of 8 nm/s and a sampling frequency of 2 kHz.

The macro scheduler typically recorded 50 *I(s)* traces per tip position and then moved the tip to a random, new position within the current 200 nm by 200 nm scan area. Within a scan area, typically 10 tip positions were randomly selected, resulting in 500 *I(s)* traces per scan location. After recording 500 traces, the macro would randomly select a new scan area on the substrate and record another 500 traces. The process was repeated automatically until a predefined number of *I(s)* traces was reached. Bias voltage and setpoint current were varied during experiments between -1 V to 1 V and 0.1 nA to 80 nA, respectively.

Start and finish position could be reversed to measure in the approach mode. In the approach mode, the initial distance was typically set to 7 nm and the final distance was set to – 1 nm, relative to setpoint distance s_0. Approaching the tip to a distance of – 1 nm typically resulted in currents larger than 100 nA, at which the preamplifier saturated.

Cleaning procedure

The *I(s)* spectroscopy experiments were run largely unsupervised. Since tip contamination can never be fully avoided in ambient STM experiments, it was necessary for the macro to detect when traces associated with tip contamination occurred. Tip contamination is thought to happen when the sharp exponential decay associated with a "clean" junction is replaced with noisy slowly decaying or non-exponential traces, Figure 2-12. The macro scheduler had the following cleaning procedure in place to automatically detect contamination of this deviation from clean exponential traces: After recording an *I(s)* trace, the macro scheduler checked a subsection of the *I(s)* trace for a certain image (Insets in Figure 2-12).

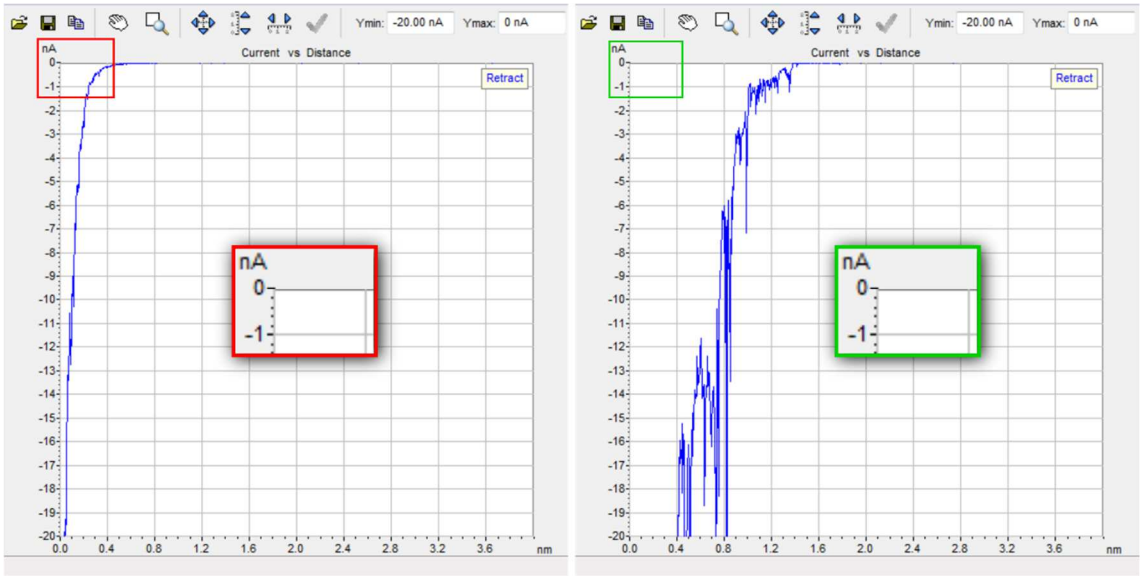


Figure 2-12: Spectroscopy window of PicoView 1.14.4 during data collection. Left side of the image shows a typical "clean" I(s) trace. Subsection of window is not found because it is obscured by I(s) trace. This is indicated by the red inset. In the case of a "dirty" trace (right hand side), the subsection is found by the macro scheduler, as indicated by green inset.

In the presence of a clean *I(s)* trace, this subsection of the graph would be partially obscured by the trace. If the tip got contaminated, several *I(s)* traces in a row would show a slow and noisy decay. In this case, the subsection of the graph would be exposed, the macro would find the subsection in the image and the trace would be counted as noisy. If the macro scheduler recognized five noisy traces in a row, it would assume the tip was contaminated. It would then increase the *I* gain to 100 % and thereby shake the tip to try to get rid of the contamination. If the contamination persisted, *e.g.* tip shaking would be repeated three times without going back to fast decaying *I(s)* traces, the macro scheduler would apply a voltage pulse of 9 V for a period of 0.1 ms to sputter off any contamination.[10,11,12] In case the traces did not go back to clean exponential decays within the application of two voltage pulses, the scan

location was assumed to be contaminated and the tip was sent to a different position. Note that independent of the presence of tip contamination, all *I(s)* traces were saved. The described cleaning process was developed by M. Inkpen.

Imaging

During the *I(s)* spectroscopy experiments, the macro scheduler recorded images at every scan location, before starting any spectroscopy. Images were recorded in constant current mode, typically with the following parameters: I and P gain 10 % and 15 %, respectively. Setpoint current 0.5 nA and bias voltage 0.3 V.

2.2.5 Simulated *I(s)* Data

To develop data analysis techniques, it was desirable to simulate *I(s)* traces. MATLAB (MATrix LABoratory, Natick, USA) scripts were written to generate simulated data. All relevant scripts are included in the appendix. The simulated data was designed to resemble experimental data closely. Current-distance traces were generated at different noise levels. Different noise levels were generated by pairwise multiplying a clean exponential trace with a "random vector". The "random vector" consisted of elements distributed with a certain standard deviation (STDEV) around one. A large STDEV would thus result in a noisy *I(s)* trace (see examples in Figure 2-13). By multiplying the clean trace with random numbers distributed around one, the noise was proportional to current level, analogue to what was observed in experimental *I(s)* traces. The setpoint current and decay coefficient could be distributed around a certain value, see Figure 2-14. The distribution was included by adding a random number distributed around zero with a certain STDEV to the setpoint current.

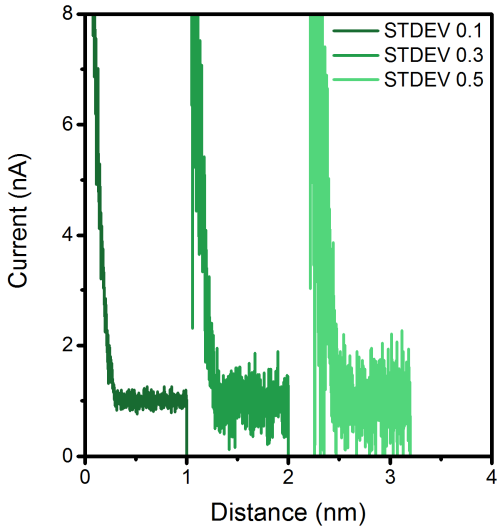


Figure 2-13: Simulated I(s) trace with different noise levels of 0.1, 0.3 and 0.5. Dark to light green, traces were offset for clarity.

The presence of a molecular junction could be simulated by interrupting the exponential decay at a plateau current I_p and keeping it constant up to a break-off distance s_b. After break-off, the trace would go back to the original exponential decay. Both break-off distance and plateau current could be varied by adding a random number, distributed around zero with a certain STDEV as indicated with black arrows and increasingly faint traces in Figure 2-14.

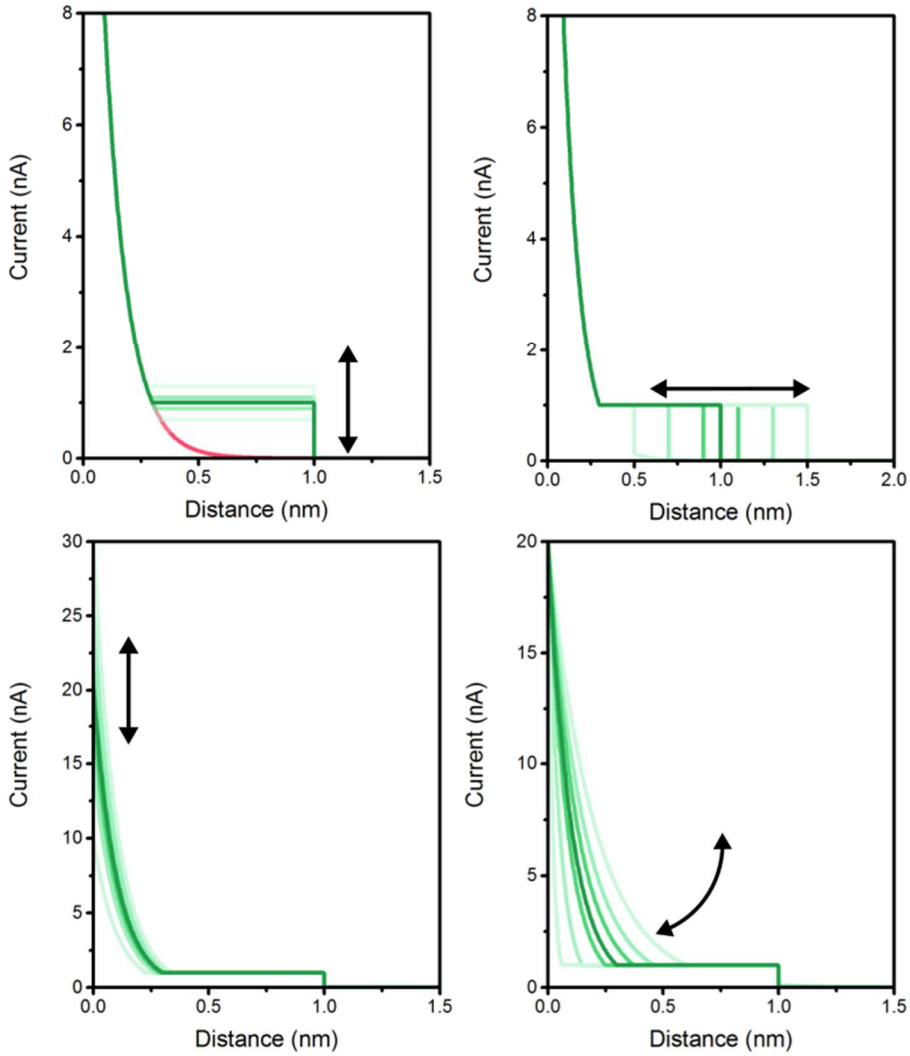


Figure 2-14: Simulated data, without plateau (top left in red) and plateaus at various plateau currents (top left in green). Different break-off distances (top right), I_0 (bottom left) and decay factors (bottom right).

Different junction formation probabilities (JFP) could be simulated by changing the ratios between traces with and without plateaus. The formation of molecular junctions with more than one molecule was simulated by multiplying the plateau current with the number of molecules K. The probability of K molecules forming a junction would decrease with $(JFP)^K$.

The plateau current was modulated with different additional signals to mimic plateau-shapes that were observed in the experimental data. Telegraphic noise was simulated by randomly switching between two values during the plateau. As an approximation to non-linear plateaus, a sine-shaped signal with a certain amplitude and frequency was added to the plateau signal (See Figure 2-15).

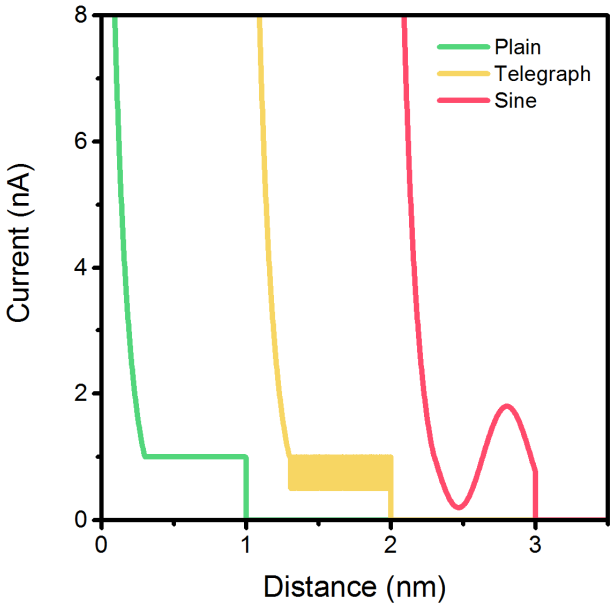


Figure 2-15: Noise-free simulated data with plain plateau, telegraphic noise, and sine-shaped plateau, from left to right. Traces were offset for clarity.

2.2.6 Data Analysis Considerations

The method for data analysis evolved during the process of the work. Initially, *I(s)* traces with plateau features longer than 0.1 nm were selected by hand. Further into the project, a conventional data analysis algorithm based on individual current histograms was developed by M. S. Inkpen *et al.*[9] Later on, MPVC[1], a vector-based approach, was implemented. Data analysis with MPVC will be discussed in detail in Chapter 3.

In some cases, spectroscopy data was filtered before analysis. Those instances are indicated in the main text. The filtering proved helpful, especially with short and low-conductance plateaus. The filter parameters used were as follows: The pass-band start frequency was 10^{-6} Hz, the stop-frequency 5 Hz, Apass 0.5 and the attenuation in the stop band was 40 db.

2.3 Electrochemistry

2.3.1 General

All electrochemical experiments were performed using either one of the following bipotentiostats: Gamry Reference 600 potentiostat (Gamry Instruments, Warminster, USA) or CHI760C potentiostat (CH Instruments, Austin, USA).

Before use in electrochemical experiments, any glass ware was rigorously cleaned by boiling in 20 % nitric acid (68 %, VWR International, Radnor, USA) for a period of at least 30

min. After boiling, glassware was rinsed three times with deionised water and sonicated in deionised water for 10 min (Ultrasonic cleaning bath 1200 TH, VWR, Radnor, USA). Glassware was then rinsed again three times with deionized water and dried in a drying oven for at least 3 h.

2.3.2 Procedure for Solution CV

Platinum wires, (Diameter 1 mm, Goodfellow, Huntingdon, England) rolled into a coil shape, were used as counter electrode and pseudo reference electrode. Pt electrodes were cleaned by annealing with a hydrogen flame until red glowing. Red glow was kept for 10 s and wires were subsequently rinsed with deionised water and annealed again for a duration of 10 s. A glassy carbon working electrode was cleaned by polishing with a felt disc and alumina slurry (0.5 µm) for 15 min. After polishing, the glassy carbon electrode was thoroughly rinsed with deionized water.

Solution electrochemical experiments were performed on 1 mM to 10 mM solutions of the analyte molecule in DCM. 0.1 M tetrabutylammonium hexafluorophosphat ($^{n}Bu_4NPF_6$, for electrochemical analysis, ≥ 99.0%, Sigma Aldrich, St. Louis, USA) was used as supporting electrolyte. Experiments were performed under the absence of oxygen in custom built electrochemical cells. To remove oxygen, cells were flushed with Argon (zero grade, BOC, Guildford, UK) for 15 min prior to any experiment. Measurements were performed with platinum pseudo reference electrode. Ferrocene ($[Cp_2Fe]^+/[Cp_2Fe]$, 98 %, Sigma Aldrich, St. Louis, USA) or Decamethylferrocene ($[Cp^*_2Fe]^+/[Cp^*_2Fe]$, 0.495 V versus $[Cp_2Fe]^+/[Cp_2Fe]$, synthesized by M. S. Inkpen following published procedures) were used as external reference.

Open circuit potential, cyclic voltammograms at different scan rates, and potentiostatic impedance spectroscopy were recorded to assess reversibility of oxidation/reduction processes and potentials of the redox processes.

2.3.3 Procedure for Surface Immobilized CV

Electrochemical experiments on surface immobilized molecules on polycrystalline gold electrodes were performed by incubating gold bead electrodes in 1 mM solutions of the analyte in THF or DCM. After incubation for various times, the electrodes were rinsed with ethanol and dried in a stream of nitrogen. Gold bead electrodes were then used as working electrodes in an otherwise identical manner as for the solution voltammetry.

2.3.4 Method for Single Crystal Voltammetry

After spectroscopy and imaging, (sub)-monolayers of molecular adsorbate on Au(111) single crystals were investigated by electrochemistry. To facilitate measurements, the STM cell containing the single crystal was cast into mouldable silicone rubber (Ecoflex 5, Smooth-on, Macungie, USA). After cleaning procedure described above, platinum wires were used as pseudo reference electrode and counter electrode in the STM cell. The gold single crystal functionalized with the surface immobilized analyte was used as working electrode as shown in Figure 2-16.

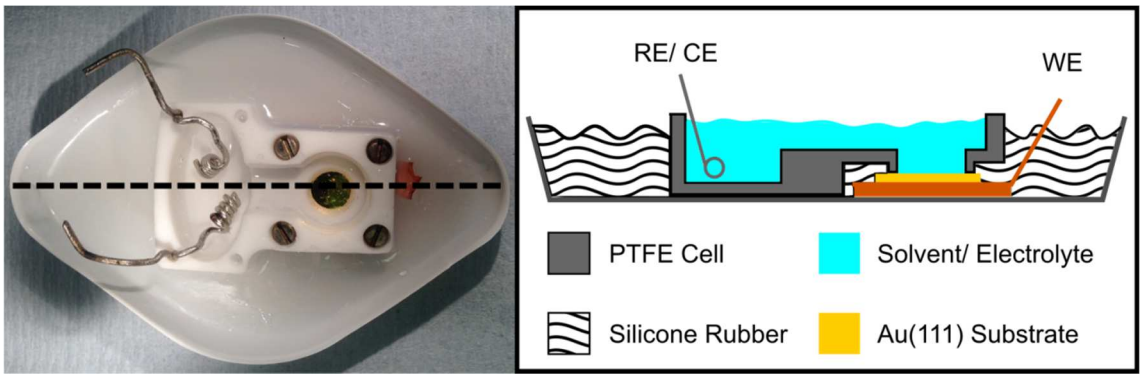


Figure 2-16: Image of the STM cell setup for surface immobilized cyclic voltammetry on Au(111) single crystals (left). Right hand side of the figure is a schematic of the cross section of the STM cell along the dashed line.

0.1 M $^{n}Bu_4NPF_6$ in DCM was degassed by bubbling dry Argon (Zero Grade, BOC, Guildford, UK) through the solution before adding it to the cell. The STM cell was kept in the environmental chamber of the STM, which was flushed with Argon before for at least 20 min and during the measurement to ensure minimal oxygen in the setup.

After experiments on the surface immobilized analyte, reductive desorption could be probed in 0.1 M sodium hydroxide as electrolyte. For this purpose, the reference electrode was exchanged with a saturated Ag/AgCl reference electrode. The STM cell was rinsed with deionized water and filled with degassed 0.1 M sodium hydroxide solution.

To estimate the exposed electrochemical surface area, gold oxide reduction could be measured following a procedure detailed in reference 13. Therefore, a 0.5 M solution of sulphuric acid was degassed and used as electrolyte. A saturated Ag/AgCl electrode was used as reference and a platinum coil as counter electrode.

2.4 References

1 M. Lemmer, M. S. Inkpen, K. Kornysheva, N. J. Long and T. Albrecht, *Nat. Commun.*, 2016, **7**, 17–19.

2 M. S. Inkpen, A. J. P. White, T. Albrecht and N. J. Long, *Dalton Trans.*, 2014, **43**, 15287–90.

3 M. S. Inkpen, T. Albrecht and N. J. Long, *Organometallics*, 2013, **32**, 6053–6060.

4 L. E. Wilson, C. Hassenrück, R. F. Winter, A. J. P. White, T. Albrecht, N. J. Long, C. Hassenruck, R. F. Winter, A. J. P. White, T. Albrecht and N. J. Long, *Chem. - A Eur. J.*, 2016, **2017**, 1–10.

5 S. Scheerer, M. Linseis, E. Wuttke, S. Weickert, M. Drescher, O. Tröppner, I. Ivanović-Burmazović, A. Irmler, F. Pauly and R. F. Winter, *Chem. - A Eur. J.*, 2016, **22**, 9574 – 9590.

6 J. Tang, M. Petri, L. A. Kibler and D. M. Kolb, *Electrochim. Acta*, 2005, **51**, 125–132.

7 D. Porath, E. Bar-Sadeh, M. Wolovelsky, A. Grayevsky, A. Goldstein and O. Millo, *J. Vac. Sci. Technol. A*, 1995, **13**, 1165–1170.

8 X. Wang, Z. Liu, M. De Zhuang, H. M. Zhang, X. Wang, Z. X. Xie, D. Y. Wu, B. Ren and Z. Q. Tian, *Appl. Phys. Lett.*, 2007, **91**, 101105-1-101105–3.

9 M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. Costa-Milan, R. J. Nichols, N. J. Long and T. Albrecht, *J. Am. Chem. Soc.*, 2015, **137**, 9971–9981.

10 C. S. Chang, W. B. Su and T. T. Tsong, *Phys. Rev. Lett.*, 1994, **72**, 574–577.

11 C. X. Guo and D. J. Thomson, *Ultramicroscopy*, 1992, **42–44**, 1452–1458.

12 H. J. Mamin, P. H. Guethner and D. Rugar, *Phys. Rev. Lett.*, 1990, **65**, 2418–2422.

13 H. Angerstein-Kozlowska, B. E. Conway, A. Hamelin and L. Stoicoviciu, *Electrochim. Acta*, 1986, **31**, 1051–1061.

Chapter 3

Multivariate Pattern Analysis

3.1 Motivation for Vector-Based Analysis 68
3.2 Analysis with MPVC 68
3.3 MPVC Analysis of Simulated Data 71
3.4 MPVC Analysis of ODT 74
3.5 MPVC Analysis of OPE 79
3.6 Summary and Conclusion 82
3.7 References 83

Synopsis: *In this Chapter, a vector-based analysis method for single-molecule conductance data is presented. After pointing out the need for the method, mathematical details are given. This section is followed by the application of the method on simulated I(s) data. Finally, the algorithm is applied to two well-studied molecular systems.*

3.1 Motivation for Vector-Based Analysis

In Chapter 1 it was described that typically many current-distance traces are recorded in single molecular conductance measurements. To increase the signal-to-noise ratio, some form of data selection is typically applied.[1,2] The data selection is usually guided by some expectation of the *I(s)* trace. For example, the current is expected to plateau, while through-molecule transport is dominant. The selected plateau traces are then typically compiled into current or conductance histograms, where peaks are interpreted as the conductance of the molecule.[3,4]

However, this approach comes with problems. First, the reliance on an expected signal shape can influence the result. If the through-molecule current is not constant but more complex for example, the expected outcome is *a priory* unknown.[4] Second, the conventional histogram analysis has a strong preference for the "majority species", meaning most abundant class of molecular event. However, even for simple molecules like 1,8-Octanedithiol (ODT) it is now well-documented that different conductance classes can be observed.[5,6,7,8,9] Slanted and non-linear plateaus as well as plateaus with telegraph noise have also been reported on for different molecules, black traces Figure 3-1.[10,11,12,13,14,15,16] Different plateau shapes or conductance classes may contain information about binding dynamics, conformational changes, multiple molecules bridging the junction or other processes on the nanoscale. Utilizing as much information of the recorded *I(s)* traces as possible is important to learn about charge transport processes on the nanoscale.[4]

In the following sections, a new, multi-parameter vector-based classification (MPVC) approach to single molecule conductance data analysis is presented. The method does not make any assumptions about presence or type of molecular signal. MPVC offers the possibility for largely unsupervised analysis of single molecular conductance data. The method can find charge transport events beyond the majority species. In the following sections, MPVC is first demonstrated on simulated conductance data and then applied to two well-published reference systems, namely ODT and OPE.

3.2 Analysis with MPVC

It was discussed in Chapter 1, that a molecular conductance data set with M *I(s)* traces and N data points per *I(s)* trace can be considered as an N by M matrix. Or phrased differently, each of the M *I(s)* traces can be thought of as a vector in N-dimensional space. Vector operations can be performed on these.

Figure 3-1 gives a selection of *I(s)* traces. Shown in blue (red) is a simulated exponentially decaying trace without (with) plateau. Yellow and green traces are simulated traces with sine-

shape plateaus as an example for non-linear plateaus. Shown in black are experimental *I(s)* traces of ODT. The dashed magenta line is a simulated, noise-free exponentially decaying trace that will be used as reference (*vide infra*).

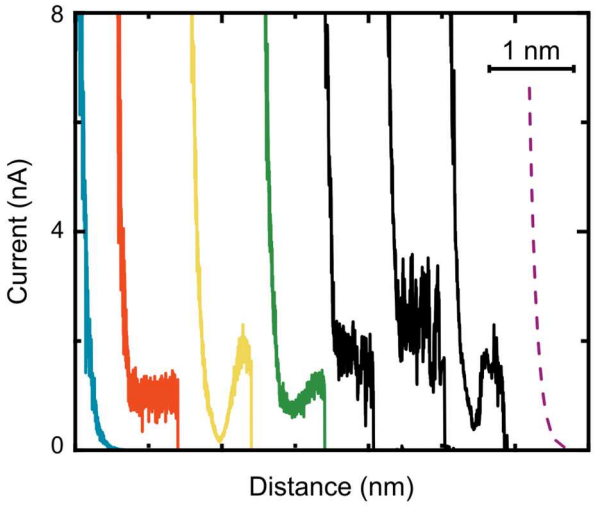


Figure 3-1: Simulated I(s) traces with different plateaus are shown in different colours. Black traces are experimental traces with ODT in the junction. The dashed, magenta trace is a typical reference trace as described below. Traces were offset for clarity The figure was redrawn with permission from Lemmer et al., 2016.[4]

I(s) traces can be considered to be vectors with N components. Specifically, N = 2000 in this case, since the current was sampled 2000 times for the traces in Figure 3-1. For illustration purposes, let us now consider *I(s)* traces with N = 2, meaning that the current was only sampled twice during each withdrawal of the STM tip. In this case, only two current values describe each *I(s)* trace, namely I_n and I_{n+1}. Each *I(s)* trace can thus be represented by a vector in two dimensions, $\mathbf{I}_m$ Figure 3-2 **A**. The same concept can be applied to a simulated reference vector, **R**.

In two dimensions as well as more generally in N dimensions, the distance vector $\mathbf{Y}_m$ can be calculated between **R** and any trace $\mathbf{I}_m$. The concept of MPVC relies on calculating the following three vector properties of every distance vector with respect to the above-mentioned reference vector:

i) The norm ΔX_m of $\mathbf{Y}_m$

ii) The angle θ_m between $\mathbf{Y}_m$ and -**R**

iii) The reduced Hamming distance, h_r, between $\mathbf{Y}_{m,reduced}$ and $\mathbf{R}_{reduced}$

The length of the distance vector, ΔX, is here defined by the Euclidian norm, equation 3.1. This is a measure of similarity between the reference vector and the current vector, $\boldsymbol{I}_m$. If the traces are identical, the distance is zero. If they are very different, the distance is large.

$$\Delta X = || \boldsymbol{Y}_m || = \sqrt{\sum_{k=1}^{N} | \boldsymbol{Y}_k |^2} \tag{3.11}$$

The second parameter, θ, was introduced because two traces can have the same ΔX without being similar. This is indicated in Figure 3-2 **A**. The red and yellow traces have a similar ΔX, even though the traces are not very similar. The angle, θ, was thus defined as follows: [4]

$$\cos(\theta_m) = -\frac{\boldsymbol{R} \cdot \boldsymbol{Y}_m}{| \boldsymbol{R} | \cdot | \boldsymbol{Y}_m |} \tag{3.12}$$

The third parameter is h_r. The reduced Hamming distance is the Hamming distance of the "reduced" distance and reference vectors, $\mathbf{Y}_{m,reduced}$ and $\mathbf{R}_{reduced}$. Reduced meaning here that the components of the vectors were divided by their absolute value. This transformed vectors into vectors containing only 1, -1 and 0. The Hamming distance is the number of component changes necessary to transform $\mathbf{Y}_{m,reduced}$ into $\mathbf{R}_{reduced}$.[17]

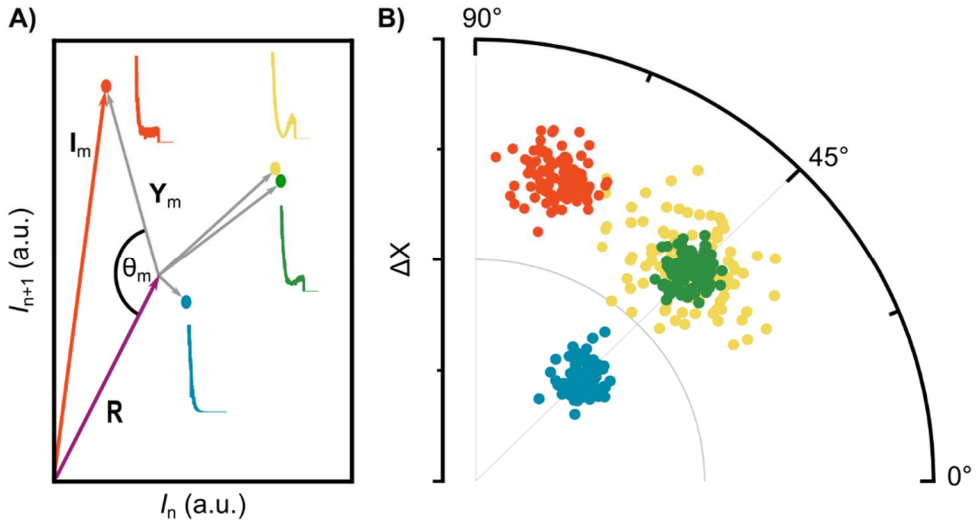


*Figure 3-2: Principle of MPVC - **A**) For illustration purposes I(s) traces contain only two current values, I_n and I_{n+1}. Every I(s) trace can thus be represented by a vector in two dimensions, $\boldsymbol{I}_m$. Similar traces have a similar distance, $|\boldsymbol{Y}_m|$ and angle, θ_m, to a common reference vector, **R**. Four different types of I(s) traces are shown in the graph in different colours. **B**) Illustrative polar plot showing that different types of I(s) traces form different clusters when ΔX and θ of many traces are plotted. The figure was redrawn with permission from Lemmer et al., 2016.*[4]

By calculating these three vector properties for every *I(s)* trace with respect to the reference vector, the whole data set is transformed into three-dimensional space while retaining information about the conductance measurement. Since θ is an angle, plotting in polar or

cylindrical coordinates is most convenient. This is indicated in Figure 3-2 **B**. The blue, exponentially decaying trace without plateau is similar to the reference vector. ΔX is therefore small. The red trace contains a plateau and is dissimilar to the reference. ΔX is large. Red and yellow traces have a similar ΔX, but their different plateau shapes lead to different angles. Yellow and green traces both have similar shapes, but in this illustration, can be distinguished with the third parameter, h_r.

When the data set is shown in three dimensions, clusters in the data can become obvious. Cluster meaning here point clouds of increased density, where similar traces are part of the same cluster and dissimilar traces are part of different clusters.[18] In the introduction chapter it was explained that in this representation, a number of different methods are available to cluster the data set into meaningful subgroups.

After splitting up the data set into meaningful clusters, the individual clusters can be analysed with 1D or 2D current or conductance histograms.[3] It is noteworthy that this is a conceptual departure from conventional histogram analysis because similar traces that are thought to describe similar or identical configurations on the nanoscale are grouped together. The 1D or 2D histogram analysis after the cluster process is thus focused on single configurations, rather than a convolution of all observed processes with a strong focus on the most probable process. MPVC thus has the capability of also analysing conductance events that occur with a lower probability than the majority species.

3.3 MPVC Analysis of Simulated Data

To validate the new analysis approach, simulated data sets were created. Details of the simulated data can be found in the Methods Chapter. To summarize, 1000 *I(s)* traces were created. Samples of each type of *I(s)* trace are given in Figure 3-3 **A**. Of the 1000 traces 400 did not contain plateaus. 200 each contained plain (red), telegraphic (magenta) and sine-shaped (green) plateaus. **B** shows a 2D current histogram of the whole simulated data set. The histogram illustrates that multiple charge transport processes can be present in one data set. Each process corresponds to one plateau shape in this example. The superposition of all these processes can prevent conclusive analysis even for very clean data.

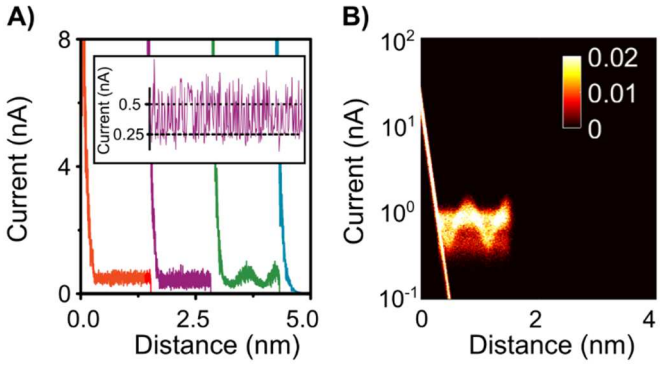


*Figure 3-3: **A**) Simulated I(s) traces with plain, telegraphic, sine-shaped and without plateau. Inset shows a magnification of the telegraphic plateau, emphasizing the binary switching. Traces are offset for clarity. **B**) Logarithmic 2D current histogram of all traces in the simulated data set. The figures were reproduced with permission from reference 4.*

The MPVC analysis process was then applied to the simulated data set. As reference vector, an exponentially decaying trace with a decay factor of 1.0 $Å^{-1}$ was used, eq. 1.1. When plotting ΔX and θ of the whole data set, four distinct clusters become apparent, Figure 3-4. The third parameter, h_r, was omitted here because the different types of traces formed distinct clusters and clustering was straightforward without the Hamming distance parameter.

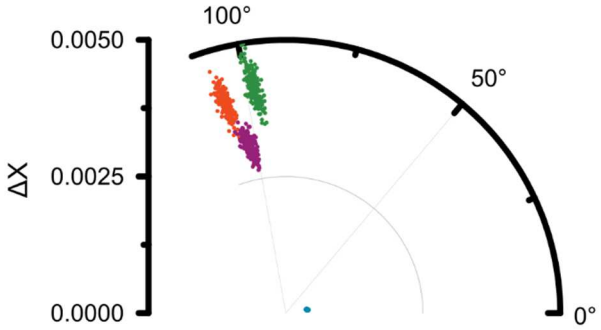


Figure 3-4: Polar plot of the whole data set. Dimensionality was reduced with the MPVC approach. Data forms four distinct clusters corresponding to the four different types of I(s) traces. Colour assignment was done based on parameter distribution (vide infra). The figure was redrawn with permission from Lemmer et al., 2016.[4]

The cluster assignment was performed by fitting distributions to the parameters, Figure 3-5. In the first step, the plain exponential traces were separated from the other traces by fitting a Gaussian distribution to ΔX of all traces. All traces within a 2 σ cut-off (dashed blue lines) were considered to be part of the distribution, **A**. In a second step the telegraphic traces were isolated from the rest by fitting a Gaussian distribution to ΔX, **B**. Those two fits to ΔX were

performed separately, because the distribution widths were vastly different. The remaining sine-shaped and plain plateau trace are part of the same distribution and are shown in black. Those two types of traces are not distinguishable from the ΔX parameter. However, they form two distinct distributions when plotting the second parameter, θ (Panel **C**). **D** shows the distribution of the reduced Hamming distance for completeness. However, this third parameter was not required in this instance. Almost all traces were correctly assigned to their respective cluster. Only 3 traces were assigned to the wrong cluster, which results in a misclassification ratio of 0.3 %. 20 traces were outside 2 σ cut-offs and were not assigned to any cluster (2 %).

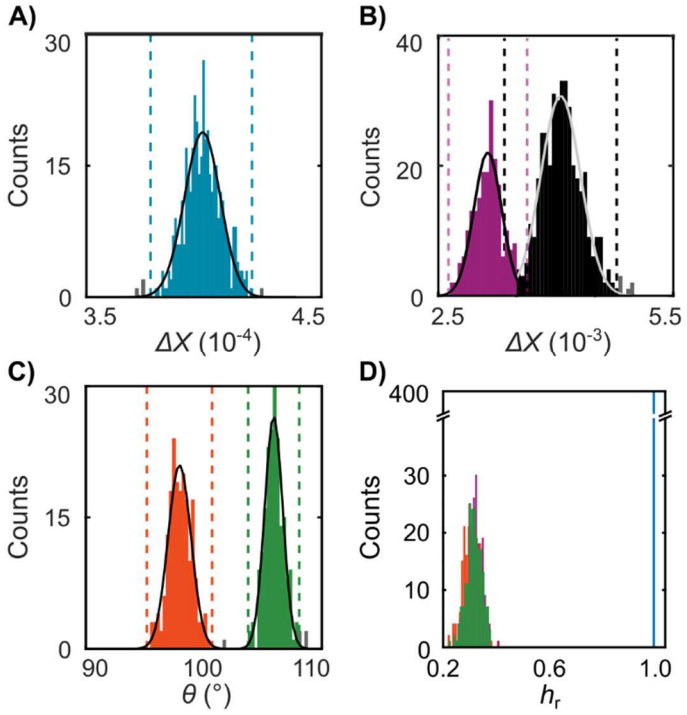


*Figure 3-5: Distribution-based clustering of the simulated data set. **A**) ΔX was fitted with a Gaussian distribution. Dashed lines are the 95 % confidence intervals. **B**) Second fit of ΔX. Shown in magenta are all traces assigned to the telegraph cluster. Shown in black are traces not yet assigned to any cluster. **C**) In a last clustering step, traces in the black cluster are assigned to two different clusters based on a fit of the θ with two Gaussian distributions. **D**) Histogram of h_r is shown for completeness. The figure was redrawn with permission from Lemmer et al., 2016.*[4]

After dimensionality reduction and clustering the simulated data set, the histogram analysis is much improved. Individual logarithmic 2D current histograms of all clusters are shown in Figure 3-6. **A** contains plain exponential traces, **B** traces with a plateau with telegraphic signal. **C** and **D** contain traces with plain plateau and sine-shaped plateau, respectively.

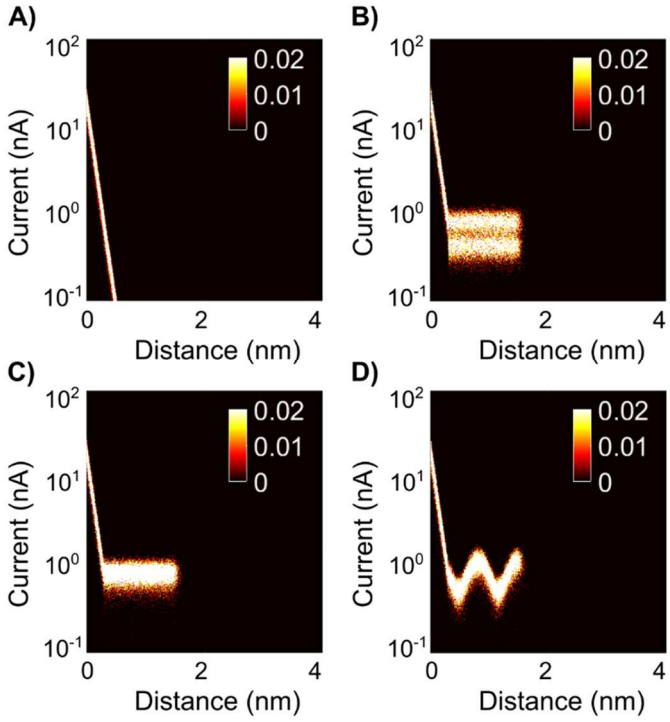


*Figure 3-6: Logarithmic 2D current histograms of the four individual clusters. **A**) The plain exponential cluster. **B**) The cluster with traces with telegraphic signal in the plateau. **C**) Traces with plain plateau. **D**) Sine-shaped plateau. The figure was redrawn with permission from Lemmer et al., 2016.*[4]

3.4 MPVC Analysis of ODT

After verifying the new analysis approach on simulated data, experimental conductance data of the well-published Au-ODT-Au junction was analysed, Figure 3-7. 70000 traces were recorded and published by Inkpen *et al.*[11] The setpoint current was 20 nA and the applied bias voltage was 0.3 V. Details of sample preparation and data recording can be found in the reference. The sheer number of traces recorded for this molecule yet again emphasizes the need for automated, ideally unsupervised data analysis routines. 1D and logarithmic 2D current histograms of the whole data set are shown in Figure 3-8 **A** and **B**, respectively. In **A**, a shoulder, meaning a peak partially obscured by the exponential background, can be seen. The 2D histogram in **B** shows a plateau, but data selection is necessary to increase clarity of the conductance feature.[4]

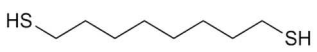


Figure 3-7: Structure of ODT.

The data set was then analysed using the MPVC approach. A region of interest (roi) was defined between 0.25 nm and 2 nm. This was done to increase the signal-to-noise ratio. No signal related to the molecule was expected far beyond the molecular length of 1.52 nm (minus

the initial tip-substrate separation of 0.93 nm).[11] The first 0.25 nm of each trace were also excluded to limit the influence of initial current fluctuation around the setpoint current.[4] As a reference vector, a noise-free, plain, exponentially decaying trace with a decay factor of 1.0 $Å^{-1}$ was used. After application of the MPVC dimensionality reduction, the whole data set can be seen in the cylinder plot of Figure 3-8 **C**.

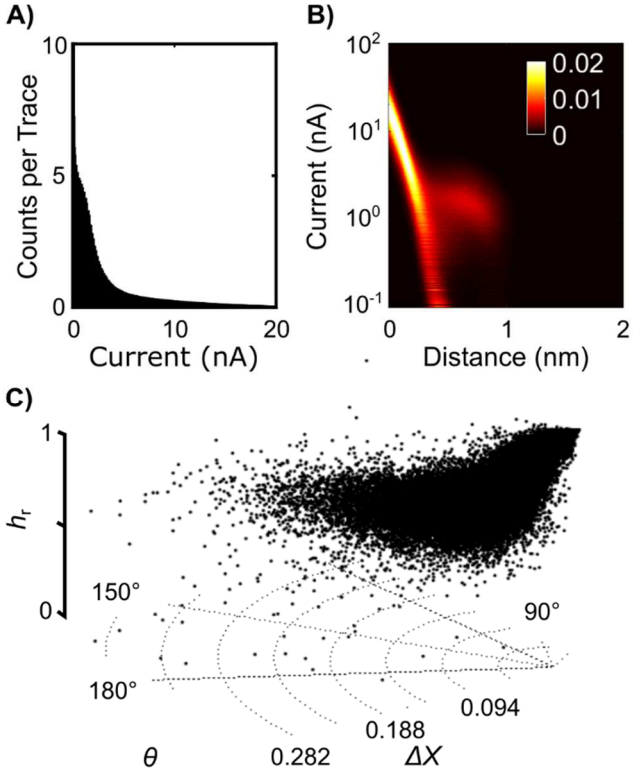


*Figure 3-8: 1D **A**) and **B**) 2D current histograms without any data selection. **C**) shows the whole data set, reduced to three parameters per trace with MPVC. Figure was reprinted from Lemmer et al., 2016.[4]*

To increase the signal-to-noise ratio in the current histograms, the data set was then clustered with Fuzzy c-means using the Gustafson-Kessel algorithm. Two clusters were chosen because in a first instance to separate traces that describe the formation of a molecular junction from those that do not. The clustering process divides the data set into two subsets. 29334 (42 %) traces were assigned to one cluster which is shown below in red and 40666 (58 %) to a second cluster shown in blue in Figure 3-9. **A** and **C** show the 1D current histograms of the two clusters. The peak in the red histogram, which was previously covered by the exponential background, is now very well-defined. A peak current of 1.0 nA, which corresponds to a conductance of $4.3 \cdot 10^{-5}$ G_0, was calculated. This conductance is in good agreement with the medium conductance group reported before at $4.9 \cdot 10^{-5}$ G_0.[5] The blue cluster on the other hand does not show any peak or shoulder as would be expected from a cluster that contains only plain

exponentially decaying traces without any plateaus. The cylinder plot in panel **E** shows the whole data set where colour coding was performed according to cluster membership.

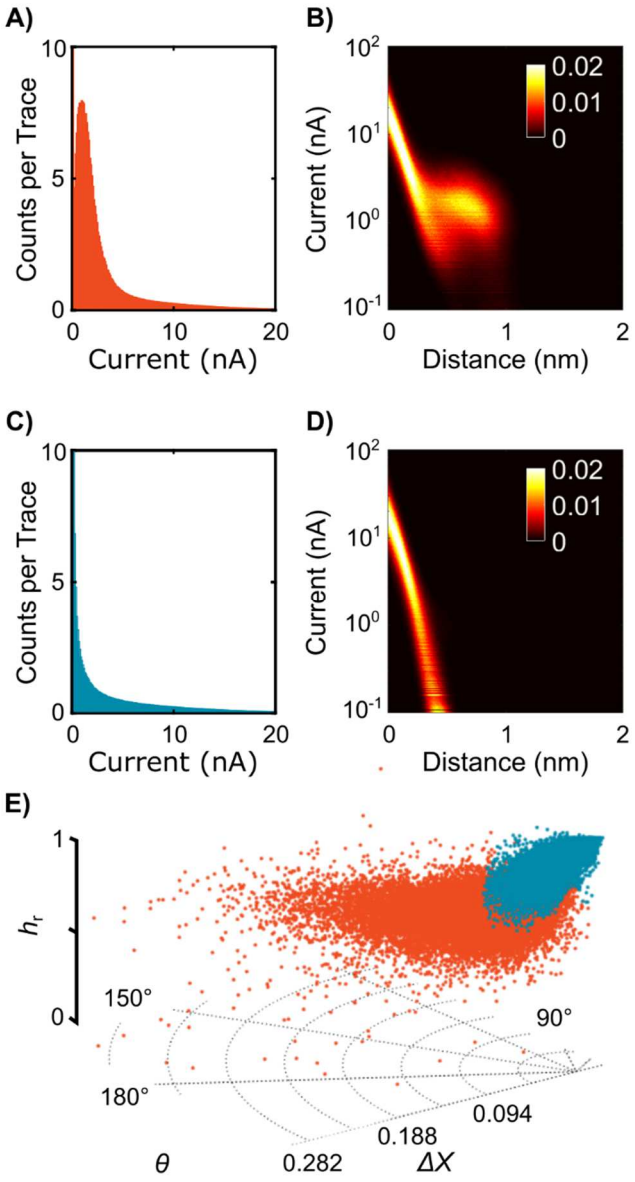


*Figure 3-9: 1D current histogram of the red **A**) and blue cluster **C**). **B**) and **D**) show the logarithmic 2D histograms of the red and blue cluster, respectively. **E**) is the cylinder plot after cluster assignment with FCM using the GK algorithm. The figure was redrawn with permission from Lemmer et al., 2016.*[4]

Traces that contain plateaus, and those that do not, actually form two well-distinguished clusters. In Figure 3-10 the same cylinder plot as in Figure 3-9 **E** is shown. The view direction is along the h_r-axis. Cluster assignment appears imperfect in this representation, because the Hamming distance was used as third parameter and is not showing when viewed along this axis.

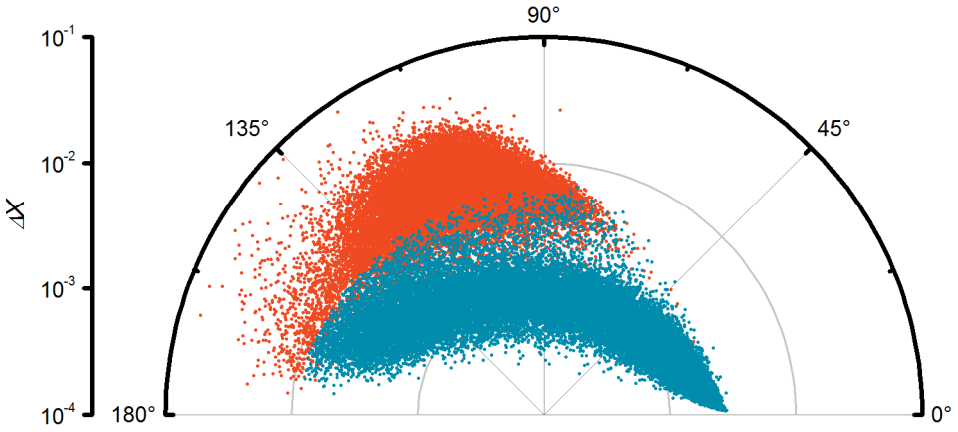


Figure 3-10: Cylinder plot of the whole data set viewed along the h_r-axis. Colour coding was done according to cluster membership. Cluster assignment was performed with FCM using the GK algorithm. Figure was reprinted from Lemmer et al., 2016.[4]

Only two distinct clusters, corresponding to plain exponential and plateau-containing traces, are seen in the cylinder plot. It is, however, well established that three conductance groups can be observed for ODT.[6,7,8,9] This raises the questions if these three conductance groups are present in the data and if they can be distinguished with the MPVC approach. 0.9 nS and 17 nS would be the expected conductance for the low and high conductance group of ODT, which correspond to $1.2 \cdot 10^{-5}$ G_0 and $22 \cdot 10^{-5}$ G_0.[5] Compared to the reference vector, it would be reasonable to expect the high conductance group to be at greater ΔX than the medium conductance group. The low conductance group is expected to be closer to the reference vector than the medium conductance group. Indication that the low and high conductance groups are present in the data can be found in the individual *I(s)* traces even though they do not form distinct clusters in the MPVC representation and clustering with four clusters proved unsuccessful.[4]

To explore the presence of the high conductance class in the data, the 5 % of plateau-containing traces with the highest ΔX (1467 traces) were selected. Note that by doing this, the algorithm is no longer unsupervised, but the additional insight into the conductance data justifies this. The 1D and 2D current histogram of those traces are shown in Figure 3-11 **A** and **B**. In **A**, a well-defined peak is observed at 3.15 nA ($\triangleq 14 \cdot 10^{-5}$ G_0). The discrepancy between the observed and the reported high conductance group is currently unknown.[4] A clear plateau corresponding to the conductance peak in the 1D histogram can be seen in the 2D histogram in **B**.

Similarly, when the traces with the lowest 5 % of ΔX are selected, predominantly low conductance traces are selected. The corresponding 1D and 2D histograms are displayed in Figure 3-11 C and D, respectively. A peak current of 0.4 nA was determined ($\triangleq 1.7 \cdot 10^{-5}$ G_0). This is in good agreement with the reported low conductance class of $1.2 \cdot 10^{-5}$ G_0.

The reason why only two and not four distinct clusters are present in the MPVC representation is likely that in our data, an *I(s)* trace rarely adheres to a single conductance class but rather switches between classes are observed within one trace. This would make the conductance classes inherently inseparable.[4]

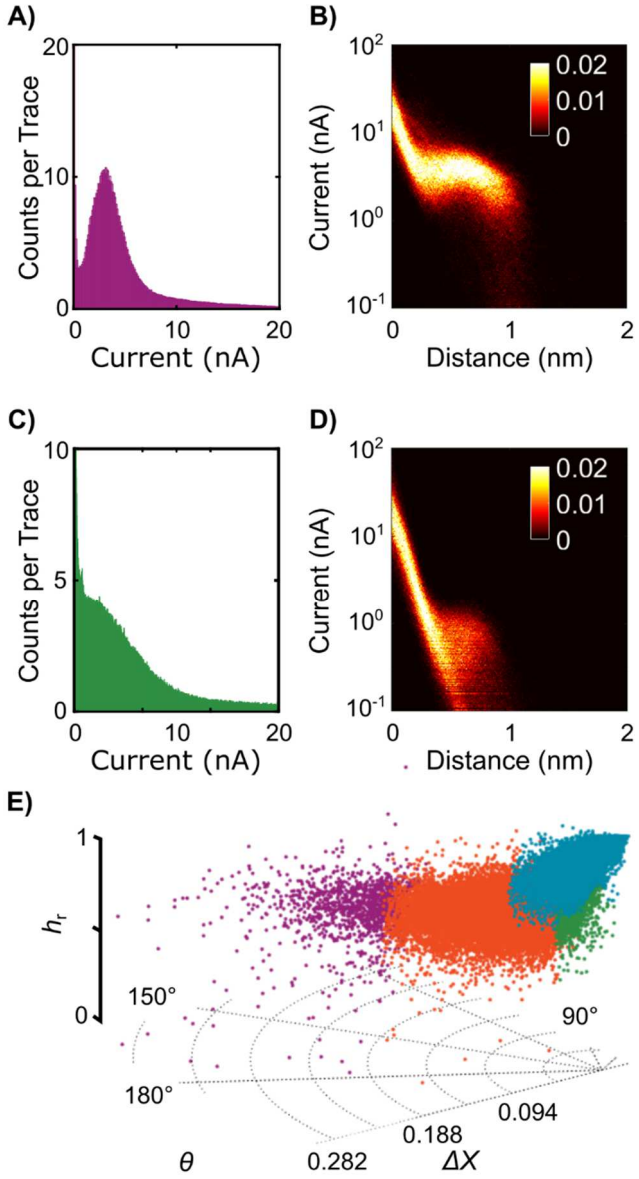


*Figure 3-11: Left column - 1D current histogram of the high **A**) and low **C**) conductance fraction of traces. Right column - 2D current histogram of the high **B**) and low **D**) conductance fraction of traces. **E**) Cylinder plot where the traces with 5 % highest (lowest) ΔX are shown in magenta (green). Figure was reprinted from Lemmer et al., 2016.[4]*

3.5 MPVC Analysis of OPE

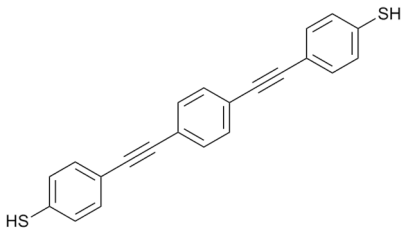


Figure 3-12: Structure of OPE.

Next, the MPVC analysis was applied to an experimental data set of OPE, Figure 3-12. OPE is a reference system for molecular electronics which has frequently been studied from the early years of experimental molecular electronics.[19] Typically conductance values between $1.2 \cdot 10^{-4}$ G_0 to $2.9 \cdot 10^{-4}$ G_0 are reported.[20,21,22,23,24,25,26] Sometimes a lower conductance value is also observed below 10^{-5} G_0.[24,27,28,29] This low conductance group is attributed to two molecules forming a joint junction by π-π stacking.[4,20]

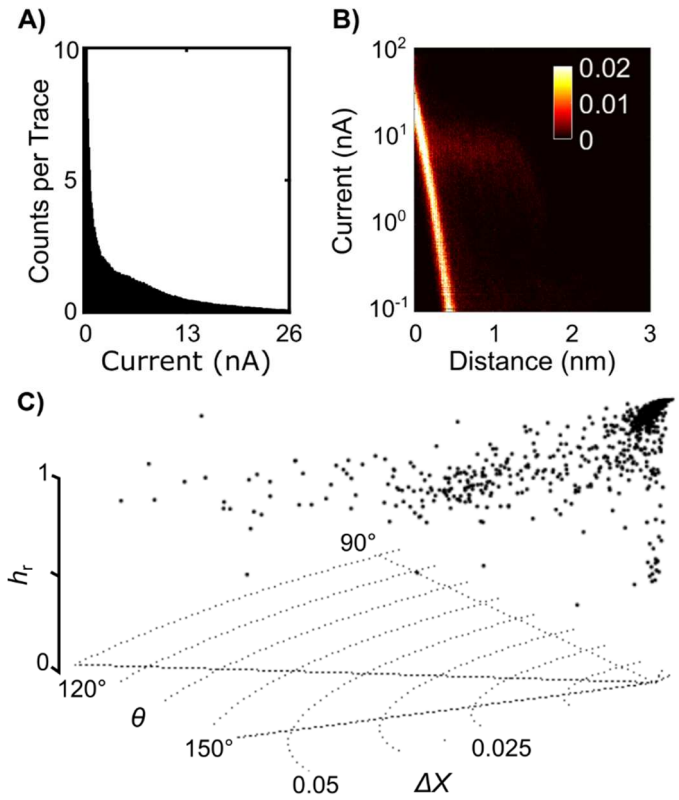


*Figure 3-13: **A**) 1D and **B**) 2D current histogram of the whole data set without any data selection. **C**) Cylinder plot of the whole data set in the MPVC representation. Figure was reprinted from Lemmer et al., 2016.[4]*

The data was previously published by Lemmer *et al.*[4] 2000 *I(s)* traces were recorded at 0.4 V and 26 nA setpoint current. Details of the sample preparation and data recording can be found within the reference. In this section, MPVC analysis will be used to gain additional insights which were not seen with a conventional plateau finder algorithm.

1D and 2D current histograms of the whole data set without data selection are displayed in Figure 3-13 **A** and **B**. No clear peak is observed in the 1D histogram. A faint plateau is visible in the 2D histogram below 10 nA. However, analysis is difficult due to the low signal-to-noise ratio. Subsequently, MPVC was applied to the data set. A region of interest was defined between 0.4 nm and 4 nm to exclude regions far beyond the molecular length and regions of strong current variation. An exponentially decaying reference with a decay factor of 1.0 $Å^{-1}$ was selected. After MPVC dimensionality reduction, every trace is defined by three parameters. The cylinder

plot in Figure 3-13 **C** displays the three parameters for every trace of the data set. In the cylinder plot a high point density can be seen close to the origin at large h_r. Furthermore, there are two other distinct regions. One is extending to large ΔX and large h_r. The other one is at low h_r and low ΔX. To investigate those three regions in detail, FCM was applied to the data set using the GK algorithm.

After running the cluster algorithm, the data set is divided into three clusters. This is indicated by the colour assignment in the cylinder plot in Figure 3-14. The blue cluster contains most traces (74 %), the green and red cluster contain 11 % and 14 % of all traces, respectively.

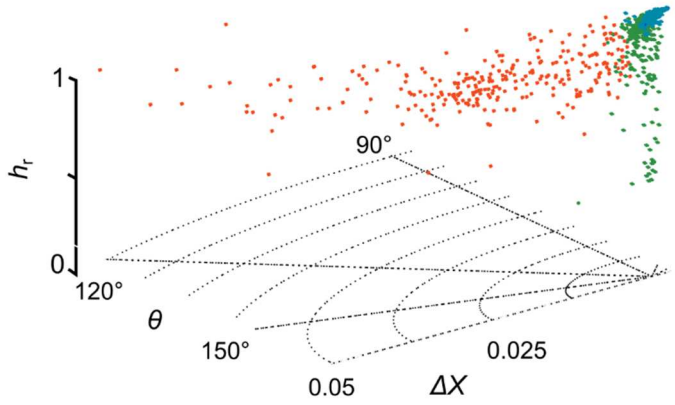


Figure 3-14: Cylinder plot of the whole data set where cluster membership is indicated by colour coding. Cluster assignment was determined with FCM using the GK algorithm. Figure was reprinted from Lemmer et al., 2016.[4]

The cluster assignment in the cylinder plot matches the optical intuition and it will be shown below that the cluster assignment in the MPVC representation groups the *I(s)* traces into meaningful subgroups. For this purpose, the 1D and 2D current histograms of the three clusters are plotted in Figure 3-15. The blue cluster contains only plain exponential traces. Accordingly, no peak is observed in the 1D current histogram in **A** and in the logarithmic 2D histogram in **B**, only a straight line is observed. The red cluster on the other hand contains plateau traces. From the 1D current histogram in **C**, a peak current of 5.5nA was calculated (≙ $1.8 \cdot 10^{-4}$ G_0). This is in excellent agreement to the conductance typically reported in literature ($1.2 \cdot 10^{-4}$ G_0 to $2.9 \cdot 10^{-4}$ G_0).[20-26] In accordance with the peak in the 1D histogram, the 2D histogram in **D** shows a clear plateau. The third, green cluster contains traces with low current plateaus. A peak at 0.11 nA was observed in the 1D histogram in **E**. This corresponds to a conductance of $3.6 \cdot 10^{-6}$ G_0, which is in good agreement with the case of two molecules coupled through π-π stacking. Wu *et al.* reported a conductance of $5.9 \cdot 10^{-6}$ G_0.[20] In the logarithmic 2D current histogram in **F**, a plateau can be seen at the same current of 0.11 nA.

It is noteworthy that within the green cluster only 22 individual traces, or 1 % of the total traces of the data set, show these low current plateaus. The remainder of the green cluster consists of traces which show ill-defined plateau.

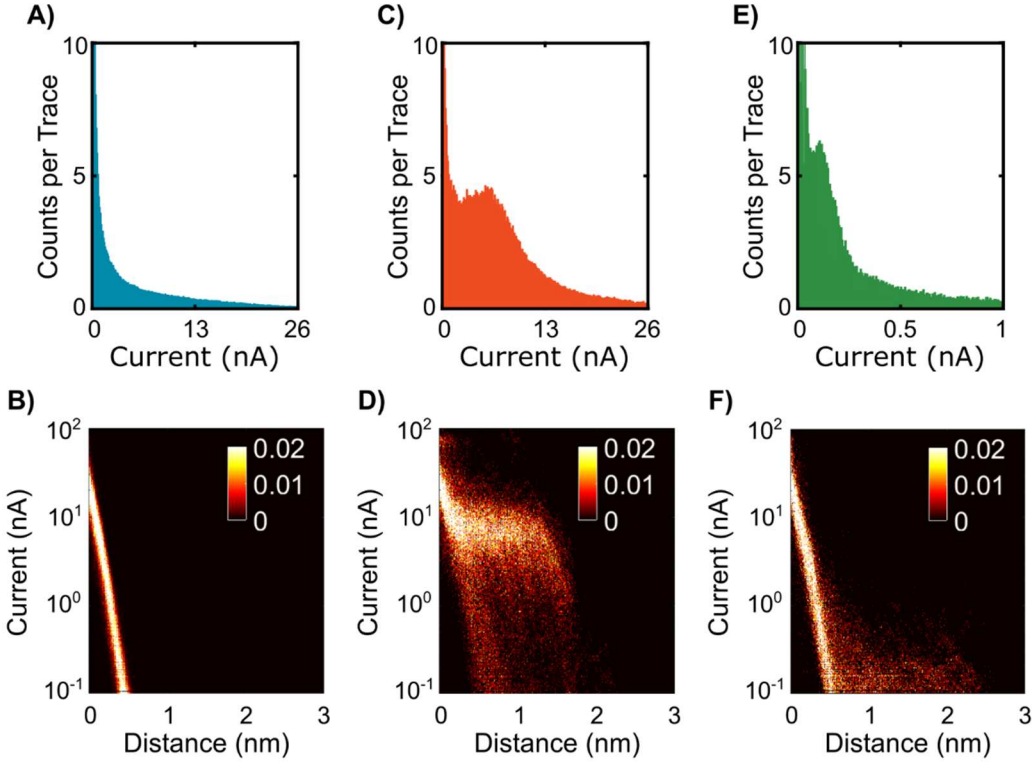


*Figure 3-15: 1D and logarithmic 2D current histograms of the blue cluster (**A** and **B**), red cluster (**C** and **D**) and green cluster (**E** and **F**). The blue cluster contains only plain exponential traces, red contains traces with high conductance plateaus. The green cluster contains I(s) traces with low conductance plateaus. Figure was reprinted from Lemmer et al., 2016.*[4]

3.6 Summary and Conclusion

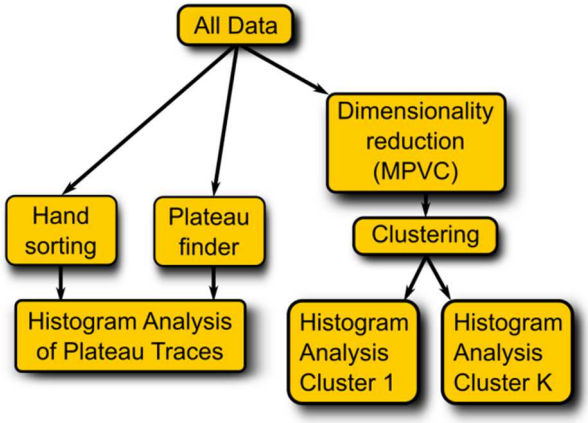


Figure 3-16: Flow chart of analysis processes. Typically, data sets are sorted by hand or automated by some criterion and the plateau-containing traces are analysed in histograms (left two branches). MPVC analysis is a conceptual departure, where traces are considered to be vectors. Dimensionality reduction is applied and then clustering algorithm split up the data into K meaningful subgroup. Individual subgroups are then analysed individually. Adapted with permission from reference 4.

To summarize, a novel, vector-based analysis approach to analyse single molecule conductance data was presented. The method works unsupervised and does not make any assumptions about the presence and form of signatures of the molecular junctions. It does not rely on assumed criteria of the current distance trace, *e.g.* a constant current during molecular conduction. MPVC is also capable of identifying charge transport events beyond the majority species. It is less dependent on the relative abundance of transport events. MPVC is therefore a conceptual departure from conventional hand-sorting or sorting by automatic detection of features in the *I(s)* trace, which are attributed to molecular conduction.

This novel approach was first verified on simulated *I(s)* traces. It was demonstrated here that the method can be used to identify qualitatively different types of plateau shapes and cluster those traces according to similarity. The MPVC analysis process was then applied to experimental data of two well-studied molecules, ODT and OPE. For ODT, the approach proved that large data sets could be easily clustered into traces associated with molecular conductance and plain exponential traces. Within the molecular cluster, regions of higher and lower conductance were identified. For OPE, the analysis approach proved to be an effective way to identify two fundamentally different charge transport situations, which were not readily identified with a conventional plateau finder algorithm. Generally, good agreement was seen with conductance reported in literature.

It is therefore concluded that the MPVC analysis process is an effective contribution to the field on molecular electronics.

3.7 References

1 S. Y. Jang, P. Reddy, A. Majumdar and R. a. Segalman, *Nano Lett.*, 2006, **6**, 2362–2367.

2 I. V. Pobelov, RWTH Aachen, 2008.

3 B. Xu and N. Tao, *Science*, 2003, **301**, 1221–1223.

4 M. Lemmer, M. S. Inkpen, K. Kornysheva, N. J. Long and T. Albrecht, *Nat. Commun.*, 2016, **7**, 17–19.

5 R. J. Nichols, W. Haiss, S. J. Higgins, E. Leary, S. Martin and D. Bethell, *Phys. Chem. Chem. Phys.*, 2010, **12**, 2801–2815.

6 C. Li, I. Pobelov, T. Wandlowski, A. Bagrets, A. Arnold and F. Evers, *J. Am. Chem. Soc.*, 2008, **130**, 318–326.

7 W. Haiss, S. Martín, E. Leary, H. Van Zalinge, S. J. Higgins, L. Bouffier and R. J. Nichols, *J. Phys. Chem. C*, 2009, **113**, 5823–5833.

8 X. Li, J. He, J. Hihath, B. Xu, S. M. Lindsay and N. Tao, *J. Am. Chem. Soc.*, 2006, **128**, 2135–2141.

9 M. Frei, S. V. Aradhya, M. S. Hybertsen and L. Venkataraman, *J. Am. Chem. Soc.*, 2012, **134**, 4003–4006.

10 T. Albrecht, S. F. L. Mertens and J. Ulstrup, *J. Am. Chem. Soc.*, 2007, **129**, 9162–9167.

11 M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. Costa-Milan, R. J. Nichols, N. J. Long and T. Albrecht, *J. Am. Chem. Soc.*, 2015, **137**, 9971–9981.

12 K. Yoshida, I. V Pobelov, D. Z. Manrique, T. Pope, G. Mészáros, M. Gulcur, M. R. Bryce, C. J. Lambert and T. Wandlowski, *Sci. Rep.*, 2015, **5**, 9002.

13 Y. Kihira, T. Shimada, Y. Matsuo, E. Nakamura and T. Hasegawa, *Nano Lett.*, 2009, **9**, 1442–1446.

14 O. Adak, E. Rosenthal, J. Meisner, E. F. Andrade, A. N. Pasupathy, C. Nuckolls, M. S. Hybertsen and L. Venkataraman, *Nano Lett.*, 2015, **15**, 4143–4149.

15 T. A. Su, H. Li, M. L. Steigerwald, L. Venkataraman and C. Nuckolls, *Nat. Chem.*, 2015, **7**, 215–220.

16 M. Fujihira, M. Suzuki, S. Fujii and A. Nishikawa, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3876–3884.

17 R. W. Hamming, *Bell Syst. Tech. J.*, 1950, **XXIX**, 147–160.

18 A. Ng, Coursera - Machine Learning, www.coursera.org/learn/machine-learning, (accessed 1 April 2017).

19 J. M. Tour, L. Jones, D. L. Pearson, J. J. S. Lamba, T. P. Burgin, G. M. Whitesides, D.

L. Allara, A. N. Parikh and S. Atre, *J. Am. Chem. Soc.*, 1995, **117**, 9529–9534.

20 S. Wu, M. T. González, R. Huber, S. Grunder, M. Mayor, C. Schönenberger and M. Calame, *Nat. Nanotechnol.*, 2008, **3**, 569–574.

21 R. Huber, M. T. González, S. Wu, M. Langer, S. Grunder, V. Horhoiu, M. Mayor, M. R. Bryce, C. Wang, R. Jitchati, C. Schönenberger and M. Calame, *J. Am. Chem. Soc.*, 2008, **130**, 1080–1084.

22 X. Xiao, L. A. Nagahara, A. M. Rawlett and N. Tao, *J. Am. Chem. Soc.*, 2005, **127**, 9235–9240.

23 Y. Xing, T. H. Park, R. Venkatramani, S. Keinan, D. N. Beratan, M. J. Therien and E. Borguet, *J. Am. Chem. Soc.*, 2010, **132**, 7946–7956.

24 V. Kaliginedi, P. Moreno-Garcia, H. Valkenier, W. Hong, V. M. García-Suárez, P. Buiter, J. L. H. Otten, J. C. Hummelen, C. J. Lambert and T. Wandlowski, *J. Am. Chem. Soc.*, 2012, **134**, 5262–5275.

25 V. Kolivoska, P. Moreno-Garcia, V. Kaliginedi, W. Hong, M. Mayor, N. Weibel and T. Wandlowski, *Electrochim. Acta*, 2013, **110**, 709–717.

26 R. Frisenda, M. L. Perrin, H. Valkenier, J. C. Hummelen and H. S. J. Van der Zant, *Phys. Status Solidi*, 2013, **250**, 2431–2436.

27 M. T. Gonzalez, E. Leary, R. Garcia, P. Verma, M. A. Herranz, G. Rubio-Bollinger, N. Martin and N. Agrait, *J. Phys. Chem. C*, 2011, **115**, 17973–17978.

28 S. Martin, I. Grace, M. R. Bryce, C. Wang, R. Jitchati, A. S. Batsanov, S. J. Higgins, C. J. Lambert and R. J. Nichols, *J. Am. Chem. Soc.*, 2010, **132**, 9157–9164.

29 R. Frisenda, V. A. E. C. Janssen, F. C. Grozema, H. S. J. van der Zant and N. Renaud, *Nat. Chem.*, 2016, 1–6.

Chapter 4

Ferrocene Compounds

4.1 Introduction Ferrocene Compounds 86
4.2 Linear Ferrocene Compounds..... 87
4.3 Double-Branched Ferrocene 104
4.4 Biferrocene Compounds 105
4.5 Summary and Conclusion..... 125
4.6 References 127

Synopsis: *In this chapter, the spectroscopic and electrochemical results for three families of ferrocene compounds are presented. After an introduction, linear ferrocenes will be discussed, followed by the results for double-branched ferrocene. In the third part of the chapter, the findings for biferrocenes will be presented, followed by a summary at the end of the chapter.*

4.1 Introduction Ferrocene Compounds

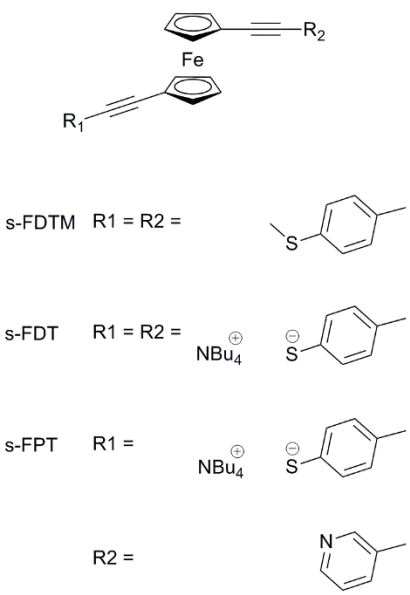


Figure 4-1: Structures of linear ferrocenes s-FDTM, s-FDT and s-FPT.

In the Introduction Chapter, it was pointed out that ferrocene-containing molecules are of special interest to the molecular electronics community. Among the interesting features that ferrocenes exhibit is a low attenuation of conductance with distance and redox activity. These attributes make them interesting candidates for molecular wires and switches. In this chapter, the experimental results for three different types of ferrocene compounds will be discussed.

In the first section of the chapter, results for three linear ferrocene compounds are presented: s-FDTM, s-FDT and s-FPT. The structures of these three molecules are shown in Figure 4-1. They consist of an identical molecular backbone, but the anchor groups are varied. s-FDTM has two thiomethyl and s-FDT two thiol anchors. In the third of the three ferrocene molecules, the symmetry was broken by introducing a meta-substituted pyridine anchor group in addition to one thiol anchor.

In the second part of the chapter, the results for a double-branched ferrocene molecule are shown. The structure is drawn in Figure 4-2. This molecule is a double-branched version of s-FPT and was designed to be compared to its linear pendant. The comparison with the linear ferrocenes was supposed to give indication about the conductance pathway and mechanism.

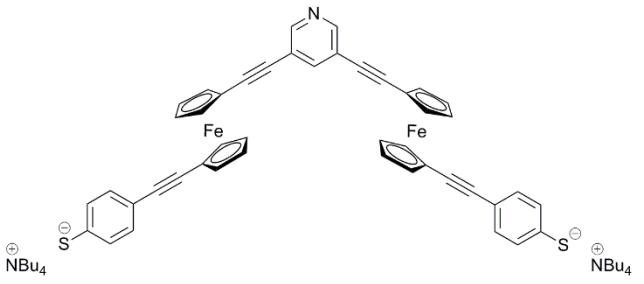


Figure 4-2: Structure of the double-branched ferrocene, b-FPT.

In the last experimental section of this chapter, the findings for two biferrocene compounds are presented: BFTM and BFP. The structure of those compounds is summarised in Figure 4-3. They consist of an identical biferrocene backbone, differing only in their anchor groups. BFTM has two para-substituted thiomethyl anchor groups. BFP has two meta-substituted pyridine anchor groups.

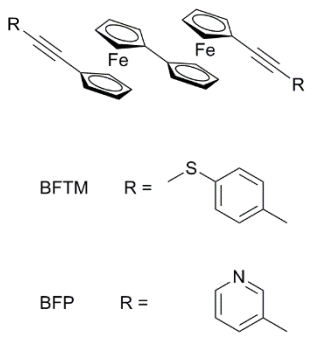


Figure 4-3: Structure of biferrocene compounds, BFP and BFTM.

4.2 Linear Ferrocene Compounds

4.2.1 Conductance of the Au-(s-FDTM)-Au Junction

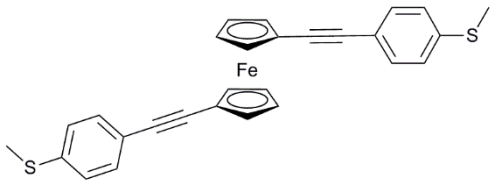


Figure 4-4: Structure of s-FDTM.

Spectroscopy measurements were performed on the linear ferrocene compound with thiomethyl linkers in Figure 4-4, s-FDTM. It was found that *I(s)* traces showed plateaus at various conductance values but often without statistical significance. Despite the low reproducibility of plateau events, a molecular conductance could be estimated for a data set recorded at -0.1 V and 6.5 nA setpoint current. The results for this data set are presented below.

The data set consisted of 2000 *I(s)* traces, which was recorded as part of an experiment during which 10 voltages were probed. Only one of these voltages produced the results shown below. Despite a considerable amount of experimental data, the conductance phenomena described below was not observed under any other condition. The presumed molecular features were found in the data after conclusion of the experimental work, therefore it was not feasible to reproduce the results using the same experimental conditions. All 2000 traces are shown in the two histograms in Figure 4-5. **A** is the logarithmic 2D conductance histogram. Some deviation from the plain exponential decay is observed here. **B** is the linear one-dimensional conductance histogram, showing no obvious peak.

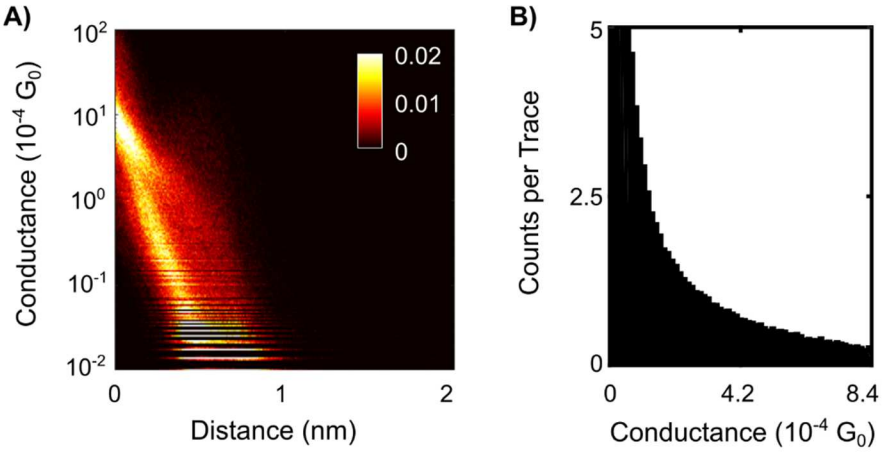


*Figure 4-5: **A**) Log 2D conductance histogram. **B**) Linear 1D conductance histogram of the whole data set without any selection. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) 0.13·10^{-4} G_0.*

MPVC analysis was applied to the data set. A region of interest was defined between 0.4 nm and 8 nm to exclude the first 100 data points. This was done to alleviate the strong relative influence that small tip-substrate distance variations have on the measured current at small distances. A simulated, exponentially decaying trace with a decay factor of 1 Å^{-1} was chosen as reference. Plotting the logarithmic point density of the data set after MPVC analysis reveals two distinct clusters (Figure 4-6). The two clusters are separable along the *ΔX*-axis.

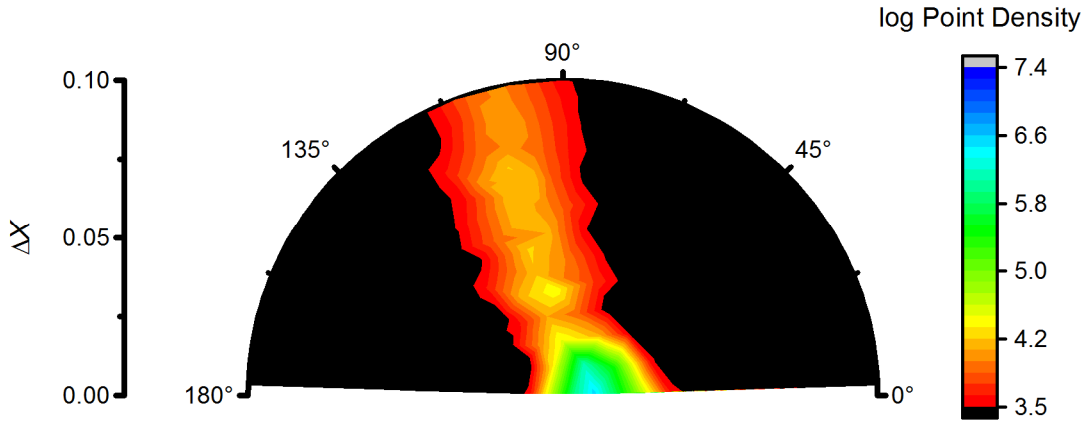


Figure 4-6: Logarithmic point density of the data set. The point cloud from MPVC analysis was projected onto the ΔX/θ-plane and the point density was calculated with binning increments of 3.6° and 0.02 in angular and radial direction, respectively. The colour scheme corresponds to the exponent of the point density.

Cluster assignment was then performed with fuzzy clustering, using the GK algorithm. The algorithm separated the data set into two clusters in agreement with the density plot (compare Figure 4-6 with Figure 4-7). The red cluster in Figure 4-7 contains 118 traces and the blue cluster the remaining 1882 traces (JFP 6 % and 94 %, respectively).

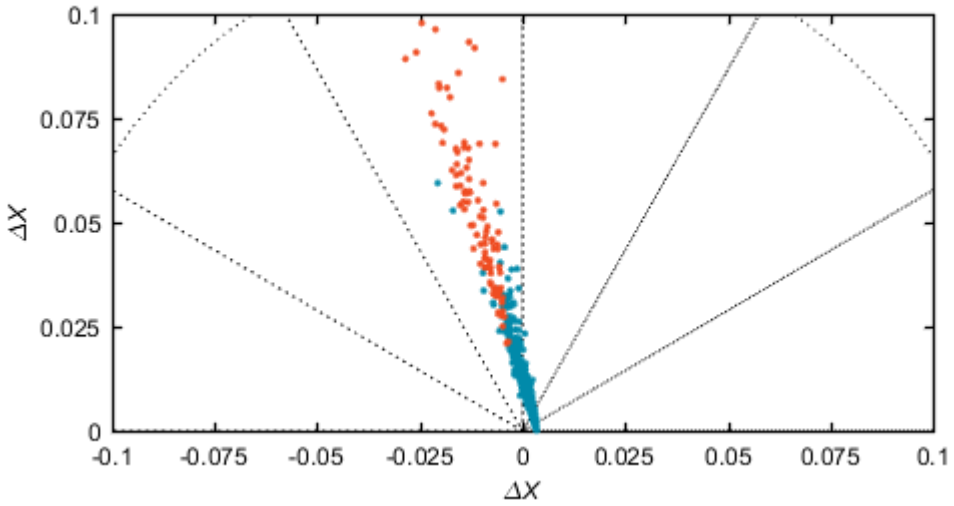


Figure 4-7: Cylinder plot of the whole data set viewed along the h_r-axis. Colour assignment was done per cluster membership. Clustering was performed with FCM using the GK algorithm.

After cluster assignment, the 2D conductance histogram of the red cluster in Figure 4-8 **A** shows a distinct deviation from the exponential decay. A slow exponential decay (small exponential decay factor) is interrupted by a sudden increase in conductance, followed by a fast decay.

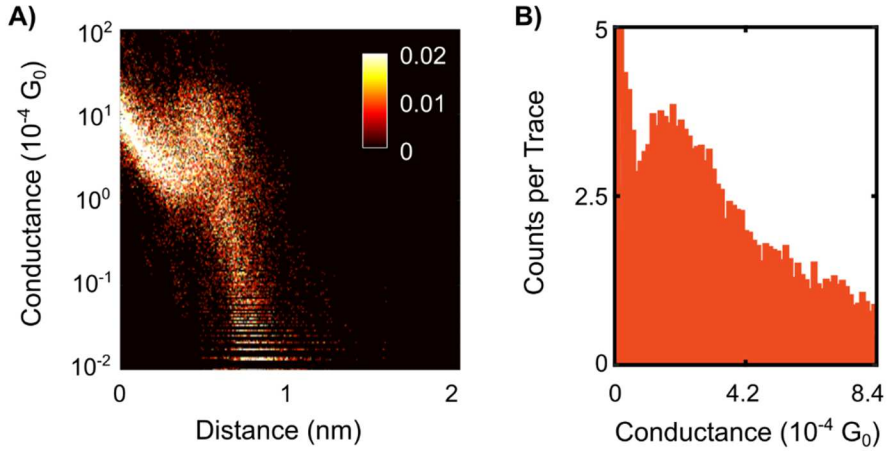


*Figure 4-8: **A**) Logarithmic 2D conductance histogram of the red cluster. **B**) Linear 1D conductance histogram of the traces that contain conductance features. Plateau traces were selected per MPVC analysis using GK FCM. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) 0.13·10⁻⁴ G_0*

The corresponding linear conductance histogram in **B** shows a clear peak at a conductance of $(1.83 \pm 2.55)\cdot10^{-4}$ G_0 or (14.2 ± 19.8) nS. Note that the large error associated with the measurement is one sigma of the Gaussian fit and reflects the confidence interval for a single measurement and is not indicative of the standard deviation of the conductance value.

The individual *I(s)* traces generally do not exhibit traditional plateaus, but rather conductance "spikes". A representative selection is given in Figure 4-9, left. The right panel contains all cluster members overlaid and their median in red.

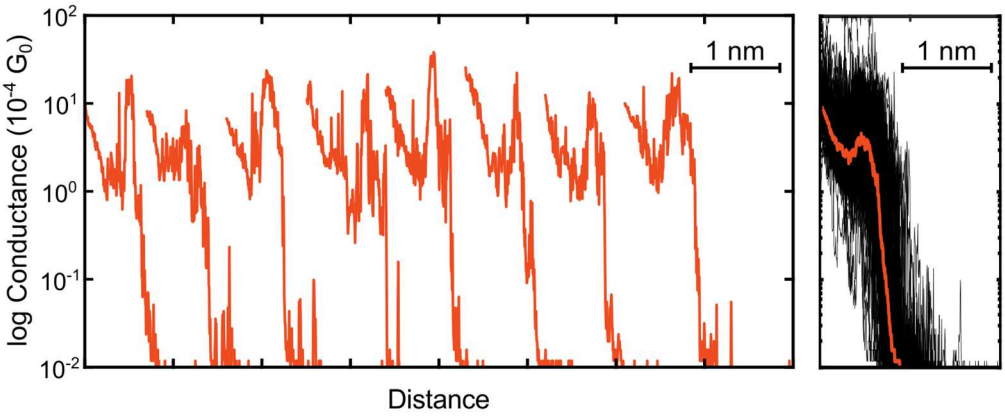


Figure 4-9: Representative selection of conductance traces in the red cluster. Traces are offset in the left panel and overlaid in the right panel with their median shown in red.

The "event distribution" of the red cluster is shown in Figure 4-10 **A**. "Event distribution" meaning here the cumulative counts of the red cluster over the course of the experiment. The graph reveals that two-thirds of plateau traces are collected in the first 15 % of scans. Towards the end of the experiment this conductance feature becomes less likely. The reason for this is currently unknown.

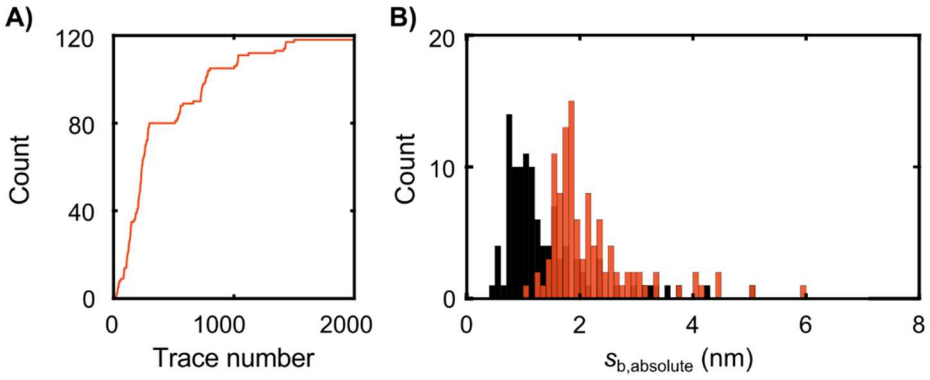


*Figure 4-10: **A**) Event distribution of the plateau traces over the course of the experiment. **B**) Absolute (tip-substrate) break-off distance histogram of the red cluster. Shown in black is the initial tip-substrate separation. Bin width 1 Å.*

Robert Stadler from TU Wien and his group estimated a S-S distance for the molecule of 2.00 nm. Taking the length of the anchor groups into account (0.30 nm each[1]), this gives a maximal expected junction extension $s_{b,exp}$ of 2.60 nm. The experimental break-off distance of 1.82 nm ± 0.45 nm (Figure 4-10 **B**) is considerably shorter than the expected length for the maximally extended junction. This could indicate that the molecule is bound at an angle relative

to the normal of the substrate and that the junction ruptures before the molecule can be straightened out. Comparing the measured with the expected length would suggest a tilt angle of 46° versus the normal of the substrate.

The second, blue cluster, which was found with MPVC, was also analysed in more detail. It can be seen from the logarithmic 2D conductance histogram in Figure 4-11 **A** that there is still a deviation from the plain exponential decay. No peak can be seen in the linear histogram in **B**.

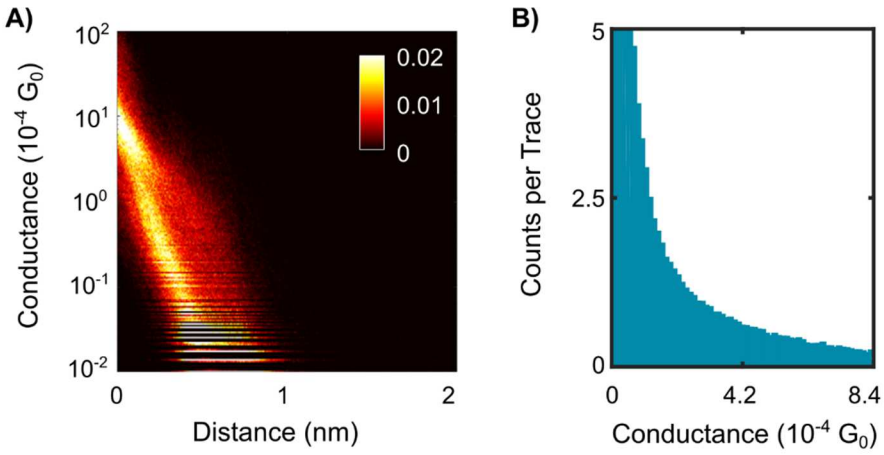


*Figure 4-11: **A**) Logarithmic 2D conductance histogram of the predominantly plain exponential cluster. **B**) Linear 1D conductance histogram of the traces in the blue cluster. The members of the blue cluster contain predominantly plain exponential traces. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) $0.13 \cdot 10^{-4}$ G_0*

To investigate the nature of traces that showed deviation from plain exponential behaviour, but were not found by MPVC analysis, the whole data set was analysed with the conventional linear plateau finder algorithm (PF), described in Chapter 1. The plateau finder marks 113 traces as plateau-containing (JFP 6 %, similar JFP as found by MPVC analysis). All plateau-traces were compiled into histograms in Figure 4-12. The 2D histogram in **A** is qualitatively different than the one obtained with MPVC analysis. A more conventional plateau shape is observed. The logarithmic 1D histogram, **B**, shows a peak and the conductance was determined to be $(1.4 \pm 2.55) \cdot 10^{-4}$ G_0, similar to $1.83 \cdot 10^{-4}$ G_0 with MPVC analysis.

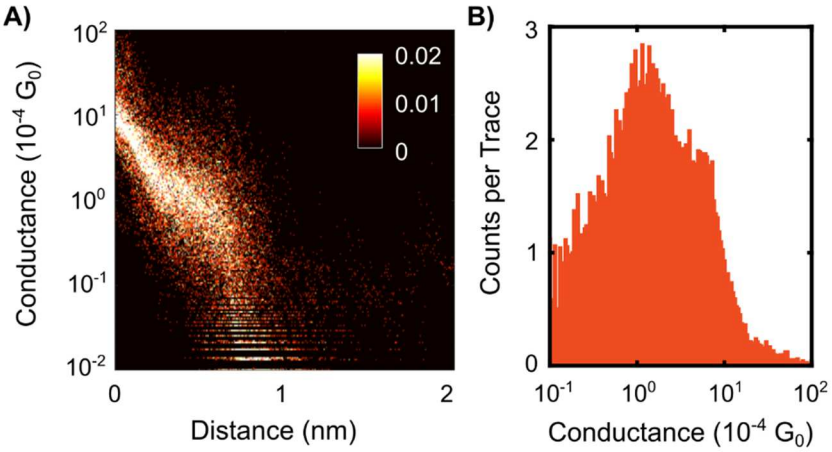


*Figure 4-12: **A**) Logarithmic 2D conductance histogram. **B**) Linear 1D conductance histogram of the traces that contained conductance features. Plateau traces were selected by plateau finder. Bin widths: **A**) 0.04 Å and 50 bins per decade. **B**) 0.13·10^{-4} G_0*

The two analysis approaches (MPVC and PF) select a similar number of traces. However, they do not select the same traces. Both analysis techniques share only 10 % of the selected traces (12 traces of 118). Plateau traces found by MPVC analysis often peak at the end of the plateau, plateau traces found by plateau finder analysis do not, Figure 4-13.

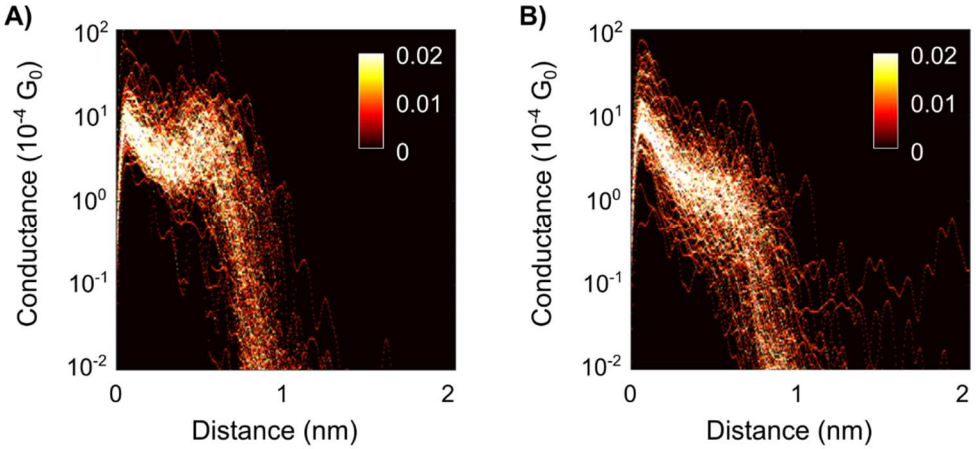


*Figure 4-13: Logarithmic 2D conductance histogram of the plateau traces identified by **A**) MPVC and **B**) plateau finder. Data in both histograms was filtered with a low-pass Bessel filter to improve visibility of the described effect. Binning: **A**) and **B**) 0.04 Å and 50 bins per decade.*

The mean break-off distance of the traces selected by the plateau finder is very similar to those found by MPVC analysis (1.82 nm ± 0.45 nm compared to 1.75 nm ± 0.57 nm). Also, the event distribution of the traces selected by the plateau finder is similar to the one for MPVC traces. Most plateau-traces are collected in the beginning of the experiment. Both graphs are provided in the appendix.

When plotting the decay factor as a function of distance, it can be seen that the conductance spike at the end of the plateau is also present in the traces found by the plateau finder (compare Figure 4-14 **A** to **B**).

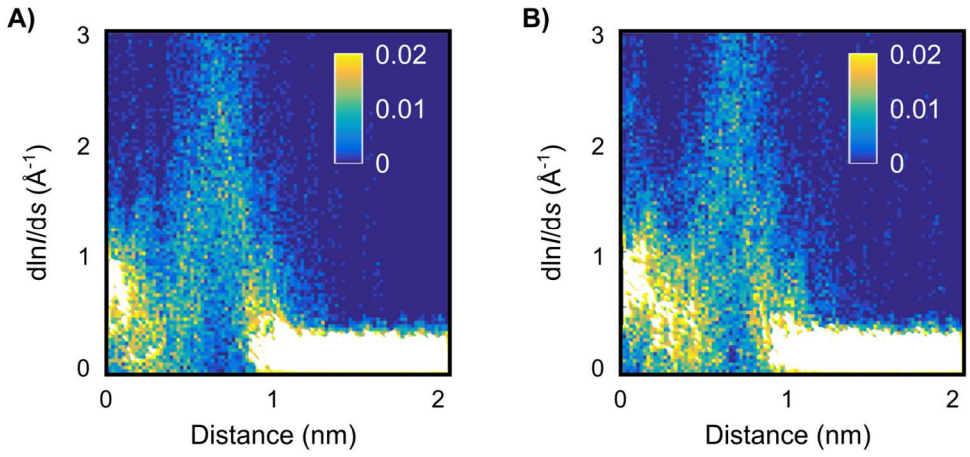


*Figure 4-14: Decay factor as a function of distance of the plateau traces. Plateau traces were found by **A**) MPVC and **B**) plateau finder. Bin widths: 0.2 Å and 0.03 Å^{-1}.*

Unfortunately, electrochemical measurements after spectroscopy could not confirm that the molecule was indeed surface bound and stayed intact over the course of the experiment. It was hypothesised that the molecule was only weakly bound to the surface and quickly disappeared into solution when the monolayer came in contact with the electrolyte. Chen *et al.* observed the same effect for molecules with thioacetyl anchor groups.[2] This suggests that the analyte layer would only be destroyed after the spectroscopy and that the spectroscopy results would still stand.

4.2.2 Discussion and Summary s-FDTM

Spectroscopic results for a linear mono-ferrocene molecule with thiomethyl anchors were presented. Two qualitatively different types of junction elongation traces were found with two analysis techniques. The two types of *I(s)* traces showed similar conductance values of $1.8 \cdot 10^{-4}$ G_0 and $1.4 \cdot 10^{-4}$ G_0 (≙ 14 nS and 11 nS, MPVC and PF). Reasonable agreement was seen with literature. Lu *et al.* reported a conductance of $3.7 \cdot 10^{-4}$ G_0 (1.45 nA at 0.05 V) for the same system with NH_2 anchor groups.[3] For OPE type molecules, NH_2 and thiomethyl anchors showed similar conductance.[4] A higher conductance would be expected for the system in literature since the break-junction was used by Lu, which can favour the high conductance group.[5,6] Another factor to consider is that Lu *et al.* performed the experiment under argon. In our system, the ambient humidity is likely to increase the conductance and thus mitigate the above-mentioned effect to some degree.[7]

Similar break-off distances, conductance values, and decay factor evolutions lead to the conclusion that both types of traces describe the same junction configuration and each algorithm finds a fraction of the plateau traces. The two analysis techniques predominantly select traces which show large spikes (MPVC) or small/no spikes (PF). The plateau finder cannot find traces that have large spikes in the plateaus, because the spikes spread out over several bins in the single-trace histogram and the algorithm thus does not recognise them as plateaus. The MPVC on the other hand should be able to find both traces, but with the selected reference vector and cluster algorithm it does not find all of them. A different choice of cluster parameters could improve classification results here.

The origin of the conductance spike might be that in some cases the junction transforms into a high conductance state just before it ruptures. This has been reported before by Vankataraman *et al.* on aliphatic Si wires, as well as by Kametska *et al.* for 1,4-bis(dimethyl-phosphino)-butane.[8,9] Tao *et al.* attributed an increase in junction conductance within single plateaus by more than an order of magnitude to force induced resonances.[10]

Break-off distances for traces found with either method, were smaller than expected for a maximally extended molecular junction. One interpretation of the short break-off distance is that the molecule is bound at an angle of about 46° and the junction ruptures before the theoretical maximum extension is reached. However, a different and possibly more likely explanation is that the molecules are connected at entities other than the anchor groups, thus creating a reduced current pathway (*vide infra*).

4.2.3 Conductance of the Au-(s-FDT)-Au Junction

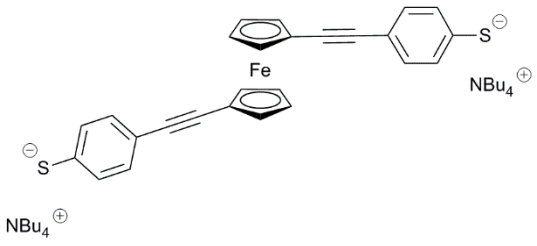


Figure 4-15: Structure of s-FDT.

To increase reproducibility compared to s-FDTM, the same molecule with thiol anchor groups was investigated next, Figure 4-15. Voltages between -0.5 V and +0.5 V were probed during several experiments. Good reproducibility was found between experiments and results are presented below. Details of the results are provided for one data set recorded at -0.3 V and 20 nA setpoint current. 3000 *I(s)* traces were recorded per bias voltage during this experiment.

The whole data set without data selection does not show any conductance features of the analyte. By applying the PF algorithm to the data set, the signal-to-noise ratio is enhanced. Of

the 3000 traces, 1340 traces contained plateaus (JFP 45 %). The 2D and 1D histograms of those plateau-traces are shown in Figure 4-16 **A** and **B**, respectively. No obvious features can be seen in the 2D histogram. The 1D histogram on the other hand has a strong "setpoint peak" and multiple smaller peaks that are thought to be connected to the analyte (*vide infra*). The origin of this "setpoint peak" is currently unknown, but is attributed in some way to the molecular adsorbate, because it does not appear to the same degree in the control experiments (*vide infra*). The same 1D conductance histogram is shown again in **C**. The dashed lines correspond to the most likely conductance of four different conductance classes and the coloured intervals indicate their error intervals. Red colour-coding of the histogram was dropped here to increase clarity of the figure.

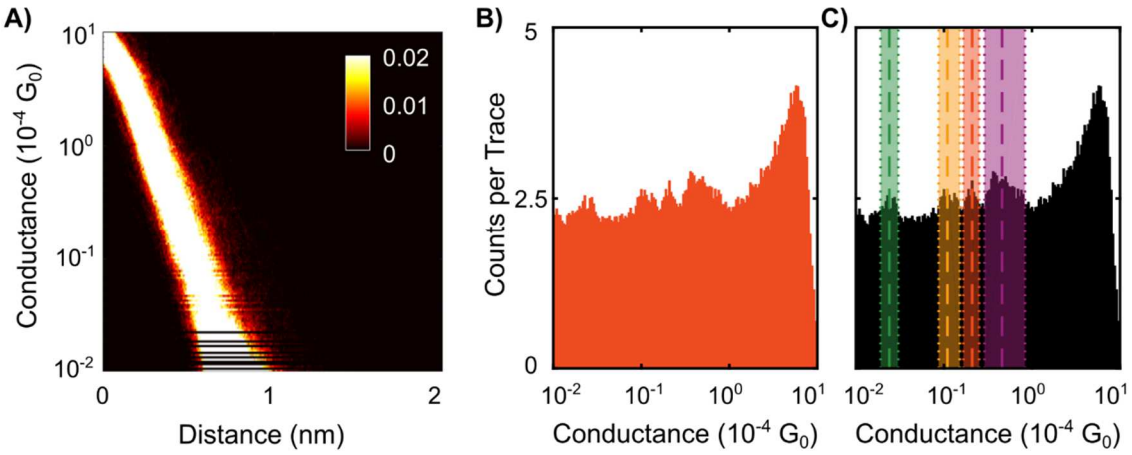


*Figure 4-16: Logarithmic **A**) 2D and **B**), **C**) logarithmic 1D conductance histogram of the data set recorded at -0.3 V. Data selection was made with the PF algorithm. Bin widths: **A**) 0.16 Å and 50 bins per decade in **A**) and **B**). Colours in **C**) correspond to estimated peak conductance with error intervals.*

The estimated peak conductance and their error margins are summarised in Table 4-1.

Table 4-1: Summary of peak conductance and error margins of the experiment conducted at -0.3 V.

	Conductance (10^{-4} G_0)	Error Interval (10^{-4} G_0)
Lowest	0.024	0.016 to 0.030
Low	0.100	0.079 to 0.151
Medium	0.210	0.166 to 0.251
High	0.460	0.302 to 0.759

In Figure 4-17, a representative selection of the traces that make up the histogram, is shown. The most probable conductance values of the different conductance classes and their respective errors are indicated by dashed lines and coloured intervals. Plateaus can be seen at the different conductance values. Often traces exhibit plateaus at more than one plateau within

a single trace. It is noteworthy, that the plateaus are very short compared to ODT for example, despite the considerable length of the molecule. (For comparison the length of Au-ODT-Au is 1.59 nm)[11] Presumably, this is one reason why no plateau feature can be seen in the 2D conductance histogram in Figure 4-16 **A,** despite the considerable JFP.

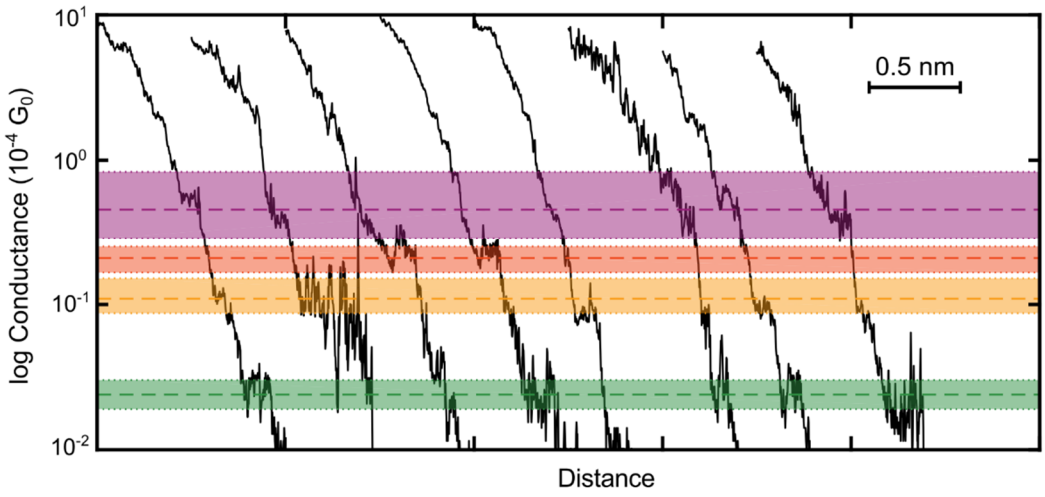


Figure 4-17: Representative sample traces with estimated conductance (dashed lines) and error margins (coloured intervals). Traces were offset for increased clarity.

The event distribution of the traces, which contain plateaus over the course of the experiment, shows a uniform, steady incline (Figure 4-18). This is to be expected with a JFP this high. The high junction formation probability and the steady incline of cumulative counts suggest that junction formation was achieved throughout the sample and over the whole course of the experiment. This indicates uniform coverage and a steadily high probability to create a molecular junction.

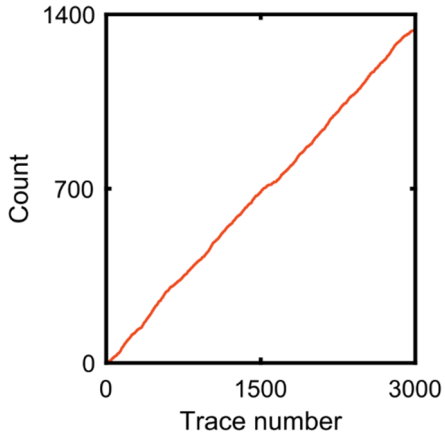


Figure 4-18: Event distribution of the plateau traces over the course of the experiment.

Similar conductance peaks were seen for most experiments and bias voltages. An overview of the peak conductance found during different experiments is shown in Figure 4-19. Conductance classes were colour-coded, analogous to the previous figures. It is noteworthy here that peak conductance seems to decrease with increasing absolute bias voltage for all conductance classes. The individual histograms are shown in the appendix.

Conductance at positive and negative bias voltage are very similar for the three higher conductance classes. Ratios of 0.99 ± 0.12, 0.98 ± 0.11 and 0.99 ± 0.04 were calculated (purple,

red and orange respectively), meaning no rectification within experimental error. The green, lowest conductance class seems to show some rectification. A rectification ratio of 0.68 ± 0.16 was calculated.

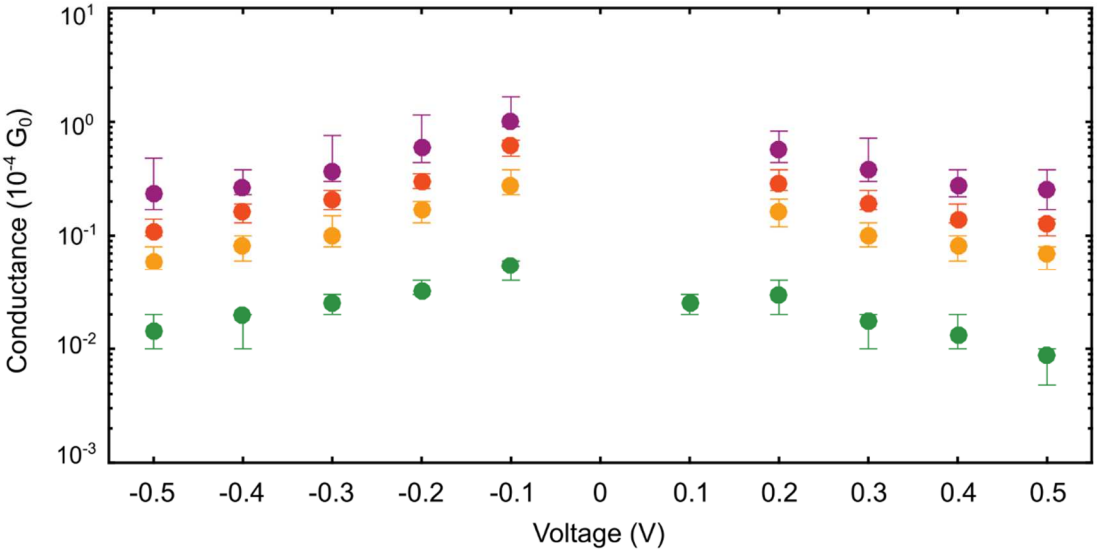


Figure 4-19: Conductance of s-FDT as function of applied bias voltage. Different colours represent different conductance classes.

The break-off distance was calculated for each data set. Because of the shortness of the plateaus, the current did not drop to noise level following the plateaus. The break-off distance was thus not defined as distance at which the noise level is reached. It was defined as the last point in the lowest single-trace histogram bin above the plateau-determining bin count (PDBC).[11] The distribution of the break-off distances was then fitted with a lognormal distribution and the modus of the distribution was given as break-off distance. The experimental break-off distances are shown in Figure 4-20. All s_b histograms are provided in the appendix.

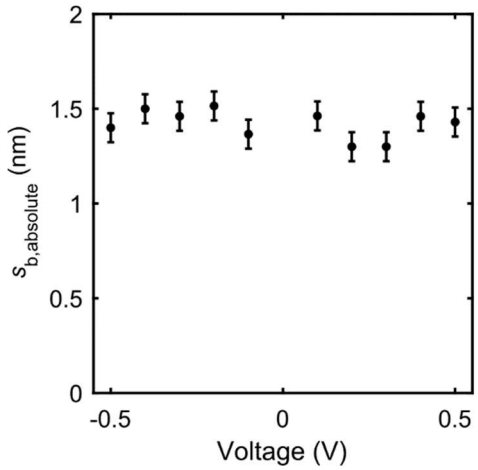


Figure 4-20: Break-off distances as a function of bias voltage for s-FDT.

Robert Stadler and his group kindly simulated break-off distances for this molecule and found an S-S extension of 2.00 nm and a gold-gold distance of 2.37 nm. The experimentally found mean break-off distance was 1.42 nm ± 0.08 nm. The experimental break-off distances are well below the expected junction extension for the case of one thiol anchor coupled to either electrode. Stadler *et al.* also simulated a junction where both anchors couple to one electrode

and the other electrode is contacted with the ferrocenyl moiety. Here the expected junction extension is 1.43 nm. Details of the simulations are currently being prepared for publication.

4.2.4 Discussion and Summary s-FDT

s-FDT showed four distinct conductance classes with excellent reproducibility between experiments. Generally, the junction formation probability was very high throughout all experiments. The conventional, PF algorithm by Inkpen *et al.* was used to for data selection. This was done because the observed plateaus were very short. The vector-based analysis struggles to identify short plateaus with the current set of parameters. MPVC is not expected to give deeper insight into the data set, since the observed conductance peaks often corresponded to multiple plateaus at different conductance within single traces. The different conductance classes are thus inherently indivisible. (*cf.* ODT section Chapter 3)

When considering the conductance as a function of bias voltage, it was observed that conductance was reduced at increased absolute bias voltages. A possible reason behind this could be local heating of the molecular junction induced by electron-phonon interaction.[12] Huang *et al*. reported a method that uses junction rupture force measurements to calculate the local temperature and observed a 30 K increased local temperature at 1 V for ODT.[13] For s-FDT, the expected charge transport process would be tunnelling, as reported by Qui *et al*.[3] For tunnelling no explicit temperature dependence would be expected, however, changes in the geometry caused by local heating could cause a decrease in conductance. Another observation from the G(V) plot is that G is symmetrical with respect to the bias voltage. There seems to be little to no rectification for the highest three conductance classes (0.99 ± 0.12, 0.98 ± 0.11 and 0.99 ± 0.04) and little rectification for the lowest class (0.68 ± 0.16).

Robert Stadler and his group conducted non-equilibrium Green's function (NEGF) density functional theory (DFT) simulations on this molecular junction, which are currently being prepared for publication. They calculated the transmission function and estimated a Fermi level conductance of $0.155 \cdot 10^{-4}$ G_0 for the case of a fully extended junction connected with thiols to either electrode. This is in the range of the observed conductance. For a "folded geometry", meaning a junction where both anchors contact one electrode, a conductance was calculated to be $130 \cdot 10^{-4}$ G_0. The observed conductance thus suggests that the full-length conductance was measured.

The experimental break-off distances stayed well below the simulations for the fully extended junction (1.42 nm and 2.37 nm, respectively) indicating a folded structure. For the

folded structure, excellent agreement with simulated break-off distances was observed. Apart from the longer break-off distance, it would also be expected to see a larger variation in break-off distance if the molecule was indeed contacted by the thiol groups. It has been reported before that strong thiol-gold bonds pull out strings of gold from the surface and thus give large variations in observed break-off distance, which were not observed in the presented experiments.[14]

Comparing conductance results with the experiments suggests that the conductance over the full length of the molecule was measured. Break-off distances, on the other hand, suggest that this was not the case. The origin of this discrepancy is currently unknown.

Electrochemical measurements were performed after spectroscopy. No signal related to the molecule could be observed here This could mean that the molecule did not bind to the substrate. However, this seems unlikely considering the two thiol anchor groups. A rapid desorption in the presence of the electrolyte could be the reason for the missing CV signal.

4.2.5 Conductance of the Au-(s-FPT)-Au Junction

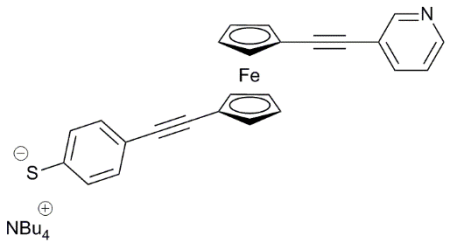


Figure 4-21: Structure of s-FPT.

To further investigate the influence of the anchor group on the molecular conductance, experiments were performed on s-FPT, Figure 4-21. One aryl thiol group was exchanged with a meta-substituted pyridine anchor and compared to the molecule discussed in the previous section. Various spectroscopy experiments were performed at voltages between -0.6 V and +0.6 V. The results are summarised below. Details of the results are shown by example of the data set recorded at +0.15 V and 20 nA setpoint current.

The PF algorithm found plateaus in 1433 of the 4000 total traces (JFP 36 %). Those plateau-containing traces were compiled into histograms (Figure 4-22). Where no peaks could be seen before data selection, four peaks and a peak at the setpoint conductance appear after data selection. A qualitatively similar event distribution was observed for this molecule compared to the one discussed in the previous section. The graph is provided in the appendix.

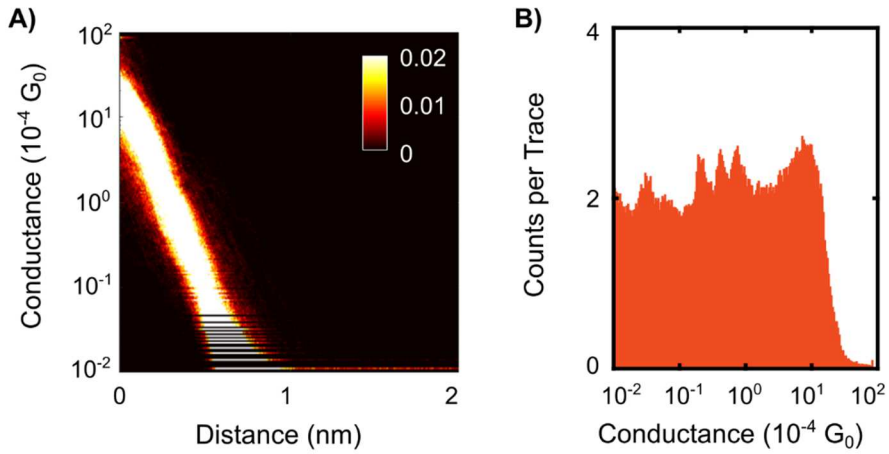


*Figure 4-22: Logarithmic **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.15 V. Data selection was made with linear plateau finder algorithm. Bin widths: **A**) 0.04 Å and 50 bins per decade in **A**) and **B**). Data in **B**) was filtered with a low-pass Bessel filter.*

The observed peaks in the 1D conductance histograms were not produced by filter or logarithmic binning effects as shown below in Figure 4-23. In panel **A** is the same histogram as in Figure 4-22 **B**, but the red colour has been removed for increased contrast. Dashed lines indicate the most probable peak conductance and the shaded areas are their estimated error margins. In panel **B** is a histogram of the same data set on a linear scale. The same peak conductance values are shown with errors. Peaks are less well-defined, but the medium and low peaks can still be seen on the linear scale and are thus not artefacts produced by logarithmic binning. Panel **C** contains the same histogram as **B**, but data is not filtered. Although peaks are less pronounced without filtering, peaks around the same positions can be seen. The peaks are therefore not produced by filtering the data. Error margins and peak conductance indications were omitted here to improve visibility.

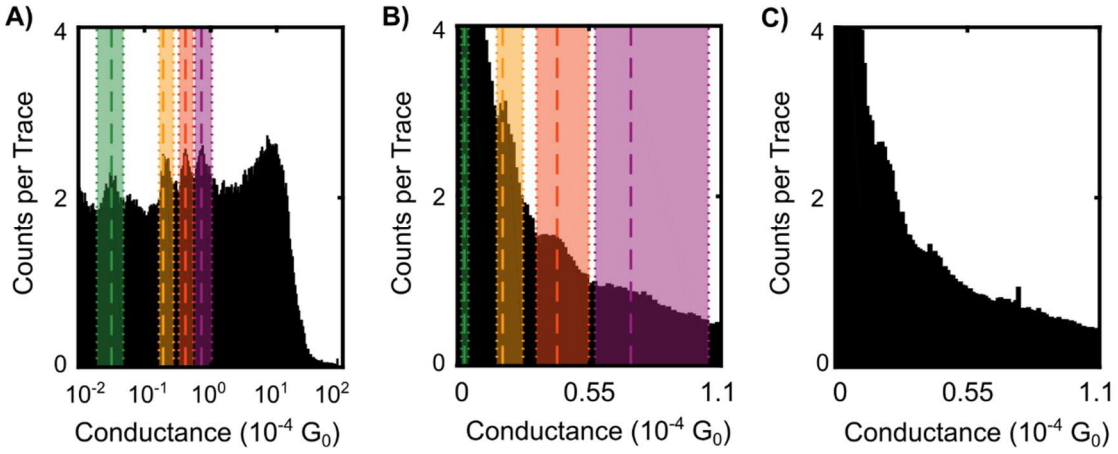


*Figure 4-23: **A**) Filtered logarithmic, **B**) filtered linear and **C**) unfiltered linear conductance histograms of plateau-containing traces. Histogram colour was set to black to improve contrast. Peak conductance is indicated by dashed lines, error margins by shaded intervals. Binning: 50 bins per decade for logarithmic histogram. Bin width for linear histogram: $0.01 \cdot 10^{-4}$ G_0.*

When looking at the individual traces that were found by the PF algorithm, plateaus can be seen at the estimated peak conductance values. A representative selection of traces with plateaus is shown in Figure 4-24. As with the s-FDT, combinations of all conductance classes can be seen within single traces for s-FPT.

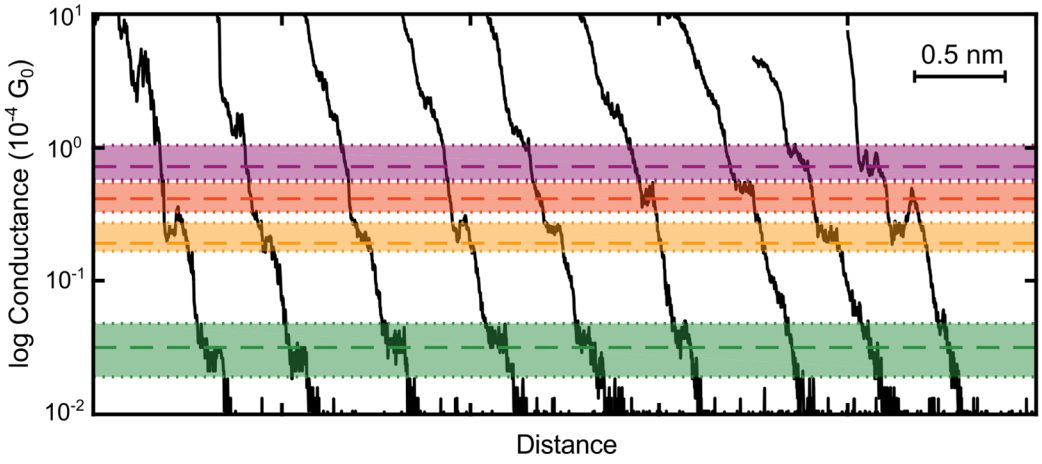


Figure 4-24: Representative sample traces. Dashed lines correspond to the most likely conductance values as estimated from the logarithmic histogram in the previous figure. Coloured intervals correspond to the error margins associated with the conductance values. Traces were offset for clarity.

Similar conductance peaks were found in experiments at different conductance values. The peak conductance as a function of applied bias voltage for all experiments performed on s-FPT is shown below, in Figure 4-25. The individual histograms are presented in the appendix.

The observed conductance peaks are remarkably similar to the ones observed for the dithiolate. Conductance ratios for positive/negative peak conductance are close to unity: 1.05 ± 0.08, 0.98 ± 0.05 and 0.94 ± 0.28 (low, medium and high), meaning that no rectification was

observed within experimental error. No ratio was calculated for the lowest class, because not enough bias positive/negative pairs showed conductance peaks.

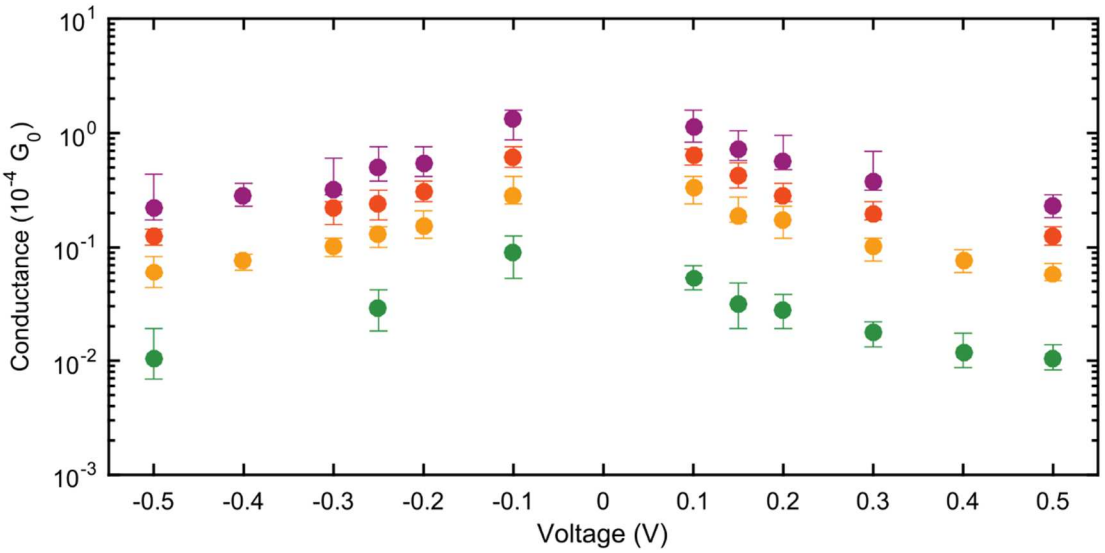


Figure 4-25: Summary of peak conductance values of s-FPT as determined from their log 1D conductance histograms. Different colours denote different conductance classes.

Robert Stadler and his colleagues calculated a molecular length of 1.77 nm (S-N) or a junction extension of 2.16 nm (Au-Au) for s-FPT. The experimentally measured break-off distances, Figure 4-26, stayed well below this expected junction elongation. The mean experimental break-off distance is 1.24 nm ± 0.10 nm. The individual histograms are shown in the appendix.

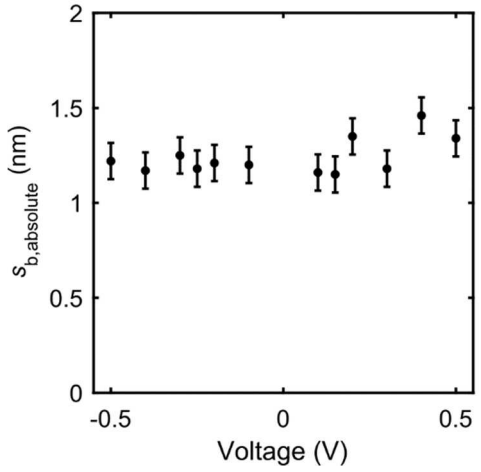


Figure 4-26: Break-off distances of s-FPT as a function of bias voltage.

Long plateaus are rarely observed for this analyte. An example of a data set that did show long plateaus is shown in Figure 4-27. The data was recorded at 0.6 V and 20 nA. Three peaks can still be seen around 10^{-5} G_0, but the peaks are rather ill-defined and were thus not included in Figure 4-26.

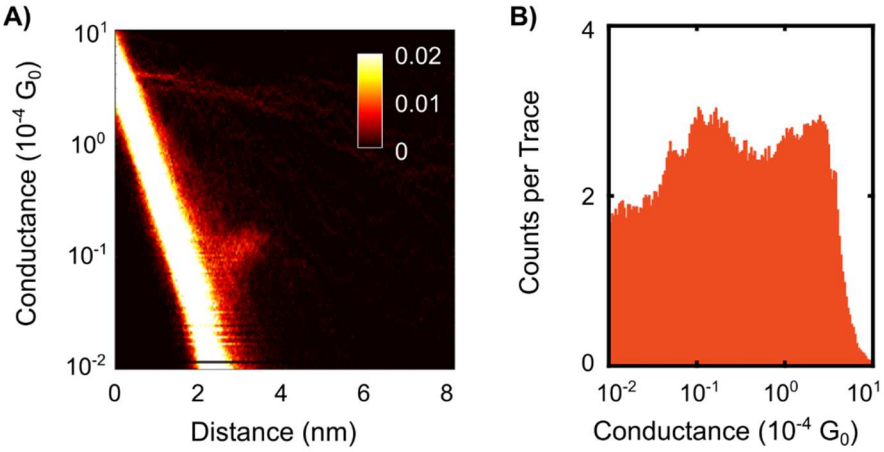


*Figure 4-27: Example data set with long plateaus. Logarithmic **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.6 V. Data selection was made with linear plateau finder algorithm. Bin widths **A**) 0.16 Å and 50 bins per decade **B**) $6.5 \cdot 10^{-7}$ G_0. Data in **B**) was filtered with a low-pass Bessel filter.*

4.2.6 Discussion and Summary s-FPT

Spectroscopy experiments have been performed on an unsymmetrical, linear, ferrocene-containing molecule. Plateaus in *I(s)* traces were reproducibly found at different experimental conditions. The observed conductance plateaus were typically very short and thus difficult to analyse with MPVC based on the current set of classifiers, but well-suited for analysis with the conventional PF algorithm. Long plateaus were sometimes, albeit rarely, observed and match the conductance pattern. Four conductance peaks were reproducibly seen at different bias voltages with high junction formation probability. The origin of these classes is currently unknown. Variations in the binding configuration, electrode rearrangements or multiple, interacting molecules bridging the junction could be explanations.[15,16,17,18] No rectification was observed within the experimental error: 1.05 ± 0.08, 0.98 ± 0.05 and 0.94 ± 0.28 for high medium and low conductance events, respectively. The conductance decreases with increasing absolute applied bias. The same effect was seen for s-FDT, where it was attributed to local heating of the molecular junction with increasing bias.

Robert Stadler and his group conducted NEGF DFT simulations for this system. They estimated a Fermi level conductance for s-FPT of $0.054 \cdot 10^{-4}$ G_0 (s-FDT: $0.155 \cdot 10^{-4}$ G_0). This conductance corresponds to the case of a molecule bridging of the gap between the electrodes, where each anchor group couples to one electrode. The conductance magnitude is in reasonable agreement with the experimentally observed conductance of the low conductance group.

The mean experimental break-off distance is 1.24 nm ± 0.10 nm. As for s-FDT, the experimental break-off distance stayed well below the expected length of 2.15 nm. No simulated junction extension is available for the case of a folded up molecular junction for s-FPT, but it is possible that the low junction extension corresponds to a molecule that is connected to one electrode with both anchor groups.

Very similar conductance trends were found for s-FDT and s-FPT. A comparison of conductance results for both molecules is shown in Figure 4-28. The reason for this similar conductance is currently unknown.

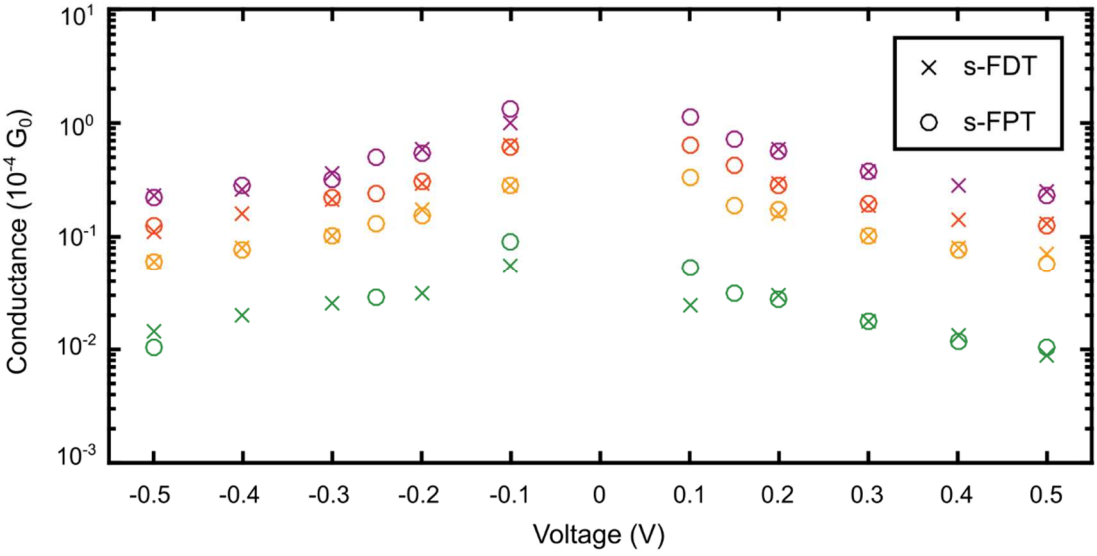


Figure 4-28: Comparison of peak conductance of s-FDT and s-FPT. s-FDT is shown as "x" and s-FPT as "o". Error bars were omitted for improved clarity.

4.3 Double-Branched Ferrocene

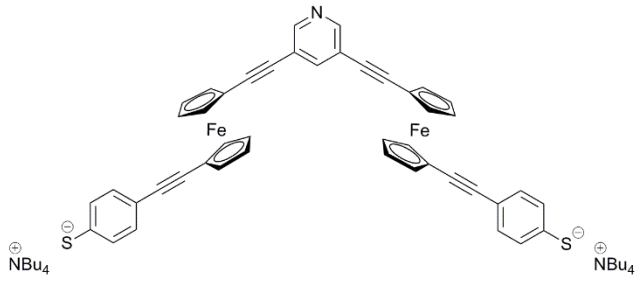


Figure 4-29: Structure of the double-branched ferrocene molecule, b-FPT.

In following experiments, it was attempted to measure the conductance of b-FPT, a branched version of the s-FPT molecule that was investigated in the previous section (see Figure 4-29 for structure). It was postulated that quantum interference could be probed with this molecule and other ring-shaped molecules similar to this one. However, despite considerable effort, no consistent conductance value could be found for b-FPT. Only in one experiment with setpoint current of 50 nA and an applied bias voltage of 0.2 V, a conductance peak could be seen. The conductance was

estimated to be $0.06 \cdot 10^{-4}$ G_0. ($0.03 \cdot 10^{-4}$ G_0 < G < $0.08 \cdot 10^{-4}$ G_0) The 2D and 1D conductance histograms are shown in Figure 4-30 **A** and **B**, respectively. However, the applied experimental conditions (low maximum extension and missing repeats and electrochemistry) prevented conclusive analysis about the nature of the junction. It is noteworthy that the conductance in this case is in close agreement with the Fermi level conductance of the single-branched version of the molecule, s-FPT ($0.054 \cdot 10^{-4}$ G_0).

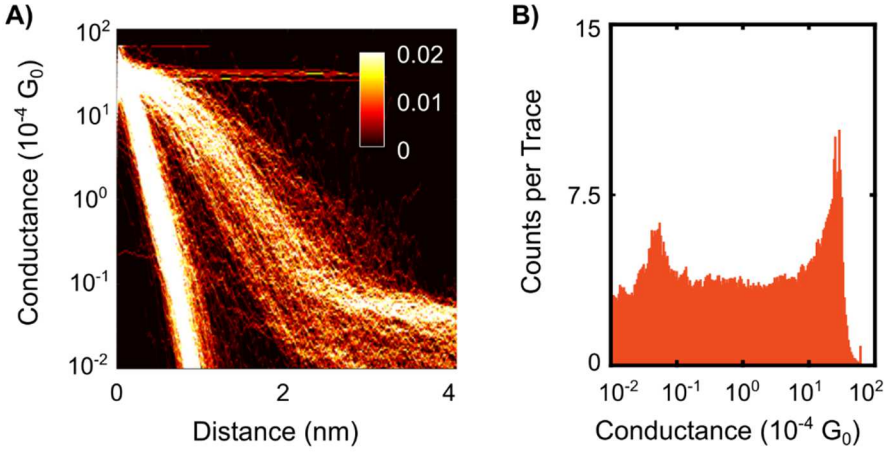


*Figure 4-30: **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.2 V and 50 nA. Data selection was made with linear plateau finder algorithm. Bin widths: **A**) 0.08 Å and 50 bins per decade in **A**) and **B**). Data in **A**) and **B**) was filtered with a low-pass Bessel filter.*

4.4 Biferrocene Compounds

4.4.1 BFTM Introduction

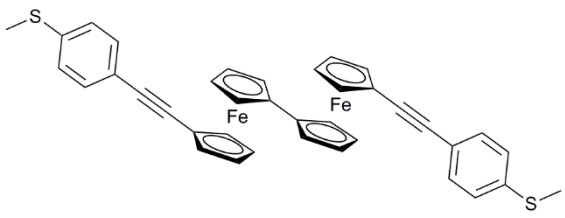


Figure 4-31: Structure of BFTM, a biferrocene molecule with thiomethyl linkers.

In this section, the results of the experiments performed on a biferrocene molecule, BFTM, are presented. The structure of BFTM is shown in Figure 4-31. It is identical to the one of s-FDTM except for the additional ferrocene centre in the backbone of the molecule. Thiomethyl linkers on either end of the molecule provide coupling sites to the electrodes. A sulfur-sulfur distance of 2.37 nm was calculated by Chen *et al.*[2] Accounting for twice the thiomethyl-gold bond length of 0.48 nm, this results in a total maximum junction extension of 2.85 nm.

4.4.2 Electrochemistry of BFTM

Electrochemical experiments were performed on the Au(111) substrate after spectroscopy experiments. Voltammograms were measured with an internal $[FeCp_2]^+/[FeCp_2]$ reference. In panel **A** in Figure 4-32, voltammograms at different scan rates are shown. The scan rate was varied between 25 mV/s up to 600 mV/s (inner to outer trace). Two separate oxidation and reduction steps were expected for this molecule. However, at the time only one transition was measured. The voltammograms are heavily affected by solution resistance and thus distorted from the expected shape.

The graph in Figure 4-32 **B** displays the peak current for the redox process. To reduce the instrumental error, the peak current of only the lowest five scan rates (25 mV to 125 mV) were fitted with a power fit. Exponents of 0.69 and 0.65 were determined for the anodic and cathodic processes, respectively. The deviation from an ideal surface immobilised process with an exponent of one was attributed to the unfavourable electrochemical/STM cell setup (see also Methods chapter).

Panel **C** shows the peak separation, ΔE, for the oxidation/reduction process (square symbols). The peak separation increases with increasing scan rate. This might also be due to the STM cell setup. The equilibrium potential (triangle symbols) stays approximately constant at higher scan rates.

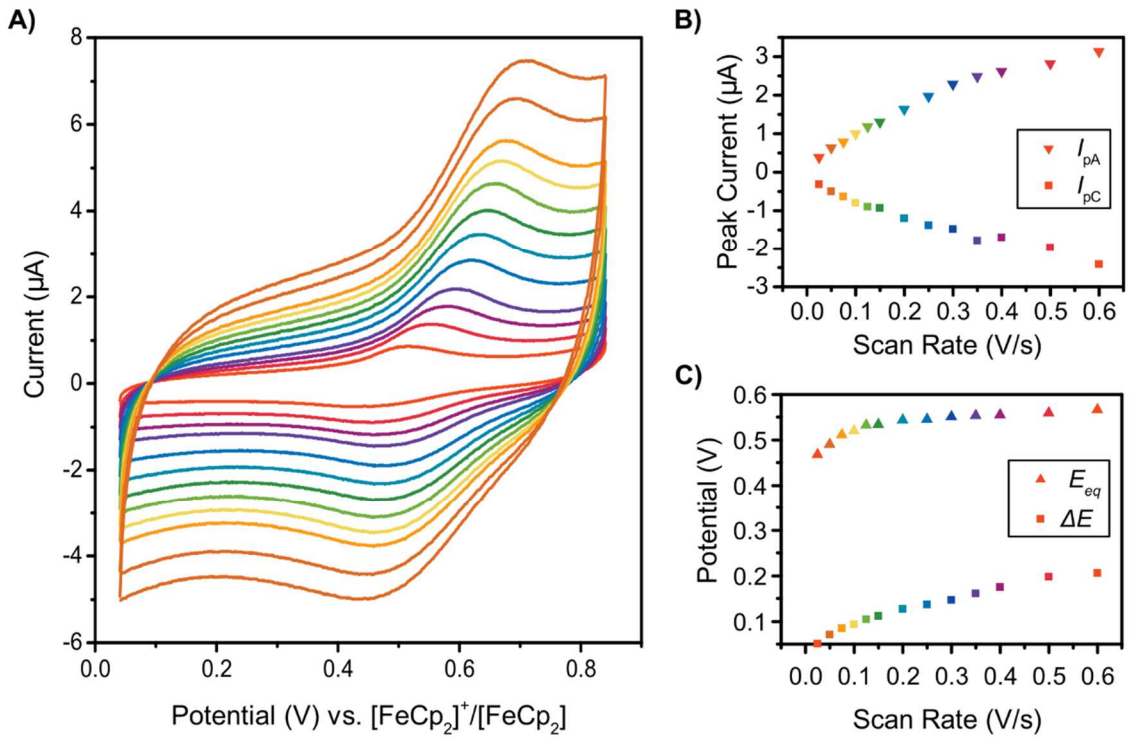


*Figure 4-32: **A**) Cyclic voltammograms with increasing scan rate (Inner to outer traces 25 mV/s to 600 mV/s). **B**) Peak current of the oxidation/reduction wave. **C**) Peak separation (squares) and equilibrium potentials (triangles) of the oxidation/reduction step. Graphs were corrected for solution resistance. Colour coding in all panels according to scan rate. Measurements were performed in degassed 0.1 M $TBAPF_6$/DCM under argon atmosphere. The same substrate was used for CV and for spectroscopy experiments.*

The surface area was estimated from the gold oxide reduction peak in 0.5 M H_2SO_4. A charge of 135.9 µC ± 6.7 µC was measured. The electrochemical surface area was calculated to be 0.31 cm^2 ± 0.02 cm^2. 222 µC/cm^2 were assumed as surface charge per area of the Au(111) crystal plane and two electrons are involved in the reduction step.[19]

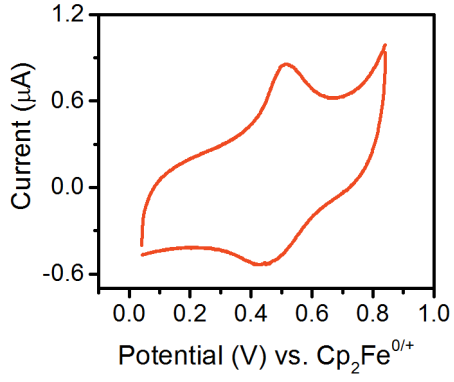


Figure 4-33: Cyclic voltammogram recorded at a scan rate of 25 mV/s.

From the voltammogram that was recorded at 25 mV/s, Figure 4-33, a mean charge of 1.15 µC ± 0.04 µC was calculated. The error associated with it was estimated by manually fitting three baselines to the capacitive current and calculating the standard deviation of the resulting peak charge. Comparing the peak charge to the surface charge gives a surface coverage of 1.8 %. This is the ratio of analyte molecules per surface gold atom. An approximate number of (7.19 ± 0.25)·10^{12} molecules were on the surface. This resulted in an approximate area of 4.3 nm^2 ± 0.23 nm^2 per molecule.

For 25 mV/s, the equilibrium potential was estimated to be 0.48 V. This is significantly higher compared to the 0/+ equilibrium potential in solution of 0.033 V, reported by Wilson *et al.* for a closely related molecule.[20] For the +/2+ process, the reported equilibrium potential is 0.382 V. The reasonable agreement between the observed equilibrium potential and the reported value could mean that the observed transition is the +/2+ process.

4.4.3 Conductance of the Au-(BFTM)-Au Junction

In two spectroscopy experiments with a substrate incubation time of one hour, no plateaus or other sign of the analyte could be detected. After increasing the incubation time to 70 hours, clear deviation from plain exponential decay was observed. Details of results and analysis are provided for a data set recorded at 0.1 V and 6.5 nA.

1000 traces were recorded and analysed with MPVC on the natural logarithm of the data set. An exponentially decaying trace with a decay factor of 0.8 $Å^{-1}$ was selected as reference vector. For the MPVC, a region of interest was defined between 0.12 nm and 1 nm. After

assigning each *I(s)* trace a 3D vector with MPVC, the whole data set can be visualised in the cylinder plot in Figure 4-34. At least two distinct clusters can be seen here: predominantly plain exponential traces (blue) and traces deviating from a plain exponential decay (red). These two types of traces were separated by GK FCM with two clusters. 8 % of traces were found to deviate from plain exponential decay.

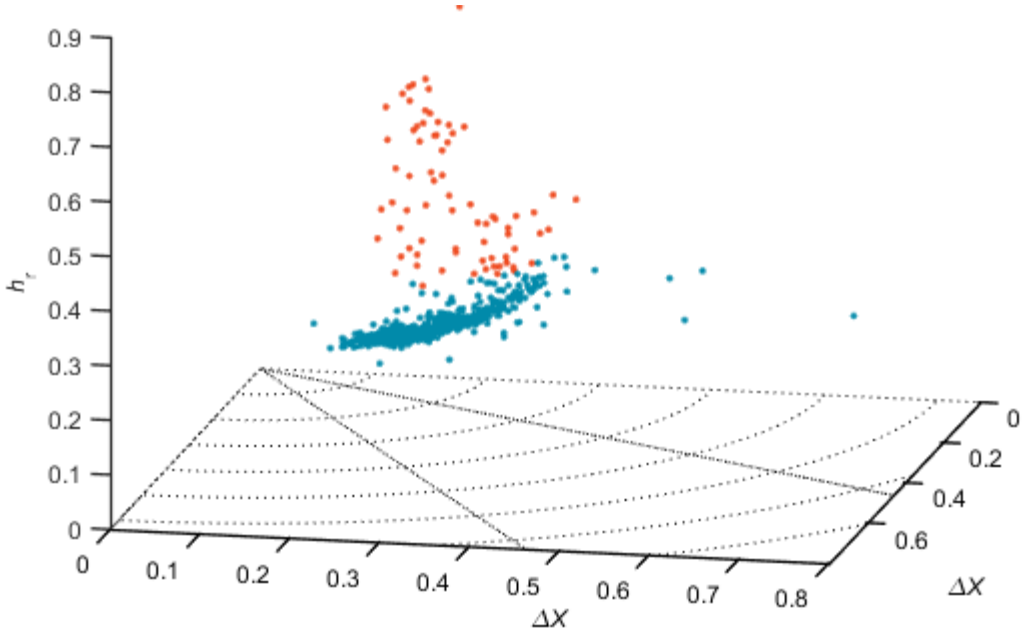


Figure 4-34: Cylinder plot of the whole data set. Colour assignment was done per cluster membership. Clustering was performed with GK FCM. Shown in red are plateau traces and data points in blue represent plain exponential traces.

The 1D and 2D conductance histograms of the red cluster were calculated and are shown in Figure 4-35. The logarithmic 2D histogram in **A** shows clear deviation from plain exponential decay. However, the plateau is not well-defined. Accordingly, the 1D histogram in **B** shows a broad peak with some substructure.

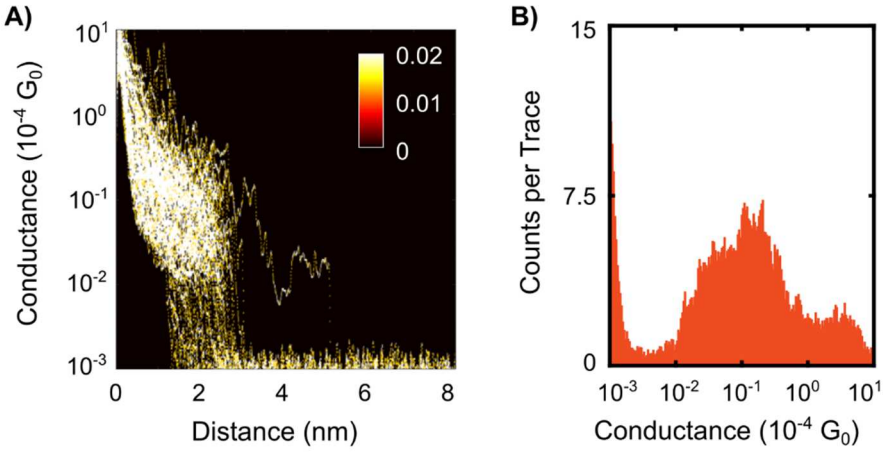


*Figure 4-35: Logarithmic **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.1 V and 6.5 nA. Data selection was made with MPVC and GK FCM. Bin widths: **A**) 0.16 Å and 50 bins per decade in **A**) and **B**). Data was filtered with a low-pass Bessel filter.*

When analysing the plateau traces in more detail, it can be seen that the data consists of several sub-clusters. See cylinder plot shown from different angles in Figure 4-36. Three clusters can be seen here. The plain exponential traces are shown in grey for improved contrast. Sub-clustering of the plateau traces was performed again with GC FCM clustering with three clusters.

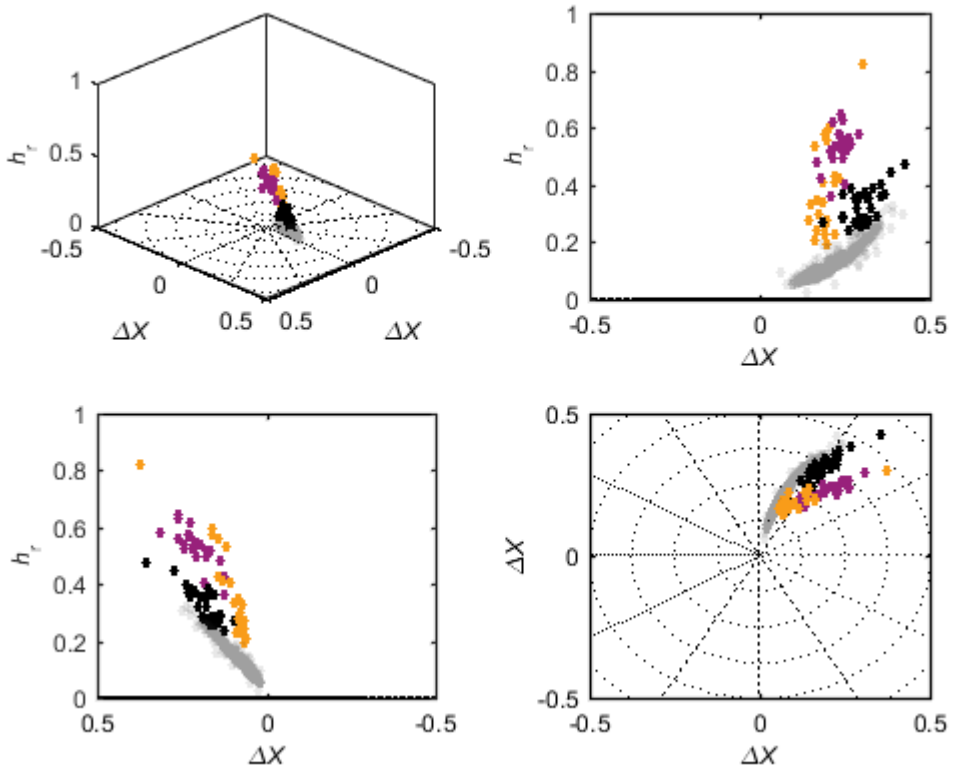


Figure 4-36: Cylinder plot of the whole data set in the MPVC representation. Colour coding was done after cluster assignment with FCM clustering. The cylinder plot is viewed diagonally, along two orthogonal ΔX directions and along Hamming axis.

After splitting the plateau traces into three clusters, the histograms are very well-defined. Figure 4-37 **A** and **C** show the log 2D and 1D conductance histograms of the orange, low-conductance cluster. **B** and **D** contain the same histograms for the purple, high-conductance cluster. Fitting the 1D histogram with Gaussian distributions gives the most probable conductance values of $0.05 \cdot 10^{-4}$ G_0 ($0.01 \cdot 10^{-4}$ G_0 < G < $0.13 \cdot 10^{-4}$ G_0) and $0.19 \cdot 10^{-4}$ G_0 ($0.08 \cdot 10^{-4}$ G_0 < G < $0.43 \cdot 10^{-4}$ G_0), respectively. The black cluster contains traces that cannot be clearly assigned to low or high conductance class as their plateaus are very short. The corresponding histograms are provided in the appendix.

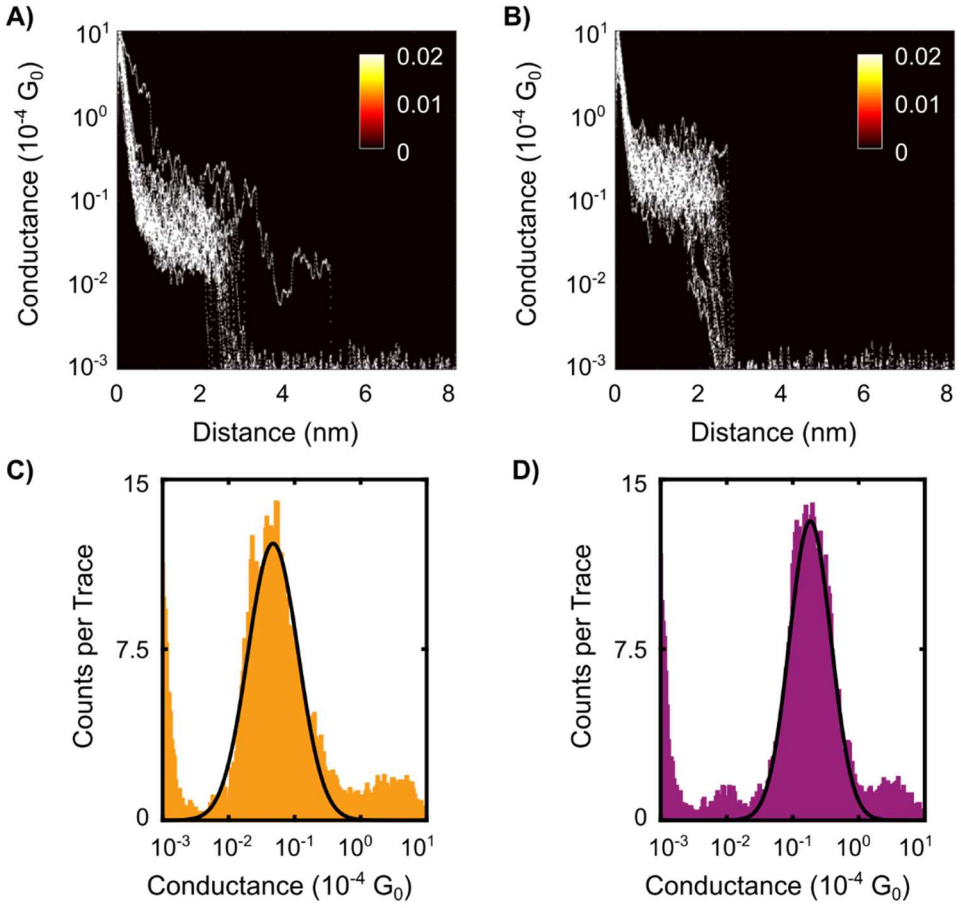


*Figure 4-37: **A**), **B**) 2D and **C**), **D**) 1D conductance histogram of the orange and purple cluster, respectively. Data selection was made with MPVC and GK FCM. Histograms in **C**) and **D**) were fitted with a Gaussian distribution. Bin widths: 0.16 Å and 50 bins per decade. Data was filtered with a low-pass Bessel filter.*

When looking at the individual traces, clear plateaus were observed around the estimated conductance values. A representative selection of traces for both conductance classes is shown in Figure 4-38.

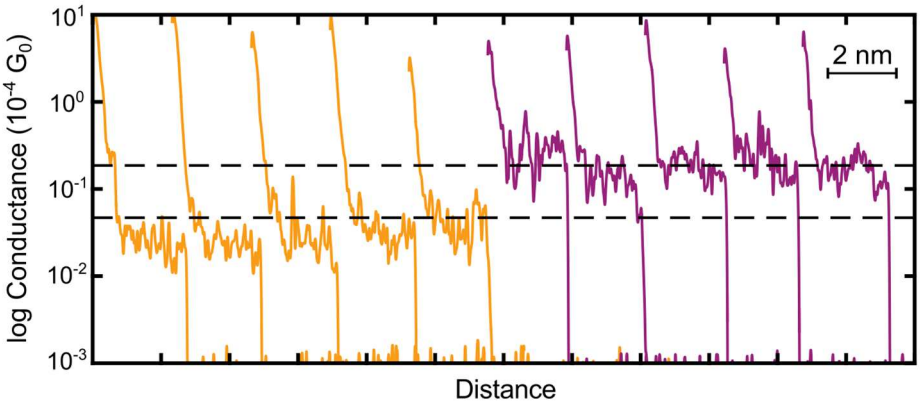


Figure 4-38: Representative sample traces of both conductance groups. Traces were filtered with a low-pass Bessel filter and offset for clarity. Dashed lines are most probable conductance of both conductance classes as determined from Gaussian fit of 1D log histogram.

The event distributions are displayed in Figure 4-39. It is obvious that traces associated with the two conductance classes are not collected uniformly during the experiment. Both for the high and low conductance class, traces are collected in the beginning of the experiment. After 123 traces, no further plateaus were observed for the remainder of the experiment.

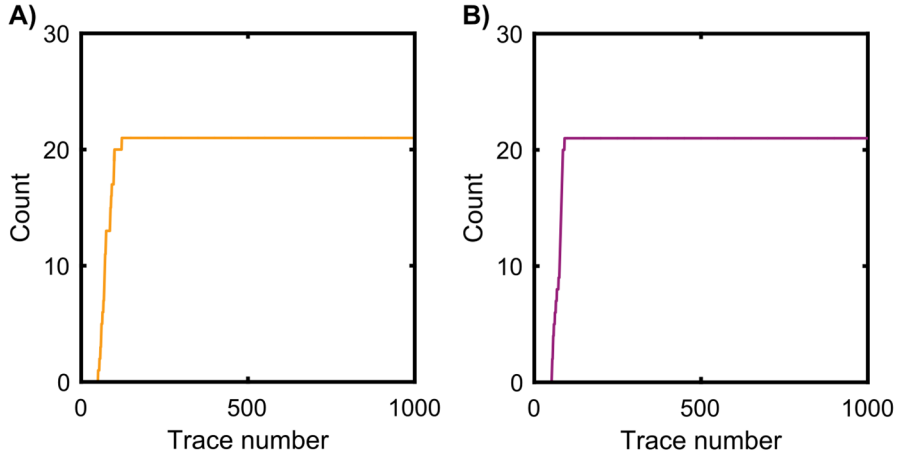


*Figure 4-39: Event distribution over the course of the experiment of the **A**) orange low-conductance class and **B**) purple high-conductance class.*

The origin of the two conductance classes was further investigated through the break-off distance. Histograms of the absolute break-off distance are shown in Figure 4-40. Absolute meaning here that the distance was corrected for the initial tip-substrate distance as calculated from linear fit of the logarithmic *I(s)* traces. For the orange cluster, a most likely break-off distance of 3.29 nm ± 0.25 nm was calculated. For the purple cluster, the most likely break-off distance was 3.24 nm ± 0.17 nm. The expected junction extension is 2.85 nm.

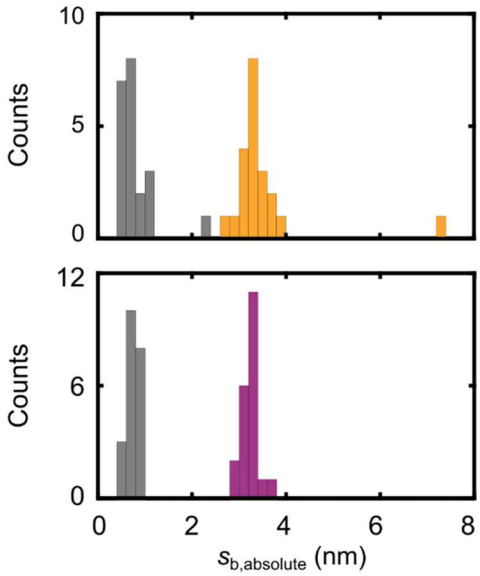


Figure 4-40: Break-off distance histograms of the orange and purple cluster. Shown in grey is the estimated initial tip-substrate distance. Bin width: 2 Å.

More voltages were probed during the experiment, however, for other conditions, the results were less clear. Plateaus were observed at different voltages, but often, a wide distribution in plateau current or short plateaus were observed. A selection of resulting histograms is shown in Figure 4-41.

Table 4-2: Approximate conductance values at different bias voltages. G_{low} and G_{high} representing the limits of the error interval given by the full width half maximum of the Gaussian fit. Note that conductance values given are only a rough estimate as no well-defined conductance peaks were observed. The corresponding conductance histograms are shown in Figure 4-41.

Bias (V)	G_{low} (10^{-4} G_0)	G (10^{-4} G_0)	G_{high} (10^{-4} G_0)
0.1	0.01	0.05	0.13
0.1	0.08	0.19	0.43
-0.2	0.003	0.1	3.72
0.4	0.08	0.5	2.82
0.5	0.001	<0.01	0.02
0.6	0.04	0.7	10

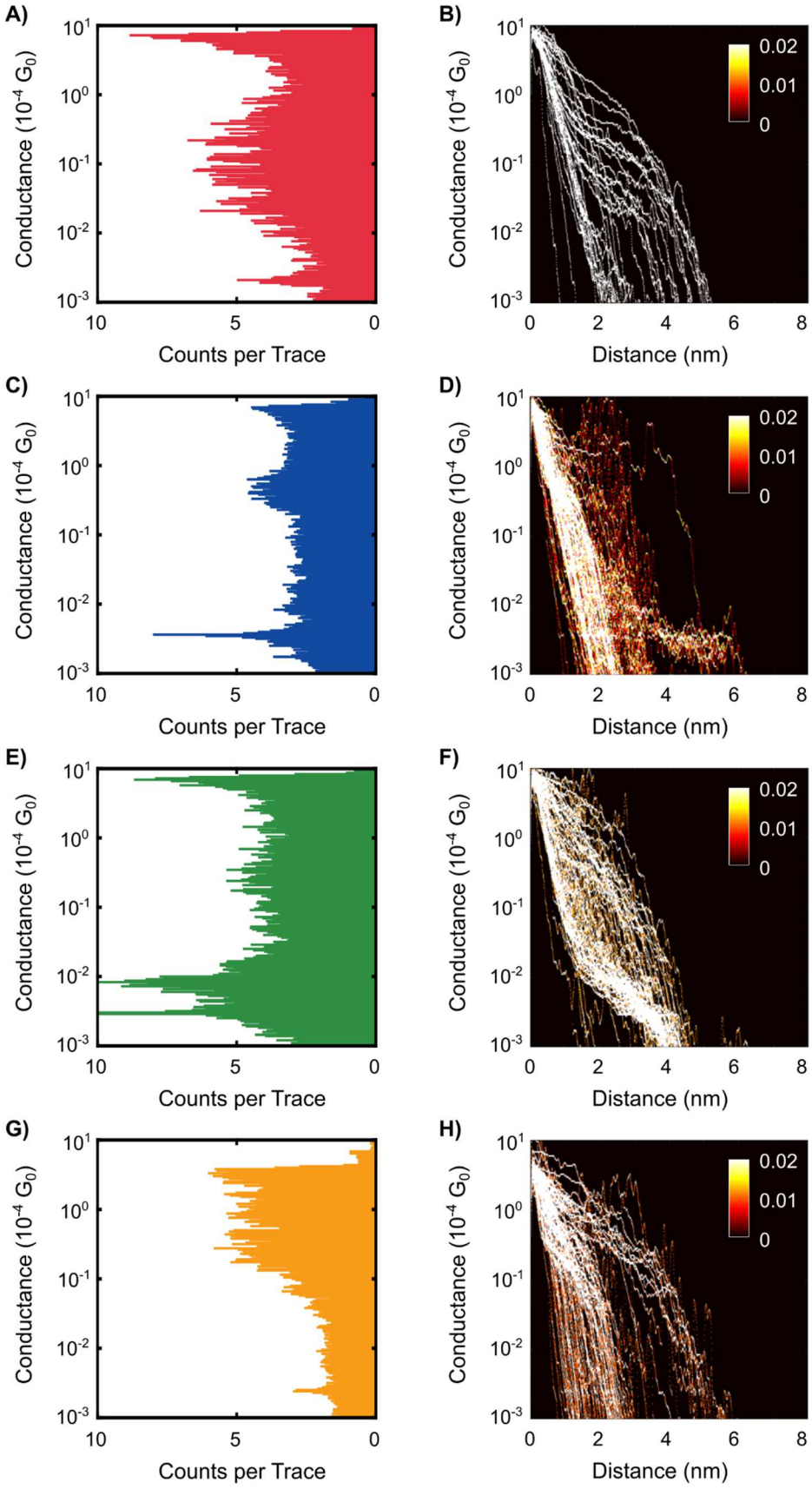


Figure 4-41: Histograms of plateau traces recorded at various voltages. The first column contains the log 1D histograms and the second column the log 2D

*histograms. Bias voltages: **A**) and **B**) -0.2 V, **C**) and **D**) 0.4 V, **E**) and **F**) 0.5 V and **G**) and **H**) 0.6 V. Data selection was performed with MPVC for **E**) and **F**) and logarithmic PF for the rest. Note that data was filtered with a low-pass Bessel filter. Binning: 0.16 Å and 50 bins per decade.*

4.4.4 Discussion and Summary BFTM

In spectroscopy experiments conducted at 0.1 V, three distinct conductance classes were identified. Two of those showed long plateaus at different conductance values. The third class showed short plateaus and is attributed to premature junction rupture. The two conductance classes with long plateaus are possibly due to different binding sites of the anchor groups, as reported before.[15,16,17,18] The conductance values were $0.05 \cdot 10^{-4}$ G_0 and $0.19 \cdot 10^{-4}$ G_0. Chen *et al.* reported a similar conductance of $0.03 \cdot 10^{-4}$ G_0 for the same biferrocene with thiol anchors in the ohmic region between -0.25 V and 0.25 V.[2]

The observed event frequency dropped to zero after about 123 traces for the two conductance classes. This drop of event frequency did not seem to be caused by a change in tip location by the macro scheduler. An explanation is that junction formation was hindered by some factor. Once contacted, the tip could have gotten functionalised with the molecule, remeasuring the same molecule in each consecutive trace.

In electrochemical experiments, it was found that the analyte was intact and on the substrate throughout spectroscopy experiments. Reversible oxidation and reduction was observed and an area of 4.3 $nm^2 \pm 0.23$ nm^2 per molecule was determined. This equals a surface coverage of (0.23 ± 0.01) molecules/nm^2

Both high and low conductance classes showed similar break-off distances of 3.2 nm and 3.3 nm, indicating that the same analyte was measured. Differences in binding configuration are likely below the experimental error. The expected junction extension of 2.85 nm was slightly exceeded, potentially caused by reorganization of gold atoms in the junction.[21]

4.4.5 Introduction BFP

Results for another molecule with biferrocene backbone will be discussed in this section. The structure for this biferrocene, BFP, is shown in Figure 4-42. The molecule is identical to the one discussed in the previous section, but the anchor groups were changed to meta-substituted pyridines.

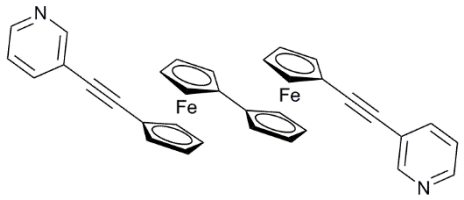


Figure 4-42: Structure of biferrocene with meta-substituted pyridine anchors, BFP.

4.4.6 Electrochemistry of BFP

Using cyclic voltammetry, it was confirmed that the molecule stayed surface bound throughout the spectroscopy experiments. To conclude this, cyclic voltammograms were recorded, Figure 4-43. Details of the sample preparation and experimental conditions are provided in Chapter 2. Two separate oxidation and reduction waves can be seen. The peak current of those redox waves rapidly decreases over the cycles indicating that the molecule desorbs or degenerates. For the first cycle, a peak separation of 0.35 V was observed. This is in excellent agreement with peak separation of BFTM, which was reported to be 0.349 V by Wilson *et al.*[20] In the same reference, an instability of the analyte with this particular electrolyte was also reported.

For the two reduction waves of the first cycle, a charge of 2.01 µC ± 0.46 µC was estimated. The charge is the mean of three integrations with different base line fits for both peaks. The error is the standard deviation. With the elementary charge, this results in $1.25 \cdot 10^{13}$ $\pm 0.29 \cdot 10^{13}$ molecules on surface.

The surface area of the gold substrate was estimated by integrating the gold oxide reduction peak in 0.5 M H_2SO_4. A mean charge of 326.2 µC ± 5.2 µC was measured in three different curves. This charge results in a surface area of 0.73 cm^2 ± 0.01 cm^2. With the surface area and the number of molecules on the surface, a total area per molecule of 5.84 nm^2 ± 1.36 nm^2 was calculated.

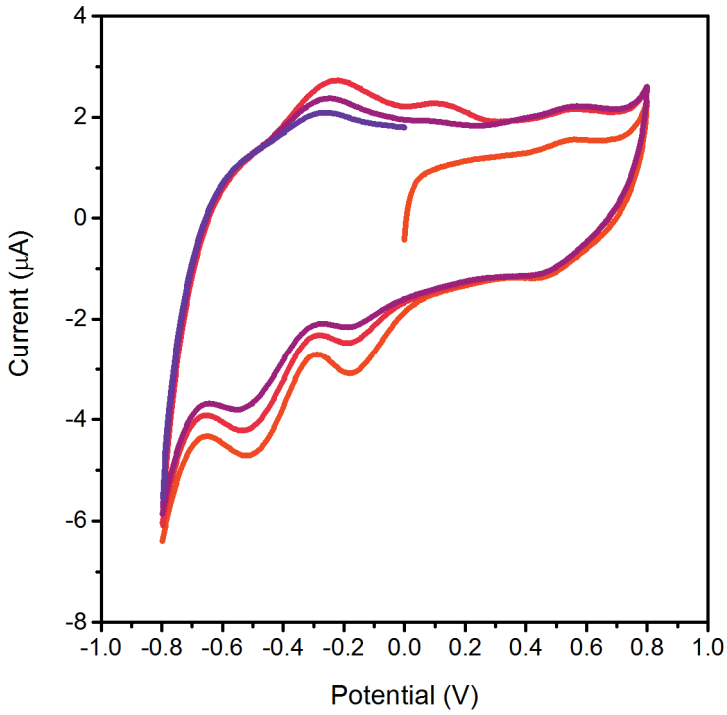


Figure 4-43: Surface immobilised cyclic voltammogram of BFP in 0.1 M TBAHFP in DCM. Peak intensity decreased over time. Note that potential was not measured against a reference since the analyte disappeared before a reference could be added.

4.4.7 Conductance of the Au-(BFP)-Au Junction

The conductance of BFP was measured with the *I(s)* technique. 20 data sets in two separate experiments were probed and 2000 traces were recorded for each data set. Two distinct conductance values were identified.

In a control experiment, before assembling the analyte on the substrate, spectroscopy was performed on the substrate. No deviation from plain exponential decay was observed in the 2D histogram in Figure 4-44 **A**. Accordingly, no peaks are observed in the 1D histogram in **B**.

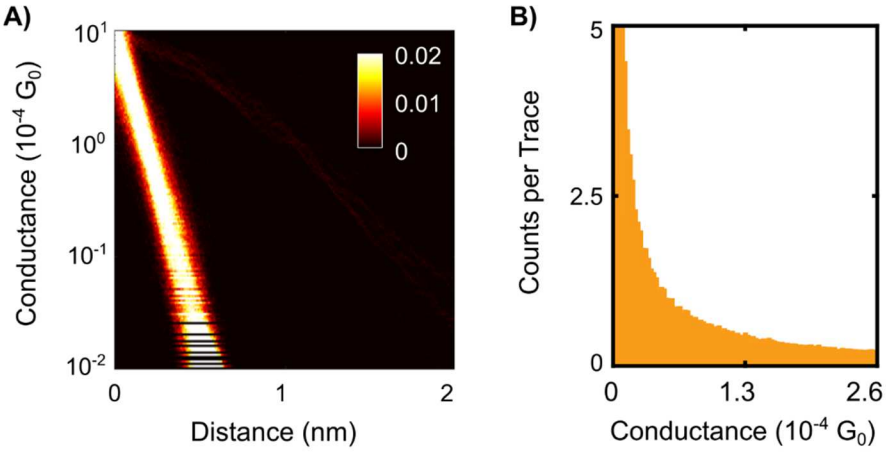


*Figure 4-44: **A**) All data log 2D and **B**) log 1D conductance histogram recorded before incubation of the substrate. 2000 traces were recorded at 0.3 V and 20 nA setpoint current.*

After incubation, spectroscopy was repeated. Shown in Figure 4-45 are the 2D and 1D conductance histogram of the whole data set recorded at 0.5 V (**A** and **B**, respectively). The 1D histogram does not show any peaks. The 2D histogram, on the other hand, shows clear deviation from the plain exponential decay. Many traces extend far beyond the molecular length. These non-exponential traces are in some form connected to the analyte as they do not appear in the control experiment, but they do not represent a well-defined junction and are therefore of limited value. They will be referred to as "noisy" traces. As shown below, MPVC can be a valuable tool in extracting meaningful data even from noisy data sets like this one.

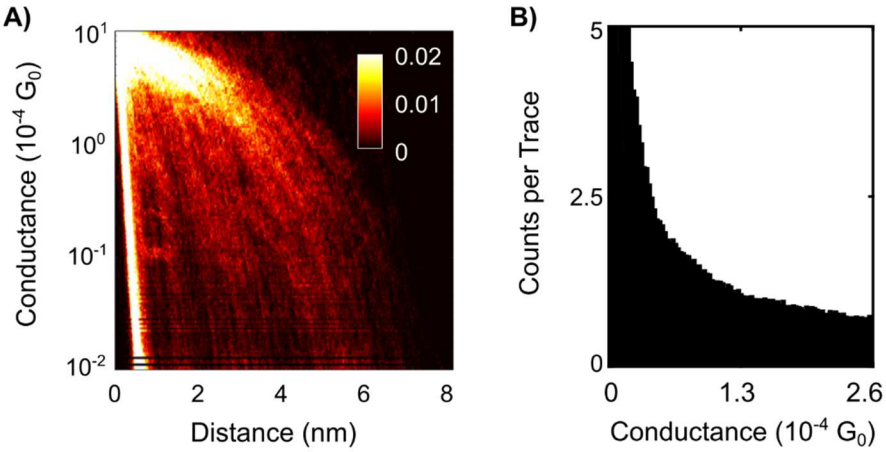


*Figure 4-45: **A**) 2D and **B**) 1D conductance histogram of the whole data set recorded at 0.5 V and 32.5 nA. Bin widths: **A**) 0.16 Å and 50 bins per decade **B**) $2.6 \cdot 10^{-6}$ G_0.*

MPVC was applied to the natural logarithm of the data. A reference vector with a decay factor of 0.8 $Å^{-1}$ was used. A region of interest was defined between 0.12 nm and 4 nm. The resulting vector representation of the data set is shown in the cylinder plot in Figure 4-46.

Colours and symbols were assigned with three cluster steps using the GK algorithm. During cluster assignment, four distinct clusters were found. Assigning all clusters in one step proved difficult because of the different sizes of the clusters and the challenging cluster shape.

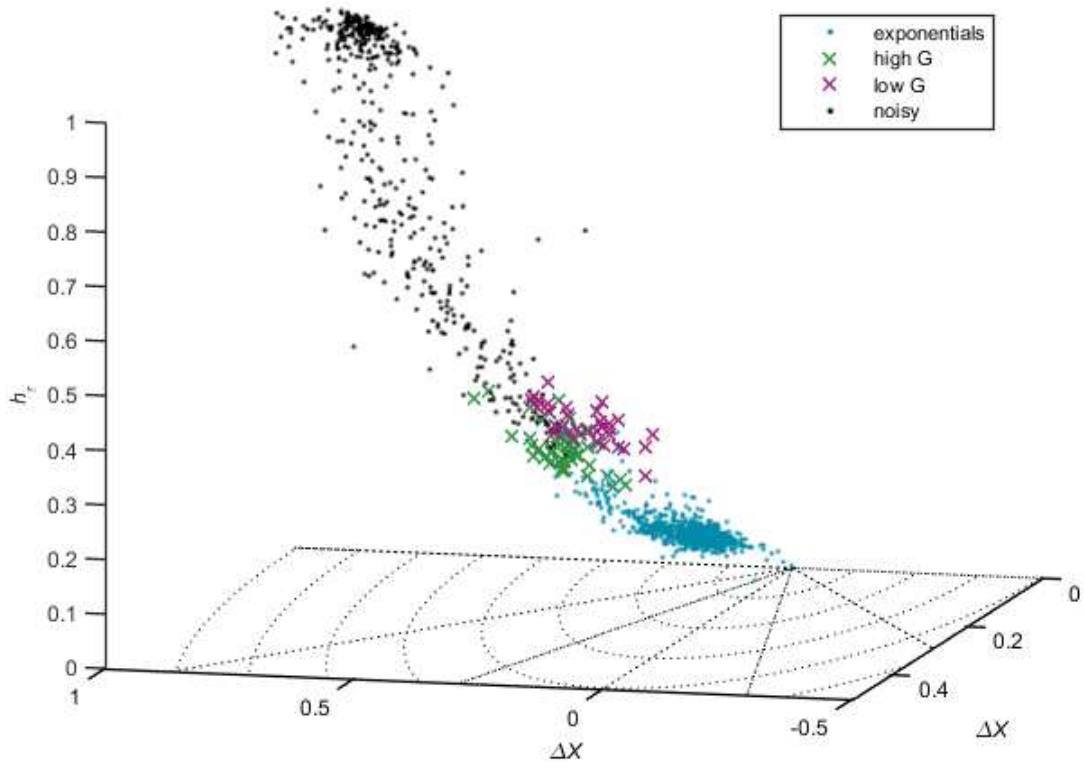


Figure 4-46: Cylinder plot of the whole data set in the vector representation. The vector representation was produced with the MPVC approach. Shown in blue are plain exponential traces, green and purple "x" are members of high and low conductance cluster, respectively. Shown in black are noisy traces that are connected to the analyte, but do not provide further insight into the data set.

Histograms of the two conductance groups are shown in Figure 4-47. In the first column are the 2D and 1D conductance histograms of the low conductance group and in the second column those for the high conductance group. In both 2D histograms in **A** and **B**, clear plateaus can be seen. In addition to the plateau traces, sub-linear traces without plateaus are observed. The origin of these traces is currently unknown, but better cluster assignment would likely exclude these traces from the clusters.

In agreement with the plateaus seen in the 2D histograms, the 1D histograms in **C** and **D** exhibit clear peaks. For the low conductance class in **C**, two conductance values were determined to be $0.11 \cdot 10^{-4}$ G_0 ± $0.02 \cdot 10^{-4}$ G_0 and $0.21 \cdot 10^{-4}$ G_0 ± $0.04 \cdot 10^{-4}$ G_0 by fitting with Gaussian distributions. The most likely conductance of the high conductance class was $0.69 \cdot 10^{-4}$ G_0 ± $0.47 \cdot 10^{-4}$ G_0.

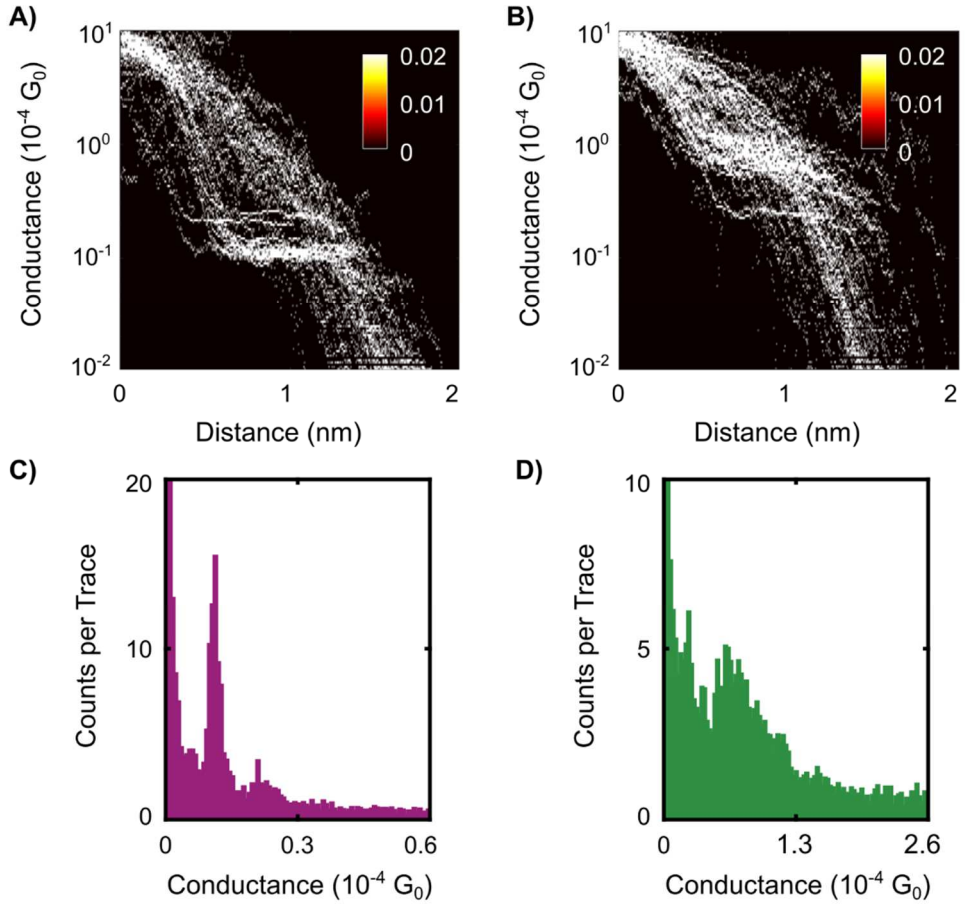


Figure 4-47: ***A****),* ***B****) 2D and* ***C****),* ***D****) 1D conductance histogram of the low conductance purple, and high conductance green cluster, respectively. Data selection was made with MPVC and GC FCM. Bin widths:* ***A****) and* ***B****) 0.04 Å and 50 bins per decade.****C****) and* ***D****) 2.6·10^{-6} G$_0$. Trace count:* ***A****) and* ***C****) 34.* ***B****) and* ***D****) 49.*

Low conductance and high conductance classes have similar, albeit low, junction formation probabilities. Only about 2 % of traces show low or high conductance features. Members of both clusters are shown in Figure 4-48. Members of the low conductance class are very well-defined and traces of the high conductance class are noisier and slanted.

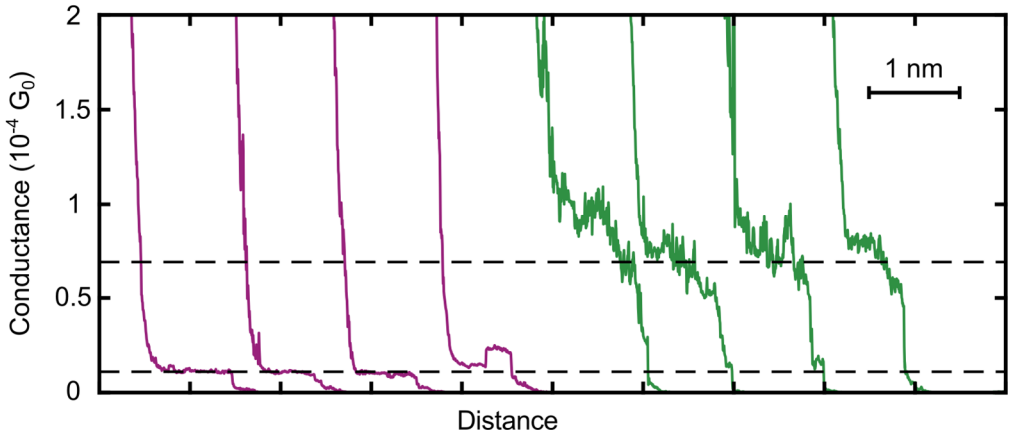


Figure 4-48: Sample traces of the low conductance (purple) and high conductance (green) cluster. Dashed lines represent most likely conductance values. Traces were offset for clarity.

The event distribution for both clusters is shown in Figure 4-49. Traces of the low conductance class were recorded in two separate bursts. Members of the high conductance class were mainly recorded in one consecutive string.

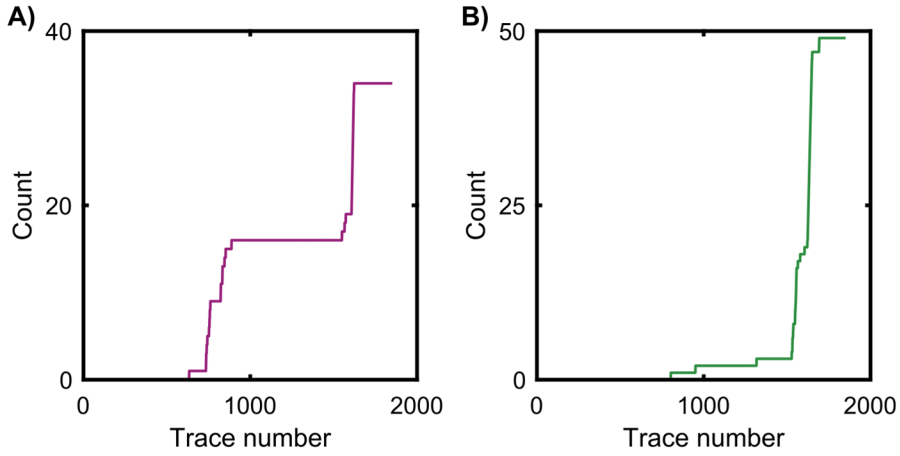


*Figure 4-49: Event distribution of the low conductance (**A**, purple) and high conductance (**B**, green) cluster over the course of the experiment.*

The break-off distances of both clusters are shown in histograms in Figure 4-50. By fitting with Gaussian distributions, the most likely break-off distances were calculated at 4.29 nm ± 1.41 nm for the low conductance class and 3.78 nm ± 1.07 nm for the high G conductance class. The error associated with these distances is large because of the low number of members in each group. Chen *et al.* reported the length of a compound with the same backbone.[2] From this length, the total expected junction extension was estimated to be 2.5 nm.

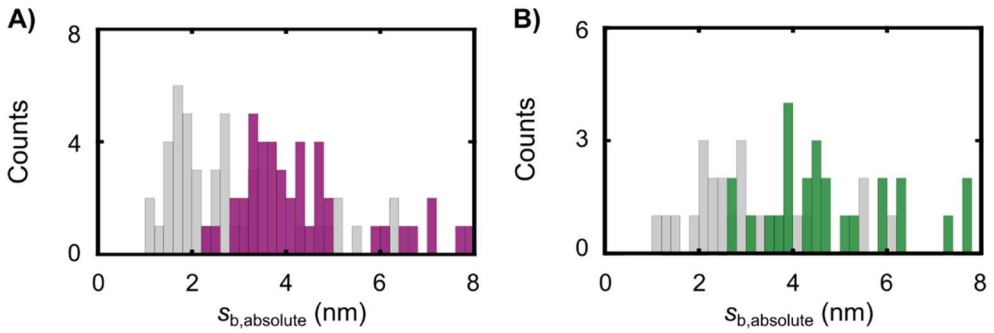


*Figure 4-50: Absolute break-off distance histograms of the **A**) low conductance and **B**) high conductance cluster. Grey bars are the initial tip-substrate distances. Bin width: 2 Å*

Repeating the conductance result proved to be difficult. Figure 4-51 shows 2D and 1D histograms of two experiments where peaks were seen in histograms. The first column shows a data set recorded at 0.3 V and 20 nA. The conductance was estimated to be $0.28 \cdot 10^{-4}$ G_0 ($0.06 \cdot 10^{-4}$ $G_0 < G < 1.35 \cdot 10^{-4}$ G_0). The second column was recorded at 0.9 V and 20 nA. Here the conductance was estimated to be $0.14 \cdot 10^{-4}$ G_0 ($0.01 \cdot 10^{-4}$ $G_0 < G < 2.45 \cdot 10^{-4}$ G_0) The large errors associated with both measurements come from the fitting on logarithmic scale.

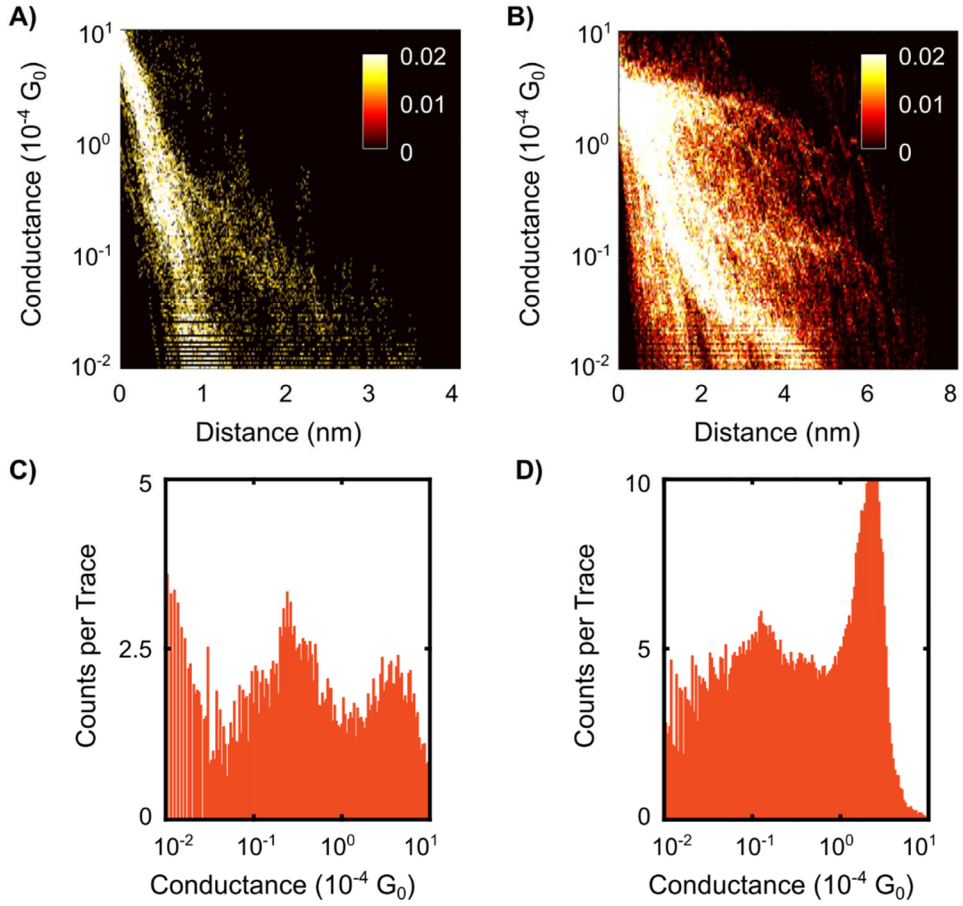


*Figure 4-51: **A**), **B**) 2D and **C**), **D**) 1D conductance histogram of data sets recorded at 0.3 V and 0.9 V, respectively. Data selection was made with linear plateau finder. Bin widths: 0.08 Å and 50 bins per decade.*

4.4.8 Discussion and Summary BFP

Cyclic voltammetry and spectroscopy were performed on a biferrocene compound with meta-substituted pyridine anchor groups. With cyclic voltammetry, two redox waves were found with peak splitting in excellent agreement with literature. Due to unfavourable interaction of the analyte with the electrolyte, the analyte assembly degraded or desorbed during cyclic voltammetry. It was confirmed that the molecule stayed intact and surface bound throughout the spectroscopy experiment. An area of approximately 6 nm^2 was estimated per analyte molecule on the gold surface. For the same molecule with thiomethyl linkers an area of about 4 nm^2 was estimated.

Spectroscopy experiments were performed and MPVC analysis was applied to the data. Comparison with the control experiments suggested that the deviations from exponential decay were in fact due to the molecule. Two distinct conductance groups were found. A very well-defined low conductance group was seen at $0.11 \cdot 10^{-4}$ G_0 ± $0.02 \cdot 10^{-4}$ G_0 with a second, minor

peak at $0.21 \cdot 10^{-4}$ $G_0 \pm 0.04 \cdot 10^{-4}$ G_0. The second peak at about twice the conductance of the first peak is likely to be two molecules bridging the gap.[22] The most likely conductance of the high conductance class was $0.69 \cdot 10^{-4}$ $G_0 \pm 0.47 \cdot 10^{-4}$ G_0. Note that the high error associated with the measurement corresponds to the error of a single measurement. The appearance of two conductance groups could be due to the pyridine anchors. Kamenetska *et al.* reported on the binary conductance switching for four para-substituted bipyridine anchored molecules.[23,24] They write that the low conductance state corresponds to the case of the molecule connecting to an apex atom on either electrode, while the high conductance state corresponds to a tilted configuration with increased coupling compared to the low conductance case. They also report that the low conductance configuration follows the high conductance configuration within single traces. This effect is also observed in the presented data, albeit not as pronounced as Kamenetska reports (Green traces, Figure 4-48). The conductance switching during trace evolution seems to be more pronounced in literature could be explained by the fact that BFP has meta-substituted pyridine anchors, which would be expected to couple to the electrodes differently than the para-substituted in the reference. The longer break-off distance of the low- compared to the high conductance group (4.29 nm ± 1.41 nm compared to 3.78 nm ± 1.07 nm) also aligns well with Kemenetska's findings, where the low conductance state is reached before break-off. Both break-off distances exceeded the expected break-off distance of 2.5 nm. Reorganisations of the electrodes could cause this effect.[21]

As a reminder, for DFTM conductance values of $0.05 \cdot 10^{-4}$ G_0 and $0.19 \cdot 10^{-4}$ G_0 were found. This means that the conductance found for BFP was about a factor of two to four higher than the one found for DFTM. Similar conductance values were found during two other experiments, albeit plateaus were not as well-defined as in the presented analysis. The high conductance of BFP compared to BFTM is surprising at first since the meta-substituted pyridine would likely incur destructive quantum interference[25,26] and even para-substituted pyridine has shown smaller conductance than thiomethyl.[4] However, it has been reported that constructive quantum interference can be observed for meta-substituted heteroatom rings.[27] It also needs to be stressed that the conductance values were recorded under different bias voltages and are therefore not necessarily comparable.

The event distribution suggested that plateau traces were recorded discretely rather than continuously throughout the experiment. This could mean that molecules rarely bind to the tip, but if binding occurs, the tip can become functionalised and multiple *I(s)* traces can get measured with the same molecule.

4.5 Summary and Conclusion

In the first section, spectroscopy and electrochemical experiments were performed on three linear ferrocene molecules. The molecules consisted of an identical backbone and three different anchor groups.

For the thiomethyl-anchored s-FDTM, low reproducibility was seen. Plateaus were rarely observed, but if observed, multiple plateaus were recorded in a row. At -0.1 V conductance values of $1.8 \cdot 10^{-4}$ G_0 (MPVC) and $1.4 \cdot 10^{-4}$ G_0 (plateau finder) were seen. Unconventionally-shaped plateaus were observed. Reasonable agreement was observed with conductance reported by Lu *et al.* for the same system with NH_2 anchor groups ($3.7 \cdot 10^{-4}$ G_0 at 0.05 V).[3] This agreement might, however, be coincidental because during the analysis of the other linear ferrocenes it became apparent that the molecules might not have been contacted across their whole length.

To increase reproducibility compared to s-FDTM, the anchor groups were changed to stronger thiols anchors.[18] Reproducibility and junction formation probability increased considerably for s-FDT. Short plateaus were observed in the data with high JFP. Four distinct conductance classes were observed at voltages between -0.5 V and 0.5 V. The conductance values were in agreement with the Fermi level conductance that Stadler *et al.* simulated for the system. Break-off distances stayed well below expected values (1.42 nm compared to 2.37 nm). The short break-off distances are even more surprising considering the strong thiol-gold interaction is expected to lead to gold strings being pulled out of the surface.[14,18] That is, if both electrodes are connected to thiols.

In the third linear ferrocene, s-FPT, one thiol anchor was changed to a meta-substituted pyridine. Again, four conductance classes were observed with high JFP and good reproducibility. Like for the previous two molecules, the break-off distances were shorter than expected and the simulated Fermi level conductance was in agreement with the lowest observed conductance class.

The results of the experiments on the linear ferrocenes suggest that there is doubt if the molecules were measured across their full length. It is possible that the STM tip contacted a Cp-ring of the ferrocene instead of the anchor group, possibly because both anchors bound to the substrate. Cyclic voltammetry proved to be difficult on the system to investigate this matter further.

In the second section, spectroscopy experiments were performed on the double-branched ferrocene b-FPT. Reproducibility was low and plateaus were only observed in one experiment

with a high setpoint current of 50 nA and a potential of 0.2 V. A conductance of $0.06 \cdot 10^{-4}$ G_0 ($0.03 \cdot 10^{-4}$ $G_0 < G < 0.08 \cdot 10^{-4}$ G_0) was estimated. Due to unfortunate experimental conditions and low reproducibility, the nature of the junction cannot be elaborated on. However, the estimated conductance is in good agreement with the simulated conductance for single-branched version of the molecules-FPT of $0.054 \cdot 10^{-4}$ G_0.

In the last section, the results of two biferrocene compounds were presented. With cyclic voltammetry, it was determined that BFTM stayed intact throughout spectroscopy. Each molecule occupied an average area of about 4 nm^2. Two conductance groups were identified at 0.1 V. One at $0.05 \cdot 10^{-4}$ G_0 and one at $0.19 \cdot 10^{-4}$ G_0. The measured break-off distances were larger than expected. The expected junction length was 2.85 nm compared to the measured length of 3.29 nm and 3.24 nm for the low and high conductance class, respectively. In addition to the large error of the experimental s_b, electrode reorganization could be responsible for the mismatch. The event distributions of the two conductance classes were discrete. Despite high coverage, junction formation seemed to be limited by some factor.

As the last molecule in this chapter, BFP, a biferrocene with meta-substituted pyridine linker, was investigated. The analyte signal quickly vanished during cyclic voltammetry, likely due to interactions with the electrolyte. It was, however, concluded that the molecule did stay intact during spectroscopy. An area of 6 nm^2 per molecule was estimated, meaning lower coverage than BFTM. MPVC was applied to the data and two conductance groups were found at $0.11 \cdot 10^{-4}$ G_0 and $0.69 \cdot 10^{-4}$ G_0. The low conductance class had a smaller "satellite peak" at double the conductance associated with it. This smaller peak was attributed to two molecules in parallel forming the junction. The measured break-off distances of 4.29 nm and 3.78 nm exceeded the expected distance of 2.5 nm.

4.6 References

1 R. Davidson, O. A. Al-Owaedi, D. C. Milan, Q. Zeng, J. Tory, F. Hartl, S. J. Higgins, R. J. Nichols, C. J. Lambert and P. J. Low, *Inorg. Chem.*, 2016, **55**, 2691–2700.

2 C.-P. Chen, W.-R. Luo, C.-N. Chen, S.-M. Wu, S. Hsieh, C.-M. Chiang and T.-Y. Dong, *Langmuir*, 2013, **29**, 3106–3115.

3 Q. Lu, C. Yao, X. Wang and F. Wang, *J. Phys. Chem. C*, 2012, **116**, 17853–17861.

4 R. Frisenda, S. Tarkuc, E. Galan, M. L. Perrin, R. Eelkema, F. C. Grozema and H. S. J. van der Zant, *Beilstein J. Nanotechnol.*, 2015, **6**, 1558–1567.

5 O. Al-owaedi, S. Eaves, M. Lemmer, B. W. Skelton, H. M. Osorio, S. J. Higgins, P. Cea, N. J. Long, C. J. Lambert, P. J. Low, S. Bock, O. Al-owaedi, S. Eaves, D. C. Milan, M. Lemmer, B. W. Skelton, H. M. Osorio, R. J. Nichols, S. J. Higgins, P. Cea, N. J. Long, T. Albrecht, S. Martin, C. J. Lambert and P. J. Low, *Chem. - A Eur. J.*, 2017, **23**, 2133–2143.

6 W. Haiss, S. Martín, E. Leary, H. Van Zalinge, S. J. Higgins, L. Bouffier and R. J. Nichols, *J. Phys. Chem. C*, 2009, **113**, 5823–5833.

7 E. Leary, H. Höbenreich, S. J. Higgins, H. Van Zalinge, W. Haiss, R. J. Nichols, C. M. Finch, I. Grace, C. J. Lambert, R. McGrath and J. Smerdon, *Phys. Rev. Lett.*, 2009, **102**, 1–4.

8 T. A. Su, H. Li, M. L. Steigerwald, L. Venkataraman and C. Nuckolls, *Nat. Chem.*, 2015, **7**, 215–220.

9 M. Kamenetska, M. Koentopp, A. C. Whalley, Y. S. Park, M. L. Steigerwald, C. Nuckolls, M. S. Hybertsen and L. Venkataraman, *Phys. Rev. Lett.*, 2009, **102**, 2–5.

10 C. Bruot, J. Hihath and N. Tao, *Nat. Nanotechnol.*, 2011, **7**, 35–40.

11 M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. Costa-Milan, R. J. Nichols, N. J. Long and T. Albrecht, *J. Am. Chem. Soc.*, 2015, **137**, 9971–9981.

12 A. . Fallis, *Introducting Molecular Electronics*, 2013, vol. 53.

13 Z. Huang, B. Xu, Y. Chen, M. Di Ventra and N. Tao, *Nano Lett.*, 2006, **6**, 1240–1244.

14 D. Krüger, H. Fuchs, R. Rousseau, D. Marx and M. Parrinello, *Phys. Rev. Lett.*, 2002, **89**, 186402.

15 J. Ulrich, D. Esrail, W. Pontius, L. Venkataraman, D. Millar and L. H. Doerrer, *J. Phys. Chem. B*, 2006, **110**, 2462–2466.

16 C. Li, I. Pobelov, T. Wandlowski, A. Bagrets, A. Arnold and F. Evers, *J. Am. Chem. Soc.*, 2008, **130**, 318–326.

17 X. Li, J. He, J. Hihath, B. Xu, S. M. Lindsay and N. Tao, *J. Am. Chem. Soc.*, 2006, **128**, 2135–2141.

18 M. Frei, S. V. Aradhya, M. S. Hybertsen and L. Venkataraman, *J. Am. Chem. Soc.*, 2012, **134**, 4003–4006.

19 H. Angerstein-Kozlowska, B. E. Conway, A. Hamelin and L. Stoicoviciu, *Electrochim. Acta*, 1986, **31**, 1051–1061.

20 L. E. Wilson, C. Hassenrück, R. F. Winter, A. J. P. White, T. Albrecht, N. J. Long, C. Hassenruck, R. F. Winter, A. J. P. White, T. Albrecht and N. J. Long, *Chem. - A Eur. J.*, 2016, **2017**, 1–10.

21 C. R. Arroyo, E. Leary, A. Castellanos-Gomez, G. Rubio-Bollinger, M. T. Gonzalez and N. Agrait, *J. Am. Chem. Soc.*, 2011, **133**, 14313–14319.

22 B. Xu and N. Tao, *Science*, 2003, **301**, 1221–1223.

23 M. Kamenetska, S. Y. Quek, a. C. Whalley, M. L. Steigerwald, H. J. Choi, S. G. Louie, C. Nuckolls, M. S. Hybertsen, J. B. Neaton and L. Venkataraman, *J. Am. Chem. Soc.*, 2010, **132**, 6817–6821.

24 S. Y. Quek, M. Kamenetska, M. L. Steigerwald, H. J. Choi, S. G. Louie, M. S. Hybertsen, J. B. Neaton and L. Venkataraman, *Nat. Nanotechnol.*, 2009, **4**, 230–4.

25 C. J. Lambert, *Chem. Soc. Rev.*, 2015, **44**, 875–888.

26 D. Z. Manrique, C. Huang, M. Baghernejad, X. Zhao, O. a Al-owaedi, H. Sadeghi, V. Kaliginedi, W. Hong, M. Gulcur, T. Wandlowski, M. R. Bryce and C. J. Lambert, *Nat. Commun.*, 2015, **6**, 1–8.

27 S. Sangtarash, H. Sadeghi and C. J. Lambert, *Nanoscale*, 2016, **8**, 13199–205.

Chapter 5

Ruthenium Macrocycles

5.1 Introduction 130
5.2 Conductance of the Au-(T-Ru_2M)-Au Junction 131
5.3 Conductance of the Au-(sPy-Ru_4M)-Au Junction 137
5.4 Conductance of the Au-(lPy-Ru_4M)-Au Junction 146
5.5 T-Ru_4M 155
5.6 Ruthenium Macrocycles: Summary and Discussion 168
5.7 References 170

Synopsis: *After a brief introduction, the findings for spectroscopic measurements on a diruthenium macrocycle will be discussed. After that, the single molecule conductance results for a tetraruthenium macrocycle with three different anchor groups will be presented. All outcomes will be discussed and summarised in the last section of the chapter.*

5.1 Introduction

Organometallic molecules have tremendous potential for the use in molecular electronics. It was explained in Chapter 1 that amongst other things, organometallic compounds generally have well-aligned frontier orbitals with the electrodes. This makes charge injection into the molecule efficient. Furthermore, attenuation of conductance with length is low in organometallics compared to their purely organic counterparts. These attributes make them interesting candidates for molecular wires. Apart from their excellent charge transport properties, they offer the advantage of being electrochemically active. This enables manipulation of the redox state during the measurement.

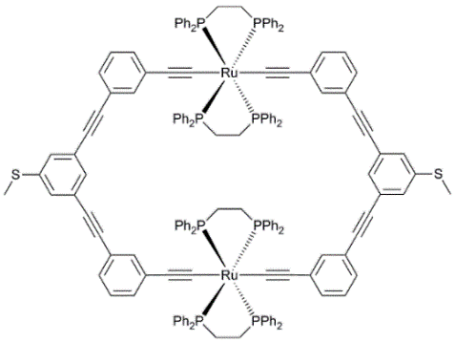


Figure 5-1: Structure of the ruthenium macrocycle with thiomethyl linkers. (T-Ru_2M)

Another class of molecules that show potential for molecular electronics are ring-shaped systems. They offer the possibility to probe quantum interference effects (see Chapter 1 for details). Possible applications include, but are not limited to single-molecule transistors and sensors.

Through the synthetic effort of members of the Albrecht/Long group and the Winter group in Konstanz, for the first time, a combination of both classes became available for single molecule conductance experiments. Macrocyclic organometallic molecules potentially merge the benefits of both classes of molecules, mentioned above, and are therefore of special interest to the molecular electronics community.

In this chapter, the findings for four macrocyclic ruthenium compounds will be discussed. Firstly, the experimental results for the diruthenium macrocycle shown in Figure 5-1 will be presented. Next, three structurally similar, tetraruthenium macrocycles will be analysed (Figure 5-2). The influence of the anchor groups on the molecular conductance is of special interest here.

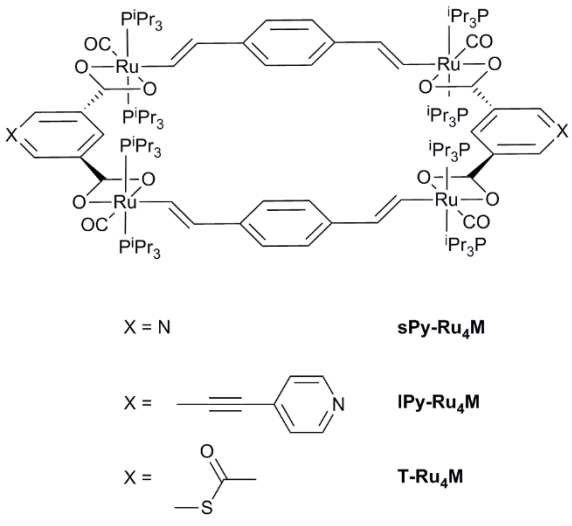


Figure 5-2: Overview of tetraruthenium macrocycle and the variations in anchor groups.

5.2 Conductance of the Au-(T-Ru$_2$M)-Au Junction

5.2.1 T-Ru$_2$M General

Multiple data sets using the *I(s)* technique and break-junction technique were recorded for the ruthenium macrocycle shown in Figure 5-3. Each data set typically consisted of 2000 *I(s)* traces. Experimental conditions used for the experiments on T-Ru$_2$M are described in detail in the Methods Chapter. Although not all the recorded data sets showed peaks in conductance histograms after MPVC analysis, a conductance value could be estimated at around $3 \cdot 10^{-6}$ G_0 ± $2 \cdot 10^{-6}$ G_0. Detailed results for T-Ru$_2$M are shown below by example on the data set recorded at 0.5 V and 8 nA setpoint current. Results for a different data set are shown in the appendix.

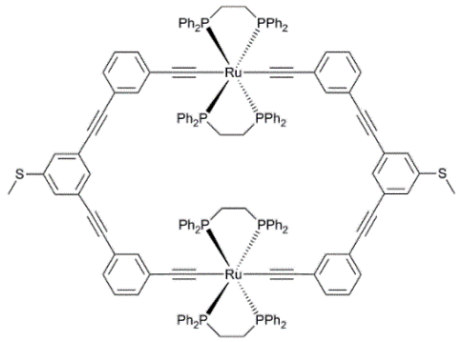


Figure 5-3: Structure of the diruthenium macrocycle with thiomethyl linkers

5.2.2 T-Ru$_2$M at 0.5 V and 8 nA

1950 *I(s)* traces were recorded at 0.5 V and 8 nA. In the all data log 2D conductance histogram, shown in Figure 5-4 **A**, no plateau is visible without data selection. The log 1D conductance histogram, in Figure 5-4 **B**, has a peak just below the setpoint conductance ($2 \cdot 10^{-4}$ G_0). The origin of this "setpoint peak" is currently unknown, but is attributed in some way to the molecular adsorbate, because it does not appear to the same degree in the control experiments (*vide infra*). The absence of plateau or peaks aside from the "setpoint peak" and exponential tail in the conductance histograms is presumably due to the low junction formation probability.

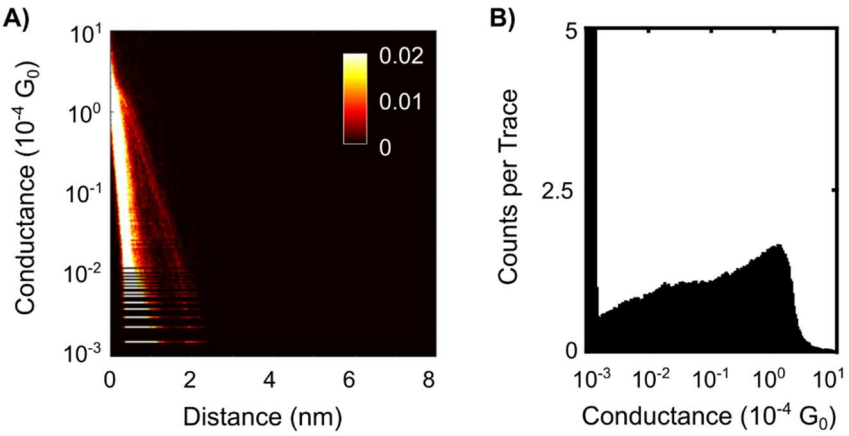


*Figure 5-4: All data log 2D (**A**) and log 1D (**B**) conductance histogram, consisting of all 1950 recorded I(s) traces. Note that data in the log 1D histogram was filtered with a low-pass Bessel filter. Bin width: 0.16 Å and 50 bins per decade.*

MPVC was applied to the logarithm of the data set. The reference vector decay factor, was chosen to be 2.0 $Å^{-1}$, compared to the decay factor of the data set, which was 1.14 $Å^{-1}$ ± 0.65 $Å^{-1}$ (1 σ). A region of interest was defined between 0.5 nm and 2 nm. This was done to eliminate any setpoint effects and to increase the resolution of MPVC by excluding parts of the curve far beyond the molecular length. In the MPVC representation, two distinct clusters become apparent. In the point density plot (Figure 5-5), one cluster has semi-circular shape and extends to the origin of the graph. *I(s)* traces close to the origin are similar to the reference vector. To the right of the semi-circular first cluster, a second area of increased point density is located. This second cluster contains the traces attributed to molecular junction formation.

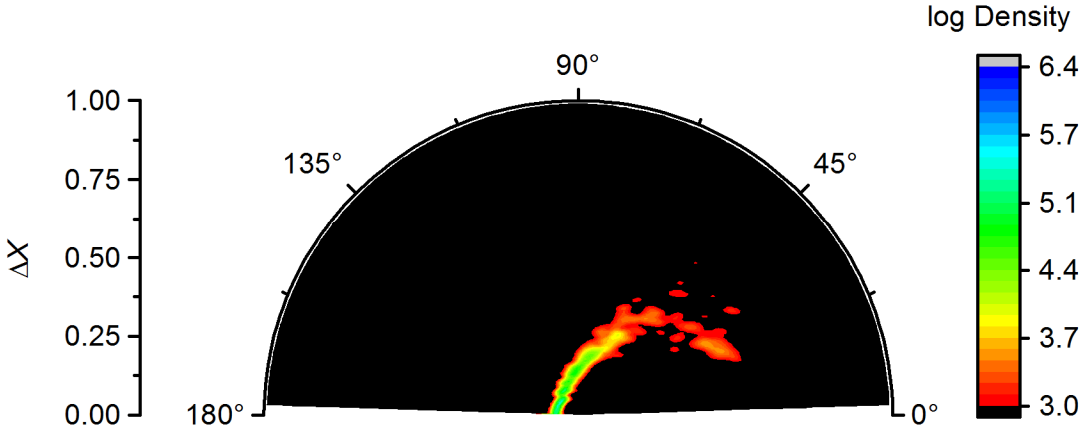


Figure 5-5: Logarithmic point density of the cylinder plot viewed along the h_r-axis. ΔX was normalised for this plot. The point cloud from MPVC analysis was projected onto the ΔX/θ-plane and the point density was calculated with binning increments of 3.6° and 0.02 in angular and radial direction.

Figure 5-6 shows the MPVC representation of the whole data set viewed along different angles. Although the plot in Figure 5-5 suggests that the two clusters are well separable by angle, the cylinder plot in Figure 5-6 indicates that separability might be best over the Hamming distance. (Compare **B** and **C** with **D**) The data points in this plot were assigned to red and blue clusters by fitting the Hamming distance histogram with two Gaussian distributions with 2 σ cut-offs. Cluster assignment based on h_r distribution was chosen over fuzzy clustering here, because of the uneven cluster sizes caused by the low JFP of the experiment (red to blue cluster: 2 % to 98 %) and the good separability along h_r-axis.

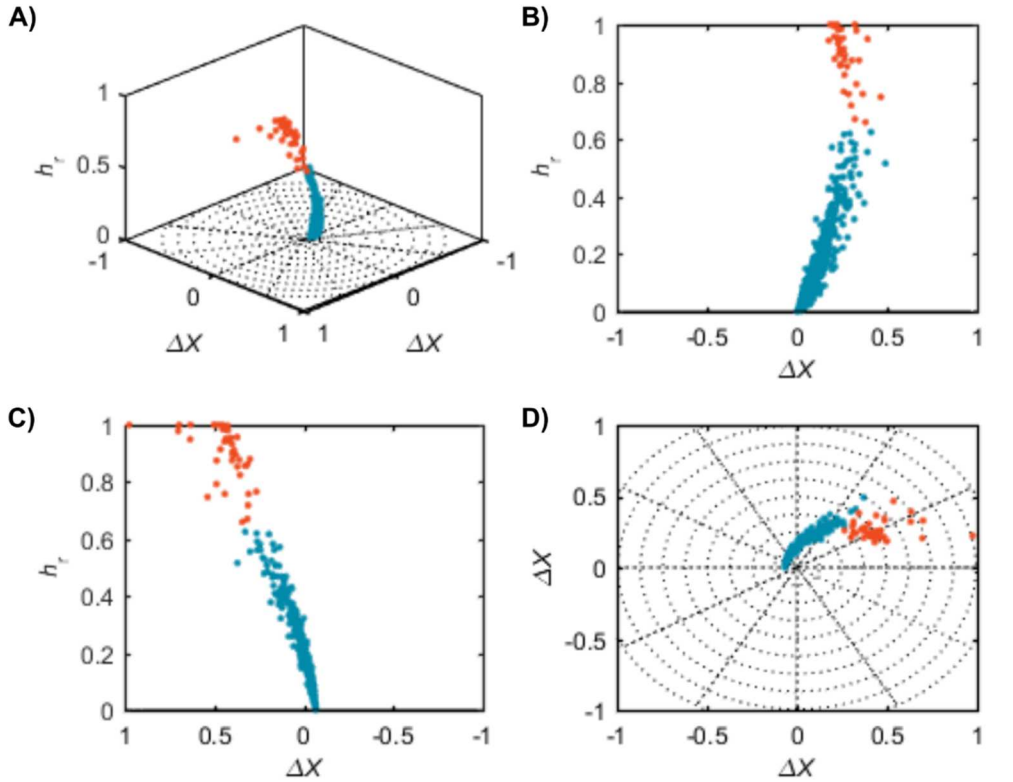


*Figure 5-6: Cylinder plot of the whole data set in the MPVC representation. Viewed diagonally, along 0°-axis, along 90°-axis and along h_r-axis. (**A** to **D**) Members of the plateau-containing cluster are coloured in red, members of the plain exponential cluster are shown in blue.*

The red cluster contains predominantly plateau traces and the blue cluster contains mostly plain exponential graphs without plateau features. Compiling the traces of the red cluster into a logarithmic 1D conductance histograms reveals a peak conductance of $4 \cdot 10^{-6}$ G_0 ± $3 \cdot 10^{-5}$ G_0 (Figure 5-7, **B**). Note that the data in the log 1D histogram had been filtered with a low-pass Bessel filter. The rather large error associated with the conductance peak reflects the low plateau currents and the high noise level associated with the experiments on this macrocycle.

No clear plateau can be seen in the 2D conductance histogram in Figure 5-7 **A**, which is also attributed to the low current and high noise of the plateau features.

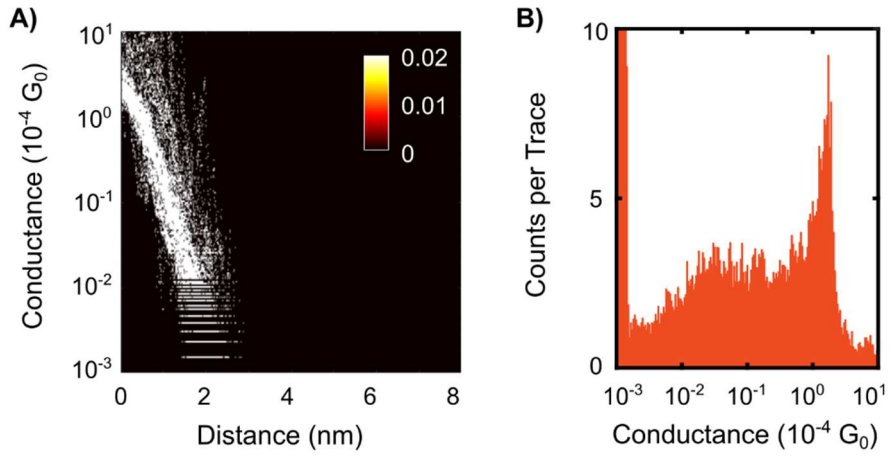


*Figure 5-7: Log 2D (**A**) and log 1D (**B**) conductance histogram of the plateau cluster after MPVC analysis. Note that the data in the log 1D current histogram was filtered with a Bessel low-pass filter to account for the low plateau currents. Bin widths: 0.16 Å and 50 bins per decade*

Further indication that the red cluster contains traces describing molecular junctions rather than stochastic noise events comes from the event distribution in Figure 5-8 **A**. Event distribution meaning here the cumulative counts of the plateau cluster as a function of trace number during the experiment. It is obvious that most traces containing plateaus are not collected uniformly over the course of the experiment, but rather discretely, as one batch of consecutive traces. This indicates that plateau traces are not a stochastic event. A possible interpretation is that the probability to contact a molecule is very low, but when a molecule is connected, the molecule can functionalise the tip and it can be measured multiple times. The discrete event distribution fits well with the general observation that not all experimental data sets seem to contain molecular events.

One could argue that the red cluster and its event distribution might be due to tip contamination of some sort. Although tip contamination cannot be fully ruled out, further evidence that the cluster is indeed caused by the formation of a molecular bridge, is the narrow distribution of the break-off distance. The histogram in Figure 5-8 **B** shows the break-off distance of the red cluster. Fitting with a Gaussian distribution gives a most likely s_b of 1.92 nm ± 0.35 nm (1 σ). s_0 was calculated to be 0.74 nm (0.47 nm to 1.73 nm, 1 σ of the distribution mean). This gives an absolute break-off distance $s_{b,abs}$ of 2.66 nm (2.22 nm to 3.71 nm). The expected total junction length, accounting for the S-Au bond length, is 3.3 nm (2.84 nm sulfur-

sulfur distance). The experimental break-off distance is smaller, but within error of the expected length.

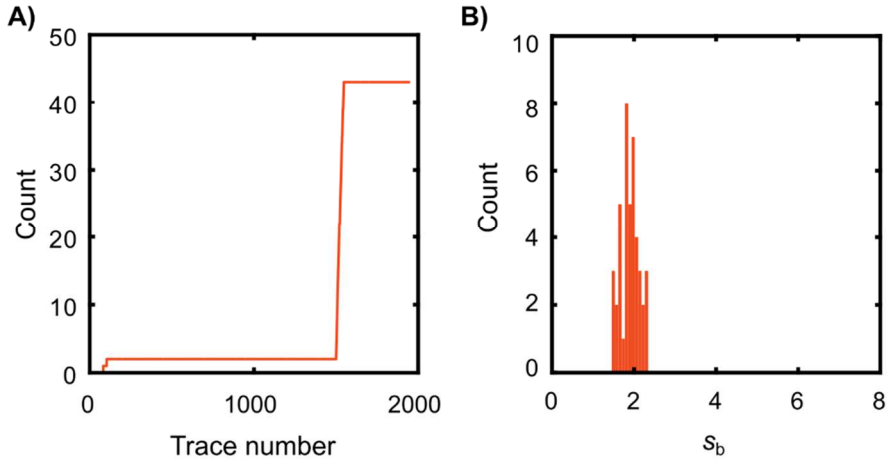


*Figure 5-8: **A**) Event distribution and **B**) break-off distance histogram of the plateau-containing cluster. Bin width: 1 Å.*

Typically traces decay with a decay coefficient of about 1.14 $Å^{-1}$ (determined through fit of the first 100 data points). However, in the presence of a molecule, the decay factor is reduced below 0.5 $Å^{-1}$.

5.2.3 Cross Validation with Plateau Finder

Validation of the data selection comes from a comparison of the MPVC analysis with the plateau finder algorithm. The plateau finder was applied to the natural logarithm of the conductance data. When the blue cluster, *i.e.* the plain exponential cluster, is analysed with the PF the plateau finder marks only 43 of the 1907 *I(s)* traces as plateau-containing. From the perspective of the plateau finder, these 43 traces are "false negative". The plateau finder agrees with the rest of the selection. This means a 98 % agreement with MPVC.

For the red, plateau-containing cluster, the plateau finder agrees with 95 % of the classification results. Or in other words, only 5 % of the traces found by the vector analysis to contain plateaus are "false positives" from the plateau finder point of view (*i.e.* the plateau finder does not find plateaus in those 5 % of traces).

5.2.4 Control Experiment

Further indication that the plateau traces are indeed related to the analyte comes from comparison with data collected for the same substrate with the same STM tip before incubation with the molecule. 1600 traces were recorded at 0.3 V and 20 nA setpoint current. Log 2D and

log 1D conductance histograms of the whole data set are shown in Figure 5-9. Nothing but plain exponential decays are can be seen in the log 2D histogram and no peak is present in the log 1D histogram. Only one cluster is visible in the cylinder plot after running the MPVC algorithm. Upon running the plateau finder algorithm on the data set, only 1 of 1599 traces is marked as plateau-containing. This suggests that the events observed above are due to the molecule.

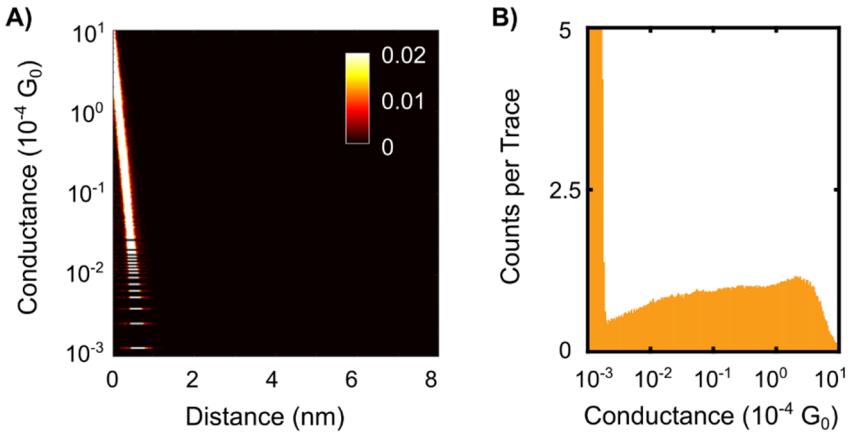


*Figure 5-9: **A**) All data log 2D and **B**) log 1D conductance histogram recorded before incubation of the substrate. 1600 traces were recorded at 0.3 V and 20 nA setpoint current. Note that the data in the log 1D current histogram was filtered with a Bessel low-pass filter for comparison with the previous histograms. Bin widths: 0.16 Å and 50 bins per decade*

5.2.5 Discussion and Summary

The synthesis of macrocyclic molecules is very challenging.[1,2,3] To the author's best knowledge, this is the first time that spectroscopy on this class of molecules is reported on. During the analysis of two data sets, a conductance of $3 \cdot 10^{-6}$ $G_0 \pm 2 \cdot 10^{-6}$ G_0 was found with the *I(s)* technique. The analysis was performed with both a vector-based data selection method and a conventional PF algorithm. Good agreement of data selection was found for both analysis methods. Control experiments on the substrates before analyte assembly confirmed that the observed signal was connected to the macrocycle.

Although a reproducible conductance was found for the molecule, it is currently unclear whether it corresponds to the full through-molecule conductance. The observed break-off distance was smaller than expected, albeit within experimental error. When comparing the conductance of $3 \cdot 10^{-6}$ G_0 with results of a linear system of comparable length, the conductance is in the expected range: A linear system which compares to T-Ru_2M was published by Al-Owaedi *et al.*[4] The ruthenium molecule in this study was slightly shorter (2.65 nm compared to 2.84 nm of T-Ru_2M) and had a higher conductance of $1.8 \cdot 10^{-5}$ G_0 compared to the $3 \cdot 10^{-6}$ G_0 of T-Ru_2M.

However, preliminary results from simulations by Ian Grace in Lancaster suggest that the expected conductance of T-Ru_2M should be on the order of 10^{-9} G_0 and thus below the resolution limit of the experimental setup. In simulations, the low conductance of the macrocycle originated in destructive quantum interference in the multiple meta-substituted benzene rings along the molecule. A coherent charge transport process was assumed in the simulations. However, considering the length of the molecule, incoherent hopping might contribute significantly to the charge transport. A hopping dominated charge transport is unlikely to be affected by destructive interference. This could mean that the experimental result is in a reasonable range.

Another explanation for the high measured conductance compared to simulations could be that the molecule is reproducibly being contacted at points other than the anchor group. Benzene moieties could couple to the electrode.[5] This reduced effective path length could result in the measured conductance. This interpretation matches with the reduced break-off distance, which is observed in the experiment compared to the expected junction length.

To conclude, a conductance value associated with a macrocyclic ruthenium complex has been measured for the first time. While further experiments and simulations are still required to determine whether this conductance is indeed the full-length conductance, this is an intriguing result.

5.3 Conductance of the Au-(sPy-Ru_4M)-Au Junction

A macrocyclic tetraruthenium compound was kindly provided by the Winter group in Konstanz, Germany.[6] Pyridine moieties on either end of the macrocycle were anchor groups, connecting the macrocycle to the macroscopic electrodes of the STM. Crystallography data for the macrocycle suggested a total N to N length of 1.93 nm.[6] Taking the gold-Pyridine distance into account, a total break-off distance of 2.43 nm would be expected in the *I(s)* spectroscopy experiments.

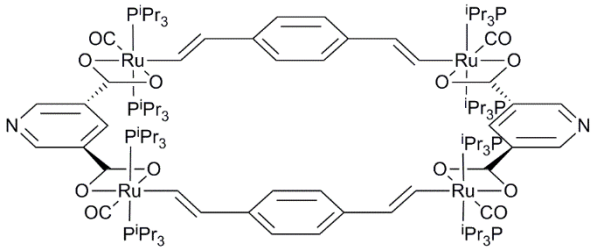


Figure 5-10: Structure of the short tetraruthenium macrocycle (sPy-Ru_4M), which was kindly provided by the group of Rainer Winter in Konstanz.

5.3.1 Single Molecule Conductance at 0.5 V

Spectroscopy data was recorded for a range of different voltages from -0.5 V to 1 V. For a data set with 2000 *I(s)* traces recorded at 0.5 V and a setpoint current of 32.5 nA, the results are shown in detail in this section. Details of the sample preparation are provided in the Methods Chapter.

The logarithmic 2D conductance histogram in Figure 5-11 **A** shows some deviation from the straight line that would be expected for purely exponential decay. However, signal-to-noise ratio is too low without data selection. The all data logarithmic 1D conductance histogram in **B** does not show any obvious peaks.

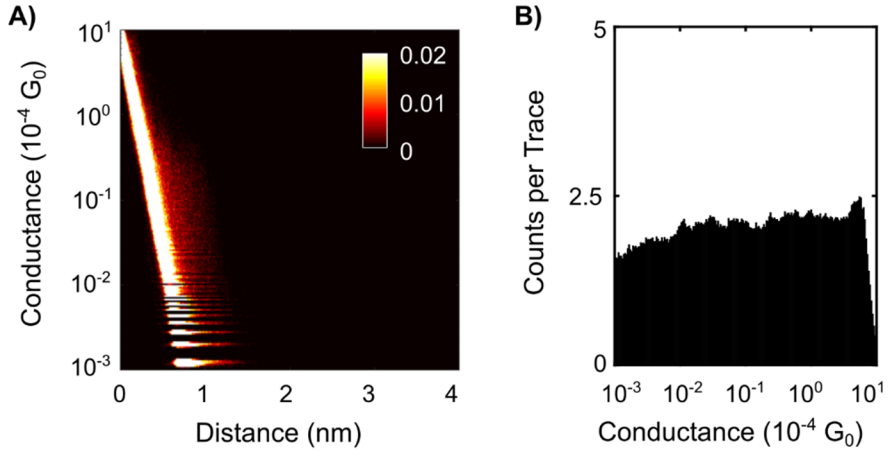


*Figure 5-11: **A)** Log 2D conductance and **B)** log 1D conductance histograms of the whole data set recorded at 0.5 V bias after incubation of the substrate with the analyte. The data in the log 1D histogram was filtered with a low-pass Bessel filter. Bin widths: 0.08 Å and 50 bins per decade*

The low signal-to-noise ratio demands some form of data selection to extract any conductance information. MPVC was hence applied to the logarithm of the data (β = 2 Å^{-1}, roi 0.2 nm to 3 nm). The resulting three-dimensional point cloud was projected onto the *θ/ΔX*-plane and the point density was calculated. The result in Figure 5-12 shows two clusters (white ovals in figure), one above approximately 45° and one below.

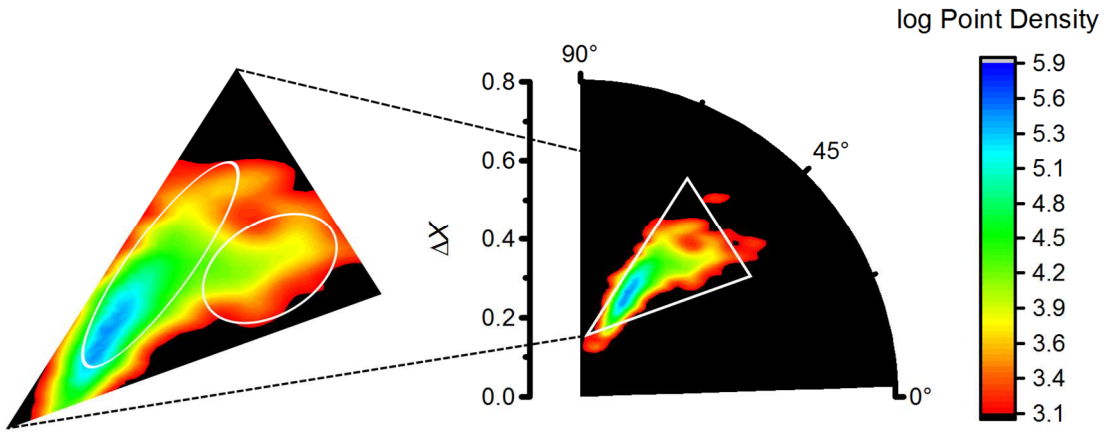


Figure 5-12: Logarithmic point density of the whole data set in the MPVC representation, projected onto the θ/ΔX-plane. Triangle is an enlarged view of the white triangle in the polar plot.

Applying fuzzy c-means clustering using the GK algorithm with two clusters separates the data set as shown in the polar plot in Figure 5-13. The blue cluster contains predominantly plain exponential traces. The red cluster consists mainly of traces that show some deviation from plain exponential decay. 370 traces or 18.5 % are members of the red cluster.

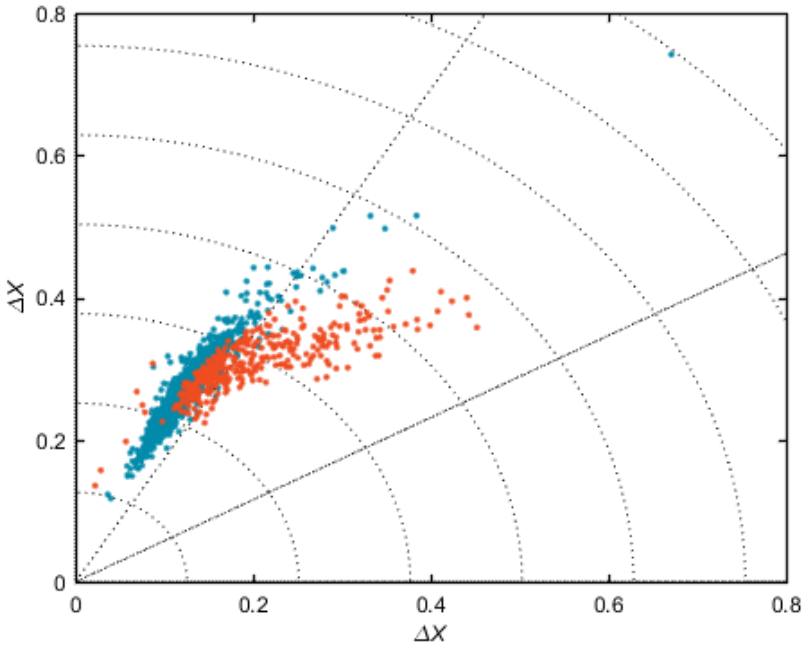


Figure 5-13: Cylinder plot of the data set in the MPVC representation containing the whole data after cluster assignment with FCM GK. The plot is viewed along the h_r-axis

Plotting only the red cluster in histograms shows a greatly improved signal to noise ratio, when compared to the all data histograms (see Figure 5-14). In **A,** a clear deviation from the exponential decay is visible. In **B**, a broad conductance peak becomes apparent. Fitting this

peak with a Gaussian distribution gives a most probable conductance of $(9 \pm 3)\cdot10^{-6}$ G_0. The error was estimated from the standard deviation of six experiments described below, the fit is provided in the appendix.

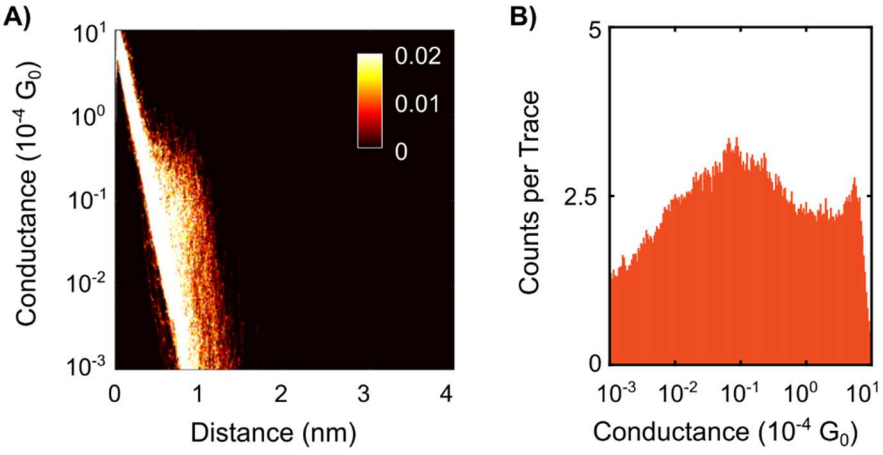


*Figure 5-14: Logarithmic **A)** 2D and **B)** 1D conductance histogram of the red cluster. This cluster contained predominantly traces with plateaus as determined by MPVC FCM with the GK algorithm. Data for both histograms were filtered with a low-pass Bessel filter. Bin widths: 0.08 Å and 50 bins per decade*

A representative selection of members of the red clusters, Figure 5-15, shows clear deviations from exponential decay around the peak conductance value (indicated by dashed line). The observed plateau features are remarkably noisy. It is noteworthy that such noise is not observed for the same current amplitudes in the absence of plateau features. The high plateau noise appears to be the reason for the absence of a clear plateau and for the broadness of the peak in the 2D and 1D histograms in Figure 5-14.

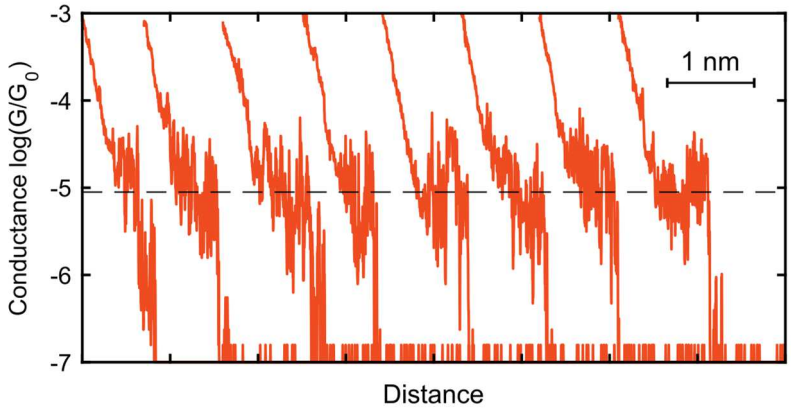


Figure 5-15: Representative sample traces of the red, plateau-containing cluster. Traces are offset for clarity. The dashed line represents the conductance calculated from the log 1D conductance histogram in the previous figure.

When investigating the red cluster in more detail, a steady incline is observed in the cumulative counts (Figure 5-16). The incline is qualitatively different to the rather discrete "jumps" observed for the biruthenium macrocycle that was discussed in the previous section.

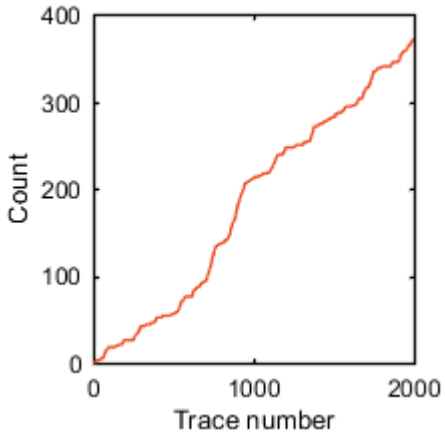


Figure 5-16: Cumulative count of the plateau cluster as a function of trace number during the experiment.

The break-off distance of the junction can serve as an indication of the nature of the junction. As a reminder, the experimentally observed break-off distance, s_b, was calculated as the distance of the first current value below 3 σ of the noise level. The experimentally observed distance in the STM is a relative distance to the initial setpoint distance s_0, meaning the distance between STM tip and substrate, when the setpoint current was measured. This is visualised in the schematic of Figure 5-17.

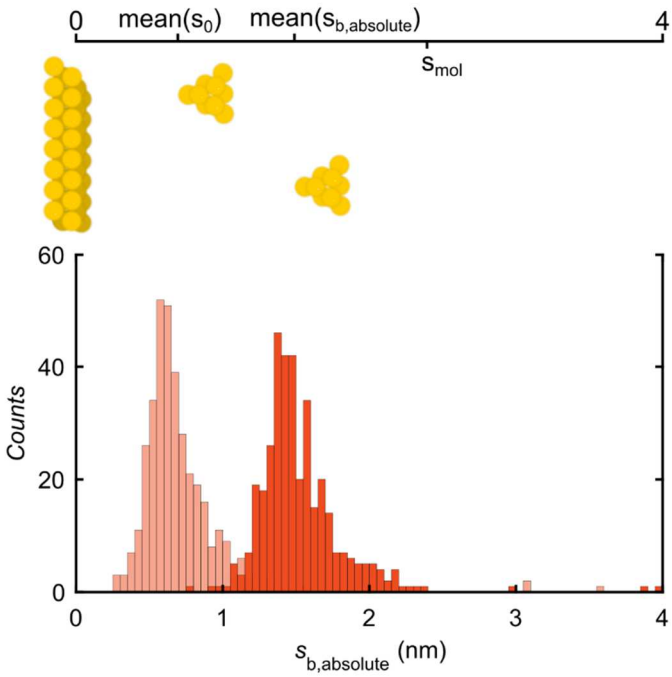


Figure 5-17: Histogram of the break-off distance on an absolute scale with zero being the substrate surface. The light red histogram is the initial distance at the setpoint current. The schematic on top of the histogram illustrates the relative and absolute scale. Bin width: 0.5 Å.

Below the schematic, the light-red histogram is the calculated experimental initial distance between tip and substrate for the red cluster. The red histogram is the absolute break-off distance histogram. "Absolute" meaning here that the distance is relative to the substrate rather than the initial distance.

When comparing the mean of the red absolute break-off distance distribution with the expected molecular length, s_{mol} (1.46 nm ± 0.23 nm to 2.43 nm respectively), it is obvious that the break-off distance is considerably shorter than the distance that would be expected for a maximally extended molecular junction.

5.3.2 Plateau Finder Analysis

Plateau data selection can be performed by fuzzy clustering, as shown above. Alternatively, traces containing plateaus can be selected with the plateau finder algorithm. Once the plateau data has been selected, the data can still be transformed into MPVC representation to compare the outcomes. This has been done in Figure 5-18.

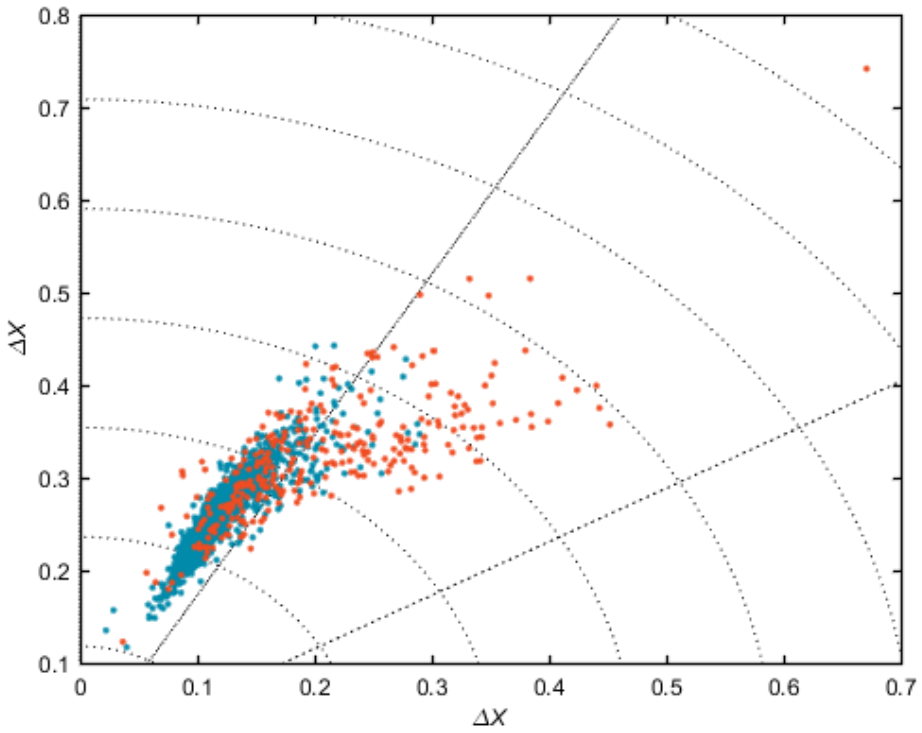


Figure 5-18: Cylinder plot of the whole data set. Clustering was performed by plateau finder algorithm. The polar plot is viewed along the h_r-axis. Red traces contain traces with plateaus, blue traces plain exponentially decaying I(s) traces.

370 traces have been found to contain plateaus by MPVC cluster analysis. Of those the plateau finder agrees with 168. 202 traces, on the other hand, are not found by the plateau finder. These 202 traces are "false positives" from the perspective of the plateau finder algorithm. Among the false positives are traces with very noisy plateaus, traces with plateaus that are very short and slanted as well as plain exponential traces that might have ended up in the plateau cluster by imperfect cluster assignment.

In contrast, of the 1630 traces that were selected as being plain exponential decays by cluster analysis, the plateau finder marked 130 as plateau-containing traces. The traces are "false negatives" by plateau finder standards. By visual inspection, false negatives are noisy traces, or plateaus at low conductance values. Also, some plateaus that might have slipped through cluster analysis by imperfect cluster assignment can be found.

Despite some disagreement between the plateau finder and MPVC cluster data selection, quantitatively similar results are obtained for both methods. The corresponding histograms are provided in the appendix.

5.3.3 Control Experiment

A control experiment was performed on the same substrate before analyte assembly. For the control experiment, the substrate was only incubated with pure DCM without any analyte. 2000 *I(s)* traces were recorded at 0.3 V. The all-data log 2D histogram in Figure 5-19 **A** only shows a straight line corresponding to a plain exponential decay. No deviation from plain exponential behaviour of the sort seen in Figure 5-11 **A** is visible. A little kink at conductance just above 10^{-7} G_0 is attributed to digitizer quantisation at currents close to the scanner resolution limit. Accordingly, in the log 1D conductance histogram in **B**, a sharp peak can be seen. Apart from this artefact, no signal comparable to the one observed after analyte assembly is observed.

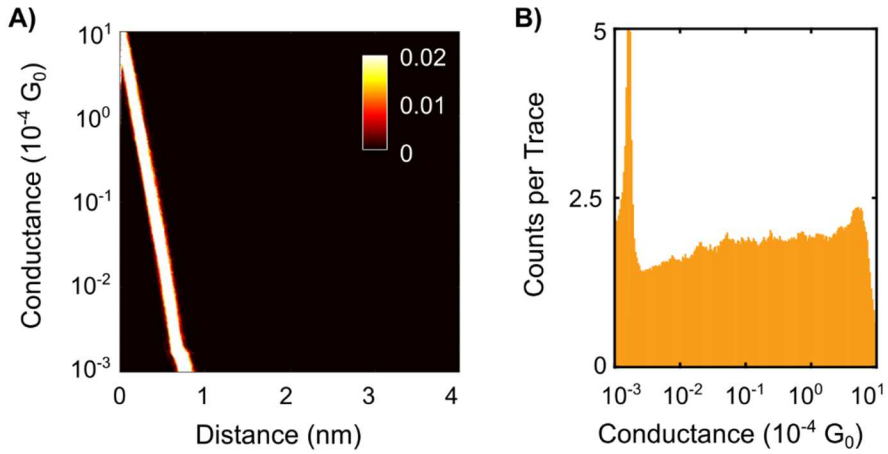


*Figure 5-19: **A)** log 2D and **B)** log 1D conductance histogram of the control experiment. Both histograms have been filtered with a low-pass Bessel filter. Bin widths: 0.08 Å and 50 bins per decade*

When transforming the data set of the control experiment into the MPVC representation, only one oval cluster can be seen (Figure 5-20). The plateau finder algorithm finds only 18 traces with plateaus in the data set (JFP 0.9 %). This is a significantly lower JFP than what was observed after analyte assembly (JFP 15 %, MPVC 19 %).

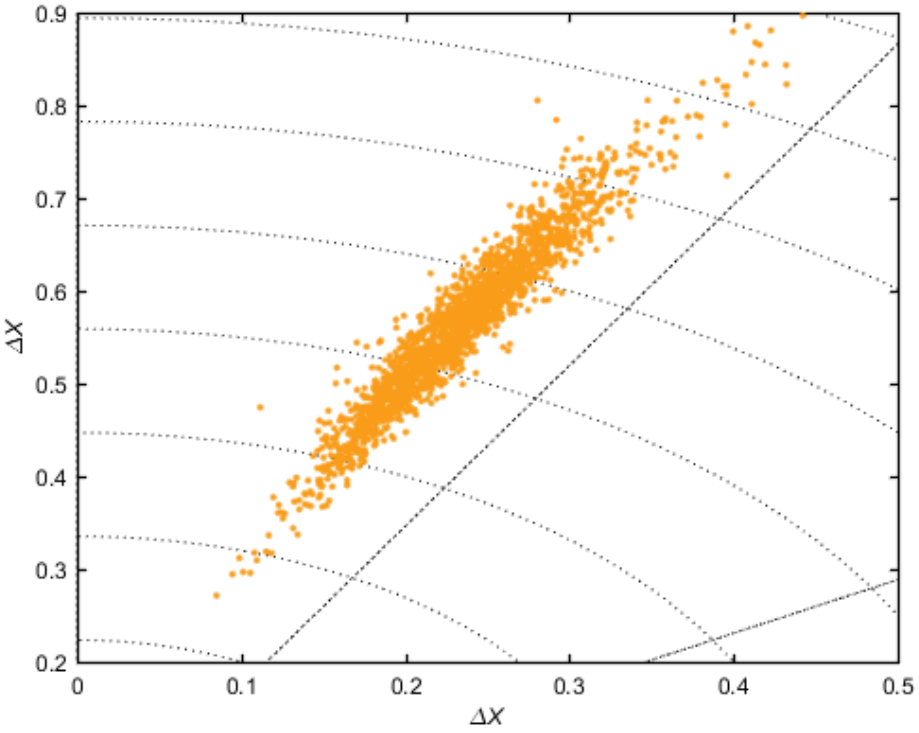


Figure 5-20: Cylinder plot of the MPVC representation of the whole data set recorded in the control experiment. The plot is viewed along the h_r-axis.

5.3.4 Conductance Series

Further experiments have been performed on the substrate after analyte assembly. Not all of them gave conclusive results, but most of them did show some conductance peak in the log 1D conductance histograms after data sorting. The results are summarised in Figure 5-21. Panel **A** displays all log 1D conductance histograms that showed a peak after data selection. Even though the peaks are often not well-defined, there seems to be a trend emerging. In **B**, the peak conductance is plotted against the applied bias voltage. Low absolute bias voltages seem to result in low peak conductance.

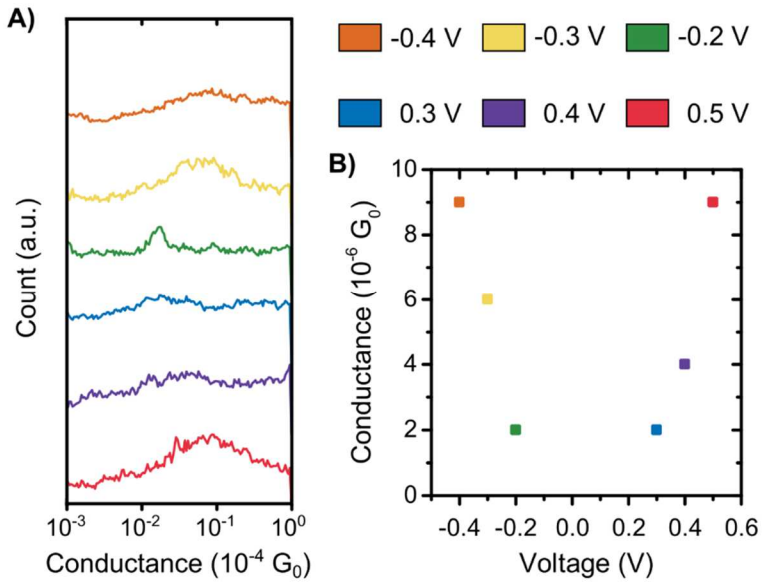


*Figure 5-21: **A)** Logarithmic 1D conductance histograms for data recorded at bias voltages from 0.5 V to -0.4 V. Histograms are offset for clarity and filtered with a low-pass Bessel filter. **B)** Peak conductance values of log 1D conductance histograms. Bin width: 50 bins per decade*

5.3.5 Discussion and Summary

Spectroscopy experiments were performed on a short tetraruthenium macrocycle. It was shown by example on a data set recorded at 0.5 V that by applying suitable MPVC analysis, traces with and without plateaus could be separated in good quantitative agreement with a conventional plateau finder algorithm. When analysing the plateau-containing traces in more detail, it was found that plateau traces were recorded continuously throughout the experiment with a high JFP of 18.5 %. This could suggest that homogenous, low coverage or monolayer coverage was obtained during incubation. A control experiment without analyte indicated that the observed signal was connected to the analyte.

When analysing the break-off distance of the data set recorded at 0.5 V, it became apparent that the measured s_b, stayed below the expected break-off distance. Only the tail end of the break-off distribution reached the expected length. This could indicate that the molecule was connected at an angle. From the s_b distribution mean, an average angle of 53° relative to the normal of the substrate could lead to the observed break-off distances.[7] Steric effects could have prevented a full extension of the junction. A different explanation of the low break-off distance could be that the molecule was contacted through moieties other than the anchor groups with a higher probability. The iPr_3P ligands or benzene moieties could be possible coupling sites.

In six spectroscopy experiments, a conductance of $(5 \pm 3) \cdot 10^{-6}$ G_0 was calculated. The junction conductance seemed to decrease towards low absolute bias voltages.

5.4 Conductance of the Au-(lPy-Ru$_4$M)-Au Junction

5.4.1 lPy-Ru$_4$M General

The ruthenium macrocycle with long pyridine linkers (see structure in Figure 5-22) was synthesised by the Winter group in Konstanz, Germany with the intention to increase the junction formation probability and reproducibility compared to the macrocycle with short pyridine linkers. It was hypothesised, at the time, that junction formation in STM experiments was sterically hindered by the bulky ruthenium cage.

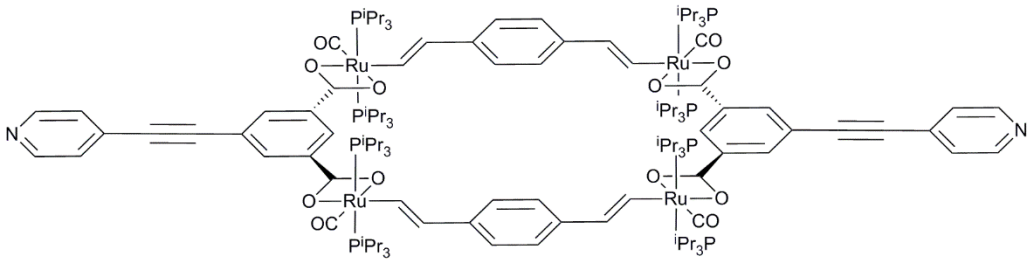


Figure 5-22:Structure of the tetraruthenium macrocycle with extended pyridine linkers. The molecule was kindly provided by the group of Rainer Winter in Konstanz, Germany.

Two separate spectroscopy experiments were performed with three different setpoint currents. The resulting *I(s)* traces revealed a complex conductance landscape, as one would expect from a molecule as complex as this one. Two conductance classes were identified.

5.4.2 *I(s)* Spectroscopy

The findings for this molecule will be discussed by example of an experiment performed at 0.3 V and 20 nA with a total *I(s)* trace count of 5000. The experiments performed under the same or higher incubation concentration showed qualitatively similar results.

When displaying all recorded traces in histograms, as done in Figure 5-23, several features become apparent. In the log 2D conductance histogram in panel **A**, a faint plateau is visible between 10^{-1} G_0 and 10^{0} G_0. It is also noted that well below 10^{-6} G_0, conductance values are quantised. In the log 1D conductance histogram in **B**, three peaks are visible. A sharp setpoint peak, as described above, can be seen. A second peak lies between 10^{-5} G_0 and 10^{-4} G_0 is attributed to the analyte and a third peak is close to 10^{-7} G_0. This faint peak at low conductance is likely an artefact caused by the quantised nature of current output by the analogue-digital converter.[8]

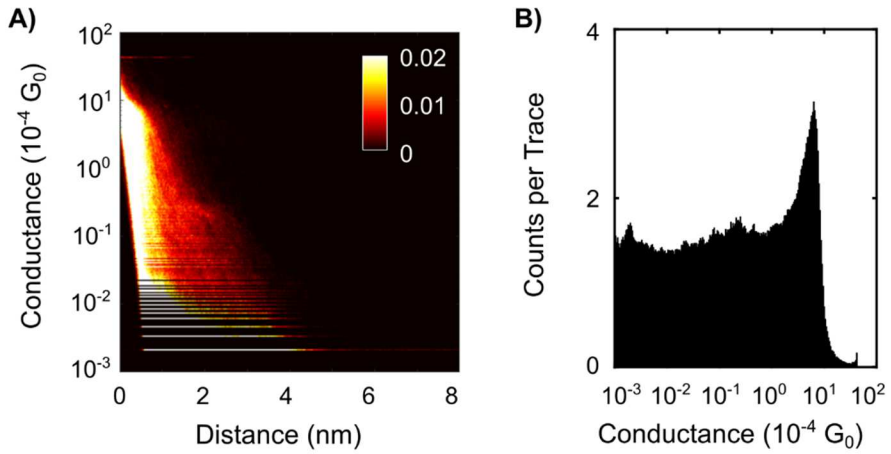


*Figure 5-23: **A**) Log 2D and **B**) log 1D conductance histogram of the whole data set without any data selection. The log 2D conductance histogram is unfiltered to visualise the quantised nature of the current on this scale, the log 1D conductance histogram is filtered with a low-pass Bessel filter. Bin widths: 0.16 Å and 50 bins per decade*

To shed light onto the conductance landscape and to resolve the nature of these conductance peaks, MPVC was applied to the natural logarithm of the data. For MPVC, a reference vector decaying with a decay coefficient of 2 $Å^{-1}$ was used. A region of interest was defined between 0.2 nm and 4 nm. The MPVC reduces the dimensionality of the data set to a 3-dimensional point cloud. The diagonal view of the point cloud in Figure 5-24 shows two distinct clusters. The data points were coloured according to their cluster membership. Clustering was performed with FCM using the GK algorithm with two clusters. The red cluster contained predominantly plateau traces (15 %), the blue one mainly plain exponentially decaying traces (85 %).

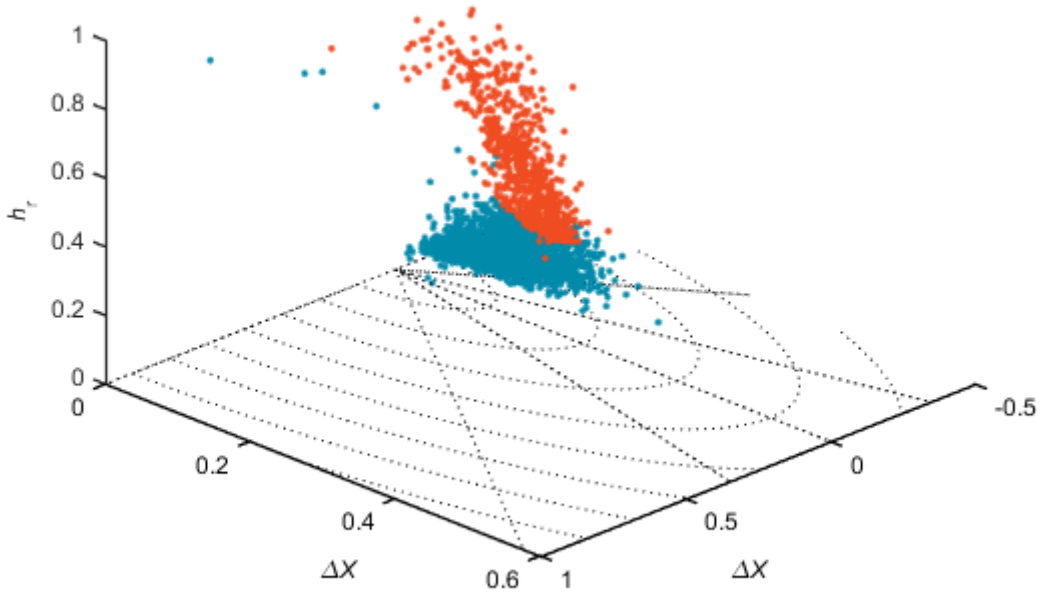


Figure 5-24: Cylinder plot of the whole data set in the MPVC representation. Colour assignment was performed per cluster membership. Cluster membership was determined by FCM with the GK algorithm. All 5000 traces of the data set are displayed in the graph.

When plotting the conductance histograms of the red cluster, the plateau feature is more dominant (compare Figure 5-25 **A** to Figure 5-23 **A**). Also, the corresponding peak in the log 1D histogram (**B**) is enhanced. The conductance was determined to be $2 \cdot 10^{-5}$ G_0. The error interval from HWFM of Gaussian fit was $(0.5\text{-}8) \cdot 10^{-5}$ G_0, the fit is provided in the appendix. The setpoint peak remains unchanged.

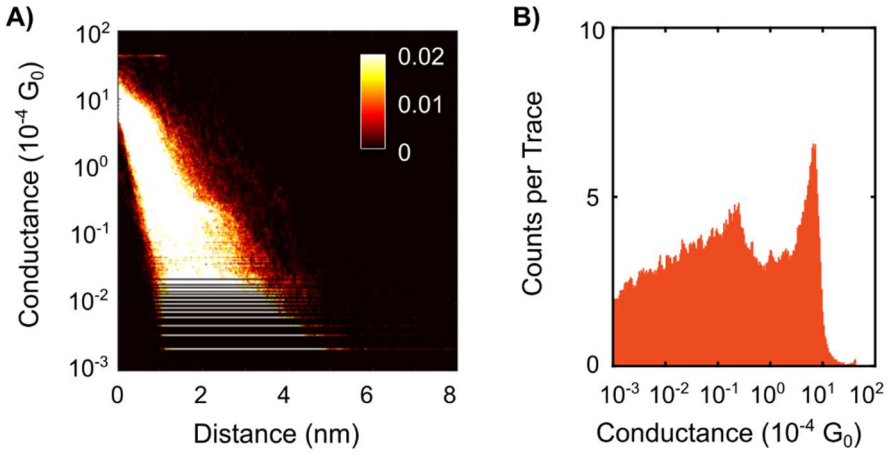


*Figure 5-25: **A**) 2D and **B**) 1D conductance histograms of the red cluster. This cluster contains predominantly traces with plateaus, as determined by MPVC. Data in **B**) was filtered with a low-pass Bessel filter. Bin widths: 0.16 Å and 50 bins per decade*

The plot of the cumulative counts of the red cluster, Figure 5-26 **A**, shows a shape deviating from a uniform incline. This event distribution suggests that molecular bridges only form sporadically, but if they form they can be measured repeatedly.

The molecular length was determined by the Winter group to be 3.37 nm (N to N). Together with twice the Au-N bond length of 0.25 nm[7], this gives a total expected junction extension of 3.9 nm. The initial distance was estimated to be 0.91 nm ± 0.35 nm. An observed break-off distance of 3.0 nm would therefore correspond to a junction extended to the full molecular length. The break-off distance histogram is shown in Figure 5-26 **B**. The graph does not resemble a monomodal Gaussian. A skewness of 0.16 was calculated for the distribution (mean 2.5 nm, median 2.4 nm STDEV 0.8 nm). A bimodal Gaussian distribution on the other hand describes the data very well. Distribution means were calculated to be 1.8 nm ± 0.3 nm and 2.7 nm ± 0.9 nm. The expected molecular length of 3.0 nm is well within the range of the larger break-off distance distribution 2.7 nm ± 0.9 nm

As a reason for this bimodal shape, an imperfect cluster assignment (*i.e.* the errorneous presence of plain exponential traces in the red cluster) could be ruled out. This was done by visual inspection of the positions of short s_b traces in the cylinder plot of Figure 5-24. The origin of the skewed distribution shape was suspected to be the low current and high fluctuation of the plateau currents, which lead to an underestimation of the break-off distance of a part of the traces (*vide infra*).

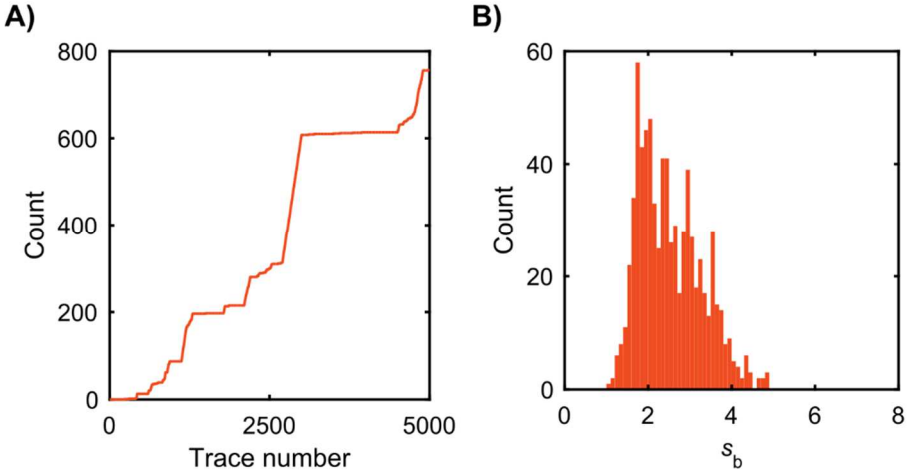


*Figure 5-26: **A**) Cumulative counts of the red cluster as a function of the trace number of the experiment. **B**) Break-off distance histogram of the red cluster. Bin width: 1 Å.*

Manual inspection of the members of the red clusters indicated that there were two distinct conductance classes present. Examples of those both classes are shown in Figure 5-27.

The red traces represent members from the high-conductance class. The dashed line is the most probable conductance, as determined by fitting the histogram in Figure 5-25 **B**. Examples of a low-conductance sub-class are displayed in green.

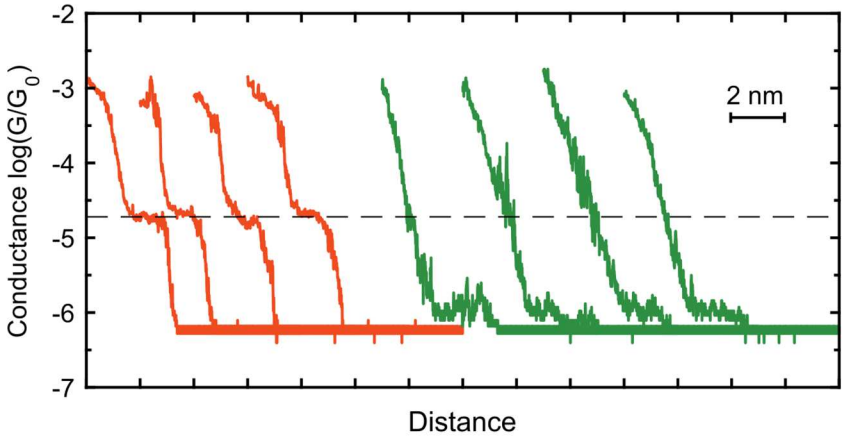


Figure 5-27: Representative examples of high (left four traces in red) and low (green traces) conductance traces. The dashed line marks the conductance of the high conductance class ($2 \cdot 10^{-5}$ G_0). Traces were off-set for clarity.

The position of the manually selected traces displayed in Figure 5-27 can be shown in the MPVC representation. This is done in Figure 5-28, where the "x"s denote traces selected in the previous figure. The red, plateau-containing point cluster is shown in semi-transparent red for improved visibility. The four manually selected high and low conductance traces lay on different sides of the light-red plateau point cloud, they do not form separate clusters with the choice of MPVC parameters. Accordingly, further sub-clustering with GK does not separate traces into low and high conductance traces.

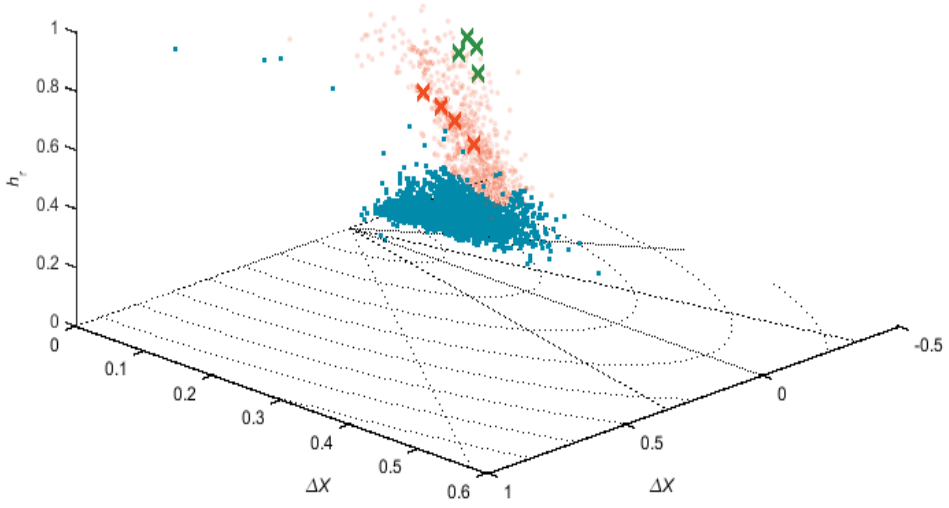


Figure 5-28: Cylinder plot of the whole data set in the MPVC representation. Coloured in blue are the plain exponential traces, in semi-transparent red are plateau traces. Red "x" are the high conductance traces and green "x" are the low conductance traces selected in the previous figure.

To investigate further, if high and low conductance traces are two distinct classes, the red plateau cluster was hand-sorted into high and low conductance traces. Figure 5-29 displays the log 2D (**A** and **B**), the log 1D (**C** and **D**) and the break-off distance (**E** and **F**) histograms for those hand-sorted groups. The first column contains the high conductance traces and the second column, the low conductance traces. The 2D histograms show well-defined plateaus in both conductance groups (**A** and **B**). Their corresponding log 1D histograms (**C** and **D**) have clear peaks at $0.2 \cdot 10^{-4}$ G_0 (0.03 to $0.7 \cdot 10^{-4}$ G_0) and at $1 \cdot 10^{-6}$ G_0 (<1 to $5 \cdot 10^{-6}$ G_0). Error intervals were estimated from the FWHM of Gaussian fits (provided in appendix). The histograms are made up from 231 and 109 traces, respectively. (JFP 5 % and 2 %)

The break-off distance histogram of the "high conductance class" had a mean of 2.64 nm ± 1.01 nm and a low skewness of 0.02. The "low conductance class" had a mean of 2.32 nm ± 1.40 nm with a higher skewness of 0.10. The higher skewness of the "low conductance class" is attributed to the low plateau current of these traces: The break-off distance was calculated as the distance of the first current value below 3 σ of the noise level. Because the plateau current was so low in these traces, this 3 σ threshold was easily reached with fluctuation of the plateau current before the junction ruptured. The break-off distance is thus assumed to be underestimated leading to this skewed shape. By filtering the low G data before calculating the break-off distance histogram, this effect is suppressed. The filtered break-off distances had a mean(s_b) of 3.0 nm ± 1.0 nm.

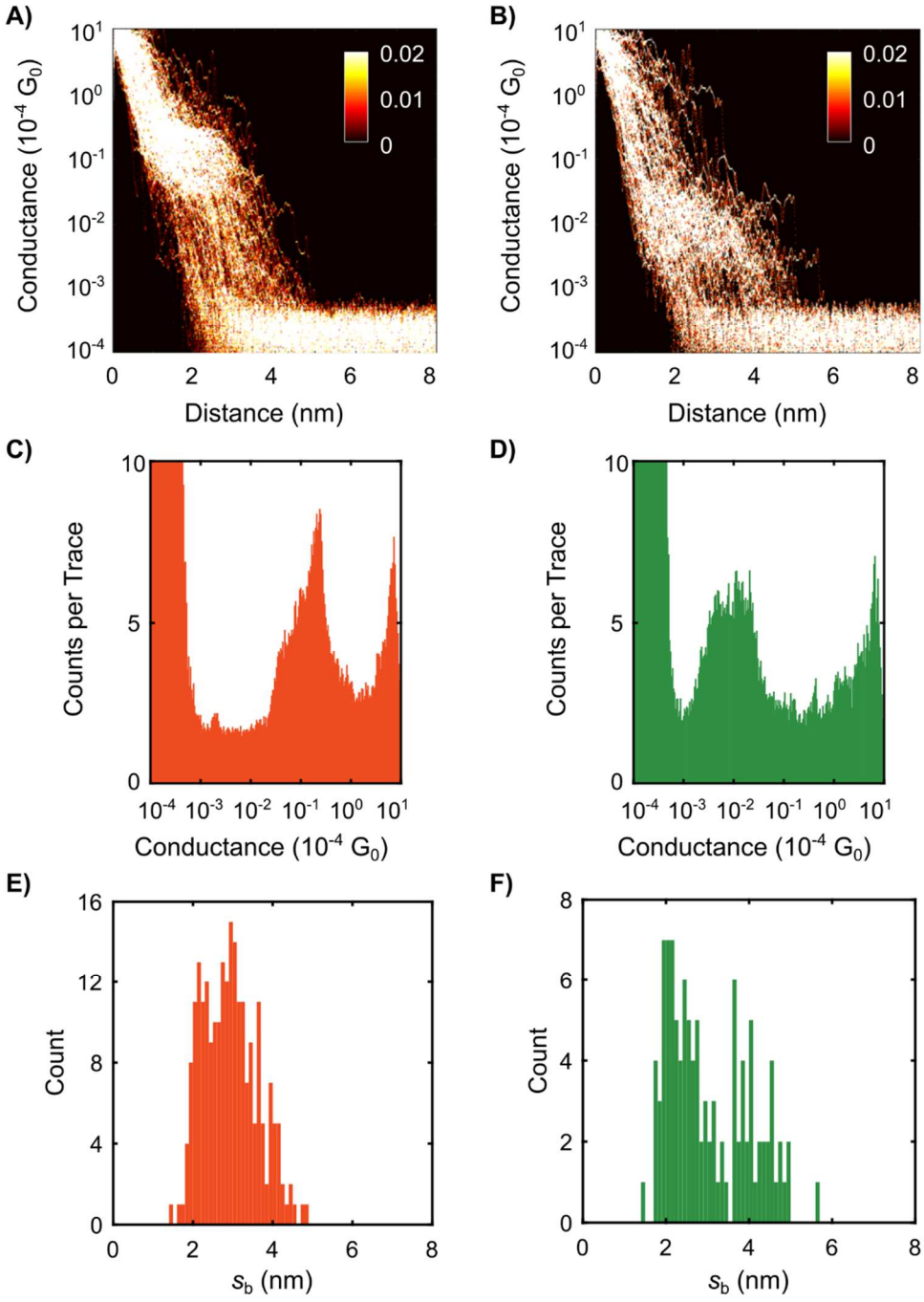


*Figure 5-29: Log 2D histogram of the presumed high (**A**) and low (**B**) conductance group. Log 1D conductance histogram of the high (**C**) and low (**D**) conductance group as well as their break-off distance histograms (**E** and **F**). Bin width: 1 Å. Data was filtered with a low-pass Bessel filter.*

5.4.3 Control Experiment

To confirm that the effects described above are due to incubation with the analyte, control experiments were conducted. This means that *I(s)* spectroscopy was performed on the substrate after incubation in pure DCM.

1997 traces were recorded at 0.3 V. The corresponding log 2D and log 1D conductance histogram of the whole data set is shown in Figure 5-30. About 2 % of traces in the control experiments contained plateaus (as determined by plateau finder). Accordingly, a short plateau-like feature can be seen in **A** and a faint peak in **B**. This might have been due to contamination of tip and/or substrate. The contamination signal was present before and after incubation with bare DCM, even though it was not noticed due to the small signal-to-noise ratio.

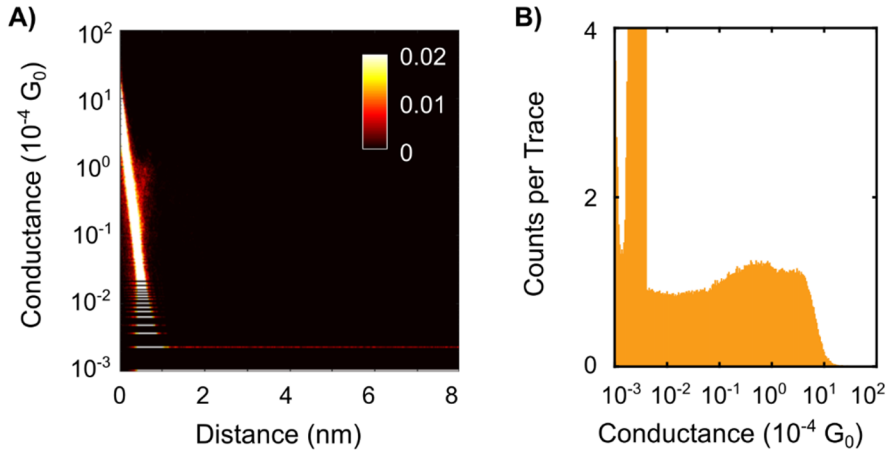


*Figure 5-30: Log 2D (**A**) and log 1D (**B**) conductance histograms of the control experiment. The Au(111) substrate was incubated in pure DCM without analyte before the experiment. Note that data in **B** was filtered with a low-pass Bessel filter to be comparable to the previous graphs.*

When plotting the event distribution of the contamination signal, it is noteworthy that plateau events occurred all throughout the experiment in a rather stochastic fashion (Figure 5-31 **A**).

The plateaus observed had a well-defined conductance about a factor of three higher than the high conductance group, measured for the molecule. The contamination signal had a sharp break-off distance distribution with a distribution mean at 0.76 nm ± 0.12 nm (Figure 5-31 **B**) and is thus considerably shorter than the events observed after incubation. The significant difference in break-off distance between contamination and molecule makes the two well distinguishable. The measurement is thus believed to be valid despite the present contamination.

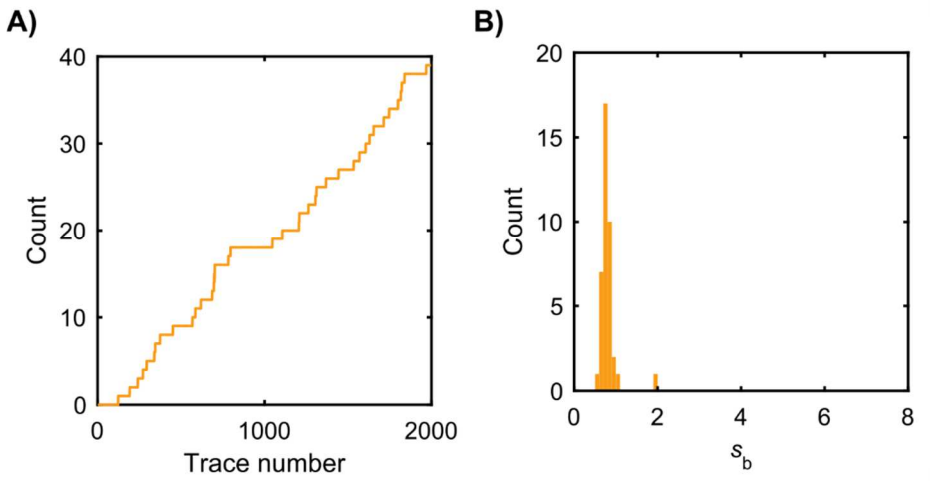


*Figure 5-31:**A**) Event distribution of the control experiment. Plateaus were selected by the PF algorithm. **B**) Break-off histogram of plateaus in the control experiment. Bin width: 1 Å.*

5.4.4 lPy-Ru$_4$M Discussion and Summary

A macrocyclic tetraruthenium compound with long pyridine linkers was synthesised by the Winter group to improve junction formation characteristics compared to the same macrocycle with shorter linkers. Spectroscopy experiments were performed on the molecule. Utilizing MPVC, plain exponential traces were separated from traces with plateaus. A junction formation probability of 15 % was observed.

By further hand sorting the plateau traces, a high and a low conductance class were found with conductance values of $0.2 \cdot 10^{-4}$ G_0 (0.03 to $0.7 \cdot 10^{-4}$ G_0) and $1 \cdot 10^{-6}$ G_0 (<1 to $5 \cdot 10^{-6}$ G_0), respectively. Control experiments without analyte incubation confirmed that the observed plateaus were connected to the molecule.

The high conductance class had a break-off distance of 2.6 nm ± 1.0 nm. The low conductance class had a skewed break-off distance histogram, which was attributed to the low plateau currents. A mean break-off distance of 3.0 nm ± 1.0 nm was calculated, in excellent agreement with the expected junction extension of 3.0 nm.

It is therefore concluded that the single molecular conductance of the Au-(lPy-Ru$_4$M)-Au was successfully measured. The low conductance group corresponded to the full-length conductance. The high-conductance class was attributed to a junction of the molecule, coupling to a site on the molecule other than the anchor group.

5.5 T-Ru$_4$M

5.5.1 T-Ru$_4$M General

In total, eight spectroscopy experiments were performed on the ruthenium macrocycle shown in Figure 5-32. During those experiments 35 voltages were probed under different experimental conditions. Details can be found in the Methods chapter. For each voltage, typically 2000 *I(s)* traces were recorded, adding up to approximately 70000 traces. This large number was necessary because the junction formation probability varied significantly, and the observed conductance landscape was very complex. After analysis of all experiments, two conductance groups at $(33 \pm 8)\cdot10^{-6}$ G_0 and $(3 \pm 1)\cdot10^{-6}$ G_0 were found reproducibly. The analysis is shown in detail below, by example of an experiment performed at a bias of 0.3 V and a setpoint current of 20 nA under argon. Afterwards, results for experiments performed under ambient conditions will be shown. Conductance results are summarised in Table 5-1.

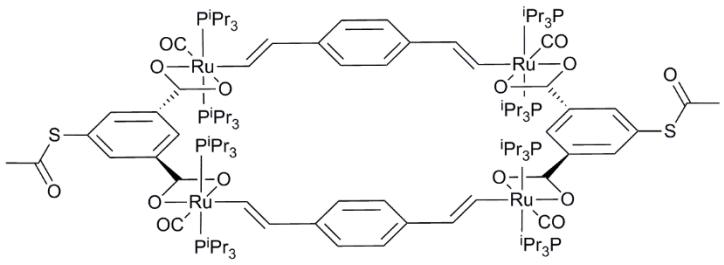


Figure 5-32:Structure of T-Ru$_4$M (thioacetyl tetraruthenium macrocycle). A sample of the molecule was kindly provided by the group of Rainer Winter in Konstanz, Germany. Structure was used with permission from P. Anders.

*Table 5-1: Summary of experimental conditions, conductance results and their junction formation probabilities for T-Ru$_4$M. *Experiment was performed under argon atmosphere. **JFP does not distinguish here between low and high conductance group found in the same data set as they were not clearly separable.*

Bias (V)	I_0 (nA)	Conductance (G_0)	JFP (%)
0.3*	20	$4\cdot10^{-6}$	12**
		$25\cdot10^{-6}$	
0.2	13	$2\cdot10^{-6}$	3
-0.4	26	$4\cdot10^{-6}$	6
0.04	20	$35\cdot10^{-6}$	25
0.05	20	$40\cdot10^{-6}$	14

5.5.2 T-Ru$_4$M Electrochemistry

Electrochemical experiments were performed on the Au(111) substrate in the STM cell after spectroscopy experiments. Voltammograms were measured with $[FeCp_2^*]^+/[FeCp_2^*]$ as

internal reference and are reported versus $[FeCp_2]^+/[FeCp_2]$ (495 mV versus $[FeCp_2^*]^+/[FeCp_2^*]$).[9] Details are provided in the Methods Chapter.

Panel **A** of Figure 5-33 shows cyclic voltammograms of the substrate used for the spectroscopy experiments under argon (see below). The scan rate was varied between 25 mV/s and 5000 mV/s (inner to outer traces).

Peak splitting was measured to be 248 mV compared to the 250 mV for the 1,4-divinylphenylene diruthenium subunit as reported by Scheerer *et al.*[6] According to their findings, in the DCM/$TBAPF_6$ supporting electrolyte all tetraruthenium macrocycles are oxidised in two consecutive and chemically reversible but slightly broadened waves, indicating that the first and the second oxidations of each individual 1,4-divinylphenylene diruthenium subunit are not resolved. The half-wave potentials $E_{1/2}$ versus $[FeCp_2]^+/[FeCp_2]$ for the two, two-electron redox transitions were -135 mV and 112 mV. This compares well to -75 mV and 175 mV for the diruthenium precursor.

The electrochemical surface area of the substrate was estimated from the charge of the gold oxide reduction peak in 0.5 M H_2SO_4. The charge Q_{AuRed} was measured at 127.4 µC ± 15.0 µC. Gold oxide reduction is a two-electron process and the electrochemical surface area of the substrate was thus calculated to be 0.29 cm^2 ± 0.03 cm^2. 222 µC/cm^2 was assumed as surface charge per area of the Au(111) crystal plane.[10]

The charge for the two-electron oxidations of the tetraruthenium macrocycle was extrapolated to zero scan rate and calculated to be 2.39 µC ± 0.29 µC per wave. Per single electron oxidation, this gives a surface charge of 4.12 µC/cm^2 ± 1.80 µC/cm^2. Comparing the surface charges of Au(111) with the surface charges of the single-electron process of the molecule, this gives a coverage of only about 2 % of molecules per gold atom. This coverage equals $(2.58 \pm 1.13)\cdot 10^{13}$ molecules/cm^2. Per molecule this results in an area of about 4 nm^2. Assuming an approximately rectangular molecule with a length of 2.29 nm (*vide infra*), and width of 1.61 nm, (from crystallography data of Scheerer *et al.* plus twice H-H distance of the benzene units on either side) this also results in an area of about 4 nm^2.

Panel **B** in Figure 5-33 displays the peak current for the 0/+2 redox process. The peak current scaled sub-linearly with the scan rate. Power fit gave exponents of 0.71 and 0.69 for scan rates from 25 mV/s to 100 mV/s for cathodic and anodic processes, respectively. Sub-linear behaviour at is attributed to the small counter electrode size compared to the working electrode.

Panel **C** displays the peak separation *ΔE* for the 0/+2 oxidation/reduction process (square symbols). The peak separation increases with increasing scan rate. This might be due to the STM cell setup, or due to the fast evaporation of solvent. The half-wave potential (triangle symbols) stays approximately constant with scan rate. The discontinuity at 0.5 V/s was caused by refilling the solvent between measurements. This was necessary because DCM evaporated so fast that the cell ran low during the experiment.

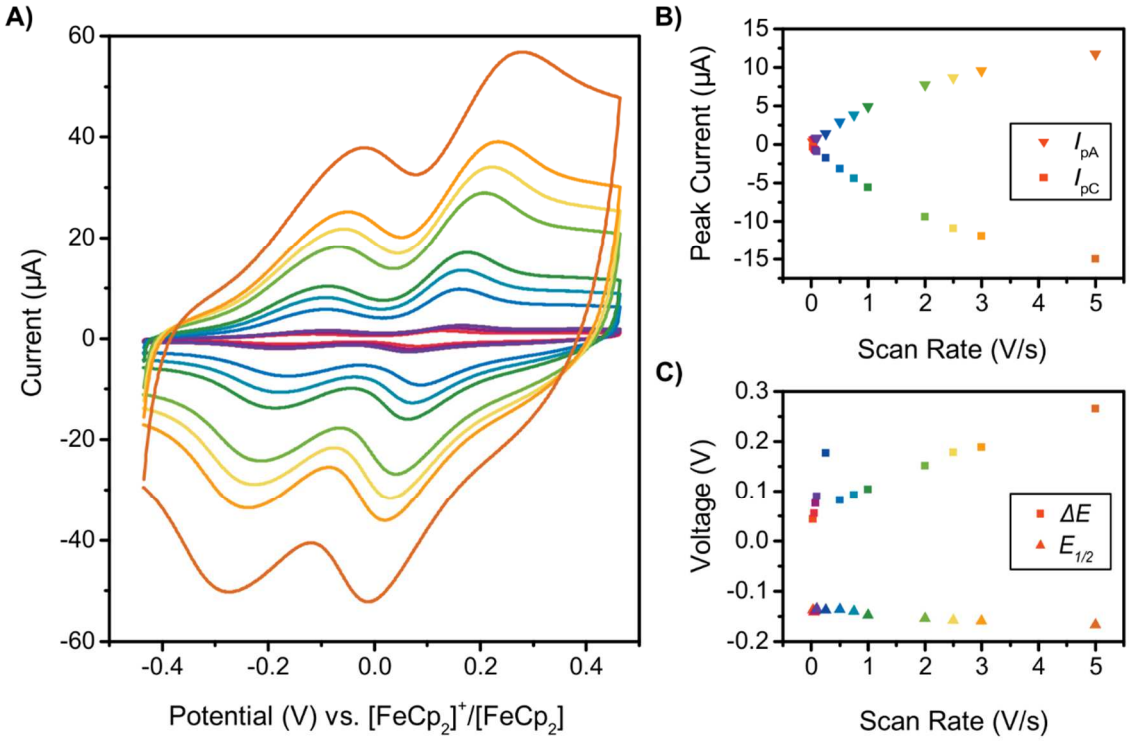


*Figure 5-33: **A**) Cyclic voltammograms with increasing scan rate (Inner to outer traces). **B**) Peak current of the 0/+2 oxidation/reduction step. **C**) Peak separation (squares) and half-way potentials (triangles) of the 0/+2 oxidation/reduction step. Corrected for solution resistance. Colour coding in all panels according to scan rate. Measurements were performed in degassed 0.1 M TBAPF₆/DCM under argon atmosphere.*

5.5.3 T-Ru₄M Spectroscopy under Argon

MPVC Analysis

The whole data set without any data selection does not show any plateaus in the log 2D conductance histogram in Figure 5-34 **A**. The log 1D conductance histogram in **B** shows the same setpoint peak described in the previous section. The absence of peak and plateau features in the histograms without data selection yet again emphasises the need for analysis techniques beyond conventional histogram analysis.

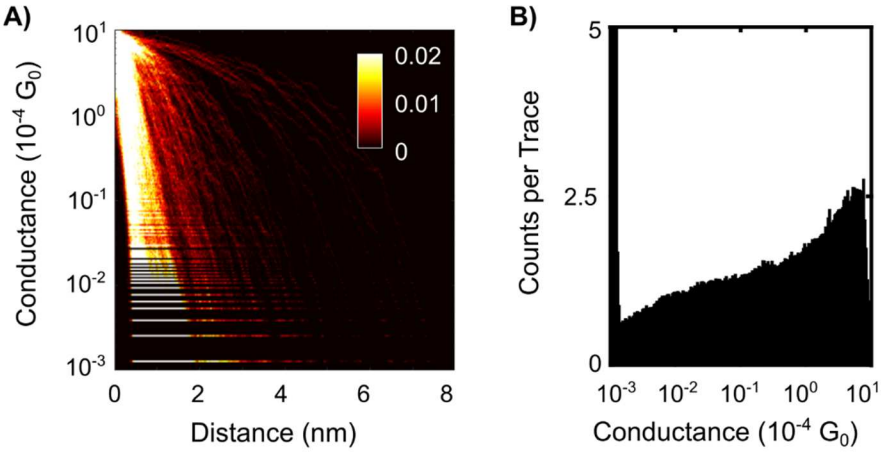


*Figure 5-34: Log 2D **A**) and log 1D **B**) conductance histograms of the whole data set without any data selection. Both histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm. Data in **B**) was filtered with a low-ass Bessel filter.*

MPVC analysis was performed on the whole data set. The reference vector decay factor was chosen to be 2 $Å^{-1}$ (roi: 0 nm to 4 nm). The vector representation of the data set is shown in Figure 5-35. At least two clusters are visible here. One cluster close to the origin of the plot, and a second distinct cluster with semi-circular shape further from the origin. The colour coding comes from a cluster assignment step that was performed with FCM using the GK algorithm with two clusters and h_r as the feature. Two clusters were chosen to accommodate for traces with and without deviations from plain exponential decay. The red cluster contained 242 traces or 12 % of traces, the blue cluster 88 %. Plateau-containing traces were predominantly assigned to the red cluster, plain exponential traces to the blue cluster.

It can be seen in the figure that part of the blue point cloud was assigned to the red cluster and the other way around. The reason for the sub-optimal cluster assignment is the cluster shape and orientation in conjunction with the selection of clustering algorithm. Additional dimensionality reduction procedures like Local Linear Embedding (LLE)[11] could improve clustering here, but are beyond the scope of this work. None of the algorithms used in this thesis resulted in perfect cluster separation (distribution based, density based, FCM with different features, plateau finder). The choice of clustering algorithm with different point distributions is an area of research by itself and interested readers are referred to the extensive body of literature available on this topic.[12,13] Although perfect cluster separation is desirable, the cluster assignment shown in Figure 5-35 is sufficient for data analysis (*vide infra*).

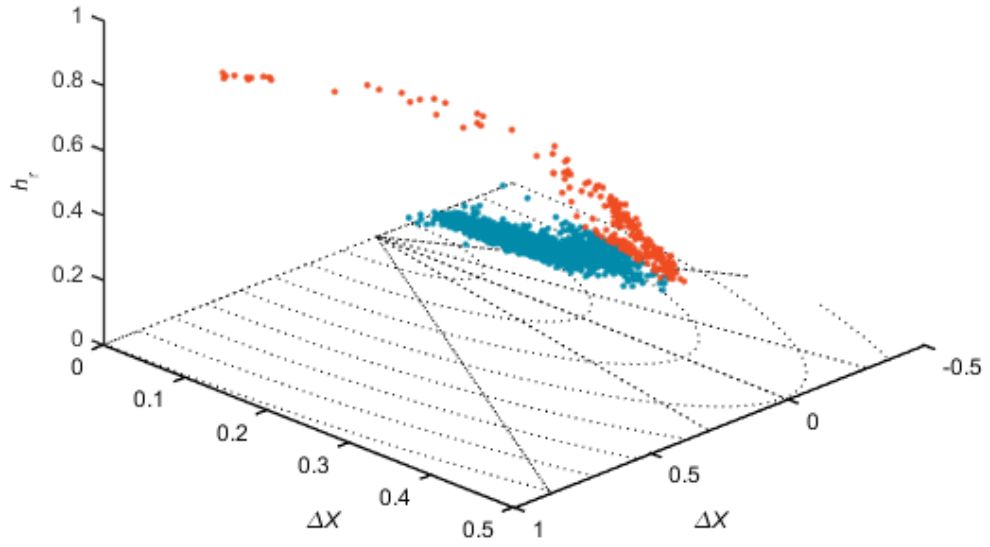


Figure 5-35: Cylinder plot of the whole data set. Coloured in blue are predominantly plain exponential decays, coloured in red are mainly plateau-containing traces as determined by MPVC with FCM GK.

All members of the red cluster were compiled into histograms. Log 2D and log 1D conductance histograms are shown in Figure 5-36. The log 2D conductance histogram does not show a clear plateau in the classical sense, however, small "kinks" can be observed histogram. Kinks meaning here small plateau-like deviations from the plain exponential decay.

In the log 1D conductance histogram shown in Figure 5-36, those kinks add up to two faint conductance peaks. Fitting with two Gaussian distributions gives a most likely conductance of $(4 \pm 1)\cdot10^{-6}$ G_0 and $(25 \pm 8)\cdot10^{-6}$ G_0 (errors are one STDEV of all conductance results presented below). The shortness of the kinks as well as the low currents associated with these conductance events make it necessary to filter the data. Also, to limit the influence of the setpoint effects, the first 10 % of all traces were disregarded from the log 1D conductance histogram. Although the conductance values found for the molecule are low, it should be stressed that the noise level conductance is still more than one order of magnitude below this conductance. The noise level conductance was estimated to be $0.2\cdot10^{-6}$ G_0 by calculating mean 3 σ of the last 10 % of data points of all traces.

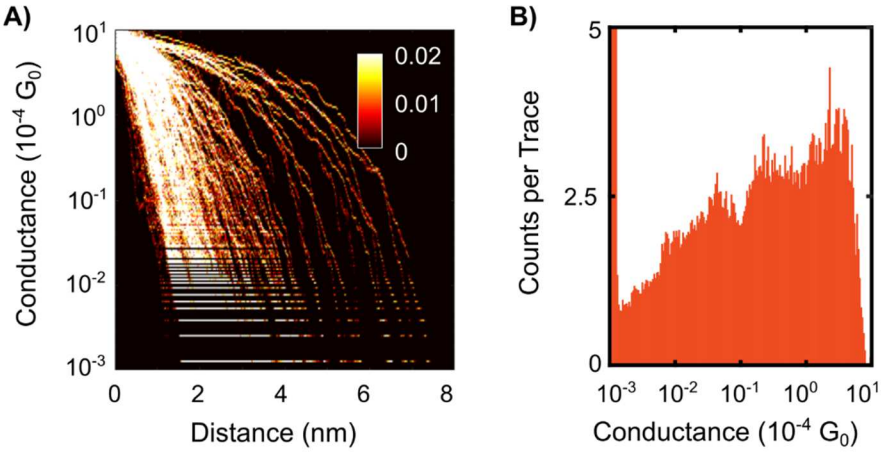


*Figure 5-36: **A**) Log 2D and **B**) log 1D conductance histogram of the red cluster. Note that the log 1D histogram has been filtered with a low-pass Bessel filter and the first 10 % of each trace were excluded to limit the magnitude of the setpoint peak. Binning: 50 bins per decade for conductance and 500 distance bins per 8 nm*

A selection of traces contained in the red cluster is displayed in Figure 5-37. The dashed lines represent the most probable conductance peaks from the log 1D conductance histogram. The traces are offset for clarity. *I(s)* traces show kinks, or short plateaus around the most likely conductance values, as estimated from the histograms. The first trace shows a plateau close to the high conductance group, traces two and three have plateaus at the low conductance value. Traces four and five are examples of traces that show both high and low conductance group within one trace.

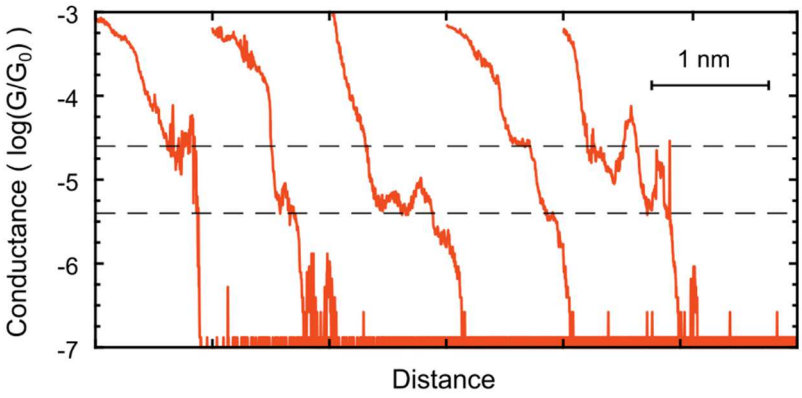


Figure 5-37: Sample G(s) traces of the red cluster on a semi-logarithmic scale. Traces are offset for clarity. Dashed lines indicate the peak conductance values from the log 1D conductance histogram.

The event distribution is shown in Figure 5-38 **A**. The graph shows discrete jumps in count rather than a steady incline.

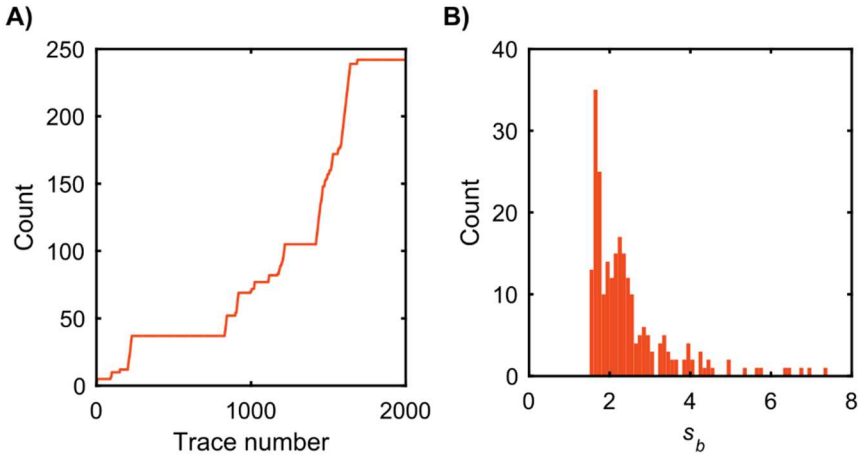


*Figure 5-38: **A**) Event distribution of the red cluster over the course of the experiment. **B**) Break-off distance histogram of the red cluster. Bin width: 1 Å.*

The break-off distance histogram in **B** indicates that the observed traces are due to the analyte. When considering the histogram in Figure 5-38 **B**, it is obvious that there are two distributions present. One narrow distribution with a distribution mean μ of 1.67 nm ± 0.11 nm and one wider distribution with a μ of 2.22 nm ± 0.53 nm. The distribution at shorter s_b consists of plain exponentially decaying traces without plateaus and can be traced back to the imperfect cluster assignment (*vide supra*). Accounting for the initial tip-substrate distance this gives a $s_{b,abs}$ of 2.87 nm ± 0.53 nm for the wider distribution. This observed absolute break-off distance is in excellent agreement with the expected junction length of 2.77 nm (estimated from the length of closely related compound[6] and twice the S-Au distance of 0.24 nm[14]).

Plateau finder analysis

In a validation step, the same data set was analysed using the logarithmic plateau finder algorithm instead of the MPVC analysis. Of the 2000 traces collected during the experiment, the plateau finder marks 621 as plateau-containing (JFP 31.1 %). In Figure 5-39, those traces are shown in a log 2D and log 1D conductance histogram. No plateau-type features can be seen in the 2D histogram, but the 1D histogram has two peaks. Fitting those peaks with Gaussian distributions gives most likely distribution means of $(4 \pm 1)\cdot10^{-6}$ G_0 and $(19 \pm 8)\cdot10^{-6}$ G_0. This is in good agreement with the results found by the vector analysis of $4\cdot10^{-6}$ G_0 and $25\cdot10^{-6}$ G_0.

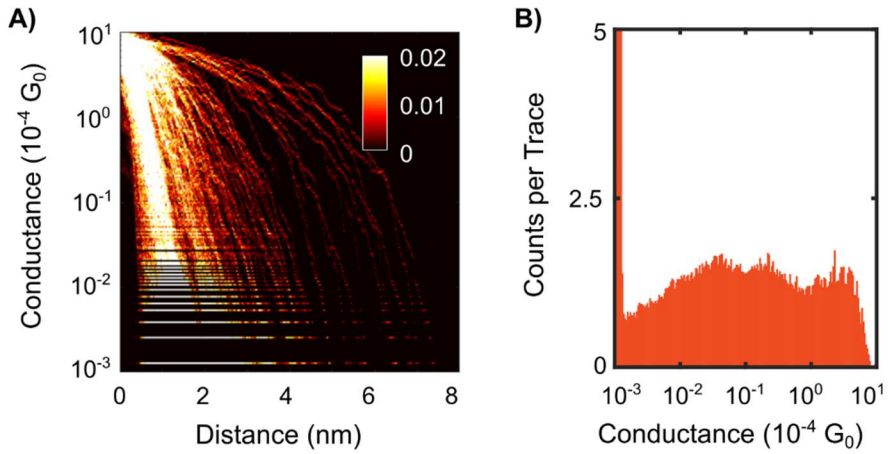


*Figure 5-39: Log 2D (**A**) and log 1D (**B**) conductance histograms of data found by plateau finder. Note that the log 1D histogram has been filtered with a low-pass Bessel filter and the first 10 % of each trace were excluded to limit the magnitude of the "setpoint peak". Both histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm*

Comparing MPVC with PF reveals that 229 of the 621 traces found by the plateau finder were also found by MPVC (37 % agreement). 63 % of traces were not found by MPVC (false negatives) and 13 were found by MPVC and not by plateau finder (2 % false positives) The cylinder plot for all traces is shown in Figure 5-40. Red data points are traces found by plateau finder. Blue traces are plain exponential decays as determined by the plateau finder algorithm.

The large number of false negatives is rather concerning at first. When looking at those traces in detail, it can be seen that many traces do indeed contain plateaus. Mainly fast-decaying traces are misclassified by the MPVC approach. This suggests that for this type of data set, where there is a large variation in decay factor, the MPVC approach is not working as well as it should under the selected set of parameters. By changing the parameter selection, for example by introducing the individual decay factor as a parameter, the MPVC could be improved. Nonetheless, despite the disagreement in classification, the resulting conductance values show good agreement.

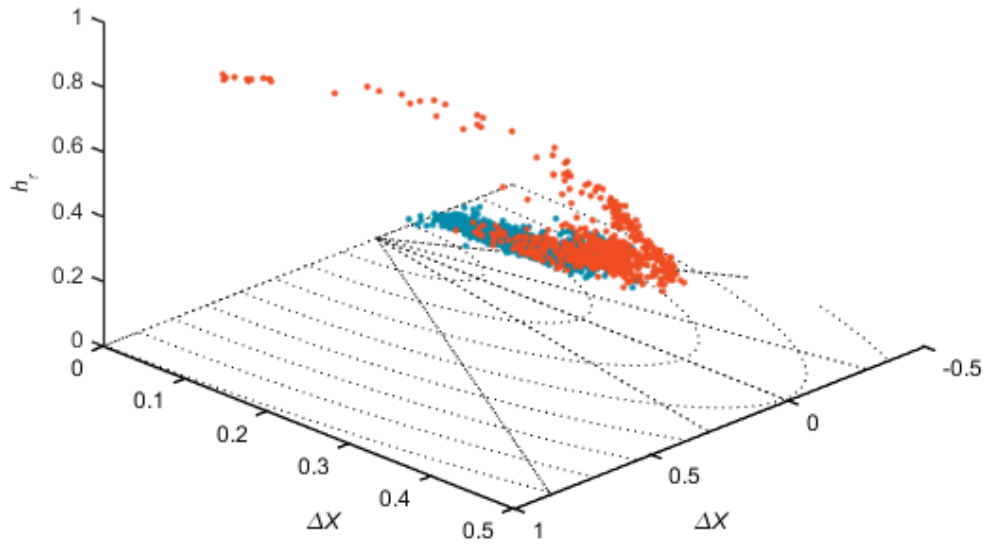


Figure 5-40: Cylinder plot of all traces in the data set. Colours are per cluster membership as determined by the plateau finder algorithm.

Control Experiment

A control experiment was conducted with the same tip and substrate under argon before the monolayer was assembled. The plateau finder algorithm was applied to the data set and 86 traces contained "plateaus" (JFP 4 %). "Plateau" is used in quotes here, because they seem to be random noise/contamination events. Random, because they do not give a peak in the 1D log conductance histogram after data selection (*c.f.* Figure 5-41). The event distribution of the control experiment is provided in the appendix.

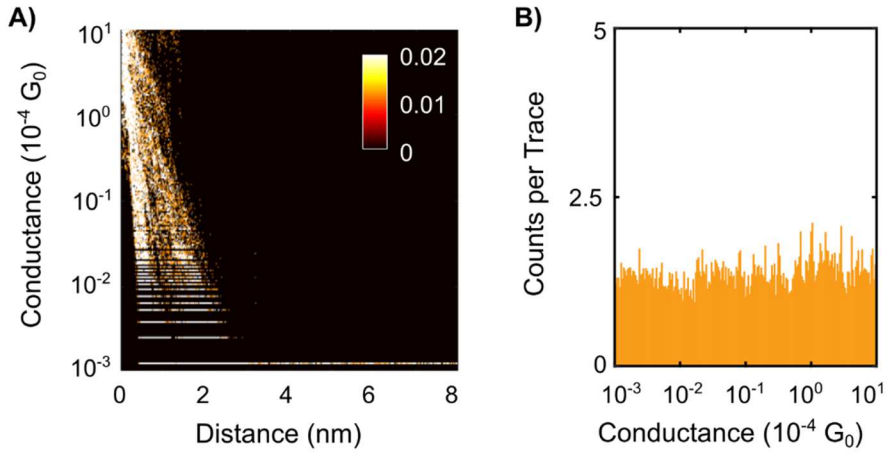


*Figure 5-41: Log 2D (**A**) and log 1D (**B**) conductance histograms of the control experiment after application of the plateau finder algorithm. 4 % of all traces were selected by the algorithm. Recorded at 0.3 V and 20 nA. Both histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm*

5.5.4 T-Ru$_4$M in Air

Low Conductance Class

Analogue to the spectroscopy experiments performed under argon, similar conductance results have been found for experiments performed in ambient conditions. The results are summarised below.

Figure 5-42 shows two experiments reproducing the low conductance group. Panel **A** and **B** are the log 2D and log 1D conductance histogram of a data set recorded at bias voltage of 0.2 V and setpoint current of 13 nA. Of the 2000 recorded *I(s)* traces, 56 contain plateaus as determined by the plateau finder (JFP 3 %). The conductance was determined to be $(2 \pm 1)\cdot10^{-6}$ G_0 by fitting with a Gaussian distribution. Panel **C** and **D** contain the log 2D and log 1D conductance histograms for experiments performed under -0.4 V and 26 nA setpoint current. 121 of the 2000 traces were marked by the plateau finder (JFP 6 %). The conductance was determined to be $(4 \pm 1)\cdot10^{-6}$ G_0 by fitting the histogram with one Gaussian distribution.

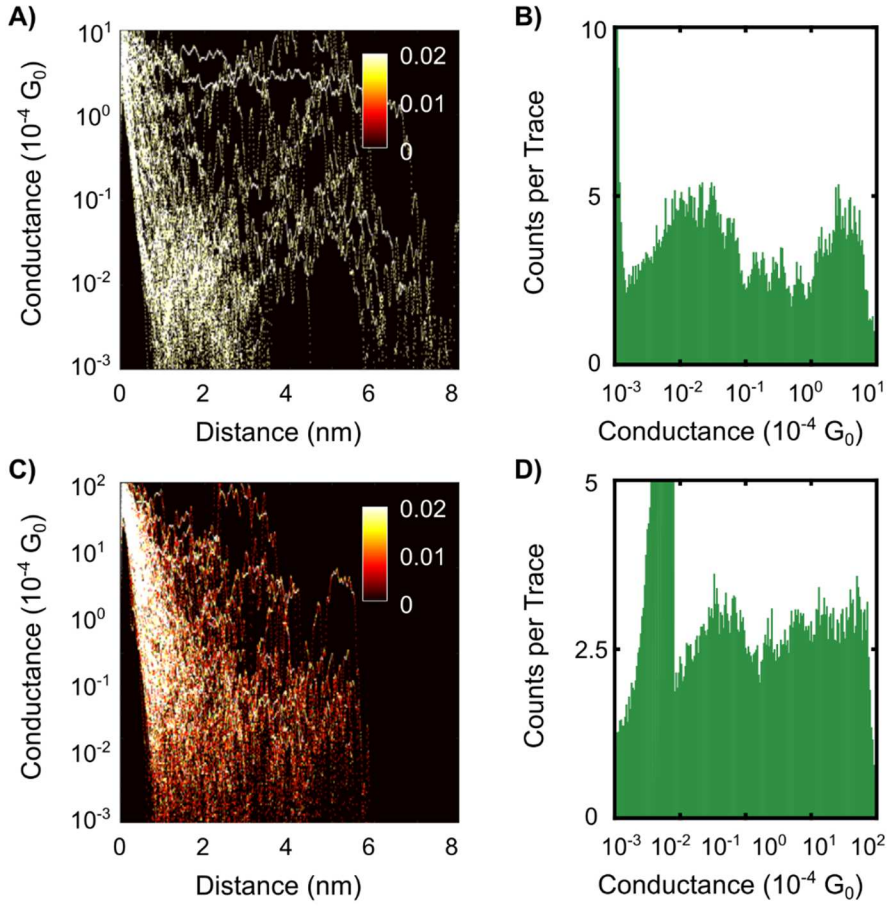


*Figure 5-42: Log 2D and log 1D conductance histograms of two different experiments where the low conductance group was observed. 2000 traces each were recorded at 0.2 V (**A** and **B**) and -0.4 V (**C** and **D**). Plateau traces were selected by the PF algorithm. Data in all histograms was filtered with a low-pass Bessel filter. All histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm.*

Break-off distance histograms of the plateau-containing traces did not show well-defined break-off distance distributions. This is attributed to the low hit-rate and to the strong thiol-gold interaction which could have caused the pull-out of gold strings from the surface and thus leading to undefined break-off distances.[15]

High Conductance Class

The high conductance group was also confirmed in several experiments. Figure 5-43 contains log 2D and log 1D conductance histograms for two experiments showing high conductance signatures. Panel **A** and **B** contain data for an experiment conducted at 0.04 V and 20 nA. 490 plateaus were found, as determined by the plateau finder (JFP 25 %). A conductance of $(35 \pm 8) \cdot 10^{-6}$ G_0 was determined by fitting the 1D conductance histogram with one Gaussian distribution.

The high conductance peak was also confirmed in an experiment conducted at 0.05 V and 20 nA, **C** and **D**. Here, 284 traces contained plateaus according to the plateau finder (JFP 14 %). The conductance was determined to be $(40 \pm 8) \cdot 10^{-6}$ G_0. Curiously, this conductance feature is dwarfed by a very sharp peak around $700 \cdot 10^{-6}$ G_0. Accordingly, there is a well-pronounced plateau in the log 2D conductance histogram in panel **C**. The origin of this very-high conductance feature is presently unknown. As it is much higher than the expected conductance of the molecule, this result is not attributed to the through-molecule conductance, but might be related to a different, shorter current pathway along the molecule. A conductance feature of similar magnitude has been observed in one other experiment.

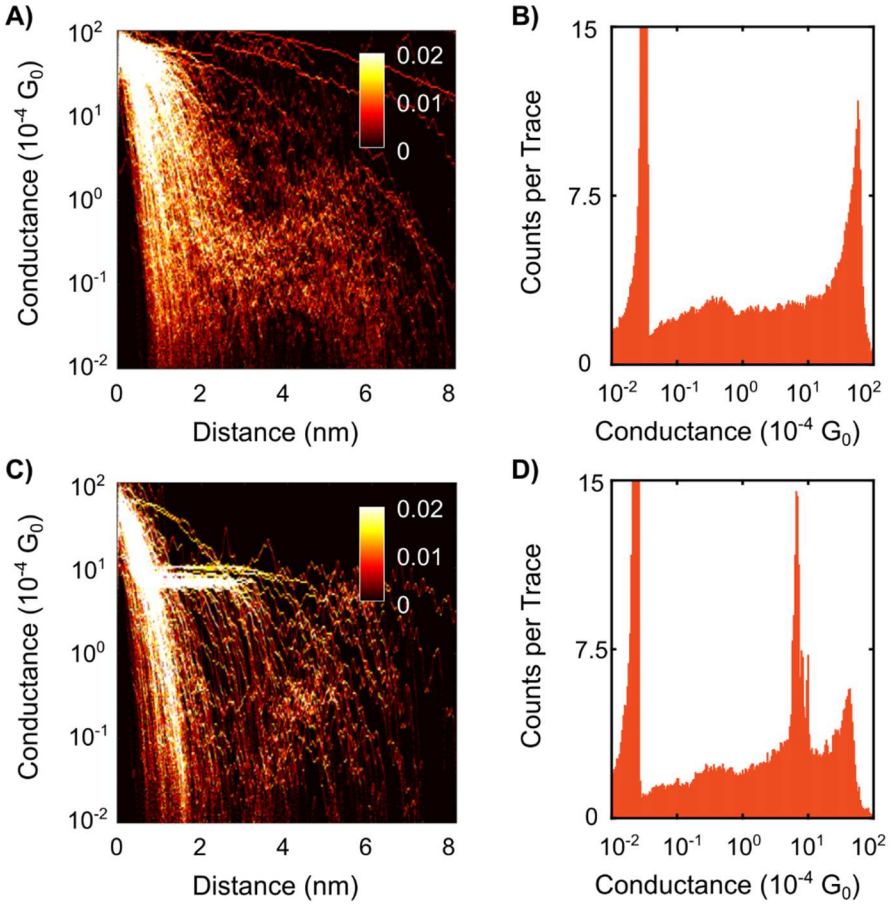


*Figure 5-43: Log 2D and log 1D conductance histograms of two different experiments where the high-conductance group was observed. 2000 traces each were recorded at 0.04 V (**A** and **B**) and 0.05 V (**C** and **D**). Plateau traces were selected by the PF algorithm. Data in all histograms was filtered with a low-pass Bessel filter. All histograms have 50 bins per decade for conductance and 500 distance bins per 8 nm.*

The break-off distances for this macrocycle were generally not well-defined. Low JFP and large scatter in s_b lead to ill-defined distributions. Only the data set recorded at 0.04 V showed a conclusive break-off distance histogram (Figure 5-44), albeit the scatter is large. Two

distributions were present in the histogram. The two distributions were related back to imperfect cluster assignment, meaning that plain exponentially decaying traces were present in the plateau cluster accounting for the peak at small distances. Fitting with two Gaussian distributions gives a most likely break-off distance for the plateau traces of 2.6 nm ± 1.3 nm. This experimental break-off distance is in good agreement with the expected junction length of 2.77 nm.

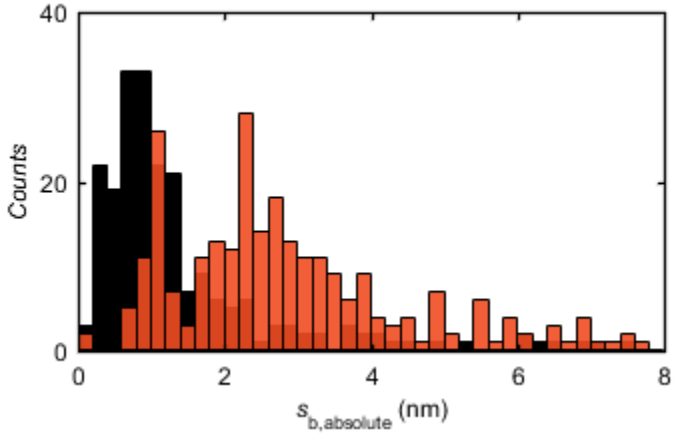


Figure 5-44: Break-off distance histogram of the high conductance class shown in the previous figure (0.04 V). Black bars correspond to the initial distance. Red bars to the absolute break-off distance.

5.5.5 T-Ru$_4$M - Discussion and Summary

By now, it is well-established in the molecular electronics community that even "simple" molecules like ODT show complex conductance behaviour when connected to a macroscopic circuit. Nonetheless, at least two distinct conductance groups were successfully identified for the first time for this rather complex macrocyclic tetraruthenium molecule. Control experiments showed that the observed signal was related to the analyte. Electrochemical experiments after *I(s)* spectroscopy suggested that the molecule was surface-bound and stayed intact over the course of the experiment. The quantitative agreement of the theoretical area per molecule and measured area per molecule suggested a full monolayer coverage of T-Ru$_4$M.

For spectroscopy experiments under argon, two conductance groups were identified in one experiment at $(4 \pm 1)\cdot10^{-6}$ G_0 and $(25 \pm 8)\cdot10^{-6}$ G_0. The results of both applied analysis techniques (MPVC and PF) showed good agreement. Both conductance groups were reproduced under ambient conditions. The overall conductance was estimated to be $(3 \pm 1)\cdot10^{-6}$ and $(33 \pm 8)\cdot10^{-6}$ G_0 for the low and high conductance class, respectively. The origin of the two observed conductance groups could be the binding configuration, similar to what has been observed for ODT.[16] Different conformations of the molecule are also possible as described

here.[6] Apart from the two conductance groups, a less frequent very-high conductance group was observed.

Break-off distances under argon were close to the expected junction extension. 2.87 nm ± 0.53 nm were seen in the experiment and 2.77 nm was the expected junction extension. Under ambient condition, break-off distances were typically ill-defined. This effect was partially attributed to the strong Au-S bond, which is known to cause undefined break-off distances. The observation of this effect suggests that the molecule coupled to both electrodes with the thiol anchor groups, indicating that the full-length conductance was indeed measured.

5.6 Ruthenium Macrocycles: Summary and Discussion

For the diruthenium complex T-Ru_2M, a reproducible conductance value of $(3\pm2)\cdot10^{-6}$ G_0 was found. It could, however, not be confirmed that this conductance corresponds to the full-length conductance of the molecule. The observed break-off distance was smaller, albeit within experimental error, compared to the expected length. According to preliminary simulation results, the expected conductance could be as low as 10^{-9} G_0 and thus below the resolution limit of the experimental setup. That is, if the assumed charge transport mechanism is coherent tunnelling. For molecules of this length it is reasonable to expect a significant hopping contribution as well. Hopping is more efficient over large distances, which could mean that the simulated conductance underestimates the conductance.

For sPy-Ru_4M, conductance values between $(5\pm3)\cdot10^{-6}$ G_0 were found at different bias voltages. The conductance seemed to be lower when the absolute of the applied bias was low. The break-off distance that was measured for sPy-Ru_4M stayed below the expected junction extension. This could mean that the molecule was contacted at entities other than the anchor, or that the molecule was bound at an angle.

The lPy-Ru_4M showed two distinct conductance classes at $20\cdot10^{-6}$ G_0 and $1\cdot10^{-6}$ G_0. Excellent agreement between the observed and expected break-off distance of the low conductance class lead to the conclusion that the full-length conductance was measured. The second, higher conductance class could correspond to a higher conductance binding configuration of the pyridine anchor or a reduced effective current path.[7]

For T-Ru_4M, two conductance classes were found at $(33\pm8)\cdot10^{-6}$ G_0 and $(3\pm1)\cdot10^{-6}$ G_0. A third conductance class appeared sporadically at very high conductance values around $700\cdot10^{-6}$ G_0. Under argon, good agreement between experimentally observed and expected junction extension was seen. In air, no well-defined break-off distributions formed. It has been reported that thioacetyl converts to strong thiol bonds on the gold surface.[17] These gold-thiol bonds are

approximately the same strength as gold-gold bonds.[18,19] It was shown in the literature that thiol bonds can lead to undefined break-off distance by pulling out strings of gold from the surface.[15,20] This lead to the conclusion that the full-length conductance was measured. The two, and sometimes three, conductance classes were attributed to different binding configurations or conformations of the molecule.

For all molecules, generally good agreement for data analysis with the PF algorithm and the MPVC approach was seen. When comparing the lowest conductance values of the three tetraruthenium compounds, a few points are noteworthy: The conductance of the pyridine-anchored sPy-Ru_4M is higher than the one of the thiol-anchored T-Ru_4M, $(5\pm3)\cdot10^{-6}$ G_0 compared to $(3\pm1)\cdot10^{-6}$ G_0. Albeit the conductance values are within experimental error of each other. lPy-Ru_4M showed the lowest conductance, albeit also within experimental error of the two other tetraruthenium macrocycles, $1\cdot10^{-6}$ G_0. This is the expected outcome since lPy-Ru_4M is longer than sPy-Ru_4M (3.37 nm compared to 1.93 nm N to N distance). It is rather surprising though that within error, all tetraruthenium macrocycles showed similar conductance values. Especially so for lPy-Ru_4M, a lower conductance would be expected. This could mean that the ruthenium cage largely governs the charge transport.

To conclude, the single molecule conductance of four macrocyclic ruthenium complexes was investigated. While not in all cases, it could be confirmed that the full-length conductance was measured, a comparison of the conductance of three tetraruthenium macrocycles could be made.

5.7 References

1 M. S. Inkpen, Imperial College London, 2013.

2 M. S. Inkpen, S. Scheerer, M. Linseis, A. J. P. White, R. F. Winter, T. Albrecht and N. J. Long, *Nat. Chem.*, 2016, 1–6.

3 M. S. Inkpen, T. Albrecht and N. J. Long, *Organometallics*, 2013, **32**, 6053–6060.

4 O. A. Al-Owaedi, D. C. Milan, M.-C. Oerthel, S. Bock, D. S. Yufit, J. A. K. Howard, S. J. Higgins, R. J. Nichols, C. J. Lambert, M. R. Bryce and P. J. Low, *Organometallics*, , DOI:10.1021/acs.organomet.6b00472.

5 S. T. Schneebeli, M. Kamenetska, Z. Cheng, R. Skouta, R. A. Friesner, L. Venkataraman and R. Breslow, *J. Am. Chem. Soc.*, 2011, **133**, 2136–2139.

6 S. Scheerer, M. Linseis, E. Wuttke, S. Weickert, M. Drescher, O. Tröppner, I. Ivanović-Burmazović, A. Irmler, F. Pauly and R. F. Winter, *Chem. - A Eur. J.*, 2016, **22**, 9574 – 9590.

7 M. Kamenetska, S. Y. Quek, a. C. Whalley, M. L. Steigerwald, H. J. Choi, S. G. Louie, C. Nuckolls, M. S. Hybertsen, J. B. Neaton and L. Venkataraman, *J. Am. Chem. Soc.*, 2010, **132**, 6817–6821.

8 I. V. Pobelov, RWTH Aachen, 2008.

9 M. S. Inkpen, A. J. P. White, T. Albrecht and N. J. Long, *Dalton Trans.*, 2014, **43**, 15287–90.

10 H. Angerstein-Kozlowska, B. E. Conway, A. Hamelin and L. Stoicoviciu, *Electrochim. Acta*, 1986, **31**, 1051–1061.

11 L. K. Roweis Sam T.Saul, *Science*, 2000, **290**, 2323.

12 a. K. Jain, M. N. Murty and P. J. Flynn, *ACM Comput. Surv.*, 1999, **31**, 264–323.

13 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *J. Mach. Learn. Res.*, 2012, **12**, 2825–2830.

14 Å. Johansson and S. Stafström, *Chem. Phys. Lett.*, 2000, **322**, 301–306.

15 L. Venkataraman, J. E. Klare, I. W. Tam, C. Nuckolls, M. S. Hybertsen and M. L. Steigerwald, *Nano Lett.*, 2006, **6**, 458–462.

16 W. Haiss, S. Martín, E. Leary, H. Van Zalinge, S. J. Higgins, L. Bouffier and R. J. Nichols, *J. Phys. Chem. C*, 2009, **113**, 5823–5833.

17 H. Valkenier, E. H. Huisman, P. A. Van Hal, D. M. De Leeuw, R. C. Chiechi and J. C. Hummelen, *J. Am. Chem. Soc.*, 2011, **133**, 4930–4939.

18 C. Nef, P. L. T. M. Frederix, J. Brunner, C. Schönenberger and M. Calame, *Nanotechnology*, 2012, **23**, 365201.

19 M. Frei, S. V. Aradhya, M. S. Hybertsen and L. Venkataraman, *J. Am. Chem. Soc.*, 2012, **134**, 4003–4006.

20 D. Krüger, H. Fuchs, R. Rousseau, D. Marx and M. Parrinello, *Phys. Rev. Lett.*, 2002, **89**, 186402.

Chapter 6
Conclusion and Outlook

Synopsis: *In this chapter, the work described in the thesis will be summarised. Key results as well as challenges will be addressed and possibilities for future directions will be presented.*

One focus of the presented work was to apply machine learning techniques to single molecular conductance data. A new analysis method was designed to find a more objective means to analyse data in greater detail. The developed analysis method was then applied to single molecule conductance data, with a special focus on linear and cyclic organometallic molecules.

In **Chapter 3**, the need for an objective, unsupervised means for the analysis of single molecular conductance data was established. A fundamentally new analysis approach was developed. The principle of MPVC analysis was demonstrated and it was shown that the MPVC analysis does not make any assumptions about the presence and type of molecular signatures. The approach was subsequently applied to simulated conductance data. Here it was found that MPVC analysis is capable of separating plateau-containing traces from exponential decays without plateaus. Furthermore, it was shown that different plateau shapes could be successfully separated. Clustering a data set into subgroups consisting only of similar traces was shown to aid analysis of often complex conductance data. MPVC was then applied to two well-documented experimental conductance data sets. For ODT, it was shown that large data sets could be analysed and good agreement with literature was found. For the second experimental data set, namely OPE, it was shown that MPVC analysis was capable of identifying two fundamentally different charge transport scenarios. One of which was not observed with a conventional analysis algorithm in our dataset, but again reported in the literature.

MPVC analysis is a first approach to apply machine learning principles to the field of molecular electronics. Even though MPVC analysis worked well in many cases, the current selection of parameters may be adapted to suit a particular experimental context. Current developments in artificial intelligence, for example in the field of deep learning, could be applied to the analysis approach to automatically select parameters for an even greater depth of analysis.

In **Chapter 4**, MPVC as well as the plateau finder algorithm were then applied to a linear ferrocene, s-FDTM. The measured conductance was found to be in agreement with literature. Next, two other, closely-related molecules with thiol and pyridine anchor groups were investigated, s-FDT and s-FPT. It was found that short conductance plateaus were reproducibly observed at different bias voltages. Data analysis proved difficult with MPVC under the selected set of parameters because of the shortness of plateaus. With a conventional plateau finder algorithm, bias dependent conductance values for both molecules were identified. Four conductance classes were identified for most probed bias voltages. The origin of the classes is

currently unknown. Both molecules showed remarkedly similar conductance values. To this end, it is not fully understood if the observed conductance corresponds to the full-length conductance or if a folded structure was measured for both molecules. Collaborative efforts with the group of Robert Stadler are ongoing to resolve this question.

Preliminary experimental results for a double-branched ferrocene compound, b-FPT, showed intriguing agreement with simulations of its single-branched counterpart, s-FPT. However, more detailed experimental and theoretical studies are required here.

For a thiomethyl-anchored biferrocene, BFTM, two conductance groups were identified with MPVC analysis. MPVC proved to be a valuable tool to identify the two classes, which appeared very infrequently. Reasonable agreement was seen with literature. For the same molecule with pyridine linkers, BFP, two conductance classes were identified with MPVC. With the conductance of the two symmetrical BFTM and BFP, it would be interesting to estimate the conductance of the unsymmetrical combination of both molecules, meaning the biferrocene backbone with a pyridine and a thiomethyl anchor. According to Manrique *et al.* the resulting conductance should be on the order of $\sqrt{G_{BFTM}G_{BFP}}$.

In the last experimental chapter, **Chapter 5**, experimental results on four macrocyclic ruthenium compounds were presented. To the author's best knowledge, this is the first time that single molecule conductance experiments have been reported for macrocyclic, organometallic cyclic systems. One biruthenium macrocycle and three closely related tetraruthenium molecules were investigated. For the biruthenium system, a conductance value could be found, even though it is still unclear whether the observed conductance was measured along the full length of the molecule. Short break-off distances and preliminary simulation results suggested that a reduced current pathway along the molecule was sampled.

For a tetraruthenium macrocycle with short pyridine linkers, sPy-Ru_4M, a conductance was measured at several bias voltages with good reproducibility. The conductance seemed to decrease with decreasing absolute applied bias voltage. The observed conductance was higher than for the same macrocycle with longer linkers, lPy-Ru_4M. This was the expected trend, since lPy-Ru_4M is significantly longer than sPy-Ru_4M. For the thiol-anchored macrocycle, T-Ru_4M, a conductance was found between sPy-Ru_4M and lPy-Ru_4M. Good agreement was seen for the expected and measured break-off distances under argon. In ambient conditions, break-off distances were often ill-defined, which can be indicative for Au-S bonds. Both of those observations lead to the conclusion that the full length of the molecule was sampled.

Throughout the project, the low reproducibility of conductance measurements proved to be challenging and it is not understood what the limiting factors for junction formation were. In future experiments the reproducibility issues could be addressed by performing complementary break-junction experiments. These BJ experiments could be STM-based or on a MCBJ setup.

Another sensible future direction for conductance studies would be to conduct spectroscopy experiments under electrochemical control. Since the molecules used throughout this thesis were all redox active, conductance dependent on the applied potential could be probed here. This is especially relevant for macrocyclic systems. With suitable functionalisation of single branches of the cycle, it could possibly be probed how charge transport is changed when single branches of the macrocycle are oxidised/reduced. This could be a convenient test bed for quantum interference studies.

The presented results in this thesis both for the novel analysis technique and for the conductance studies constitute a significant contribution to the field of molecular electronics.

Appendices

Appendices 176

Appendix I: Additional Figures 177

Appendix II: MATLAB Scripts 187

Appendix III: Permission Documents 194

Appendix I: Additional Figures

s-FDTM

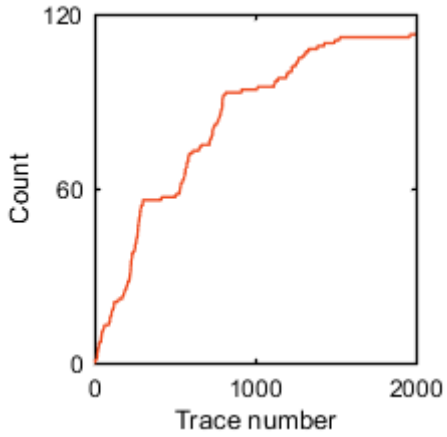


Figure App. 1: Cumulative counts of the traces found by the plateau finder algorithm.

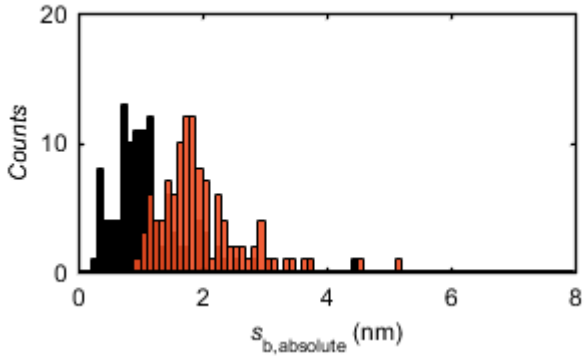


Figure App. 2: Break-off distance histogram of the traces found by the plateau finder algorithm.

s-FDT

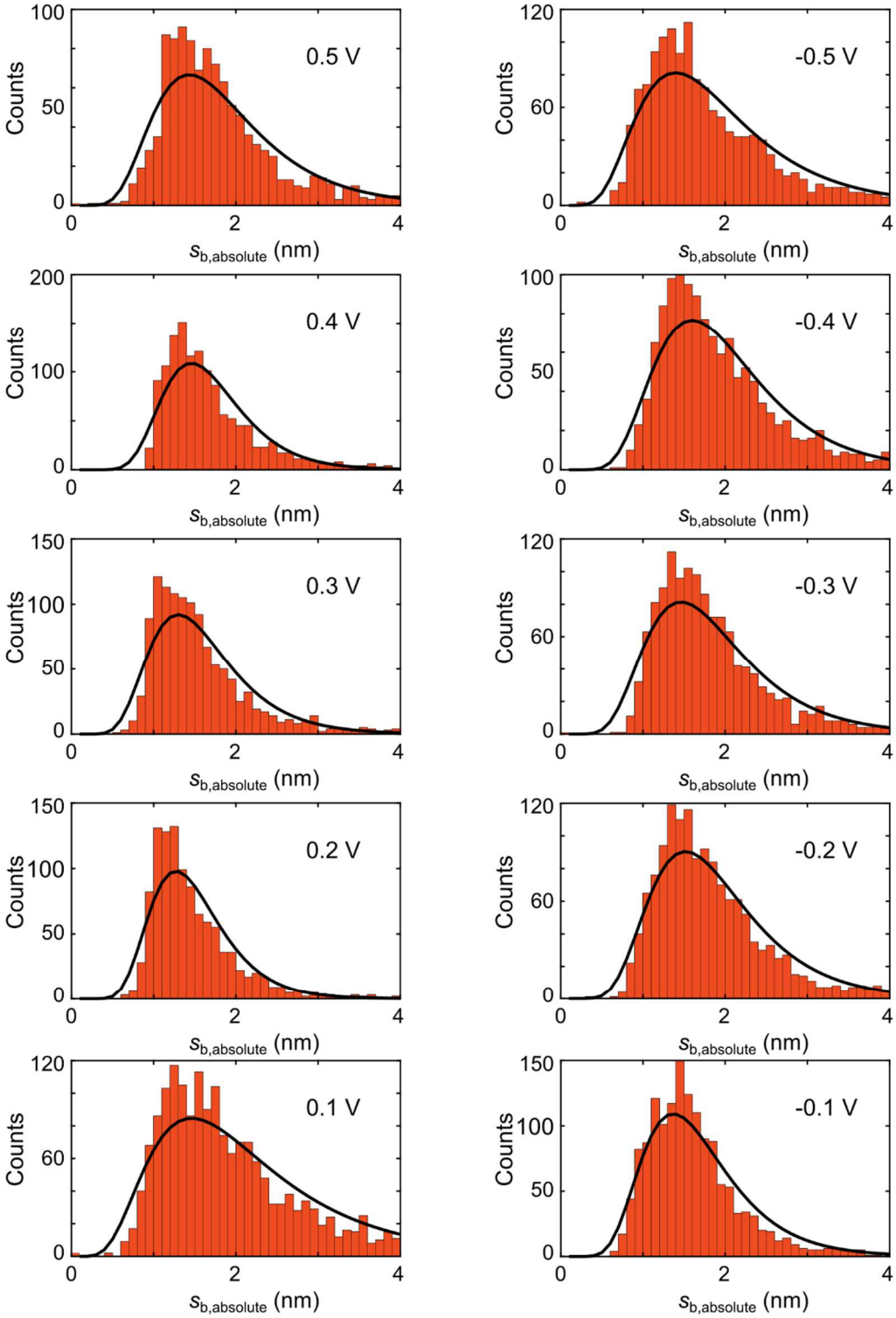


Figure App. 3: Break-off distance histograms with lognormal fit at various bias voltages for s-FDT.

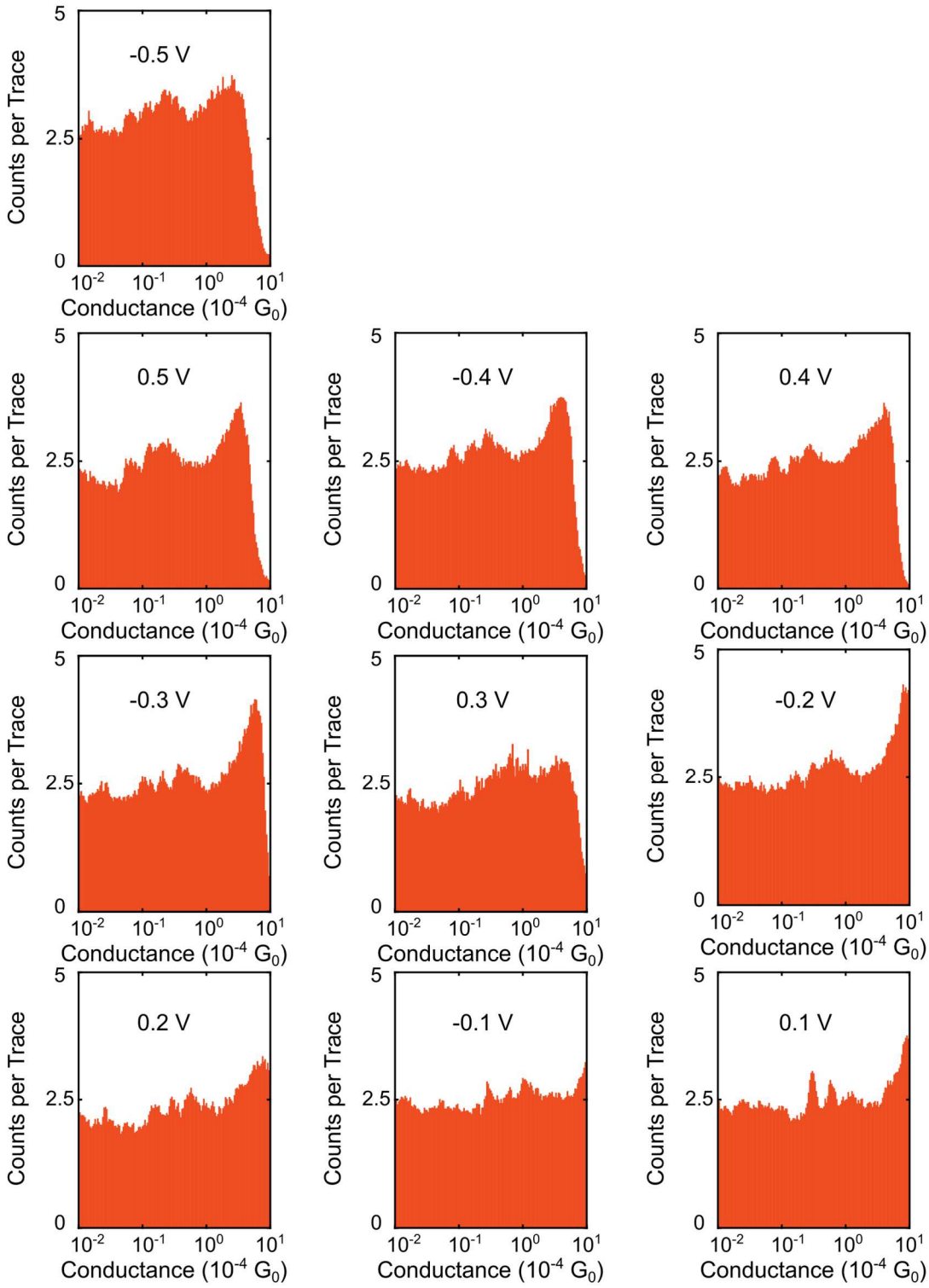


Figure App. 4: Conductance histograms of s-FDT at different bias voltages. Traces were filtered with a low-pass Bessel filter.

s-FPT

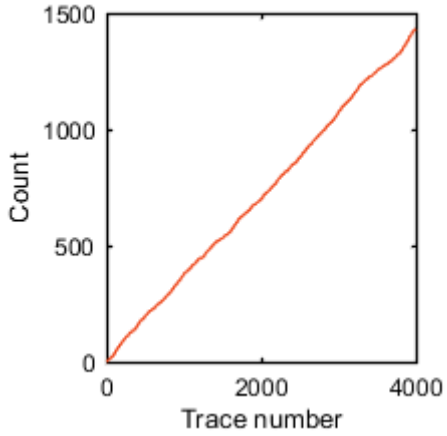


Figure App. 5: Event distribution of s-FPT.

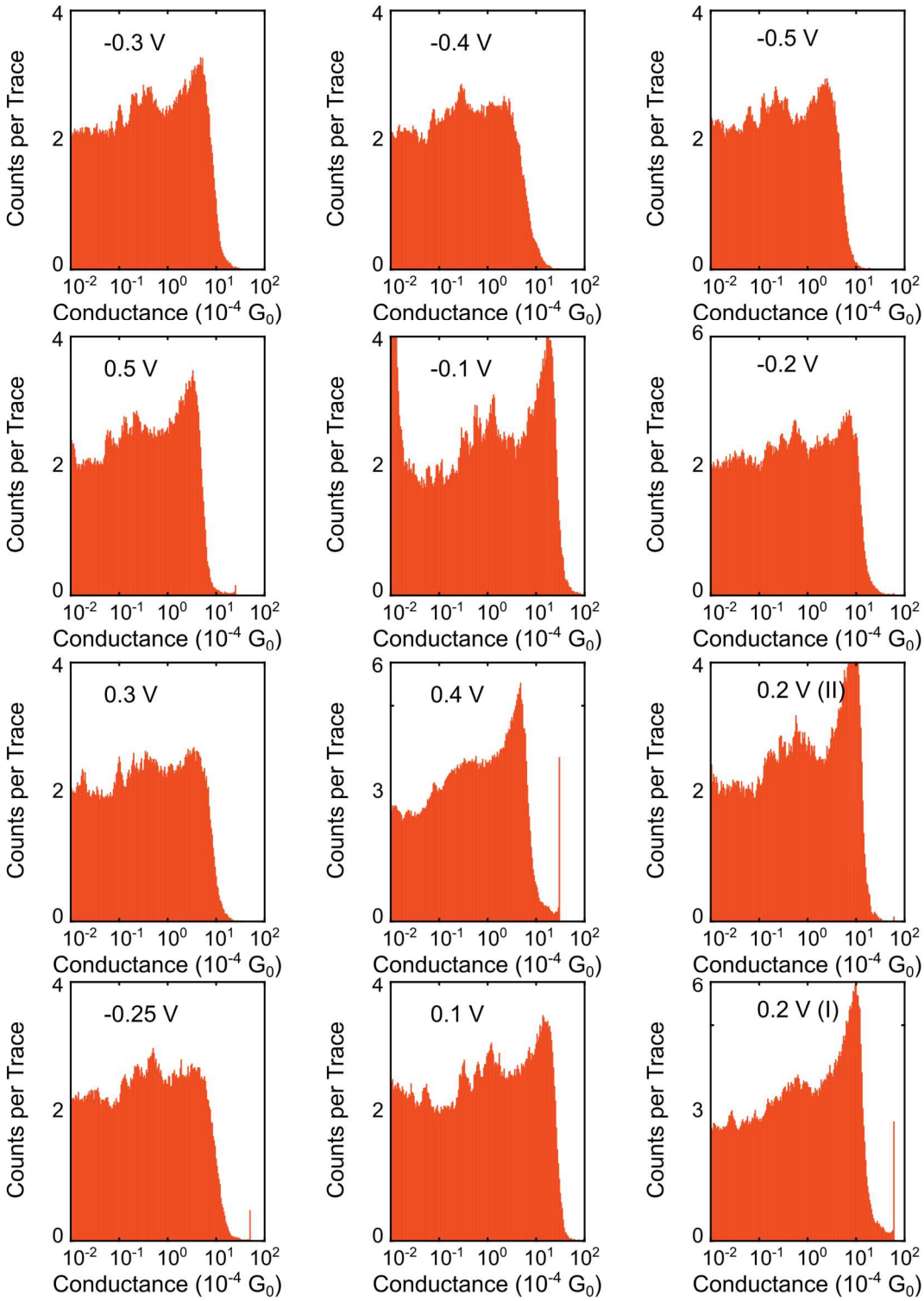


Figure App. 6: Conductance histograms of s-FPT at different bias voltages. Traces were filtered with a low-pass Bessel filter.

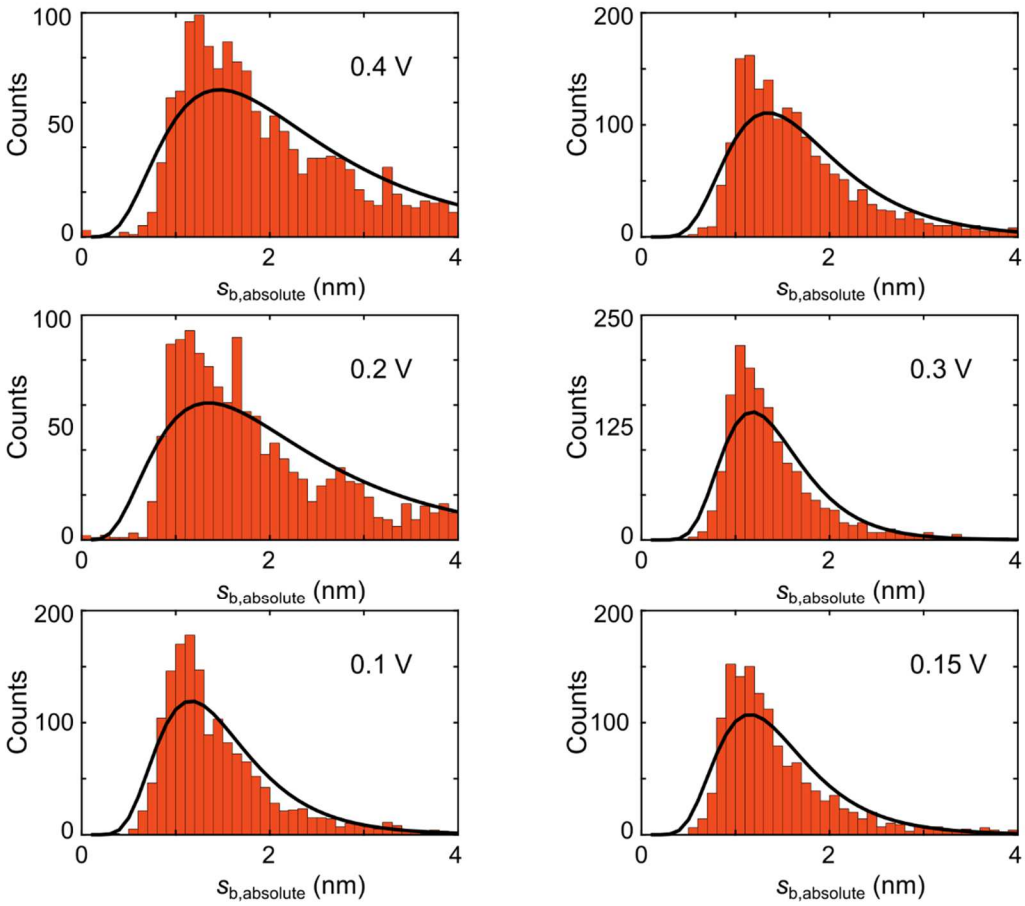


Figure App. 7: Break-off distances of experiments at positive bias with s-FPT.

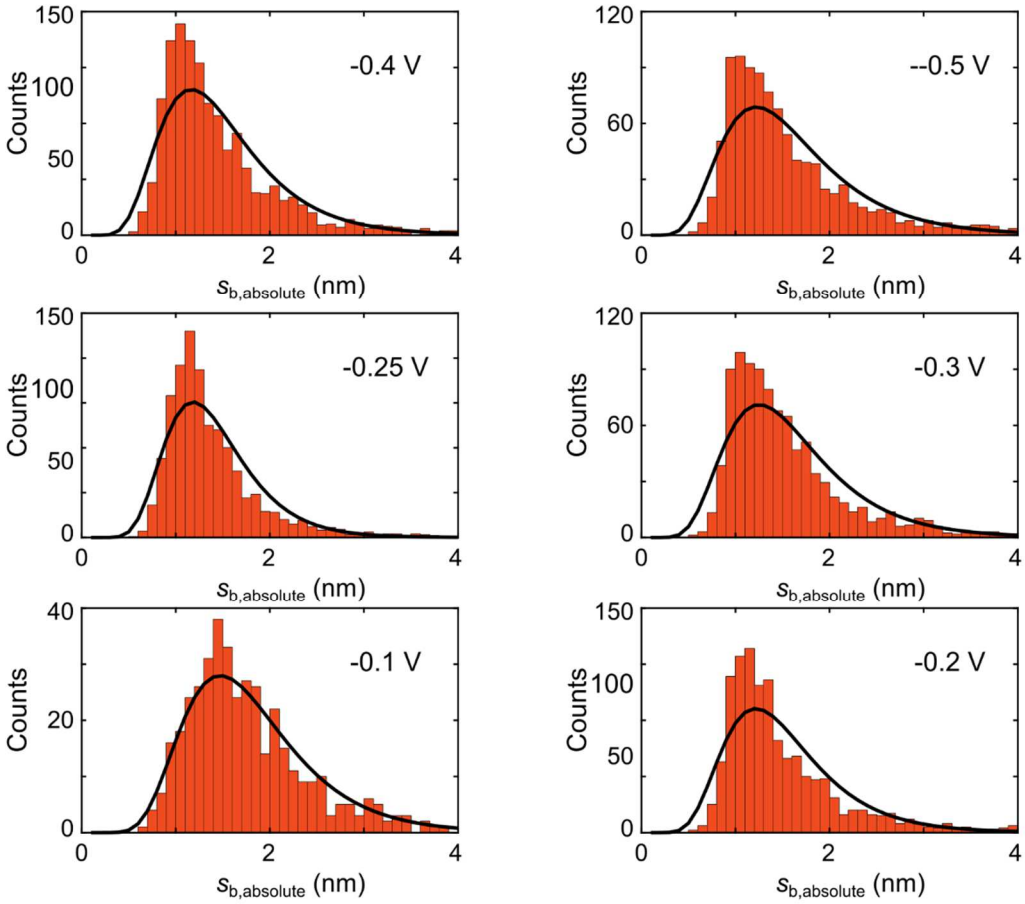


Figure App. 8: Break-off distances of experiments at negative bias with s-FPT.

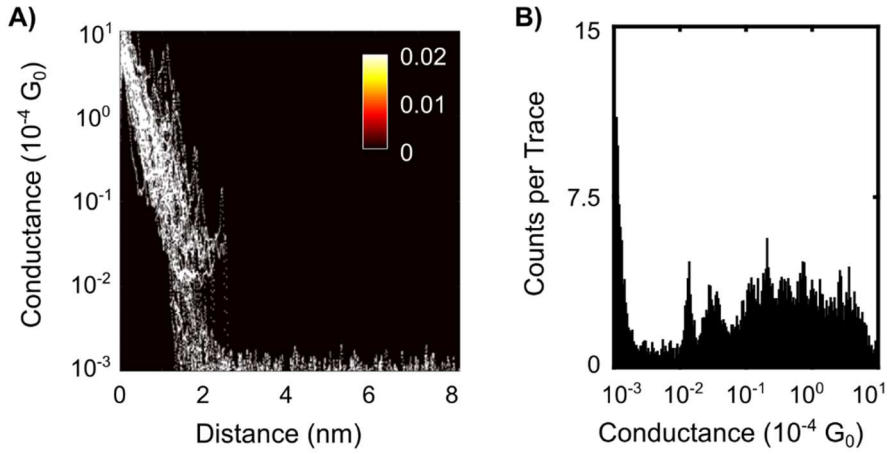


*Figure App. 9: **A**) 2D and **B**) 1D conductance histogram of the data set recorded at 0.1 V and 6.5 nA. Data selection was made with MPVC and GC FCM. Bin widths: 0.16 Å and 50 bins per decade. Data was filtered with a low-pass Bessel filter.*

T-Ru₂M

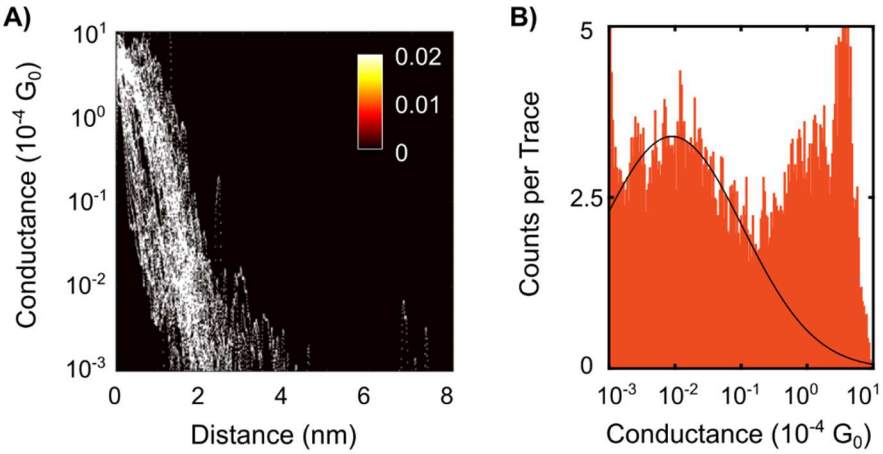


*Figure App. 10: log 2D (**A**) and log 1D (**B**) conductance histogram, consisting of all 1950 recorded I(s) traces. Note that data in the log 1D histogram was filtered with a low-pass Bessel filter. Histograms are made up of 43 of total 2000 traces.*

sPy-Ru₄M

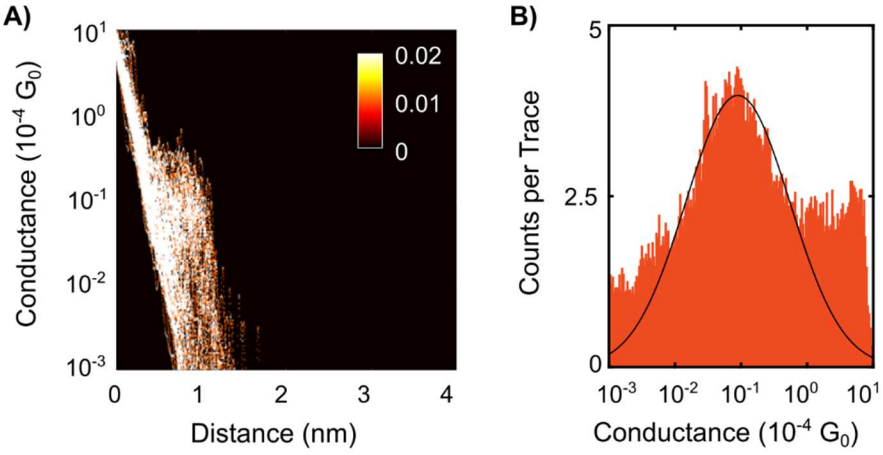


Figure App. 11: log 2D (A) and log 1D (B) conductance histogram, consisting of all 1950 recorded I(s) traces. Note that data in the log 1D histogram was filtered with a low-pass Bessel filter. Histograms are made up of 95 of total 2000 traces. Data was selected with the ln PF algorithm.

lPy-Ru₄M

The point density corresponding to a projection of this point cloud onto the $\Delta X/\theta$-plane shows two clusters, but strong overlap along this viewing direction makes it difficult to distinguish in this representation (Figure App. 12).

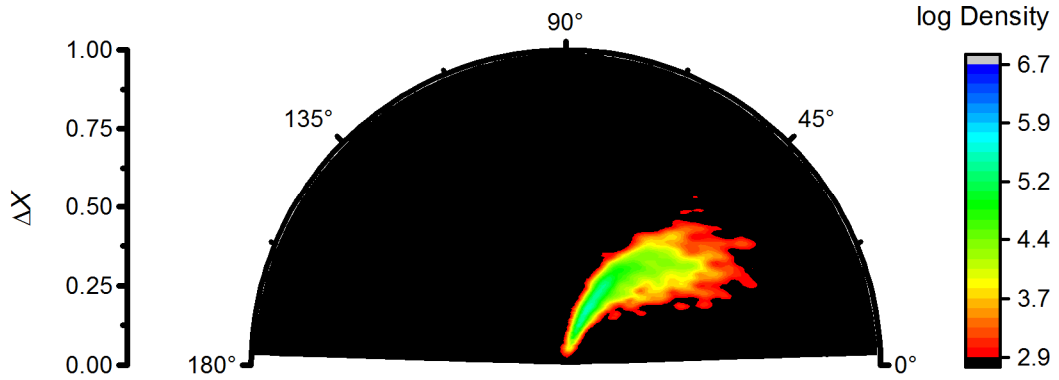


Figure App. 12: Logarithmic point density of the cylinder plot projected onto the ΔX/θ-plane. The view is along the h_r-axis. The semi-circular cluster pointing towards the origin contains plain exponentials. The right diffuse end of this cluster overlaps with the plateau cluster. (50 radial and 50 angular bins)

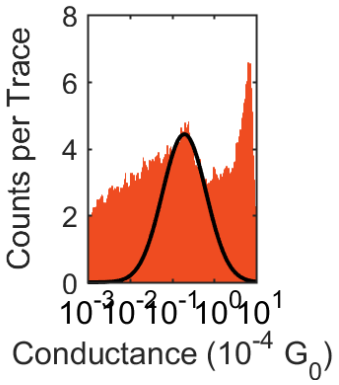


Figure App. 13: Log 1D conductance histogram with Gaussian fit of the data set presented in the main text. Data was selected with FCM GK.

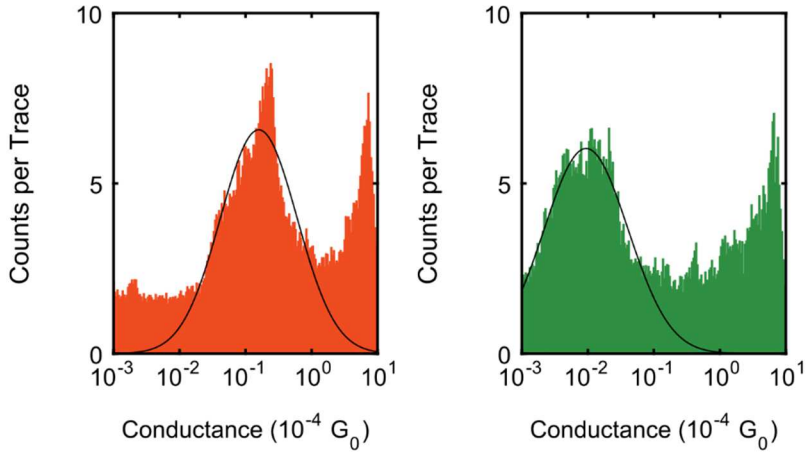


Figure App. 14: 1D log conductance histogram of the high (left) and low (right) conductance group. Data was filtered with a low-pass Bessel filter.

T-Ru$_4$M

The event distribution of the control experiment is shown in Figure App. 15. It is qualitatively different than the one after incubation in that is appears to be more uniform.

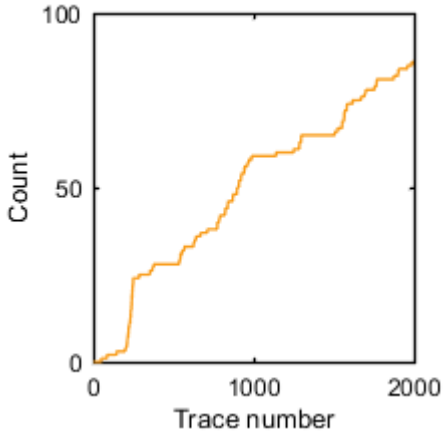


Figure App. 15: Event distribution of the control experiment. Plateaus were selected by the plateau finder algorithm.

Appendix II: MATLAB Scripts

MPVC

```
close all; clear;

%% Load The FileList that should be plotted

filename = uigetfile('*.mat','Multiselect','off');
VAR = filename(1:end-4);
% Remove ".mat" to generate the variable name
tmp_struc = load(filename,VAR);
% Loads as structure called "tmp_struc"
FileList = tmp_struc.(VAR);
% Save the FileList as FileList

%% Create Variables needed for plotting

FileList_Mat = [FileList{2,:}];
% Save all I(s) data in "FileList_Mat"
s = FileList_Mat(:,1);                          % Save the first
distance vector (Should be all the same) in m! as "s"
voltage = FileList{3,1}{1,2};                   % Extract the bias
voltage of the first trace
setpoint = FileList{3,1}{1,1};                  % Extract the setpoint
current of the first trace
I_sign = (-1)* (voltage/abs(voltage));          % Check for the sign
of the tunneling current
I = (I_sign)*FileList_Mat(:,2 : 2 : end);       % Save all current
vectors in matrix "I"
[N, M] = size(I);                               % M is the number of
Files, N is the number of Datapoints per trace
DAT = cat(2,s,I);                               % Save one distance
vector and all current vectors in "DAT"

%% Define Region of Interest if needed and normalize

ini = 1;                                    % First data point in region
of interest (e.g. 100, set 1 for full trace)
to = 1000;                                  % Last data point in
region of interst (e.g. 1000, set N for full trace)
roi = to-ini+1;                             % Define the region of
interest
I_red = I(ini:to,:);                        % Cut current matrix to
region of interest
s = s(ini:to,1);                            % Cut distance vector to
region of interest

R = 100;                                    % Maximum Scanner Current
for Normalization
normalization_factor = 1/(sqrt(roi)*R);     % Calculate Normalization
factor
strom = I_red*normalization_factor;         % Normalize current

%% Create a reference

I0 = setpoint;                          % Setpoint current
```

```
    % I0 = 30
    e = 1E10;                                % Exponential Decay factor (e.g.
1e10)
    ref = I0*exp(-e*s);                      % Create the reference vector
with parameters specified above
    ref_offset = 0;                       % Define offset for reference
vector if needed
    ref = ref+ref_offset;                    % Add offset to reference vector
if needed
    ref = ref*normalization_factor;          % Normalize reference vector

    %% Create a folder to save results

    t = datestr(now,'mmmm_dd_yyyy_HH:MM:SS');       % Take current date
and time to save results
    t = strrep(t,':','-');                           % Replace ":" with "-"
to include in foldername
    regio = sprintf('%i_to_%i',ini,to);              % Create sub_string
for region of interest
    foldername = strcat('Polar_',regio,'_',t);       % Create the
foldername
    mkdir(foldername)                                % Create folder
    cd(foldername)                                   % Enter folder

    %% Calculate the angle between every I(s) vector and Distance between
I(s) and reference vector

    [theta, Y, Y_norm,Punkt_product,Norms_norm] =
VectorDistanceAngle_V5_func(strom,ref);       % Call function to calculate
Theta and Norm

    %% Plot and save Polar Plot of norm and theta

    A_aux = polar(max(theta)+0.1*max(theta), max(Y_norm)+0.1*max(Y_norm));
% Plot an auxiliary point to scale the polar plot correctly
    hold on
    set(A_aux, 'Visible', 'Off');                    % Set auxiliary point
invisible
    A = polar(theta, Y_norm,'.');                    % Plot angles and
norms (angle needs to be in radians)
    savefig('Polar Plot.fig')                        % Save picture as .fig
    saveas(A,'Polar Plot', 'tif');                   % Save picture as .tif
    close all;                                       % Close polar plot

    %% save results

    save('theta.mat','theta');                              % Save all
thetas in radian
    theta_deg = rad2deg(theta);                             % Theta in
degrees
    save('theta_deg.mat','theta_deg');                      % Save all
thetas in degree
    save('Y_norm.mat','Y_norm');                            % Save all
norms (=DX)
    log_norm = log10(Y_norm);                               % Calculate
log of DX
    save('log_norm.mat','log_norm');                        % Save log of
norm
```

```
    save('ini.mat','ini');                                  % Save first
point of region of interest
    save('to.mat','to');                                    % Save last
point of region of interest
    save('I0.mat','I0');                                    % Save
reference setpoint current
    save('e.mat','e');                                      % Save decay
factor
    save('ref_offset.mat','ref_offset');                    % Save Ref
Offset

    %% Calculate reduced Hamming Distance to Reference Vector

    ham_dist = zeros(M,1);                                  % Preallocate
Hamming Matrix
    Y_binary = (Y./abs(Y)).';                               % Binary Vectors: Y
is the distance vector between Reference Vector and I(s) vector (Note: Not
really binary, 3 states possible)
    for i = 1:M
        tmp = cat(2,ref./abs(ref),Y_binary(:,i));           % Concatenate ith
I(s) vector with reference vector
        ham_dist(i) = pdist(tmp','hamming');                % For each binary
distance vector calculate the distance to binary reference vector
    end
    save('ham_dist.mat','ham_dist');

    %% Create a new FileList and save

    [x, y, z] = pol2cart(theta, Y_norm, ham_dist );
% Convert to Kartesian coordinates
    xyz = [x y z];
% Concatenate
    FileList_Polar_Alle_Daten = [FileList;num2cell(theta_deg');
num2cell(Y_norm');num2cell(ham_dist');num2cell(xyz')];     % Create new
FileList that also has the 3 parameters in rows
    save('FileList_Polar_Alle_Daten.mat','FileList_Polar_Alle_Daten','-
v7.3');                                                 % Save FileList (-v7.3
to enable saveing of large Lists)

    %% Plot ref vs first I vector

    plot(s*1e9,ref,s*1e9,strom(:,1))                    % Plot both reference
and first I(s) vector
    xlabel('Distance (nm)')                             % Label x axis
    ylabel('Normalized Current (a.u.)')                 % Label y axis

    savefig('Ref_vs_Is.fig')
% Save as .fig
    hgexport(gcf, 'Ref_vs_Is.png',hgexport('factorystyle'), 'Format',
'png');   % Save as .png
    close all

    %% Plot and save histograms of theta, norm and Hamming distance

    subplot(2,2,1)              % Histogram of Norm
    histogram(Y_norm,100)
    xlabel('Norm')
    ylabel('Counts')
```

```
    subplot(2,2,2)              % Histogram of log norm
    histogram(log_norm,100)
    xlabel('log Norm')
    ylabel('Counts')

    subplot(2,2,3)              % Histogram of Hamming distance
    histogram(ham_dist,100)
    xlabel('Hamming Distance')
    ylabel('Counts')

    subplot(2,2,4)              % Histogram of theta
    histogram(theta_deg,100)
    xlabel('Theta')
    ylabel('Counts')

    savefig('Histogramme.fig')  % Save histograms as .fig
    close all;                  % Close figures
```

MPVC Function

```
    function [theta, Y, Y_norm,Punkt_product,Norms_norm] =
VectorDistanceAngle_V5_func(strom,ref)
    % gives the angle theta in rad and the norm of the distance vector
    % Input is the normalised current vectors

    %% Preallocate variables for speed

    strom = strom';                 % Transpose current
    M = size(strom,1);              % Number of current vectors
    Y = zeros(M,size(strom,2));     % Preallocate distance vectors
    Y_norm = zeros(M,1);            % Preallocate Norm
    theta = zeros(M,1);             % Preallocate angle
    Norms_norm = zeros(M,1);        % Preallocate length of distance
vector x length of ref
    Punkt_product = zeros(M,1);     % Preallocate dot product

    %% Calculate theangle between each Yi and the vector from average to
ref

    refNorm = norm(ref);                                            %
Calculate norm of reference vector
    for i = 1:M
        Y(i,:) = strom(i,:) - ref';                                 % Distance
Y between current vector and reference
        Y_norm(i,:) = norm(Y(i,:));                                 %
Calculate norm of distance vector
        Punkt_product(i,1) = (-Y(i,:)*ref);                         %
Calculate dot product (-Y because vector points towards the reference
point)
        Norms_norm(i,:) = (Y_norm(i,:))*refNorm;                    %
Preallocate length of distance vector x length of ref
        theta(i,:) = acos(Punkt_product(i,1)/Norms_norm(i,1));  % Inverse
cosine in rad
    end

    end
```

Plateau Finder

```
    %% Load The FileList that should be plotted
    clear
    close all

    filename = uigetfile('*.mat','Multiselect','off');              %
Select a FileList
    VAR = filename(1:end-4);                                        %
Remove ".mat" to generate the variable name
    tmp_struc = load(filename,VAR);                                 %
Loads as structure called "tmp_struc"
    FileList = tmp_struc.(VAR);                                     % Save
the FileList as FileList

    %% Create folder to save the results in

    st = strcat('PF_lin_',VAR);    % Create string for foldername
    mkdir(st);                      % Create folder
    cd (st)                         % Enter folder

    %% Create Variables needed for plotting

    FileList_Mat = [FileList{2,:}];                 % Save all I(s) data
in "FileList_Mat"
    s = (1e9)*FileList_Mat(:,1);                    % Save the first
distance vector (Should be all the same) in nm as "s"
    voltage = FileList{3,1}{1,2};                   % Extract the bias
voltage of the first trace
    setpoint = FileList{3,1}{1,1};                  % Extract the setpoint
current of the first trace
    I_sign = (-1)* (voltage/abs(voltage));          % Check for the sign
of the tunneling current
    I = (I_sign)*FileList_Mat(:,2 : 2 : end);       % Save all current
vectors in matrix "I" (Multiply with current sign so the histograms are
comparable)
    [N, M] = size(I);                               % M is the number of
Files, N is the number of Datapoints per trace

    off_set = median(mean(I(N-round(N/10):N,:)));
    I = I-off_set;

    G = I/voltage;                  % Calculate the conductance from
positive current and bias voltage

    %% Double-checking voltage, setpoint and distance vector. There should
not change during an experiment!

    s_all = FileList_Mat(:,1:2:end-1);              % Save all distance
vectors in s_all
    s_instances = size(unique(s_all),2);            % Number of different
s vectors (just to check if the macro slipped - should be 1)
    par = cell2mat([FileList{3,:}]);                % Save all setpoints
and voltages in par
    setpoint_all = par(1,1:2:end-1);                % All setpoints
    voltage_all = par(1,2:2:end);                   % All voltages
    Sp_instances = size(unique(setpoint_all),2);    % Number of different
setpoint currents (should be 1)
```

```
    V_instances = size(unique(voltage_all),2);      % Number of different
voltages (should be 1)

    % Display waring if setpoint, bias or distance vector changed during
measurement (Due to macro slipping for example)
    if s_instances + Sp_instances + V_instances > 3
        disp('Warning! More than one instance of distance vector, voltages
or setpoint current!');
    end

    %% Define Parameters

    BinWidths = ((0.1:0.1:0.4))';          % Define Bin Widths in nS for
classification
    Counts = (30:10:140)';                              % Define Bin
Counts for classification

    %% Set the bin ranges for the histograms

    extra = setpoint/10;                                          % Add 10 %
extra Range for plotting
    rangemax = abs(setpoint+extra);  % Set the correct max and the correct
sign
    rangemin =  (0-extra);         % Set the correct min and the correct
sign

    % Calculate the bin edges
    Bin_Edges = cell(1,4);
    for j =1:4
        Bin_Edges{j} = rangemin:BinWidths(j,1):rangemax;        % Row cell
array, with cells containing the bin edges in rows
    end

    %% For each I(s) trace create a 4x12 matrix that contains 1 or 0 if
plateau is found or not

    for m=1:M   % Loop through all M files and find plateaus!
        PF_Mat = zeros(12,4);                    % PlateauFinder Matrix is
cell that contains "1" if trace contains plateau for parameters specified
in Par_Mat
        single_graph_bincount = cell(4,1);      % Preallocate cell for
single-trace histograms

        for i=1:4
            single_graph_bincount{i,1} = histc(I(:,m),Bin_Edges{1,i});  %
Calculate the single-trace histograms for each BinWidth
        end

        % Find peaks in single-trace histograms for the four binwidths
        for i=1:4
            [pkcount, pkpos] = findpeaks(single_graph_bincount{i,1},
'threshold',1);    % find all peaks, pks is count pkpos is the position

            % Disregard first peak that corresponds to exponential tail
            if size(pkpos,1) > 1;                                        %
Exclude first peak (Part of exponential)
                pkcount_wo_tail = pkcount(2:end);
```

```
                % Compare peak count to PDBCs
                max_peak = max(pkcount_wo_tail);
                for j = 1:12
                    if max_peak >= Counts(j)
                        PF_Mat(j,i) = 1;
                    end
                end
            end
        end
        FileList(7,m) = {PF_Mat};
    end

    sum_PDBC = cellfun(@sum, FileList(7,:),'UniformOutput', false);
    sum_all_Par = cellfun(@sum, sum_PDBC,'UniformOutput', false);
    sum_all_Par = cell2mat(sum_all_Par);

    index_p=find(sum_all_Par>0);
    index_no_p=find(sum_all_Par==0);

    FL_Plateau = FileList(:,index_p);
    FL_noPlateau =  FileList(:,index_no_p);

    save('FL_Plateau.mat','FL_Plateau');                % Save FileList
with Plateaus
    save('FL_noPlateau.mat','FL_noPlateau');            % Save FileList
without Plateaus
    cd ..
```

Appendix III: Permission Documents

Table App. 1: Summary of permission status of third party content.

Figure Number	Reference	Copyright	Permission	License Number
1.4	*www-03.ibm.com/press/us/en/photo/40982.wss*	*IBM*	*Granted*	*NA*
1.8	*"Electrochemically-gated single-molecule electrical devices", Guo, Shaoyin et al., Electrochimica Acta, 2013*	*Elsevier*	*Granted*	*4107360759758*
1.12	*"Transistor-like behaviour of Transition metal complexes", T. Albrecht et al., Nano letters, 2005*	*American Chemical Society*	*Granted*	*NA*
3.1-3.6 *3.8-3.11* *3.13-3.16*	*"Unsupervised vector-based classification of single-molecule charge transport data", Lemmer et al., Nat. Commun., 2016*	*Nature Publishing Group*	*Creative Commons CC-BY License*	*NA*