Computationally Interrogating Enzyme Electrochemistry for Rational Mutation of Metalloproteins



Henry Lloyd-Laney

Wadham College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

Abstract

This thesis was inspired by an interest in the metalloenzymes: profoundly complex catalysts that perform the redox-mediated small molecule conversions that underpin the biosphere. The principal methodology is computational analysis of protein-film voltammetry (PFV), an experimental technique wherein an electro-active protein (such as a metalloenzyme) is chemically linked to an electrode, and subjected to a time-varying potential input, with a corresponding current response from two processes; double-layer capacitance and electron-hopping Faradaic reactions (catalytic processes also generate current, but are beyond the scope of the thesis). It is the latter that is of biological interest — electron-hopping Faradaic current arises from the reduction and oxidation of the metal atoms within the protein. The current generated from PFV experiments has a low Faradaic-to-background ratio and is not readily interpretable — hence the need for computational analysis. By using a model that describes the current response to a potential input as a function of the biochemical reaction parameters (rates and energetics), the aim is to fit experimental voltammetry data. Inferring the values of the parameters, given this experimental data is referred to as "solving the inverse problem". Additionally, Bayesian statistical analysis methods are used to infer the likely distributions of these inferred parameters, which allows us to understand the degree of confidence we should have in our fits, as well as the degree of correlation between the various parameters. This is a computationally intensive process, requiring many simulations of the experimental current, and is complicated by the low signal-to-noise ratio and the fact that noise processes can obscure useful information. The choice of potential input is key; there are trade-offs between the amount of information that can be obtained about the (bio)chemical system under interrogation, complexity of analysis required, and simulation speed.

In chapter 1, we introduce the research problem, and provide a brief description of each chapter. In chapter 2 we lay out the necessary theory for understanding the thesis, along with a description of our approach to solving the inverse problem. In chapter 3, we detail theoretical work undertaken to validate the simulation and inference procedures described in chapter 2, and explore the impact of some assumptions made when deriving the simulation model. Chapter 4 details

a quantitative characterisation of a voltammetry technique which we term "Purely Sinusoidal Voltammetry" (PSV). Using data acquired for a model chemical system, we demonstrate that PSV is faster to simulate than more information-rich techniques, while being more sensitive to the key electron-transfer reaction parameters than simpler voltammetry approaches. In chapter 5, we use this technique to infer parameters from data generated by a bacterial cytochrome, and detail the adaptations required when dealing with the lower signal-to-noise ratio inherent to PFV experiments, ultimately suggesting a framework for inferring parameters in this context. Finally, in chapter 6, we show that the methodology developed in chapter 5 can be used to capture the electrochemical impact of mutating key residues of the bacterial cytochrome.

COVID-19 impact statement: I was recruited for this DPhil project on the basis of my background as a biochemist. I was meant to generate my own experimental data at the University of York. However, this was impeded by the Covid-19 pandemic: the University of York was not allowing access to academic visitors until August 2021, and consequently I was not able to undertake experimental work until two months before submission of this thesis (the preliminary data which was collected during this period is analysed in chapter 6). Additionally, there were several months in which I did not have experimental data to analyse, which was the result of lab-space restrictions and general lockdown. Originally, this thesis was meant to culminate with an analysis of metalloenzymes known as hydrogenases, which are highly complex. However, the slow pace of data generation has meant that we altered the focus of the biological aspects of the thesis to analysis of the bacterial cytochromes analysed in chapters 4 and 5. These proteins are much easier to work with experimentally, and undergo less complex electrochemical reactions. Additionally, I undertook several projects that could be achieved entirely computationally without the need for experimental data. This includes section 3.3 and appendix C.

For my parents.

Every sentence was written to honour them.

Even the crap ones.

Acknowledgements

Let's get two things of the way: first, doing a DPhil during a global pandemic sucks! I do not recommend it. Secondly, I'm not allowed to put any jokes in the rest of the thesis, so they're all going in the acknowledgements, undoubtedly to its detriment. With that in mind, let's get to it.

Overall I've been absurdly fortunate in my time at Oxford — I have been given a great deal of support, both intellectual and moral, by almost everyone I've met (and all they get are these lousy words of thanks — there's no justice). First and foremost are my supervisors, David Gavaghan, Alison Parkin and Martin Robinson, who took a punt on me as a decidedly un-numerical biochemist, and have variously inspired, cajoled, counselled and championed me, often simultaneously. I must thank them for introducing me to this compelling field of study, for being so engaged and invested in my work, and for reading absolutely everything I've written over the past three years. Anything that's good or informative in this thesis is down to them. It's been a mostly joyful three and a bit years, and I feel very fortunate to have the opportunity to continue working with them. I'm also deeply indebted to Alan Bond, firstly for inventing the wonder that is FTACV (and not minding too much that I stick an "r" in front of it), and for being so generous with his trove of electrochemistry knowledge. All of this was facilitated by the Oxford Synthetic Biology CDT, and the directors Robert Carlisle and Antonis Papachristodoulou. The first year of the course, with almost total (stipendiary!) academic freedom, was liberatory, and none of what came afterwards would have been possible without it. I even managed to squeeze some synthetic biology in right at the very end. Thanks are due also to Nicholas Yates and Alice Hewson who both did the fiddly, frustrating work of collecting the data in chapters 4 and 5, leaving me free to gad about in cyberspace, and weren't even a little bit resentful.

In terms of those who carried me through life outside study, Eddy Yeo has seen me at my best (getting a sticker from a DJ for good dancing), my worst (trying not to fall off a crag using the power of my thigh grip alone) and everything else besides, and remains both my friend and one of the best people I know. Mats Ahrenshop is endlessly willing to share his knowledge of economics, anthropology and experimental design, and is a very supportive climbing partner as I whimper my way up the wall. They're both such smart and good people, and I love them very much. Despite being separated by many miles, I'm also very lucky to be friends with the other two "lads on a wave", Ezra Kitson and Josh Lawrence, who besides being the first port of call for any question or observation I have, were also kind enough not to point out that "Drizal-Drizal" is only "lizard-lizard" backwards if you can't read. I'm also grateful to those who were with me for only part of my Oxford journey; Georgia Mason, who gave me a different way to see the world and everyone in it, and Claire Noble, Ciaran Guy and Alan Reed, who made those formative first six months such a joyous and cynical delight. Eddy, Mats, Josh, Ezra, Ciaran and Claire were all kind enough to proofread my thesis and provide excellent feedback, with the gold star for masochism going to Alice and Alister Dale-Evans who read the whole thing.

Lastly, thanks and endless love to my family, Robin Laney and Megan and Oscar Lloyd-Laney, firstly for being my family and secondly for being eternally engaged in my words, works and life. I look forward to telling you that there isn't an application, or in fact a point, in anything that I do for the rest of my academic career.

Contents

1 Introduction **29**

1.1 Preamble . . . 29

1.2 Background . . . 30

1.3 Thesis structure . . . 32

1.3.1 Chapter 2 — Literature review and theoretical background . . . 32

1.3.2 Chapter 3 — Theoretical results . . . 32

1.3.3 Chapter 4 — Ferrocene . . . 33

1.3.4 Chapter 5 — Cytochromes . . . 33

1.3.5 Chapter 6 — Rational mutation . . . 34

1.3.6 Chapter 7 — Conclusions . . . 34

2 Literature review and theoretical background **35**

2.1 Voltammetry . . . 35

2.1.1 Protein-film voltammetry . . . 36

2.2 Modelling . . . 37

2.2.1 Butler-Volmer kinetics . . . 39

2.2.2 Capacitance . . . 42

2.2.3 Non-dimensional constants . . . 42

2.2.4 Voltammetric waveforms . . . 43

2.2.5 Fourier transforms and voltammetry . . . 45

2.2.6 Numerical details . . . 48

2.2.7 Other modelling approaches . . . 49

2.3 The inverse problem . . . 51

2.4 Research approach . . . 53

2.4.1 Optimisation . . . 53
2.4.2 Bayesian inference . . . 55

3 Theoretical results **61**

3.1 Comparison of numerical and analytical approaches for simulating purely sinusoidal voltammetry . . . 61
3.1.1 Derivation of simulation models . . . 62
3.1.2 Derivation of the analytical approximation . . . 63
3.1.3 Comparing the analytical and numerical models of PSV . . . 66
3.2 Exploring the forwards problem . . . 69
3.2.1 Reversible potential — E^0 . . . 70
3.2.2 Rate constant — k_0 . . . 72
3.2.3 Sinusoid frequency — ω . . . 75
3.2.4 Uncompensated resistance — R_u . . . 76
3.2.5 Capacitance — C_{dl} and C_{dlEX} . . . 78
3.2.6 Surface coverage — Γ . . . 83
3.2.7 Symmetry factor — α . . . 84
3.2.8 Phases — η and C_{dl} η . . . 85
3.3 The impact of dispersion — a spotter's guide . . . 85
3.3.1 Modelling approach . . . 89
3.3.2 Results and Discussion . . . 90
3.4 Inference on synthetic noisy data . . . 110
3.4.1 Noise . . . 110
3.4.2 Total current fitting . . . 111
3.4.3 Fourier domain fitting . . . 114
3.5 Conclusions . . . 116

4 Ferrocene **117**

4.1 Introduction . . . 117
4.2 Experimental methods . . . 118
4.3 Results and discussion . . . 119
4.3.1 Inferring parameters from experimental data . . . 125
4.3.2 Model modifications . . . 126
4.3.3 Discussion of values inferred by CMAES . . . 128

4.3.4 Bayesian inference . . . 133
4.4 Conclusions . . . 139

5 Cytochromes **140**
5.1 Introduction . . . 140
5.2 Experimental methods . . . 141
5.3 Results and discussion . . . 144
5.3.1 Preliminary investigations . . . 144
5.3.2 Constraining the search space . . . 148
5.3.3 Inferring parameters from multiple films . . . 155
5.3.4 Bayesian inference . . . 166
5.3.5 DCV analysis . . . 170
5.4 Conclusions . . . 176

6 Rational mutagenesis **180**
6.1 Introduction . . . 180
6.1.1 Heme properties - thermodynamics . . . 180
6.1.2 Heme properties - kinetics . . . 183
6.2 Rational mutants . . . 183
6.2.1 M66H . . . 183
6.2.2 R44N . . . 185
6.3 Experimental methods . . . 185
6.3.1 Fitting time . . . 185
6.4 Results and discussion . . . 186
6.4.1 M66H . . . 188
6.4.2 R44N . . . 189
6.5 Conclusions . . . 191

7 Conclusions **192**
7.1 Summary . . . 192
7.2 Future work . . . 194
7.2.1 Unresolved questions . . . 194
7.2.2 More complex metalloproteins . . . 195
7.3 Final thoughts . . . 197

Appendices **209**

A The discrete Fourier transform of Gaussian noise **210**

B The CMAES algorithm **213**

B.1 Overview . . . 213
B.2 Updating the mean . . . 216
B.3 Updating the covariance matrix . . . 217
B.4 Rank-μ update . . . 217
B.5 Rank-one update . . . 218
B.6 Step-size control . . . 219

C Assessing the impact of noise on heuristic parameter inference approaches **221**

C.1 Approach . . . 222
C.1.1 Heuristic inference . . . 222
C.1.2 Markov-chain Monte Carlo . . . 223
C.1.3 Noise . . . 224
C.2 DCV . . . 225
C.2.1 Surface coverage . . . 225
C.2.2 Rate constant . . . 225
C.3 SWV . . . 226
C.3.1 Reversible rate constants . . . 227
C.3.2 Symmetry factor . . . 229
C.3.3 Quasi-reversible rate constants . . . 230
C.4 Simulation information . . . 231
C.5 Results and Discussion . . . 231
C.5.1 DCV: Surface coverage . . . 231
C.5.2 DCV: Kinetic parameter . . . 234
C.5.3 SWV: Kinetic parameter, reversible case . . . 235
C.5.4 SWV: Symmetry factor . . . 236
C.5.5 SWV: Kinetic parameter, irreversible case . . . 237
C.6 Conclusions . . . 237

D Phase in electrochemical impedance spectroscopy **240**

D.1 Introduction 240

D.2 Impedance components 240

D.3 Equivalent circuit analysis 243

D.4 Equivalent circuit models for a redox process 243

List of Figures

2.1 3 MCMC chains for 6 parameters as listed on the y-axis, collected when fitting synthetic data in the Fourier domain. The $\hat{R}$ parameter is listed on the top of each graph, without burn-in, and after discarding of the first 2500 samples (indicated with a black dotted line). 60

3.1 Current simulated using analytical and numerical methods against time. The value of the input frequency is shown in the title of each plot, and all other parameters are as in table 3.1, with $\tilde{k}_0$ set to 10 s^{-1}. 67

3.2 Percentage difference between the analytical and numerical currents as function of the kinetic parameter k_0 and the input frequency. The kinetic value $\tilde{k}_0$ in units of s^{-1} is shown in the legend, with the rest of the simulation parameters reported in table 3.1. 68

3.3 Parameter scan for E^0, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the reversible potential E^0 which is as indicated in the legend. 71

3.4 Parameter scan for k_0, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the rate constant k_0 which is as indicated in the legend. . 74

3.5 Parameter scan for the effect of altering the potential frequency ω on PSV and r-FTACV only, as this parameter is not relevant for DCV 75

3.6 Parameter scan for R_u, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the uncompensated resistance R_u which is as indicated in the legend. 77

3.7 Parameter scan for C_{dl}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the linear double-layer capacitance C_{dl} which is as indicated in the legend. 79

3.8 Parameter scan for C_{dlE1}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the higher-order capacitance term C_{dlE1} which is as indicated in the legend. 80
3.9 Parameter scan for C_{dlE2}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the higher-order capacitance term C_{dlE2} which is as indicated in the legend. 81
3.10 Parameter scan for C_{dlE3}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the higher-order capacitance term C_{dlE3} which is as indicated in the legend. 82
3.11 Parameter scan for Γ, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the surface coverage Γ which is as indicated in the legend. 83
3.12 Parameter scan for α, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the symmetry factor α which is as indicated in the legend. 84
3.13 Parameter scan for η, for PSV and r-FTACV. Simulation parameters are as in table 3.2, except the Faradaic phase η which is as indicated in the legend. The r-FTACV plots incorporate both the absolute harmonics as a dotted line, and the imaginary harmonics, as the latter shows the impact of the phase parameter. . . 86
3.14 Parameter scan for $C_{dl}\eta$, for PSV and r-FTACV. Simulation parameters are as in table 3.2, except the capacitance phase $C_{dl}\eta$ which is as indicated in the legend.The r-FTACV plots incorporate both the absolute harmonics as a dotted line, and the imaginary harmonics. 87
3.15 A representation of numerical integration of a Gaussian distribution 90
3.16 Plots of simulated r-FTACV harmonics vs time , with dispersion parameters for E^0 — mean, standard deviation and skew $(E^0_\mu, E^0_\sigma, E^0_\kappa)$ and k_0 — shape and scale $(\log(k_0)\mu, \log(k_0)\sigma)$ where the parameter being varied is in the top left hand corner. The resulting distribution, where f(parameter) is the assumed distribution of the parameter, is shown at the top of the plot, and the 3^{rd}, 5^{th} and 7^{th} harmonics shown. The colour of the harmonic corresponds to the colour of the distribution. Simulation parameters are found in table 3.3. For all cases, the kinetics of the reaction are in the quasi-reversible to reversible regimes. 94

3.17 Plots of simulated r-FTACV harmonics vs time, where the parameter distributions under investigation are plotted in the top panel of each plot subgroup. The remaining simulation parameters can be found in table 3.3. These simulations were generated using irreversible kinetic values. 95
3.18 Parameter scans for a PSV current-potential plot, where the parameter distributions under investigation are plotted in the top panel of each plot subgroup. The remaining simulation parameters can be found in table 3.3. 98
3.19 Parameter scans for a PSV current-potential plot, where the parameter distributions under investigation are plotted in the top panel of each plot subgroup. These simulations were generated using irreversible kinetic values. 99
3.20 Parameter scans for a DCV current-potential plot (left), and trumpet plots (right) which plot the potential-values of the maximum and minimum current peaks against the base 10 logarithm of the scan rate. The inset plot in the right hand column is a current(μA)-potential(V) plot of a DCV simulation using a scan rate from the middle of the range defined in the x-axis of the containing plot. The remaining simulation parameters can be found in table 3.3. 100
3.21 Parameter scans for a DCV current-potential plot (left), and trumpet plots (right) which plot the potential-values of the maximum and minimum current peaks against the base 10 logarithm of the scan rate. The inset plot in the right hand column is a current(μA)-potential(V) plot of a DCV simulation using a scan rate from the middle of the range defined in the x-axis of the containing plot. These simulations were generated using irreversible kinetic values. 101
3.22 Comparison of different forms of dispersion for r-FTACV, PSV and DCV. Each column represents a different form of dispersion, compared against the non-dispersed case (labelled as "None"). From left to right, the dispersed parameters are E^0, k_0 and a combination of the two, as listed in the legend of the middle row. All parameters are as in table 3.3. Harmonics were generated without the usage of the Hann window. 103

3.23 Comparison of different forms of dispersion for r-FTACV, PSV and DCV. Each column represents a different form of dispersion, compared against the non-dispersed case (labelled as "None"). From left to right, the dispersed parameters are E^0, k_0 and a combination of the two, as listed in the legend of the middle row. For this plot, the values for k_0 in the non-dispersed case, and $\log(k_0)\mu$ in the dispersed case have been set to 1000 s^{-1}, i.e. totally reversible at the input frequency in table 3.3. 104
3.24 Comparison of different forms of dispersion for r-FTACV, PSV and DCV in the quasi-irreversible kinetic regime. Each column represents a different form of dispersion, compared against the non-dispersed case (labelled as "None"). From left to right, the dispersed parameters are E^0, k_0 and a combination of the two, as listed in the legend of the middle row. k_0 and $\log(k_0)\mu$ are both set to 1 s^{-1} as appropriate 105
3.25 The dimensionless root mean squared difference observed between simulated PSV current in the kinetically dispersed and non-dispersed cases, as a function of varying frequency, for varying values of the shape and scale parameters of the log-normal distribution of k_0, as shown in the legend. 106
3.26 (i) Single-electron redox process, as modelled throughout this thesis. (ii) EC model, incorporating Faradaic and catalytic processes, as modelled in equation 3.28. 108
3.27 Simulated current for a r-FTACV, PSV and DCV experiment from left to right, simulated according to an EC model. The forward and backward first-order catalytic rates constants (k_2 and k_{-2}) were set at 80 s^{-1}, and all other parameters as in table 3.3. 108
3.28 Comparison between r-FTACV, PSV and DCV experiments simulated using a log-normal and a log-uniform distribution, with the two distributions shown on the left-most figure. Parameters are as in table 3.3. For the log-uniform distribution, the upper and lower values chosen were the upper and lower values used to approximate the log-normal distribution. 109
3.29 Simulated and experimental harmonics 3-6 from r-FTACV current. The left hand side was simulated using dispersion (using the r-FTACV parameters described in chapter 4). On the left, simulations were generated by setting the $E^0\mu$ value as the E^0 value, and reducing the value of the surface concentration Γ by half. . . 110

3.30 Parameter distributions inferred from synthetic noisy data for DCV, PSV and r-FTACV, where the magnitude of the noise added was 0.5% of the maximum current value. The true parameter value is indicated by an arrow on each histogram. 111
3.31 Parameter distributions inferred from synthetic noisy data for DCV, PSV and r-FTACV, where the magnitude of the noise added was 4% of the maximum current value. The true parameter value is indicated by an arrow on each histogram. . . 112
3.32 Parameter distributions inferred from fitting to PSV current harmonics 1-10 and 4-10 with 0.5% noise added to the total PSV current, where the true parameter values are shown as black dotted lines. 115
3.33 Parameter distributions inferred by fitting to synthetic r-FTACV current harmonics 1-10 and 4-10, with 0.5% noise added to the total r-FTACV current, where the true parameter values are shown as black dotted lines. 116

4.1 Cartoon representing the diazonium functionalisation procedure used to generate a ferrocene-coated glassy carbon electrode surface. 118
4.2 Simulation vs experimental data for experiments 1 and 2, with parameter values as reported in table 4.1. The data and simulation are plotted as a current vs. time plot, semi-logarithmic plots of the real and imaginary parts of the absolute Fourier spectrum, and the 1-7th harmonics. 122
4.3 Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 3 and 4, using parameters as reported in table 4.1 123
4.4 Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 5 and 6, using parameters as reported in table 4.1 123
4.5 Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 7 and 8, using parameters as reported in table 4.1 124
4.6 Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 9 and 10, using parameters as reported in table 4.1 124
4.7 Inferred normal distributions for E^0 (left) and α (right) for each experiment, as determined by the values obtained by the CMAES fitting process (shown in table 4.1) 127

4.8 Blank electrode data (blue) overlaid with fitted capacitive current, when the phase was fitted explicitly as a parameter (orange), and when the phase was fixed at the experimentally set phase of $\frac{3\pi}{2}$ (green). Harmonics 1-5 of the three current-voltage plots are shown on the right. 128
4.9 DCV current-voltage plot, showing the total DCV current (black), the background subtracted current (dashed), the polynomials used to obtain the background-subtracted current, and the fitted waveshape. 130
4.10 Comparison of harmonics 2-6 for r-FTACV experimental data (orange) with the output of simulations (blue). The DC ramp was applied over the potential range of -0.18 to +0.62 V vs Ref at a scan rate of 29.8 mV s^{-1}. The superimposed sine wave had an amplitude of 150 mV and frequency of 8.94 Hz. The simulation parameters used for the left hand side plots were as for experiment 10 in table 4.1, except for ω and phase, which were set at 8.885Hz and 0 rads respectively, as obtained by fitting these parameters for r-FTACV data. The parameter values used to simulate the harmonics on the right hand side are provided in table 4.1. and were inferred by CMAES fitting of the 4-6th harmonics of the experimental data (shown in orange). 131
4.11 Left — current-potential harmonics from PSV experiment 10, plotted against PSV harmonics simulated using the parameters in table 4.4. Right — current-time harmonics from r-FTACV experiment 1 plotted against simulated r-FTACV harmonics simulated using the same parameters (except for the input frequency and phase, which were set to 8.885Hz and 0 rads respectively) 134
4.12 2D scatter plots for individual parameters used to infer parameter distributions from PSV experiment 10, which is plotted in figure 4.6 on the bottom row. The diagonal represents the 1D histograms for each parameter. The MCMC used to generate these chains were run for 100000 samples, with the first 50000 discarded as burn-in. 136
4.13 2D scatter plots showing individual Markov chains for each parameter plotted against each other, with their individual histograms being shown on the diagonal. Chains were run for 100000 iterations and the first 50000 of these samples were discarded as burn-in. 138
4.14 Inferred distributions of parameters when $\alpha\mu$ is fitted as a parameter vs when it is fixed at the value inferred by the CMAES. 138

5.1 DCV (left), PSV (centre) and r-FTACV (right) experimental data. A-C: Potential inputs for the three experiments. D-F: Total current response of *Cj*X183 (blue) and a bare PGE electrode (red) to the three experimental inputs. G: DCV peak position data from *Cj*X183. H and I: magnitude of the Fourier spectrum corresponding to harmonics 1-10 for *Cj*X183 (blue) and a bare electrode (red) for PSV and r-FTACV. J and K: PSV and r-FTACV harmonics 1-6 from *Cj*X183 (blue) and a bare electrode (red) 143

5.2 A: PSV total current, and two simulations with highly differing parameter values obtained by fitting to PSV total current (referred to as ϕ_1 and ϕ_2, the parameter values of which can be found in table 5.1). When observing the fit in harmonics 1-6 (B), we see that the relatively good fit between simulation and experiment in the time-domain is not replicated in the higher harmonics. 145

5.3 An example of how a relatively good agreement between PSV experimental (blue) and simulated (orange) harmonics 4-7 (A) does not translate well to an r-FTACV harmonic comparison (B), driven by a low inferred k_0 value. 148

5.4 R-FTACV harmonics, as shown in figure 5.1, plotted against the linear portion of the r-FTACV potential input. The approximate mid-point of the harmonics can be read off the x-axis (in this case, we have drawn a line at -0.065V), and used to define E^0_μ boundaries 149

5.5 Comparison of two plausible parameter vectors generated by fitting to PSV harmonics 4 and above. In A, we see the very good agreement between simulated and experimental PSV harmonics 4-7 for both parameter vectors. Neither vector is clearly better than the other when compared against r-FTACV experimental harmonics in B. The "high C_{dl}" parameter was discounted on the evidence figure C, as there is a significant, unrealistic quantity of spectral bleed from the fundamental harmonic of the "high C_{dl}" simulation. 152

5.6 Single oscillation of PSV current and potential on the same time axis. The phase difference can be determined by taking the time-difference between the peaks of the two waves. As the capacitive current is proportional to the derivative of the potential, an ideal current will be shifted by 90 degrees. 155

5.7 Flowchart representation of the fitting methodology proposed in this chapter. For certain parameters we have presented graphically how the boundary values can be obtained; these figures are written in brackets as appropriate. 156

5.8 Best-fit simulations and data for harmonics 4-10 of PSV experiments 1-3 (figures A-C), using simulation parameters reported in table 5.3. Figures D-F show r-FTACV harmonics 4-7, from r-FTACV experiments 1-3 (performed using the same *Cj*X183 modified electrode as the appropriate PSV experiment), and simulations obtained using the same parameters as those used to generate the top row, except for the values of the phases (both of which were set to 0) and the input frequency (which was set to 8.96, 8.75 and 8.83 Hz respectively). Figures G-I show the same r-FTACV harmonic data as figures D-F, along with best-fit simulations, obtained using the parameters in brackets in table 5.3. 157

5.9 PSV and translated r-FTACV harmonic 6, where the upper bound of α was set at the value indicated in the r-FTACV plot. Simulation parameters can be found in table 5.4 161

5.10 Two-step fits to PSV (top row) and r-FTACV (bottom) harmonics, using the parameters listed in table 5.5, expect in for the input frequency and phases for the r-FTACV case, which are as listed in table 5.3. 163

5.11 Values inferred using two methods for PSV experiments 1-3, as in table 5.5 and table 5.3, normalised so that the lower bound (also in table 5.3) is 0 and the upper bound is one. The box plot values are taken from MCMC-inferred distributions for the simultaneous case, where this was available. As the upper bound for C_{dl} was different for the two fitting approaches, we have recorded the upper bound for the simultaneous effort as a red line. 165

5.12 Inferred parameter histograms generated by pooling three independent MCMC chains, using harmonics 4 and above of the respective PSV experiments in the likelihood function, and using the parameters in table 5.3 as the starting point. C_{dl} and α were not included in this parameter inference approach for technical reasons mentioned in the text. The DCV histograms were as inferred from running an MCMC process on the trumpet data in figure 5.18, and as such we can only infer the parameters E^0 and k_0. 167

5.13 2D histograms generated from the MCMC process for PSV experiment 1 168

5.14 MCMC-inferred histograms for PSV experiment 2. 169

5.15 MCMC-inferred histograms for PSV experiment 3. 170

5.16 The leftmost plot is an example of the background subtraction procedure which attempts to isolate purely Faradaic current from a DCV experiment. We interpolate across the region of the total DCV current (purple) using a third-order polynomial (red), and subtract this polynomial from the total current to yield the background-subtracted current (orange). In the central plot, we show the same background-subtracted current with two purely Faradaic current simulations. The DCV simulation (blue) was generated using parameters obtained by fitting to the background-subtracted current, and are recorded in table 5.6. The r-FTACV simulation uses parameters listed in table 5.3, using the values in brackets where appropriate. In the leftmost plot, the parameters in table 5.6 fitted to the DCV experiment are used to generate r-FTACV harmonics 4-7, and plotted against r-FTACV data harmonics, alongside the harmonics from the r-FTACV simulation in green. 171

5.17 Four separate attempts to infer the surface coverage parameter from a DCV experiment, using different potential window sizes. The "fitted background" current was subtracted from the total current to obtain the background subtracted current. The area under the peak was then used to estimate the surface coverage, which is shown at the top of each plot. 174

5.18 Bottom: Trumpet plot data and best fit simulation ("MCMC"). Filled circles indicate oxidative peak position, and hollow circles are for the reductive peak position. Additionally, we show the effect of altering the kinetic parameter to the value indicated for PSV 3 in table 5.3 ("PSV kinetics"). Top: r-FTACV harmonic 6, from experimental data, and using the best fit simulation parameters indicated in table 2. For the orange simulation corresponding to the "MCMC" simulation in the lower figure, the kinetic value was lowered to the one shown in table 5.6 (i.e. 109.4 s^{-1}) 175

5.19 The three experiments analysed in this chapter, direct current voltammetry (DCV) as a current vs. potential plot, purely sinusoidal voltammetry (PSV) current harmonics 4-7 vs. potential, and ramped-Fourier Transform Alternating Current Voltammetry (r-FTACV) current harmonics 4-7 vs. time. The experiments are ranked according to how interpretable they are, and the amount of Faradaic information they provide. Orange arrows indicate simulation of current using a vector of chemical parameters ϕ, and black arrows indicate that the results of these simulations are used to assess the goodness-of-fit. 177

6.1 Left: structure of the heme from *Cj*X183, with several key residues included, including the axial/distal Fe ligands H23 and M66, and the propionate coordinating residues, R44 and S50. Nitrogen atoms are in blue, carbon in green, oxygen in red, iron in brown and sulphur in yellow. Right: the cognate heme in the context of the bulk structure of *Cj*X183, represented by a cartoon in blue [78]. Structure taken from protein data bank, accession code 7B21. 181

6.2 Voltammetry data collected for three different forms of *Cj*X183: WT (blue), R44N (orange) and M66H (green). Plots A-C show the potential input as a function of time and plots D-F show the resulting current as a function of potential. Plot G shows trumpet-plot data (DCV peak-position as a function of scan rate) for R44N, and WT data, although it should be noted that the wild-type data is as collected for PSV experiment 3 in section 5.3.5.3, as shown in figure 5.1, and is not part of the WT dataset shown in the rest of the figure. Plots H and I show the absolute Fourier spectrum corresponding to harmonics 0-7 of the PSV and r-FTACV total currents. Plot J shows PSV harmonics 1-6 as a function of the potential input, and plot K shows r-FTACV harmonics as a function of the linear ramp portion of the r-FTACV potential input. 184

6.3 Fits to PSV and r-FTACV harmonics data for M66H (left column) and R44N (right column). Top row: simulated (orange) and experimental (blue) PSV harmonics 4-7, where the simulation parameters are as reported in table 6.1. Middle row: simulated (orange) and experimental (blue) r-FTACV harmonics 4-7, using the same parameters as for the PSV simulations, except the input frequency, which is written in brackets in table 6.1. Bottom row simulated (orange) and experimental (blue) r-FTACV harmonics 4-7 using parameters obtained by fitting to r-FTACV harmonics. The altered parameters are written in brackets in table 6.1 187

6.4 r-FTACV harmonics 3-5 from R44N and wild-type data, normalised so that the minimum value of each harmonic equals 0 and the maximum equals 1 (written as max(I) for clarity) 190

B.1 Iterations of the CMAES algorithm finding the minimum of the 2D Rastrigin function, which has a minimum at [0,0], shown here as a filled contour plot, with blue areas and yellow areas representing local maxima and minima respectively. Crosses in red indicate the highest scoring points that were used as part of the calculation of the covariance matrix $\mathbf{C}^{(g+1)}$, shown as a red ellipse with a scale factor of 0.95. $\mathbf{C}^{(g+1)}$ was then used to generate points for the next iteration. Crosses in yellow were discarded and not used to update the covariance. The initial mean was set at [-10, -10]. 215

B.2 Illustration of different path lengths with CMAES, where the individual steps for each figure are the same length. Left, steps are correlated, and the step size should be increased. Middle, steps are anticorrelated and cancelling each other out, and so step size should be decreased. Right, steps broadly uncorrelated, no correction of step size required. Inspiration was taken from [102] 219

C.1 Purely Faradaic DCV experiment vs. time, with black lines indicating how the integral is approximated using numerical quadrature 226

C.2 Left: Purely Faradaic DCV current vs. potential, where $v = 5\ Vs^{-1}$ and $k_0 = 100\ s^{-1}$. Right: Trumpet plot showing normalised peak positions of oxidative and reductive peaks, and the linear region through which we draw a line of best-fit (red) 227

C.3 Demonstration of the procedure for obtaining sampled current from a SWV simulation. The total current vs. the non-dimensionalised time (left) is sampled using "forwards" and "backwards" points, shown on top of a section of the total current in the central figure. These sampled points are then plotted as orange and green traces in the rightmost figure, against the potential, as well as the net current which is the backwards subtracted from the forwards current. 228

C.4 Left: net SWV current, generated using $f = 25$ Hz and $k_0 = 75\ s^{-1}$, with the peak distance metric shown in red. Right: Normalised peak distances plotted as a function of the inverse frequency (blue dots), with the best-fit line used to calculate k_0 shown in red. 228

C.5 Left: Forward and backward currents from a SWV experiment $f = 1$ Hz, $k_0 = 6$ s^{-1} and $\alpha = 0.5$, with the key features I_f and I_b indicated. Right: Peak ratios obtained from sampled current simulated with different values of α vs the values of α calculated using equation C.7 (blue) and "ideal" peak ratios calculated by rearranging equation C.7 (red) 229

C.6 Left: Net SWV current, simulated using $f = 7$ Hz, $k_0 = 0.5\ s^{-1}$ and $E_{sw} = 0.13$ V, with the maximum current indicated by the red dot. Centre: maximum current (normalised by the value of E_{sw}), plotted against the varying E_{sw} value for three different frequencies. The I_{QRM} value is indicated by the red cross. Right: I_{QRM} plotted as a function of the inverse frequency, with the best-fit line drawn in red. 230

C.7 Values of the surface coverage parameter Γ, where the true value was 10 pmol cm^{-2}, inferred using the approach described in the heuristic section, with varying amounts of noise. The noise percentage is indicated in the legend. We did not compare the results of this to MCMC in this case, because the values inferred by this method were so close to the true value. 234

C.8 Left and centre: Kinetic parameter values inferred by fitting to the linear region of a trumpet plot for varying amounts of noise, where the noise percentage is indicated by the legend, and the true value of k_0 was $100s^{-1}$. For the linear fitting, the inferred anodic and cathodic values were pooled. Right: Inferred parameter values for k_0 using MCMC, fitted to a trumpet plot. 235

C.9 Left: Values of the kinetic parameter inferred from the net sampled SWV current using the peak splitting method described in the heuristic inference section, for varying amounts of noise where the noise percentage is shown in the legend. The true value of k_0 was 75s^{-1}. Centre and right: Values of the kinetic parameter inferred using MCMC, using sampled net current (centre) and total current (right). 236

C.10 Left: α values inferred from noisy sampled forwards and backwards SWV currents using the peak ratio approach described in the heuristic inference section, for different values of noise, where the noise percentage is shown in the legend. The true value of α was 0.5. Right and centre: α values inferred from noisy sampled net current (centre) and total SWV current (right) 237

C.11 Left: Kinetic parameter values inferred from sampled net SWV current using the quasi-reversible maximum method, for different values of noise where the noise percentage is shown in the legend. The true value of k_0 was 0.5s^{-1}. Right and centre: Kinetic parameter values inferred using MCMC using noisy sampled (centre) and total (right) SWV current. 238

D.1 Top: Nyquist plots of three components, from left to right: a resistor, a capacitor and a constant phase element. The range of frequencies used was $10^{-1.5}$ - 10^{5} Hz. Bottom: The simulated time-domain current response to a sinusoidal potential with an amplitude of 5 mV and a frequency of 0.16 Hz, as shown in black. . . . 243

D.2 Equivalent circuit drawing of the Randles circuit 244

D.3 Suggested equivalent circuit for a surface-confined voltammetry experiment . . . 244

List of Tables

3.1 Simulation parameters used to generate analytical and numerical simulations as compared in figure 3.2. Parameters marked with an asterisk were not modelled in the analytical approximation, and set to zero for the numerical solution. The values of the parameters were largely taken from Bell et al. [12]. 66

3.2 Simulation parameters held constant in figures 3.3-3.14. For some experimental parameters, the simulation value is different for each technique, and is written in the following order: r-FTACV/PSV/DCV, with a * indicating that the parameter is not used in the simulation 72

3.3 Simulation parameters used to generate plots in figures 3.16-3.21. For some experimental parameters, the simulation value is different for each technique, and is written in the following order: r-FTACV/PSV/DCV, with a * indicating that the parameter is not used in the simulation 93

3.4 Standard deviations for inferred parameter distributions (such as the ones plotted in figures 3.30 and 3.31 for the 0.5% and 4% noise cases respectively), represented as a percentage of the true parameter value, which is recorded in table 3.2, except R_u which was set to 200 Ω 114

4.1 Inferred CMAES parameters for PSV experiments 1-10 and the first r-FTACV experiment. † The phase and the capacitance phase were both held at 0 during the fitting process. *The error reported for the r-FTACV experiment is the mean error across the fitted harmonics (i.e. 2-6), and so is not directly comparable to the error for the PSV experiments, which is the error for the total timeseries. . . 121

4.2 Bounds for the values drawn by the CMAES algorithm. 122

4.3 Parameters fit to background-subtracted current, using equation 4.1, where the resulting simulation is shown in purple in figure 4.9. 129

4.4 Parameters inferred by the CMAES algorithm for PSV experiment 10, incorporating kinetic dispersion (where the kinetic parameter was distributed log-normally). 132
4.5 Mean and standard deviation (in brackets) for each distribution inferred during an MCMC run with 100000 samples. The first 50000 samples were excluded as burn-in. 135

5.1 Simulation parameters for the two fits shown in figure 5.2 obtained by fitting to total PSV current. 146
5.2 Simulation parameters for simulations that were a good fit to higher PSV harmonics whilst being physically unrealistic, as in figures 5.3 and 5.5, for the "low k_0" and "high C_{dl}" cases respectively. 147
5.3 Best fit parameters for harmonics 4 and above of PSV experiments 1, 2 and 3. The resulting simulated PSV current is shown in figures 5.8(A-C). The same values were used to generate r-FTACV simulations shown in figures 5.8(D-F), and the values in brackets were used to generate the r-FTACV simulations in figures 5.8(G-I) 160
5.4 Simulation parameters used to generate the PSV simulations observed in figure 5.9, where the upper boundary for α was set to the value shown in the heading for each column. 162
5.5 Simulation parameters used to generate the PSV simulations in figure 5.10. Capacitance parameters, including R_u, were obtained by fitting to the total current. Faradaic parameters were obtained by fitting to harmonics 4-10, fixing the capacitance parameters obtained in the previous step, and bounding the uncompensated resistance to 20% of total-current inferred value. 166
5.6 Table of inferred DCV parameters. The single DCV column was used to generate the purely Faradaic simulation plotted in the central graph of figure 5.16 in blue. The parameters for the 'trumpet" column were obtained by fitting to the data in figure 5.18, where we can only fit E^0 and k_0, along with a "separation" parameter. 172

6.1 Best-fit PSV parameters as inferred by the CMAES algorithm for M66H/R44N data. 186
6.2 Parameter bounds passed to the CMAES when attempting to fit M66H/R44N PSV data. These boundaries were identical except for the E^0_μ parameter, with the values reported being in the order M66H/R44N. 188

C.1 Simulation parameters used to generate synthetic current data for both heuristic and MCMC inference approaches for the DCV case. Values in brackets refer to single-experiment simulations, and non-bracketed values refer to extracted feature simulations. 231

C.2 Simulation parameters used to generate synthetic current data for both heuristic and MCMC inference approaches for the SWV case. Values in square brackets indicate the number of values within the range. Simulation times are from a Dell laptop with 8 GB RAM and an Intel i5–8350U CPU @ 1.70 GHz. MCMC times were calculated by observing the time taken for 20 iterations with three chains and scaling for 5000, and is faster than would be expected from simulation alone as a result of parallelisation. Values in brackets refer to single-experiment simulations, and non-bracketed values refer to extracted feature simulations 232

C.3 Ratio of the standard deviation of the inferred parameter values for each inference approach to the true value of that parameter, with varying amounts of added noise. Where appropriate, the values in brackets were inferred from fitting to the total current, and the other values are from fitting to the extracted features. . . . 233

List of publications

- Henry Lloyd-Laney, Nicholas Yates, Martin Robinson, Alice Hewson, Jack Firth, Darrell Elton, Jie Zhang, Alan Bond, Alison Parkin, and David Gavaghan, 2021. Using purely sinusoidal voltammetry for rapid inference of surface-confined electrochemical reaction parameters. **Analytical Chemistry**, 93(4), pp.2062-2071,

- Henry Lloyd-Laney, Martin Robinson, Alan Bond, Alison Parkin, and David Gavaghan, 2021. A spotter's guide to dispersion in non-catalytic surface-confined voltammetry experiments. **Journal of Electroanalytical Chemistry**, p.115204 ,

- Henry Lloyd-Laney, Martin Robinson, Alan Bond, Alison Parkin, and David Gavaghan, Comparison of numerical and analytical approaches for simulating purely sinusoidal voltammetry **Chemrxiv**,

- Henry Lloyd-Laney, Nicholas Yates, Martin Robinson, Alice Hewson, Alan Bond, Alison Parkin, and David Gavaghan, 2021. A multiple-technique approach to inferring bioelectrochemical reaction parameters **Chemrxiv**,

- Henry Lloyd-Laney, Martin Robinson, Alison Parkin, and David Gavaghan, 2021. Theoretical assessment of the impact of noise on heuristic parameter inference methods for surface-confined non-catalytic voltammetry experiments (in press) **Journal of Electroanalytical chemistry**.

Additionally, I contributed to the following publication, the work in which is not included in this thesis:

- Alister Dale-Evans, Martin Robinson, Henry Lloyd-Laney, David Gavaghan, Alan Bond, and Alison Parkin. A voltammetric perspective of multi-electron and proton transfer in protein redox chemistry: Insights from computational analysis of Escherichia coli HypD Fourier transformed alternating current voltammetry, 2021. **Frontiers in Chemistry**, 9, p. 424.

Chapter 1

Introduction

1.1 Preamble

This thesis is motivated by an interest in the phenomenon of biological electron-transfer. Organisms are highly efficient electrochemists; the biological coupling of reduction-oxidation (redox) reactions underpins the entire biosphere. Human, animal and plant life is powered by such redox reactions, as in the generation of the chemiosmotic gradient in aerobic respiration [1], and the level of oxygen in the atmosphere arises from the splitting of water during photosynthesis [2]. Because these processes are so fundamental for the existence of life as we know it, understanding how they are controlled biologically is a problem of considerable scientific interest. The biological redox processes are commonly facilitated by a class of metal-containing polypeptides referred to as the metalloproteins [3], which incorporate atoms such as copper, iron and zinc into their structures. Alternative, non-metal based biological redox-active centres include reducing and oxidising organic cofactors such as quinols [4], or the sulphur atoms in cysteine residues [5]. Within an organism, redox reactions are homogeneous: occurring between two electroactive species. However, by using *heterogeneous* reactions, or electron transfer between an electrode and an electroactive species, we can record and quantify the latter's intrinsic electrochemical properties, in turn allowing for deeper understanding of *homogeneous* biological reactions [6]. To achieve this, we turn to the science of voltammetry.

1.2 Background

In voltammetry, a time-varying potential input is applied to an electro-active species of interest. The current response contains signal (referred to as Faradaic current) and noise (referred to as background current). For voltammetric data generated by proteins, the Faradaic signal of interest is swamped by both background signals arising from physical processes within the experimental system, and by experimental measurement error, arising from processes such as shot noise, or thermal electron fluctuations within the recording apparatus [7]. As we shall see throughout this thesis, maximising the Faradaic signal relative to background current is *the* great challenge of the voltammetry of proteins.

A low ratio of Faradaic to non-Faradaic current makes extracting information about the system challenging, particularly in the most widely used voltammetry technique Direct Current Voltammetry (DCV). There have been a number of attempts to design a potential input that can boost this ratio, such as Square Wave Voltammetry (SWV), and ramped Fourier Transform Alternating Current Voltammetry (r-FTACV). In particular, Bond [8] and Parkin [5, 9, 10] have successfully applied r-FTACV for the analysis of biological electron-transfer. However, whilst r-FTACV is excellent at improving the Faradaic-to-background ratio relative to DCV, its limitation is that it is very computationally expensive to simulate. The act of simulation predicts experimental current as a function of a set of parameters which describe the intrinsic properties of the species — such as the thermodynamics or the kinetics of the electron transfer process. In this thesis, we use an automatic optimisation approach with the aim of obtaining a set of parameters that generate a current simulation that is a "good fit" to the experimental data. This requires numerous simulations, which is compounded by uncertainty (and attendant iteration) over the correct model assumptions. Consequently, simulation speed is an essential consideration.

For this reason, in this thesis we describe the advantage of analysing purely sinusoidal voltammetry (PSV), which can be simulated in less time than an r-FTACV experiment — a typical simulation of 30 oscillations of a PSV experiment takes around *0.01s* on a laptop (8GB RAM and an Intel i5-8350U CPU @ 1.70GHz), which represents a 10-15 fold reduction relative to that required for a complete r-FTACV experiment. This time-saving translates to fitting runs that take tens of minutes, rather than hours. PSV also has the property of boosting the Faradaic-to-background ratio relative to DCV. It should be noted that we are not the first to describe

the technique — the original form of the PSV method on which the present work is based was invented a long time ago. As noted by Bard and Zoski, in the 1940s fast linear potential sweeps were applied at the dropping mercury electrode, known as oscillographic polarography, as reviewed in [11]. Other researchers have investigated PSV from a theoretical standpoint [12], and as a means of exploring the effects of background current [13]. The PSV method has also been used for the sensing of small molecules such as nucleotides and neurotransmitters [14, 15]. Thus, in terms of mechanistic studies, the PSV technique has been predominantly only of theoretical interest, rather than subjected to widespread studies of electrode kinetics, as with the r-FTACV method.

The case for PSV is that it retains some of the Faradaic-to-background boosting advantages of Fourier Transform voltammetry methods, whilst significantly reducing the computational burden associated with r-FTACV. The benefits of speed are tremendous — allowing for a much more comprehensive exploration of the parameter search space in the same timeframe relative to r-FTACV, which allows for assessing many different combinations of modelling choices, as we shall see in subsequent chapters.

The overall aim of this thesis is therefore to use PSV to make the computational inference of the parameters governing biological electron-transfer processes rapid, robust and reproducible. We want to be able to obtain biological insight on the order of days or weeks, to be able to do so repeatedly, and such that the information obtained is simultaneously consistent with data from DCV, PSV and r-FTACV experiments. This can be formalised into the following research questions:

- Can we use PSV to obtain information about the electrochemical properties of a system, and does this information also predict how the system responds to r-FTACV and DCV experiments?
- How fast is this inference process, relative to r-FTACV analysis?
- Can the above analysis be applied to a biological system with poor Faradaic-to-background current ratios?
- To what extent can we capture fine-scale differences between biological systems using the above approach?

1.3 Thesis structure

Having detailed our research questions, we now briefly summarise what is contained in each chapter.

1.3.1 Chapter 2 — Literature review and theoretical background

In this chapter, we lay out the key electrochemical and mathematical theory required to understand the rest of the thesis, with a focus on how we predict the current response for a particular set of electrochemical parameters, and what assumptions are required in the construction of this simulation approach. This includes a fulsome description of our automatic optimisation methodology.

1.3.2 Chapter 3 — Theoretical results

This chapter uses simulation studies to gain deeper insight into various aspects of the problem. To validate the modelling approach laid out in section 2.2 of the literature review, we compare the simulation output to another method from the literature, as published by Bell and co-authors [12]. By showing that the output from our approach and Bell's method for a given set of parameters are very similar, we can verify that our numerical implementation is correct. Having done so, we move on to a sensitivity analysis for DCV, PSV and r-FTACV, showing the effect that varying each parameter has on the appearance of the simulated current when the others are held constant. This is done to inform discussion in subsequent chapters — when we fit experimental data, much of the discussion of the credibility of a particular inferred parameter value will require an understanding of how varying that parameter affects the voltammetric current. We then extend this sensitivity analysis to a process known as dispersion, demonstrating the impact of thermodynamic and kinetic dispersion, and propose an evidential standard for when to explicitly include the different forms of dispersion in modelling. Once again, in chapters where we fit experimental data, the choice of how to model dispersion will be highly pertinent. Finally, we validate our automatic fitting methodology, comparing the amount of information that can be recovered from analysis of synthetic noisy DCV, PSV and r-FTACV data.

1.3.3 Chapter 4 — Ferrocene

With the insights obtained in chapter 3, we analyse data obtained by using PSV and r-FTACV to interrogate surface-linked ferrocene, an extremely well-understood electrochemical system, in order to troubleshoot the analysis of PSV before moving onto a biological system. The ferrocene voltammetric data has excellent Faradaic-to-background current ratios, which means we can resolve any challenges relating specifically to the analysis of PSV. This requires two modifications to the model defined in chapter 2, and once these alterations are made, it is possible to infer consistent parameter values, finding a minimum in high-dimensional parameter space for ten sequential PSV experiments by fitting to total PSV current. We also demonstrate that these parameters are comparable to those obtained by fitting to r-FTACV surface-linked ferrocene data and show that we can use Bayesian inference to obtain posterior parameter distributions.

1.3.4 Chapter 5 — Cytochromes

Having shown that it is possible to use PSV to obtain accurate parameters in the context of voltammetric data with favourable Faradaic-to-background current ratios, in chapter 5 we move on to analysing data from a surface linked protein, a bacterial cytochrome. As has been emphasised above, the Faradaic-to-background currents arising from a biological system are significantly poorer than for ferrocene. As a consequence, we have to make use of the signal-boosting properties of PSV and r-FTACV, which introduces further challenges. Additionally, as will be demonstrated in section 3.2, several of the model parameters have similar effects over the potential window of interest. This manifests itself as correlations between sets of parameters, in a phenomenon known as parameter compensation. The effect is further compounded by the impact of the background non-Faradaic current on the parameter recovery process. Overall, this means that finding the correct set of parameters is made very difficult by the presence of multiple parameter values generating good fits to the data, often in disparate parts of the overall parameter space.

Consequently, it is necessary to make use of all prior information (data) to narrow down the parameter search space by placing constraints to ensure that we include only those parameter ranges which are admissible after making use of the prior information. Additionally, we make use of the contrasting advantages and disadvantages of DCV, PSV and r-FTACV data to render the overall problem of obtaining electrochemical information tractable, and to obtain reproducible,

consistent parameter values between the three different experiments, and additionally to do so across multiple datasets. At the end of chapter 5, we formalise this process into a "recipe" for integrating insights from DCV, PSV and r-FTACV to reliably analyse protein-film voltammetry data.

1.3.5 Chapter 6 — Rational mutation

Using the methodology proposed at the end of chapter 5, integrating DCV, PSV and r-FTACV experiments, we analyse voltammetry data collected for two mutants of the bacterial cytochrome, both of which were hypothesised to have altered electrochemical properties. Although the datasets collected have even poorer Faradaic-to-background current ratios than those presented in chapter 5, we demonstrate that it is possible to obtain parameter estimates that are consistent between PSV and r-FTACV data, such that we are able to observe fine-scale shifts in the electrochemical properties of the mutants. This is a preliminary validation of our framework as one that produces results rapidly, robustly and reproducibly.

1.3.6 Chapter 7 — Conclusions

To end the thesis, we summarise the results obtained in previous chapters, and propose future work on the basis of these results.

Chapter 2

Literature review and theoretical background

This chapter describes the essential electrochemical theory of electron-transfer in a voltammetric experiment, describes how we intend to model voltammetric current as a function of Faradaic and background-current parameters, and discusses the assumptions required for this model derivation. It also surveys alternative efforts to obtain insight into biological electron-transfer using voltammetry, before moving onto a description of how we intend to perform the inference process.

2.1 Voltammetry

In this section, we will briefly outline the necessary electrochemical theory for understanding electron-transfer reactions. This theory is largely identical for simple molecules, the study of which gave birth to the field of voltammetry [16], and highly complex macromolecules such as metalloproteins.

In an idealised simple voltammetric experiment, we use three electrodes — working, counter and reference — connected in a circuit. The reference electrode is designed such that its composition (i.e. the equilibrium balance of reduced and oxidised species) does not change, and consequently its electrode potential remains a constant "reference" against which the voltage of the working electrode can be defined. At more negative potentials, electrons within the working electrode can be considered to be at higher energy, and vice versa for higher potentials. A simplified pic-

ture of redox reactions is that they occur when electrons "jump" from one molecular orbital to another orbital of lower energy. By externally controlling the potential of the working electrode in the presence of an electroactive species, whether in solution or attached to the surface of the electrode, we can observe at precisely which potentials redox reactions occur, manifesting as a flow of electrons. This electron flow, balanced by an opposing process at the counter electrode, will occur at a potential indicative of the relative energies of the molecular orbitals and the rate of electron flow, is measurable as current, in the form of charge (Coulombs (C)) per second. One mole of reactant requires the transfer of 96,485.4 C to be reduced or oxidised by one mole of electrons. This relationship between redox reactions and the resulting current is referred to as "Faraday's law", and the current as "Faradaic current". Naturally this is a simplified picture of electron transfer — there are considerations related to the distance that the electrons have to travel, the degree of tunnelling required, and so on, but for the purposes of understanding, this simplified picture is still helpful [6].

The Faradaic flow of electrons to and from electroactive species is not the only current-generating process in a voltammetric experiment. The other principal contribution to a voltammetric current arises from double-layer capacitance, which occurs when the charge within the electrode attracts oppositely-charged solvated ions towards itself. These two charges are separated by the ion solvation layer, and solvent molecules that coat the surface of the electrode (referred to as the outer and inner Helmholtz planes respectively). Two charges separated by a non-conducting layer in this manner form a capacitor, and the application of a time-varying potential, as in voltammetric experiments, results in charging and discharging cycles of this layer [6].

2.1.1 Protein-film voltammetry

It is challenging to analyse proteins using solution-phase voltammetry, in part because of their very slow rate of diffusion, and the consequent requirement for electron-shuttling mediators; as such, the electrochemical properties that can be inferred are limited by the properties of the mediators themselves. In pioneering papers in 1977, two sets of researchers showed that it was possible to obtain excellent signals from adsorbed cytochrome C on modified gold [17] or indium tin oxide [18] electrodes, indicating that direct electron-transfer to the adsorbed protein had been achieved without the need for mediators. This kick-started a wider interest in analysing proteins adsorbed onto conducting materials via electrochemistry, and the name "protein film voltammetry" (PFV) was coined to describe the technique.

It has been extensively demonstrated that proteins adsorbed to the working electrode retain their structure, as evidenced by the retention of catalytic activity. This retention has been demonstrated using both enzyme-film electrochemistry and spectro-electrochemical studies. Therefore, it is widely accepted that the electrochemical behaviour we analyse is that of a correctly folded protein, rather than that of a misfolded aggregate [19].

The prospect of using voltammetry to investigate biological phenomena is exciting, in that it offers the ability to very precisely control when a reaction occurs via fine control over the potential. Furthermore, a great deal of dynamic information can be recovered from analysis of the current. It is also possible to use other spectroscopic methods to investigate the thermodynamics of electronic transitions in proteins, such as infrared spectroscopy [20], and electron paramagnetic resonance spectroscopy [9, 21]. Spectroscopy mostly cannot provide the highly time-resolved information provided by voltammetry, although some efforts have been made using "pump-probe" spectroscopy via the medium of photo-induced electron transfer [22, 23]. Spectroscopic and voltammetric techniques should be regarded as complementary, which is reinforced by the existence of methods combining both, such as in an optically-transparent thin-layer electrochemical cell [24]. Additionally, entirely theoretical approaches, based on quantum mechanical modelling, have also been used to interrogate the electrochemical properties of metalloproteins, although this remains challenging [25, 26]. As we shall see in later chapters of this thesis, it is possible to embark on highly granular analyses of the nature of biological electron transfer using protein-film voltammetry data alone.

2.2 Modelling

Having established that it is possible to observe redox processes occurring between a protein and an electrode in section 2.1.1, it is now pertinent to discuss how the key electrochemical properties that generate the observed current can be determined. Quantitative analysis of voltammetry data usually involves inferring chemical reaction parameters using models that are proposed to mimic the electrochemical response [27]. Typically, the mathematics used to simulate the voltammetric experiment consist of a system of non-linear partial differential equations for solution-phase

electrochemistry [28], where the analyte is dissolved in the electrolyte, or non-linear ordinary differential equations for film electrochemistry [9], where the redox-active analyte is confined to the surface of the working electrode. Embedded within these models are the chemical reaction parameters that describe the Faradaic current that arises from electron transfer process(es). The simplest possible electrode process to model is a one-electron surface-confined redox reaction, shown in the chemical equation below,

$$\mathrm{Red} \underset{\tilde{k}_{\mathrm{red}}}{\overset{\tilde{k}_{\mathrm{ox}}}{\rightleftharpoons}} \mathrm{Ox} + \mathrm{e}^- \ ,$$

where Red and Ox respectively represent the reduced and oxidised forms of the surface-confined species and $\tilde{k}_{red}$ and $\tilde{k}_{ox}$ are the electron transfer rate constants for oxidation and reduction, respectively.

Throughout this chapter, we will use a tilde $\sim$ to distinguish a dimensional parameter or variable from the non-dimensional form. The Faradaic current $\tilde{I}_f$ resulting from the electron transfer process, or equivalently the rate of charge transfer $\frac{d\tilde{Q}}{d\tilde{t}}$, is described by Faraday's law, where N is the number of moles of electrochemical species, F is Faraday's constant and A is the surface area of the electrode [29],

$$\tilde{I}_f = \frac{d\tilde{Q}}{d\tilde{t}} = FA\frac{dN}{d\tilde{t}}. \tag{2.1}$$

Since we are modelling only a single-electron transfer process for a single surface-confined species, the mathematical description of the Faradaic current is a single ordinary differential equation (ODE). If we assume the redox centres are non-interacting with a Langmuir isotherm defining the surface coverage, then θ can be used to represent the relative abundance or fraction of the oxidised species. Accordingly, equation 2.1 is defined assuming mass-action kinetics

$$\tilde{I}_f = FA\tilde{\Gamma}\frac{d\theta}{d\tilde{t}} = FA\tilde{\Gamma}((1-\theta)\tilde{k}_{ox}(\tilde{E}_r) - \theta\tilde{k}_{red}(\tilde{E}_r)), \tag{2.2}$$

where $\tilde{\Gamma}$ is the total surface concentration of Ox and Red. The rate of change of θ is dependent on the electron transfer rate constants for oxidation and reduction, $\tilde{k}_{ox}$ and $\tilde{k}_{red}$, which are functions of the effective potential ($\tilde{E}_r$). $\tilde{E}_r$ is the applied potential minus the Ohmic (IR) drop, which is the total current multiplied by the resistance that is uncompensated by the actions of the potentiostat $\tilde{R}_u$, and is given by

$$\tilde{E}_r = \tilde{E}_{app} - \tilde{R}_u\tilde{I}. \tag{2.3}$$

For the expression of $\tilde{k}_{ox}$ and $\tilde{k}_{red}$, we use the Butler-Volmer equations, defined below in section 2.2.1.

2.2.1 Butler-Volmer kinetics

One of the most important equations in electrochemistry is the Nernst equation, which relates the concentration of oxidised and reduced species in a single electron process to the potential $\tilde{E}$, for the homogeneous case (shown in equation 2.4). For the heterogeneous case, this term is replaced by the applied potential $\tilde{E}_{app}$.

$$\tilde{E} = \tilde{E}^0 + \frac{RT}{F} \ln\left(\frac{[\text{Ox}]}{[\text{Red}]}\right). \tag{2.4}$$

Here $\tilde{E}^0$ is the potential at which the concentration of oxidised and reduced species is equal, under standard conditions. R is the ideal gas constant, square brackets ([ ]) represent the concentration of each species and T is the temperature in Kelvin. The reversible potential $\tilde{E}^0$ can be thought of as a measure of how much thermodynamic driving force is required to drive the reaction in a particular direction. If the electrochemical reaction as a function of potential is at equilibrium, i.e. the rates of the oxidation and reduction are equal, and we assume that the reaction rate is a function of concentration and potential only, then

$$\tilde{k}_{red}(\tilde{E})[\text{Ox}] = \tilde{k}_{ox}(\tilde{E})[\text{Red}]. \tag{2.5}$$

Rearranging this equation and taking logs of both sides results in an equation

$$\ln\left(\frac{[\text{Ox}]}{[\text{Red}]}\right) = \ln\left(\frac{\tilde{k}_{ox}(\tilde{E})}{\tilde{k}_{red}(\tilde{E})}\right). \tag{2.6}$$

The relationship between reduced and oxidised species in equation 2.6 is as in the Nernst equation. By substituting 2.6 into equation 2.4, we obtain an expression for $\tilde{k}_{ox}$ and $\tilde{k}_{red}$ that is in terms of (easily measurable) potentials rather than concentrations

$$\ln\left(\frac{\tilde{k}_{ox}(\tilde{E})}{\tilde{k}_{red}(\tilde{E})}\right) = \frac{F}{RT}(\tilde{E} - \tilde{E}^0). \tag{2.7}$$

However, for equation 2.2, we need an expression for the individual reaction rates as a function of potential. In order to do this, we differentiate equation 2.7 with respect to the potential, and

rearrange to obtain

$$\underbrace{\frac{RT}{F}\frac{d}{d\tilde{E}}\ln(\tilde{k}_{ox}(\tilde{E}))}_{1-\alpha}+\underbrace{\frac{RT}{F}\frac{d}{d\tilde{E}}\ln\left(\frac{1}{\tilde{k}_{red}(\tilde{E})}\right)}_{\alpha}=1. \tag{2.8}$$

The two terms on the left-hand side of the equation are known as symmetry factors, and are represented as α and $1-\alpha$ for the reductive and oxidative terms respectively. If we integrate the expression for α with respect to the potential such that,

$$\int \frac{d}{d\tilde{E}}\ln\left(\tilde{k}_{red}(\tilde{E})\right)d\tilde{E}=\int\frac{-\alpha F}{RT}d\tilde{E}, \tag{2.9}$$

we obtain the following expression

$$\ln\left(\tilde{k}_{red}(\tilde{E})\right)=\frac{-\alpha F\tilde{E}}{RT}+\text{constant}. \tag{2.10}$$

Substituting in E^0 from the Nernst equation into equation 2.10 and subtracting the result from equation 2.10 eliminates the constant and yields

$$\ln\left(\frac{\tilde{k}_{red}(\tilde{E})}{\tilde{k}_{red}(\tilde{E}^0)}\right)=\frac{-\alpha F}{RT}(\tilde{E}-\tilde{E}^0). \tag{2.11}$$

This process can be repeated to obtain an equivalent expression for $\tilde{k}_{ox}$. From the Nernst equation 2.4, at $\tilde{E}^0$, the concentrations of oxidised and reduced species are the same and, using equation 2.5, we can state that $\tilde{k}_{red}(\tilde{E}^0)=\tilde{k}_{ox}(\tilde{E}^0)=\tilde{k}_0$, i.e. the rate at which electrons are transferred at the reversible potential $\tilde{E}^0$ is $\tilde{k}_0$. From this and equation 2.11, we can obtain the Butler-Volmer equations [30] for the homogeneous case.

$$\tilde{k}_{red}(\tilde{E})=\tilde{k}_0\exp\left(\frac{-\alpha F}{RT}(\tilde{E}-\tilde{E}^0)\right), \tag{2.12}$$

$$\tilde{k}_{ox}(\tilde{E})=\tilde{k}_0\exp\left(\frac{(1-\alpha)F}{RT}(\tilde{E}-\tilde{E}^0)\right). \tag{2.13}$$

In the heterogeneous case, we replace $\tilde{E}$ with an applied potential $\tilde{E}_{app}$.

Having shown the link between the Butler-Volmer model and the Nernst equation, we can also derive this expression from the perspective of the standard free energy for activation of the re-

action, $\Delta\tilde{G}^{\ddagger}$. The equations for the reduction and oxidation rate constants can be expressed in terms of the free energy via the Arrhenius equation

$$\tilde{k}_{red}(\tilde{E}) = A_{red} \exp\left(\frac{-\Delta\tilde{G}^{\ddagger}_{red}}{RT}\right), \tag{2.14}$$

$$\tilde{k}_{ox}(\tilde{E}) = A_{ox} \exp\left(\frac{-\Delta\tilde{G}^{\ddagger}_{ox}}{RT}\right), \tag{2.15}$$

where $\Delta\tilde{G}^{\ddagger}_{ox}$ and $\Delta\tilde{G}^{\ddagger}_{red}$ are the free energy barriers for oxidation and reduction respectively, and A_{red} and A_{ox} are the Arrhenius factors for the two equations. We can express these values as the difference between the activation free energy at the reversible potential $\tilde{E}^0$, written as $\Delta\tilde{G}^{\ddagger}_{\tilde{E}^0_{ox}}/\Delta\tilde{G}^{\ddagger}_{\tilde{E}^0_{red}}$, and the energy of an electrode maintained at a potential $\tilde{E}$ away from $\tilde{E}^0$, which is $-F(\tilde{E}-\tilde{E}^0)$. The electron energy, in both cases, is scaled by the relative ease with which the transition between oxidised and reduced states (and visa versa) is formed — the symmetry factor α. Consequently

$$\Delta\tilde{G}^{\ddagger}_{red} = \Delta\tilde{G}^{\ddagger}_{\tilde{E}^0_{red}} + \alpha F(\tilde{E}-\tilde{E}^0), \tag{2.16}$$

$$\Delta\tilde{G}^{\ddagger}_{ox} = \Delta\tilde{G}^{\ddagger}_{\tilde{E}^0_{ox}} - (1-\alpha)F(\tilde{E}-\tilde{E}^0). \tag{2.17}$$

If we insert this into the Arrhenius equations in equations 2.14 and 2.15, we obtain

$$\tilde{k}_{red}(\tilde{E}) = A_{red} \exp\left(\frac{-\Delta\tilde{G}^{\ddagger}_{\tilde{E}^0_{red}}}{RT}\right) \exp\left(\frac{-\alpha F}{RT}(\tilde{E}-\tilde{E}^0)\right), \tag{2.18}$$

$$\tilde{k}_{ox}(\tilde{E}) = A_{ox} \exp\left(\frac{-\Delta\tilde{G}^{\ddagger}_{\tilde{E}^0_{ox}}}{RT}\right) \exp\left(\frac{(1-\alpha)F}{RT}(\tilde{E}-\tilde{E}^0)\right), \tag{2.19}$$

where the first exponential factors for $\tilde{k}_{red}(\tilde{E})$ and $\tilde{k}_{ox}(\tilde{E})$ multiplied by their respective Arrhenius constants are identical, and by comparison to equations 2.12 and 2.13 can be seen to be equivalent to $\tilde{k}_0$. This expression of electrode kinetics is useful in that it gives us a physical intuition for the value of α, and because a more complex, complete vision of electrode kinetics in the form of Marcus theory is also derived from these principles of free energy [6].

2.2.2 Capacitance

The capacitive current $\tilde{I}_c$ in the ideal case is linear, modelled as

$$\tilde{I}_c = A\tilde{C}_{dl}\frac{d\tilde{E}_r}{d\tilde{t}}, \tag{2.20}$$

where $\tilde{C}_{dl}$ is the magnitude of the capacitance, in Farads. However, as will be extensively discussed in the following chapters, the current arising from the double-layer capacitance process observed in a protein-film voltammetry experiment is often non-linear. We typically approximate this non-linearity using a third-order polynomial, such that

$$\tilde{I}_c = A\tilde{C}_{dl}(1 + C_{dlE1}\tilde{E}_r(\tilde{t}) + C_{dlE2}\tilde{E}_r(\tilde{t})^2 + C_{dlE3}\tilde{E}_r(\tilde{t})^3)\left(\frac{d\tilde{E}_r}{d\tilde{t}}\right), \tag{2.21}$$

where the higher order terms (C_{dlEX}) are dimensionless. The order of the polynomial is informed by observations of "blank" (i.e. purely capacitive) experimental data. The total current is then modelled as $\tilde{I} = \tilde{I}_f + \tilde{I}_c$.

2.2.3 Non-dimensional constants

In order to nondimensionalise the expression for the total current, we use the following substitutions,

$$E = \frac{\tilde{E}}{\epsilon_0}, \qquad t = \frac{\tilde{t}}{T_0}, \qquad I = \frac{\tilde{I}}{I_0}, \qquad \Gamma = \frac{\tilde{\Gamma}}{\Gamma_0},$$

where the units of the constants in the nondimensionalisation, ϵ_0, t_0, I_0 and Γ_0, are Volts, seconds, Amps and moles cm^{-2} respectively. Setting

$$\epsilon_0 = \frac{RT}{F}, \quad T_0 = \frac{\epsilon_0}{\tilde{v}}, \quad \Gamma_0 = [\mathrm{Ox}] + [\mathrm{Red}], \quad I_0 = \frac{FA\Gamma_0}{T_0},$$

where v is a scaling factor in units of V s^{-1}, and is dependent on the form of the potential input (i.e. E_{app}), which we will provide examples of in the following section. The value of Γ_0 is an estimated value for the "true" surface coverage, rendering Γ a dimensionless scaling factor. These substitutions yield the following equations (for the case of linear capacitance)

$$I = C_{dl}\left(\frac{dE_{app}}{dt} - R_u\frac{dI}{dt}\right) + \Gamma\frac{d\theta}{dt}, \tag{2.22}$$

$$\frac{d\theta}{dt} = (1-\theta)k_0 \exp\big((1-\alpha)(E_{app} - IR_u - E^0)\big) - (\theta)k_0 \exp\big(-\alpha(E_{app} - IR_u - E^0)\big). \quad (2.23)$$

2.2.4 Voltammetric waveforms

The model described above uses an applied potential term $\tilde{E}_{app}$. The form of the potential input has a very large impact on the ratio of Faradaic-to-background current that can be extracted from a voltammetric experiment. Numerous potential inputs have been proposed to exploit the difference between the processes that generate background (double-layer capacitance) and Faradaic current. We will detail the three techniques analysed in this thesis.

2.2.4.1 Direct current voltammetry

The simplest to perform voltammetry technique, and the most widely used, is Direct-Current (or cyclic) Voltammetry, referred to as DCV. In the simplest DCV experiment, the potential linearly changes from a starting potential $\tilde{E}_{start}$, to a switching potential $\tilde{E}_{reverse}$, and then back down to $\tilde{E}_{start}$, such that

$$\tilde{E}_{app}(\tilde{t}) = \begin{cases} \tilde{E}_{start} + \tilde{v}\tilde{t}, & \text{if } \tilde{t} < \tilde{t}_r, \\ \tilde{E}_{reverse} - \tilde{v}(\tilde{t} - \tilde{t}_r), & \text{if } \tilde{t}_r \leq \tilde{t}, \end{cases} \quad (2.24)$$

where $\tilde{t}_r = \frac{\tilde{E}_{reverse} - \tilde{E}_{start}}{\tilde{v}}$, and $\tilde{v}$ is the scan rate, in units of V s^{-1}, and is used to non-dimensionalise the time constant, such that the non-dimensional scan rate is 1.

2.2.4.2 Fourier transform alternating current voltammetry

Although DCV is simple to perform, background currents can make the extraction of the Faradaic signal challenging. Fourier transform alternating current voltammetry (FTACV) is an alternative where the Faradaic signal is boosted relative to the background current. The applied waveform is the sum of the linear ramp, $\tilde{E}_{dc}$, defined above in section 2.2.4.1 for a DC voltammetric experiment, and a sine wave $\tilde{E}_{ac}$ having a frequency f. We will refer to the technique as "ramped" (r-FTACV) throughout the rest of the thesis to distinguish it from other Fourier transform voltammetry methods.

$$\tilde{E}_{dc}(\tilde{t}) = \begin{cases} \tilde{E}_{start} + \tilde{v}\tilde{t}, & \text{if } \tilde{t} < \tilde{t}_r, \\ \tilde{E}_{reverse} - \tilde{v}(\tilde{t} - \tilde{t}_r), & \text{if } \tilde{t}_r \leq \tilde{t}, \end{cases} \quad (2.25)$$

$\tilde{t}_r$, $\tilde{v}$, $\tilde{E}_{start}$ and $\tilde{E}_{reverse}$ are defined as for the DCV case. The sinusoid is given by

$$\tilde{E}_{ac}(\tilde{t}) = \Delta\tilde{E}\sin(\tilde{\omega}\tilde{t} + \eta), \tag{2.26}$$

where η is the phase, $\tilde{\omega}$ the sinsuoidal frequency, and $\Delta\tilde{E}$ the amplitude of the sinusoid. Finally, $\tilde{E}_{app}$ is given by

$$\tilde{E}_{app}(\tilde{t}) = \tilde{E}_{ac}(\tilde{t}) + \tilde{E}_{dc}(\tilde{t}). \tag{2.27}$$

2.2.4.3 Purely sinusoidal voltammetry

Purely sinusoidal voltammetry, or PSV, also uses a large-amplitude sinusoid in the potential input, but with no linear ramp. If we define $\tilde{E}_{half} = (\tilde{E}_{reverse} + \tilde{E}_{start})/2$, where η is again the phase, $\tilde{E}_{reverse}$ and $\tilde{E}_{start}$ are the maximum and minimum of the input sinusoid respectively, $\Delta\tilde{E}$ is the amplitude and $\tilde{\omega}$ is the input frequency, then the input potential is simply

$$\tilde{E}_{app}(\tilde{t}) = \tilde{E}_{half} + \Delta\tilde{E}\sin(\tilde{\omega}\tilde{t} + \eta). \tag{2.28}$$

PSV can also be thought of as being similar to multiple sequential DCV experiments with a variable scan rate, and no discontinuity at the switching potential. The improvement in simulation speed relative to r-FTACV is because a complete PSV experiment can be undertaken with an experimentally set number of oscillations, independent of frequency. From practical experience approximately 30 oscillations is appropriate to be able to make consistent inferences from experimental data. For r-FTACV by contrast, the number of oscillations is determined by the relationship between input frequency and scan rate. For example, in chapter 4, a single r-FTACV experiment required approximately 500 oscillations at the same frequency as the PSV experiment for complete simulation.

2.2.4.4 A note on input frequency

Throughout this thesis, we have used one order of magnitude for the input frequency — approximately 10 Hz. This is partly due to experimental convenience, and partly due to the observation that the Faradaic signal appeared to decrease with increased frequency, as well as a change in the observed background current. While attempting to capture this behaviour through modelling is a topic of active research, it was possible to obtain consistent inferences in the 10 Hz regime, and consequently this is the regime we operated in.

2.2.5 Fourier transforms and voltammetry

Both r-FTACV and PSV are Fourier transform voltammetry methods, because of the large-amplitude sinusoid present in the potential input. To contextualise, we give a brief description of the theory behind Fourier series, transforms, and their computation, which will help explain how FT voltammetry methods boost the Faradaic-to-background current ratio.

2.2.5.1 Fourier Series

T-periodic functions (where $f(x+T) = f(x)$) can be represented using orthonormal sine and cosine basis functions. Taking a set of basis functions B

$$B = \left\{ \frac{1}{\sqrt{T}}, \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}x\right), \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}x\right), \ldots, \sqrt{\frac{2}{T}} \sin\left(\frac{2n\pi}{T}x\right), \sqrt{\frac{2}{T}} \cos\left(\frac{2n\pi}{T}x\right) \right\}, \tag{2.29}$$

and projecting a T-periodic function $f(x)$ onto this basis results in the expression

$$f(x) = c + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2n\pi}{T}x\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2n\pi}{T}\right), \tag{2.30}$$

which is known as the *Fourier series* representation of $f(x)$. The coefficients a_n, b_n and c are determined such that

$$a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) \cos\left(\frac{2n\pi}{T}x\right) dx, \tag{2.31}$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) \sin\left(\frac{2n\pi}{T}x\right) dx, \tag{2.32}$$

$$c = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) dx. \tag{2.33}$$

The sum in equation 2.30 will only converge under certain conditions, such as a finite number of maxima and minima, although for most functions observed in science and engineering the Fourier series will converge almost everywhere. It is also convenient to represent the sines and cosines in 2.30 using complex exponential representation defined in Euler's identity ($e^{ix} = \cos(x) + i\sin(x)$), such that

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{2\pi}{T}nx}, \tag{2.34}$$

where

$$c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) e^{-i\frac{2\pi}{T}nx} dx. \tag{2.35}$$

2.2.5.2 Fourier Transform

The Fourier series of a T-periodic function allows for the observation of how sines and cosines of various frequencies contribute to that function. This decomposition is extremely useful for analysing (for example) time-varying signals, but the Fourier series is only applicable for periodic functions. However, a similar form of analysis can still be used for any signal, designated $s(x)$, by application of the *Fourier transform*, $\mathscr{F}\{s(x)\}$, which yields the *Fourier spectrum* $\mathscr{F}(\omega)$. The signal $s(x)$ can be thought of as a T-periodic signal in the limit as T tends to ∞. The Fourier spectrum indicates the contribution of a sinusoid of frequency ω to the overall signal $s(x)$. This is defined as [31]

$$\mathscr{F}(\omega) = \int_{-\infty}^{\infty} s(x) e^{-i2\pi x\omega} dx. \tag{2.36}$$

The link to Fourier series can be made plain if we take the Fourier transform of a signal which is 0 everywhere, except for a single period T over the range $[-\frac{T}{2}, \frac{T}{2}]$. Using equation 2.35, the value of c_n is

$$c_n = \frac{1}{T}\mathscr{F}\left(\frac{n}{T}\right). \tag{2.37}$$

This result shows us that the Fourier series coefficients are equal to the Fourier transform coefficients sampled at multiples of $\frac{1}{T}$ and scaled by $\frac{1}{T}$.

It is also possible to go from the Fourier spectrum to the signal $s(x)$ using the inverse Fourier transform [31]

$$s(x) = \int_{-\infty}^{\infty} \mathscr{F}(\omega) e^{i2\pi x\omega} d\omega. \tag{2.38}$$

2.2.5.3 The discrete Fourier transform

For a real-world signal, especially one recorded digitally, the signal is not continuous, but rather discrete. For a continuous function $s(x)$ that was 0 everywhere except for a series of impulses at the N data points comprising our signal (separated by a constant sampling period T), with areas corresponding to the sampled data, $s[n]$, then the Fourier transform of such an expression would be as follows

$$\mathscr{F}(\omega) = \int_{-\infty}^{\infty} s(x) e^{-i2\pi x\omega} dx = \sum_{n=0}^{N-1} s[n] e^{-i2\pi\omega nT}. \tag{2.39}$$

By assuming that the observed signal represents a single period of an unbounded periodic function, then the frequency of the signal is $\frac{1}{NT}$ — if we evaluate the sum in equation 2.39 at this frequency at its k harmonics (up to $N-1$), we obtain

$$\mathscr{F}[k] = \sum_{n=0}^{N-1} s[n] e^{-i\frac{2\pi}{N}nk}, \tag{2.40}$$

such that $\mathscr{F}[k]$ is the discrete Fourier transform of the sequence $s[n]$. The inverse discrete Fourier transform is then [31]

$$s[n] = \frac{1}{N} \sum_{n=0}^{N-1} \mathscr{F}[k] e^{i\frac{2\pi}{N}nk}. \tag{2.41}$$

2.2.5.4 The fast Fourier transform

The calculation of each DFT coefficient $\mathscr{F}[k]$ for N points requires N^2 operations (i.e. $\mathscr{O}(N^2)$). The fast Fourier transform, as coined by Cooley and Tukey, requires only $\mathscr{O}(N\log(N))$ operations [32]. This algorithmic speed-up is driven by the realisation that there are many redundant calculations when computing the full discrete Fourier transform. The simplest case is for a signal with N samples where N is a power of 2. The expression for the discrete Fourier transform in equation 2.40 can be decomposed into sums over the even- and odd-numbered sample indexes

$$\sum_{n=0}^{N-1} s[n] e^{-i\frac{2\pi}{N}nk} = \sum_{m=0}^{\frac{N}{2}-1} s[2m] e^{-i\frac{2\pi}{N}(2m)k} + \sum_{m=0}^{\frac{N}{2}-1} s[2m+1] e^{-i\frac{2\pi}{N}(2m+1)k}. \tag{2.42}$$

The second term on the RHS of equation 2.42 can be factorised by an $e^{-i\frac{2\pi k}{N}}$ factor, which yields

$$\mathscr{F}[k] = \sum_{m=0}^{\frac{N}{2}-1} s[2m] e^{-i\frac{4\pi}{N}mk} + e^{-i\frac{2\pi k}{N}} \sum_{m=0}^{\frac{N}{2}-1} s[2m+1] e^{-i\frac{4\pi}{N}mk}. \tag{2.43}$$

The key result is that we have split the sum over N into two halves of size $\frac{N}{2}$. This can be performed recursively, progressively subdividing the signal into two. Because there are multiple combinations of m and k, the speed up in the FFT algorithm comes from removing redundant calculations of the $e^{-i\frac{2P\pi k}{N}m}$ factors, where P is the number of pairs. In terms of the time complexity, for the one-step recursion outlined in equation 2.43, the number of operations required is $2\left(\frac{N}{2}\right)^2 + N$. For P recursions, this becomes $\frac{N}{2^P} + PN$, and as $P = \log_2(N)$, this becomes $N + N\log(N)$, or $\mathscr{O}(N\log(N))$. It should be noted that this is the simplest case, and that the Cooley-Tukey algorithm can also be used for signals where N is not a power of 2 [33].

2.2.5.5 Application to voltammetry

For exponential functions of the form $e^{A\sin(\omega t)}$, the Fourier coefficients at the frequency harmonics (i.e. multiples of the input frequency ω), are proportional to the amplitude of the input sinusoid A. Because $\tilde{E}_{app}$ is present in the exponential term of the Butler-Volmer expressions for electrochemical kinetics (defined in equations 2.12 and 2.13) which are used in the model of the Faradaic current in equation 2.23, the current arising from any voltammetric experiment where $\tilde{E}_{app}$ contains a sinusoid will therefore contain harmonics that are proportional to the magnitude of $\Delta\tilde{E}$ — a non-linear response. Traditional analysis of currents obtained by "sinusoidal" voltammetry, such as in electrochemical impedance spectroscopy [19], has required an assumption of linearity, and consequently small experimental values for $\Delta\tilde{E}$, so that the assumption is valid. PSV and r-FTACV by contrast use a large amplitude sinusoid, to take advantage of the inherent non-linearity of the Faradaic process. By Fourier transforming the current from an r-FTACV or PSV experiment, we resolve the signal into a series of harmonics detected at $\omega, 2\omega, 3\omega, \ldots, n\omega$, as well as the aperiodic DC component. Access to the mechanistically sensitive higher order harmonics provides Faradaic responses hypothetically devoid of background charging current, as a result of the linearity of ideal capacitors [34]. As such the Fourier transform of current arising from an ideal capacitor will have a frequency response at the primary harmonic only. However, as indicated by our choice of a third-order polynomial in equation 2.21, we do observe capacitive current responses in the Fourier transform at higher frequency harmonics. It is likely that, for complex experimental systems, using a large amplitude potential input will also increase the non-linearity of the background current response. However, it is our experience that, because the Faradaic response is more highly non-linear, there is still a region of the Fourier spectrum that reports entirely on the Faradaic response.

2.2.6 Numerical details

2.2.6.1 Current simulations

In order to solve for the total current in equation 2.22 and 2.23 numerically, we discretise the time interval $[0, T]$ into N_t time steps and replace the time derivatives in equation 2.22 by finite difference approximations to give

$$I^{t+1} = C_{dl}\left(\frac{dE_{app}^{t+1}}{dt} - R_u\left(\frac{I^{t+1} - I^t}{dt}\right)\right) + \Gamma\left(\frac{\theta^{t+1} - \theta^t}{dt}\right), \tag{2.44}$$

where the superscript t and $t+1$ represent the value of the variable at the current and following time step respectively, for $t = 1, ..., N_t$. The expression for the rate of change of oxidised species is given by the Butler-Volmer equations (equations 2.12 and 2.13) , and rearranged for an expression of θ^{t+1} using a finite difference approximation, as above

$$\left(\frac{\theta^{t+1} - \theta^t}{dt}\right) = ((1 - \theta^{t+1})k_{ox}(E_{app}^{t+1}) - \theta^{t+1}k_{red}(E_{app}^{t+1})), \tag{2.45}$$

$$\theta^{t+1} = \frac{\theta^t + k_{ox}(E_{app}^{t+1})dt}{1 + k_{ox}(E_{app}^{t+1})dt + k_{red}(E_{app}^{t+1})dt}. \tag{2.46}$$

If equation 2.46 is substituted into equation 2.44, the results can be rearranged into a residual $F(I^{t+1}) = 0$, then we can use a minimisation or root-finding algorithm to find the value of I^{t+1} for a given I^t and set of parameters.

2.2.7 Other modelling approaches

In the above section, we discussed the necessary theory for understanding how to predict the current from a voltammetric experiment given a set of known electrochemical parameters. Assumptions about the physical behaviour of the system were made at several points, and we will briefly discuss some alternatives to these assumptions below.

2.2.7.1 Marcus-Hush theory

The Butler-Volmer model of electrode kinetics discussed in section 2.2.1 is a *macroscopic* model of electron-transfer. A deeper understanding of electrochemical kinetics can be obtained through Marcus theory, which takes into account the impact of *microscopic* effects. For the Butler-Volmer kinetics, we assume that the only electrons transferred are at the Fermi level (i.e. the highest energy level occupied by an electron at absolute zero), whereas the expression for $\tilde{k}_{red}(\tilde{E})$ and $\tilde{k}_{ox}(\tilde{E})$ under Marcus theory sums over all occupied energy levels, such that

$$\tilde{k}_{red}(\tilde{E}) = \sum_i f_i(\xi_i)H_{ab} \exp\left(-\frac{\Delta\tilde{G}_{red}^{\ddagger}(\xi_i)}{k_bT}\right), \tag{2.47}$$

$$\tilde{k}_{ox}(\tilde{E}) = \sum_i f_i(\xi_i)H_{ab} \exp\left(-\frac{\Delta\tilde{G}_{ox}^{\ddagger}(\xi_i)}{k_bT}\right), \tag{2.48}$$

where k_b is the Boltzmann constant, ξ_i is the energy of the ith orbital, H_{ab} is a measure of electronic coupling between the donor and acceptor of the electron and $f_i(\xi_i)$ is the Fermi-Dirac

probability that an electron will be in that orbital at a given potential, such that

$$f_i(\xi_i) = \frac{1}{1 + e^{\frac{\tilde{E} - \xi_i}{k_b T}}}. \tag{2.49}$$

The free energies for the oxidative and reductive reactions for a particular energy level ξ_i in the heterogeneous case (as encountered before for Butler-Volmer kinetics) are

$$\Delta\tilde{G}^{\ddagger}_{red}(\xi_i) = \frac{(\lambda + F(\xi_i - \tilde{E}^0))^2}{4\lambda}, \tag{2.50}$$

$$\Delta\tilde{G}^{\ddagger}_{ox}(\xi_i) = \frac{(\lambda - F(\xi_i - \tilde{E}^0))^2}{4\lambda}, \tag{2.51}$$

where λ is the energy required to re-organise the solvent molecules — this is following the Franck-Condon principle, which states that solvent does not move on the timescale of electronic transitions, and consequently the donor and acceptor states must be in the same solvent environment to facilitate electronic transition. Consequently, λ is a measure of the amount of energy required for the reorganisation necessary to obtain these identical environments. One of the main predictions from this construction is the existence of the so-called "inverted region". For a homogeneous reaction not involving an electrode, the $F(\xi_i - \tilde{E}^0)$ term is replaced by a $\Delta\tilde{G}^0$ (the difference between the standard free energies of the oxidised and reduced states). When $\Delta\tilde{G}^0 = -\lambda$, the activation energy for the homogeneous reduction reaction is 0 — but this is a minimum, and more negative $\Delta\tilde{G}^0$ values corresponding to a greater driving force for the reaction will have a slower rate than at the $\Delta\tilde{G}^0 = -\lambda$ point. For the heterogeneous case involving an electrode this inverted region (corresponding to more negative overpotentials) is not observed. This is because, although the electrode electrons at the Fermi level would be in the inverted region when transferring to the oxidised species, there are always electrode electrons occupying lower energies than the Fermi level (i.e. values of ξ_i) capable of transferring to the oxidised species. The gap left by the transferred electron is then filled by electrons in the electrode that are at the Fermi level. Therefore, what is observed is a plateau of the reaction rate beyond a given overpotential, corresponding to $-\lambda$. This is not predicted by Butler-Volmer theory, which expects the rate of reaction to increase exponentially and without limit as a function of overpotential [6].

2.2.7.2 Isotherms

The choice of isotherm relates to the adsorption properties of the electrochemical species under interrogation. In protein-film voltammetry, the proteins are physically immobilised on the electrode, and consequently we can neglect most of the properties of the adsorption/desorption process when constructing the model. Equation 2.23 is modelled using a Langmuir isotherm, which has a number of assumptions relevant for modelling [35].

1. Adsorbed molecules do not interact with each other (for the redox case, the only electron transfer is between the electrode and an adsorbed molecule).
2. The maximum surface coverage is a monolayer.
3. Molecule adsorption orientation is identical.

Dispensing with assumption 1 (i.e. including self-interaction terms) will predict broader oxidative and reductive peaks than one generated using a Langmuir isotherm [6], and we will show that this peak-broadening effect can also be replicated by dispensing with identical orientation assumption in chapter 3. In total, there are at least 15 isotherm models in the literature [36].

2.2.7.3 Dispersion

The Langmuir isotherm assumption of identical orientation means that every molecule linked to the surface of the electrode should have identical thermodynamic and kinetic properties. If this is not the case (as may be driven by, for example differing orientations of the molecule on the electrode), then there will be a distribution of these properties, and consequently in the parameters used to describe them, principally $\tilde{E}^0$ and $\tilde{k}_0$ [37]. These issues in the context of experimental and theoretical predictions for different surface-linked voltammetry techniques will be discussed further in section 3.3.

2.3 The inverse problem

In the preceding sections we have covered in detail how to predict a current arising from a surface-linked voltammetric experiment without the presence of additional catalytic activity, given a set of known parameters. However, when we analyse voltammetry data, we want to do the reverse — to determine the value of the governing parameters that produce an observed current. We refer to this process as solving the "inverse problem", as opposed to the "forwards problem" of

simulating the current. There is a great deal of literature (including some detailed review papers [38, 39]) related to attempts to infer parameters from PFV data, all of which in some sense require fitting a model to some observations. The mode of analysis either takes the form of a minimisation problem (between some form of the voltammetric current and simulations of the same form) or via feature extraction. The minimisation approach can be further subdivided into "manual" and "automatic" approaches — for the former, the voltammetric analyst manually alters individual parameter values to obtain a good fit "by-eye". In the automatic approach, both the parameter proposal and fit assessment is performed algorithmically, allowing for a more comprehensive and faster search through parameter space. In this thesis, we will use the automatic minimisation approach, complemented by Bayesian inference approaches which determine the level of confidence we can have in our inferred parameters. We will lay out our methodology in section 2.4. In terms of the feature extraction approach, a particular feature (such as oxidative peak-height) is extracted from multiple voltammograms recorded under different experimental conditions, such that the extracted feature has a predictable relationship to one or two parameters of interest — we compare Bayesian inference to the feature extraction approach in appendix C.

For systems with no catalytic activity, there are a variety of voltammetric techniques and concomitant methods to extract features. For DCV, the features include oxidative/reductive peak-separation, [8, 40, 41, 42], peak-width [43, 44] and background-subtracted total current (where the background current is either approximated using simulation, or assumed to be identical to the background current observed from an unmodified electrode [41, 45]). For r-FTACV, harmonic amplitudes [8, 46] and the total harmonics [8, 9, 19] have both been used.

For cases where the proteins under study are enzymatically active and catalytic redox processes are present, the overwhelmingly predominant technique is DCV, with some theoretical analyses exploring SWV [47], and the principal mode of analysis is the total current, background-subtracted or otherwise [48], sometimes interpreted as a "waveshape" [49, 50]. A large amount of effort and debate is expended on model development, as there are challenges related to distinguishing between different catalytic mechanisms [51], along with challenges related to dispersion [52] and substrate transport [53, 54]. In addition, the nature of the catalytic reactions means that boosting signal versus background by clever modulation of the input potential is not possible to the same extent as for straightforward purely non-catalytic electron-transfer reactions, and

naturally there is less innovation in this space. The variety of models, techniques and feature extraction processes referenced are indicative of the wide range of proteins that can be analysed using electrochemistry, as well as indicating something of an endless struggle to separate enough signal from background to be able to make valid inferences.

2.4 Research approach

Having provided the context of how other researchers have inferred from biological voltammetric data, we now detail the inference methodology used throughout the thesis. Our approach is twofold — instead of extracting features of interest from the total current, we attempt to use optimisation to minimise the distance between simulation and experimental voltammetry data. We start the optimisation process from random initial parameter values, and as such is performed multiple times, to ensure that a large portion of parameter space is searched. Once a vector of chemical parameters is attained that represents some underlying chemical reality, we then use Bayesian inference methods to infer a posterior distribution (i.e. given the data we have, the distribution of likely parameter values), which allows us to assess the confidence we should have in the inferred values. In the following sections, we lay out the ideas behind the algorithms used for optimisation and statistical inference.

2.4.1 Optimisation

2.4.1.1 Likelihood functions

To find the best-fit simulation for voltammetric data, we aim to maximise a Gaussian likelihood function, which is equivalent to minimising the least-squares distance between simulations (which are a function of the proposed parameter vector $\boldsymbol{\phi}$) and data. The Gaussian probability of the j^{th} current datum $I_{data}(t_j)$ being observed conditional on a set of parameters $\boldsymbol{\phi}$ (used to generate a simulated current $I_{sim}(t_j)$) and standard deviation σ is

$$p(I_{data}(t_j)|\boldsymbol{\phi}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{I_{data}(t_j) - I_{sim}(\boldsymbol{\phi}, t_j)}{\sigma^2}\right)^2}. \tag{2.52}$$

If the probability of every N_t individual current datum is independent, then the time-domain likelihood, or the probability of observing the entirety of the current I_{data} given $\boldsymbol{\phi}$, is equal to the product of the independent Gaussian probabilities, such that

$$\mathscr{L}_t(\boldsymbol{\phi}, \sigma | \boldsymbol{I_{data}}) = \prod_{j=1}^{N_t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{I_{data}(t_j) - I_{sim}(\boldsymbol{\phi}, t_j)}{\sigma^2},\right)^2}. \tag{2.53}$$

which can be simplified for the purposes of computation by taking the natural logarithm of equation 2.53 to obtain the log-likelihood

$$\log \mathscr{L}_t(\boldsymbol{\phi}, \sigma | \boldsymbol{I_{data}}) = -\frac{N_t}{2} \log 2\pi - N_t \log \sigma - \frac{1}{2\sigma^2} \sum_{j=1}^{N_t} (I_{data}(t_j) - I_{sim}(\boldsymbol{\phi}, t_j))^2. \tag{2.54}$$

Alternatively, a likelihood function can be derived to compare data in the frequency domain, filtered to contain N harmonics $h_0, h_1 ... h_N$. We again define a likelihood function $\log \mathscr{L}_f$ using squared distance, where N_ω is the length of the Fourier spectrum $\mathscr{F}$, as generated by the fast Fourier transform algorithm, and ω_k is the kth frequency. Taking the absolute value of the Fourier spectrum,

$$\log \mathscr{L}_f(\boldsymbol{\phi}, \sigma | \mathscr{F}_{\boldsymbol{data}}) = -\frac{N_\omega}{2} \log 2\pi - N_\omega \log \sigma - \frac{1}{2\sigma^2} \sum_{k}^{N_\omega} (F^{Th}(\omega_k)(|\mathscr{F}_{data}(\omega_k)| - |\mathscr{F}_{sim}(\boldsymbol{\phi}, \omega_k)|)^2), \tag{2.55}$$

such that

$$F^{Th}(\omega_k) = \sum_{n=h_0}^{h_N} F_n^{Th}(\omega_k). \tag{2.56}$$

F_n^{Th} is a top-hat filter (equation 2.57) centred on the n^{th} harmonic, and ω_{input} is the input frequency used in the calculation of E_{app} as in section 2.2.4, where w is the width of the window, with a maximum value of 0.5 (as the spacing between each harmonic has a value of ω).

$$F_n^{Th}(\omega_k) = \begin{cases} 1, & \text{if } \frac{n \cdot \omega_{\text{input}}}{w} \leq \omega_k \leq \frac{n \cdot \omega_{\text{input}}}{w}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.57}$$

By using a Gaussian log-likelihood in equation 2.55, we are assuming that the noise distribution in the frequency domain is also Gaussian. In appendix A, we show that the discrete Fourier transform of independent and identically distributed Gaussian noise is itself Gaussian distributed.

Having defined the likelihood function, we need to choose an optimisation algorithm — we have selected the Covariance Matrix Adaptation — Evolutionary Strategy (CMAES) algorithm, which was was selected for a number of reasons; it does not require the calculation of derivatives and it was explicitly designed to find a global optimum in rugged, high-dimensional parameter space. A simplified version of the algorithm is derived in appendix B.

2.4.2 Bayesian inference

The CMAES algorithm provides us with point estimates for electrochemical parameters, but using Bayesian inference we can also obtain a measure of confidence in this value — Markov chain Monte Carlo (MCMC) methods can provide information about the spread of likely values around the point estimates provided by the CMAES algorithm. Given experimental data $\mathbf{I_{data}}$, we want to infer a set of parameters $\boldsymbol{\phi}$ which accurately describe $\mathbf{I_{data}}$. We want a precise estimate of the posterior probability distribution $p(\boldsymbol{\phi}|\mathbf{I_{data}})$, which gives the probability of a particular parameter vector given our data. Bayes rule says that posterior probability is related to the likelihood, $p(\mathbf{I_{data}}|\boldsymbol{\phi})$ the probability that the data is described by the parameter vector $\boldsymbol{\phi}$, the prior $p(\boldsymbol{\phi})$ which includes our subjective belief about how the parameters are distributed, and the probability of observing the data, $p(\mathbf{I_{data}})$, such that

$$p(\boldsymbol{\phi}|\mathbf{I_{data}}) = \frac{p(\mathbf{I_{data}}|\boldsymbol{\phi})p(\boldsymbol{\phi})}{\int_{\boldsymbol{\phi}} p(\mathbf{I_{data}}|\boldsymbol{\phi})p(\boldsymbol{\phi})d\boldsymbol{\phi}}. \tag{2.58}$$

With a large set of parameters, the computation of $p(\mathbf{I_{data}})$ requires marginalising each individual parameter and integrating over all other parameters, resulting in analytically intractable multidimensional integrals. Therefore, we use the Metropolis-Hastings Algorithm, a Markov chain Monte Carlo method, to avoid computing this integral.

2.4.2.1 Markov chain Monte Carlo methods

In order to approximate the posterior probability distribution without calculating the multidimensional integrals required for the normalising factor, we use MCMC, a technique first described by Metropolis et al. [55]. The Markov chain is defined such that its stationary distribution is the posterior probability density. By drawing a sufficient number of samples, the Markov chain

will converge to the posterior. The Metropolis-Hastings algorithm performs MCMC by sampling the parameter space, and by moving accordingly towards areas of higher posterior density, where each parameter vector is a "state" in the Markov chain. Given a vector of parameter values, $\boldsymbol{\phi}$, which can include the noise standard deviation parameter σ, we wish to obtain the posterior distribution $p(\boldsymbol{\phi}|\mathbf{I_{data}})$ that best describes our data. We use MCMC methods to approximate some desired distribution $P(x)$ (for our case, this is the posterior, $p(\boldsymbol{\phi}|\mathbf{I_{data}})$). The algorithm uses an ergodic Markov chain, where the stationary distribution $h(x) = P(x)$. As the algorithm runs, the distribution of the samples should become an increasingly good approximation of $P(x)$. The transition probabilities between states of the Markov chain are calculated as the ratio of the current posterior distribution and a proposal posterior distribution. This cancels the $p(\mathbf{I_{data}})$ term, meaning we do not have to calculate the integral. This requires that the Markov chain must be reversible, so the probability of a transition $\boldsymbol{\phi} \rightarrow \boldsymbol{\phi}'$ is equal to $\boldsymbol{\phi}' \rightarrow \boldsymbol{\phi}$. This is the condition of detailed balance, such that:

$$P(\boldsymbol{\phi}'|\boldsymbol{\phi})P(\boldsymbol{\phi}) = P(\boldsymbol{\phi}|\boldsymbol{\phi}')P(\boldsymbol{\phi}'), \tag{2.59}$$

and therefore

$$\frac{P(\boldsymbol{\phi}'|\boldsymbol{\phi})}{P(\boldsymbol{\phi}|\boldsymbol{\phi}')} = \frac{P(\boldsymbol{\phi}')}{P(\boldsymbol{\phi})}. \tag{2.60}$$

For each step of the algorithm, a new state of the Markov chain is proposed according to a proposal distribution $q(\boldsymbol{\phi}'|\boldsymbol{\phi})$. Whether this transition happens or not is a function of the acceptance probability $A(\boldsymbol{\phi}', \boldsymbol{\phi})$. Therefore, the probability of the $\boldsymbol{\phi} \rightarrow \boldsymbol{\phi}'$ transition occurring is

$$P(\boldsymbol{\phi}'|\boldsymbol{\phi}) = q(\boldsymbol{\phi}'|\boldsymbol{\phi})A(\boldsymbol{\phi}', \boldsymbol{\phi}). \tag{2.61}$$

If we insert this relationship into equation 2.61, we obtain

$$\frac{A(\boldsymbol{\phi}', \boldsymbol{\phi})}{A(\boldsymbol{\phi}, \boldsymbol{\phi}')} = \frac{P(\boldsymbol{\phi}')}{P(\boldsymbol{\phi})} \frac{q(\boldsymbol{\phi}|\boldsymbol{\phi}')}{q(\boldsymbol{\phi}'|\boldsymbol{\phi})}. \tag{2.62}$$

If the proposal distribution is symmetric (i.e. $q(\boldsymbol{\phi}|\boldsymbol{\phi}') = q(\boldsymbol{\phi}'|\boldsymbol{\phi})$), equation 2.62 reduces to $\frac{A(\boldsymbol{\phi}', \boldsymbol{\phi})}{A(\boldsymbol{\phi}, \boldsymbol{\phi}')} = \frac{P(\boldsymbol{\phi}')}{P(\boldsymbol{\phi})}$. For the algorithm, we define a probability of acceptance α such that

$$\alpha = \min\left\{\frac{P(\boldsymbol{\phi}')}{P(\boldsymbol{\phi})}, 1\right\}. \tag{2.63}$$

According to Bayes rule (equation 2.58), the acceptance probability becomes

$$\alpha = \min\left\{\frac{p(data|\boldsymbol{\phi}')p(\boldsymbol{\phi}')}{p(data|\boldsymbol{\phi})p(\boldsymbol{\phi})}, 1\right\}, \tag{2.64}$$

where $\boldsymbol{\phi}'$ is a proposal parameter vector drawn from the proposal distribution such that $\boldsymbol{\phi}' \sim \mathcal{N}(\boldsymbol{\phi}, \mathbf{C})$ where $\mathbf{C}$ is the covariance matrix. For each step of the algorithm, a parameter vector is proposed, and then accepted or rejected. If the proposed vector is better than the current value, such that the acceptance ratio is greater than 1 (i.e. it has a greater posterior density), the new parameter vector is accepted. If it is lower than 1, then it may be accepted, with a probability of α [56]. The adaptive Metropolis-Hastings algorithm improves upon this by scaling the proposal distribution by a factor a such that a specific quantity of samples are accepted — usually 25%. This speeds up convergence to the entire distribution of $p(\boldsymbol{\phi}|\mathbf{I_{data}})$ because it allows the algorithm to escape areas of low-scoring parameter space when the acceptance rate is too low. It also increases the ability of the algorithm to escape local minima when the acceptance rate is too high, so the algorithm samples parameter space more coarsely [57]. The algorithm is shown in algorithm 2 without burn-in, although in practice the first few thousand samples will be discarded. Exponential smoothing is introduced in the updates to the mean of the current (t) iteration μ_t, covariance matrix $\mathbf{C}_t$ and step size a_t, using the learning parameter γ_t to incorporate information from previous generations when the number of iterations is low.

$\log(a_0) \leftarrow 0$;

$\boldsymbol{\phi}_0 \leftarrow$ initial parameter values;

$\mathbf{C}_0 \leftarrow$ a diagonal matrix;

$\boldsymbol{\mu}_0 \leftarrow \boldsymbol{\phi}_0$ $t \leftarrow 0$;

while $t < t_{final}$ **do**

  $\boldsymbol{\phi}' \sim \mathcal{N}(.|\boldsymbol{\phi}_t, a_t\mathbf{C}_t)$;

  if $p(\boldsymbol{\phi}) \neq 0$ **then**

    $\alpha \leftarrow \min\left\{\frac{p(\mathbf{I_{data}}|\boldsymbol{\phi}')p(\boldsymbol{\phi}')}{p(\mathbf{I_{data}}|\boldsymbol{\phi}_t)p(\boldsymbol{\phi}_t)}, 1\right\}$;

    $u \sim \mathcal{U}(0,1)$;

    if $u < \alpha$ **then**

      $\boldsymbol{\phi}_{t+1} \leftarrow \boldsymbol{\phi}'$;

      accepted $\leftarrow 1$;

    else

      $\boldsymbol{\phi}_{t+1} \leftarrow \boldsymbol{\phi}_t$;

      accepted $\leftarrow 0$;

    end

  else

    $\boldsymbol{\phi}_{t+1} \leftarrow \boldsymbol{\phi}_t$;

    accepted $\leftarrow 0$;

  end

  $\gamma_t \leftarrow t^{-0.6}$;

  $\mathbf{C}_{t+1} \leftarrow (1-\gamma_t) \times \mathbf{C}_t + \gamma_t \times (\boldsymbol{\phi}_{t+1} - \boldsymbol{\mu}_t)^T(\boldsymbol{\phi}_{t+1} - \boldsymbol{\mu}_t)$;

  $\boldsymbol{\mu}_{t+1} \leftarrow (1-\gamma_t) \times \boldsymbol{\mu}_t + \gamma_t \times \boldsymbol{\phi}_{t+1}$;

  $\log(a_{t+1}) \leftarrow \log(a_t) + \gamma_t \times (\text{accepted} - 0.25)$

end

Algorithm 1: The adaptive Metropolis-Hastings algorithm

2.4.2.2 Likelihoods

As a result of this algorithm, we only have to calculate the likelihood, $p(\mathbf{I_{data}}|\boldsymbol{\phi})$ and the prior, $p(\boldsymbol{\phi})$. We use an uninformative (log-uniform) prior, such that all values within the defined bounds are judged to be equally likely; the choice of bounds is specific to each parameter inference approach, and will be detailed where appropriate. For the calculation of the likelihood, the form

of the data can either be in the time- or frequency-domain, as defined in equations 2.54 and 2.55.

2.4.2.3 Burn-in

In practice, the algorithm takes some time to converge to the target distribution (if it does so at all). When inspecting the results from an inference attempt, the individual chains are inspected to assess the degree of convergence, and some number of initial samples (where the algorithm is still exploring a larger region of parameter space) are discarded as "burn-in". A metric of convergence is $\hat{R}$, which splits the (truncated) m chains in half to obtain $2m$ chains of length $\frac{n}{2}$. Two quantities are then constructed, the within chain variance W

$$W = \frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{i=1}^{n} (s_{ij} - \mu_j)^2, \tag{2.65}$$

where s_{ij} is the i^{th} sample from the j^{th} chain, and μ_j is the mean value of the j^{th} chain. The second quantity is B, or the between chain variance

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\mu_j - \mu)^2. \tag{2.66}$$

where μ is the mean of the individual chain means μ_j. The $\hat{R}$ parameter is then

$$\hat{R} = \sqrt{\frac{\frac{n-1}{n} W + \frac{1}{n} B}{W}}. \tag{2.67}$$

Consequently, the closer $\hat{R}$ is to 1, the smaller the between-chain variance is relative to the within-chain variance, which is an approximate measure of chain convergence.[58]. This is presented graphically for some example chains in figure 2.1. As can be seen, discarding the first 2500 samples from iterations where the chains were not well-mixed significantly lowers the $\hat{R}$ parameter.

In this chapter, we have discussed some general principles of surface-linked voltammetry, including how the current can be decomposed into Faradaic and capacitive generating processes, and how these processes are modelled as a function of physical and chemical parameters. We have also mentioned the assumptions behind this modelling approach, and have described some alternatives. We indicated the literature on attempting to infer these parameters from voltam-

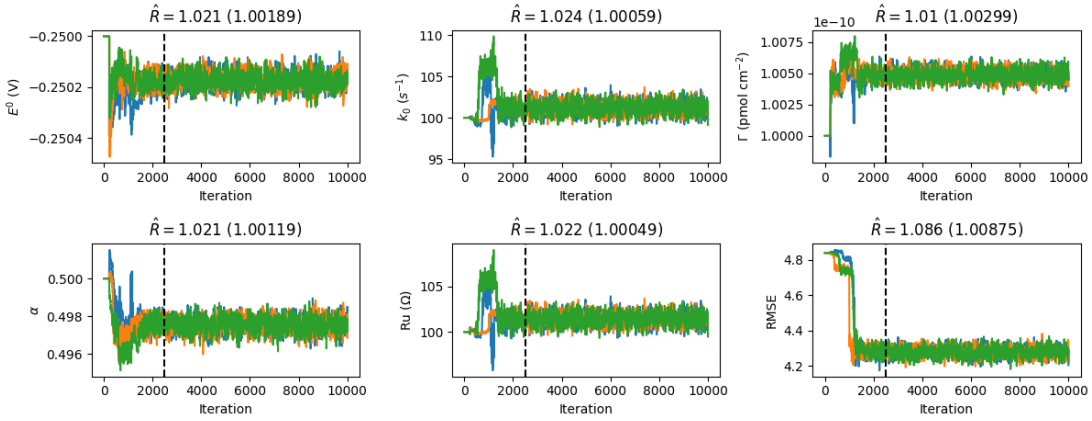


Figure 2.1: 3 MCMC chains for 6 parameters as listed on the y-axis, collected when fitting synthetic data in the Fourier domain. The $\hat{R}$ parameter is listed on the top of each graph, without burn-in, and after discarding of the first 2500 samples (indicated with a black dotted line).

metric currents in a biological setting. Subsequently, we laid out the underpinning methodology for how we intend to obtain parameters from voltammetry current data — determining a best-fit vector of parameters using multiple CMAES runs, and then using MCMC to determine the posterior distributions.

Chapter 3

Theoretical results

In chapter 2, we described the theoretical underpinnings of voltammetry, along with descriptions of how we model the experimental current arising from a surface-confined voltammetry experiment, and the methods used to infer governing chemical parameters from voltammetric data. In this chapter we use analytical and computational tools to gain a greater understanding of the forwards and inverse problem as applied to voltammetry. First, we demonstrate that our PSV simulation approach laid out in section 2.2 and 2.2.6.1 is accurate, by comparison to another method from the literature. Having demonstrated this, we then present a sensitivity analysis, graphically demonstrating the impact of the model parameters introduced in section 2.2, and further discuss complications that arise when fitting experimental voltammetric data. Finally, we infer parameters from synthetic noisy currents, generated using DCV, PSV and r-FTACV potential inputs, and assess how accurately we can recover these parameters.

3.1 Comparison of numerical and analytical approaches for simulating purely sinusoidal voltammetry

Content in this section has been re-purposed from the following reference: Henry Lloyd-Laney, Martin Robinson, Alan Bond, Alison Parkin, and David Gavaghan, Comparison of numerical and analytical approaches for simulating purely sinusoidal voltammetry ***Chemrxiv****. The paper is currently under review for the journal Chemical Results.*

An alternative method to predicting the current arising from a chemical system with known

parameters is to use an analytical approach. A unified mathematical methodology is rarely possible for analytical solutions to the forward problem, which often require approximations or constraints to achieve an expression for either the total current or some feature derived thereof, such as in the work of Bazant [59, 60]. However, a tractable analytical solution allows us to quantify numerical error, and reduces the computational burden. Prior to our interest in sinusoidal voltammetry, Bell et al. published an analytical approximation (referred to for convenience as "the analytical method") for the total current arising from a surface confined PSV experiment (which they termed "large-amplitude sinusoidal voltammetry" [12]). The effects of uncompensated resistance ($\tilde{R}_u$) was not included in the modelling approach. The effect of linear double-layer capacitance ($\tilde{C}_{dl}$) was discussed by Bell and coauthors, but we do not use it in our comparison. Consequently, the values of $\tilde{R}_u$ and $\tilde{C}_{dl}$ were set to zero when generating the numerical solution to allow for meaningful comparison.

In this section, we reproduce the derivation of the analytical approach developed by Bell et al., and demonstrate that the solutions obtained by the numerical methodology defined in chapter 2 and the analytical approach are equivalent.

3.1.1 Derivation of simulation models

Using the notation used by Bell et al. we briefly lay out the simulation approach for the numerical and analytical methods. As before in section 2.2, we use the Butler-Volmer equations to represent redox reaction rates — $\tilde{k}_{ox}$ and $\tilde{k}_{red}$,

$$\tilde{k}_{ox}(\tilde{t}) = \tilde{k}_0 e^{\left((1-\alpha)\frac{F}{RT}(\tilde{E}(\tilde{t})-\tilde{E}^{0'})\right)}, \tag{3.1}$$

using the rate constant $\tilde{k}_0$, the reversible potential $\tilde{E}^0$ and the symmetry factor α, and

$$\tilde{k}_{red}(\tilde{t}) = \tilde{k}_0 e^{\left(\frac{-\alpha F}{RT}(\tilde{E}(\tilde{t})-\tilde{E}^{0'})\right)}, \tag{3.2}$$

where the sinusoidal input potential $\tilde{E}(\tilde{t})$ is defined as

$$\tilde{E}(\tilde{t}) = \tilde{E}_{in} + \Delta\tilde{E}\sin(\tilde{\omega}\tilde{t}), \tag{3.3}$$

where $\tilde{E}_{in}$ is the average input voltage, or $\frac{E_{start}+E_{reverse}}{2}$ as long as the signal is periodic. The Faradaic current can consequently be represented as

$$\tilde{i}(\tilde{t}) = FA\frac{d\tilde{\theta}_{ox}}{d\tilde{t}} = \tilde{k}_{ox}(\tilde{t})\tilde{\theta}_{red}(\tilde{t}) - \tilde{k}_{red}(\tilde{t})\tilde{\theta}_{ox}(\tilde{t}), \tag{3.4}$$

where $\tilde{\theta}_{ox}$ and $\tilde{\theta}_{red}$ are the surface concentrations of the oxidised and reduced species on the surface of the electrode respectively. Equation 3.4 can be non-dimensionalised in multiple different ways, depending on the chosen timescale. Both the numerical and analytical methods define dimensionless E, E^0 and θ in the same way, where $\tilde{\theta}_{tot}$ is the total surface coverage

$$\tilde{E}^0 = \frac{RT}{F}E^0, \quad \tilde{E}(\tilde{t}) = \frac{RT}{F}E(t), \quad \tilde{\theta} = \theta\tilde{\theta}_{tot},$$

where $\sim$ indicates a dimensionless parameter. For the analytical case, time is non-dimensionalised using the rate constant $\tilde{k}_0$ such that

$$\tilde{t} = \frac{t}{\tilde{k}_0}, \quad \tilde{i}(\tilde{t}) = FA\tilde{k}_0\tilde{\theta}_{tot}i(t), \quad \omega = \frac{\tilde{\omega}}{\tilde{k}_0}.$$

For the numerical case, we non-dimensionalise time using the period of an oscillation of the potential input, v

$$\tilde{t} = \frac{t}{v}, \quad \tilde{i}(\tilde{t}) = FAv\tilde{\theta}_{tot}i(t), \quad \omega = \frac{\tilde{\omega}}{v}.$$

At this point, the approach to obtaining the current for the two methods diverges.

3.1.2 Derivation of the analytical approximation

This derivation follows closely that given by Bell et al. [12] and is reproduced here. Since the total concentration of the species is conserved (i.e. $\tilde{\theta}_{ox}(\tilde{t}) + \tilde{\theta}_{red}(\tilde{t}) = \tilde{\theta}_{tot}$), we can write equation 3.4 as

$$FA\frac{d\tilde{\theta}_{ox}}{d\tilde{t}} + (\tilde{k}_{ox}(\tilde{t}) + \tilde{k}_{red}(\tilde{t}))\tilde{\theta}_{ox}(\tilde{t}) = \tilde{k}_{ox}(\tilde{t})\tilde{\theta}_{tot}. \tag{3.5}$$

We can further expand this using the Butler-Volmer descriptions of the electrode kinetics defined in section 2.2.1, and non-dimensionalise as discussed, writing θ_{ox} as θ for notational convenience

$$\frac{d\theta}{dt} + \left(e^{(1-\alpha)\eta}e^{(1-\alpha)\Delta E\sin(\omega t)} + e^{-\alpha\eta}e^{-\alpha\Delta E\sin(\omega t)}\right)\theta(t) = e^{(1-\alpha)\eta}e^{(1-\alpha)\Delta E\sin(\omega t)}, \tag{3.6}$$

where $\eta = E_{in} - E^0$ is the average input voltage minus the reversible potential. This differential equation can be solved analytically using integrating factors. To be concise, we rewrite equation 3.6 in terms of $g(t)$ and $h(t)$, where

$$g(t) = e^{(1-\alpha)\eta} e^{(1-\alpha)\Delta E \sin(\omega t)} + e^{-\alpha\eta} e^{-\alpha \Delta E \sin(\omega t)}, \tag{3.7}$$

and

$$h(t) = e^{(1-\alpha)\eta} e^{(1-\alpha)\Delta E \sin(\omega t)}. \tag{3.8}$$

Equation 3.6 is then written as

$$\frac{d\theta}{dt} = h(t) - g(t)\theta(t), \tag{3.9}$$

To obtain an analytical expression for $\theta(t)$, we designate the integrating factor $\mathcal{I}$ as

$$\mathcal{I} = e^{\int_0^t g(u)du}. \tag{3.10}$$

Multiplying equation 3.6 by $\mathcal{I}$ and integrating yields

$$\theta(t) = e^{-\int_0^t g(u)du} \left(\theta_0 + \int_0^t e^{\int_0^s g(u)du} h(s) ds \right) \tag{3.11}$$

To simplify the expression in equation 3.11, we use a modified Bessel function expansion in h and g, where the modified Bessel function of order k is written as I_k [61]

$$e^{z \sin(\omega t)} = I_0(z) - 2 \sum_{k=1}^{\infty} (-1)^k \Big(I_{2k-1}(z) \sin((2k-1)\omega t) - I_{2k}(z) \cos(2k\omega t) \Big). \tag{3.12}$$

Additionally, if equation 3.12 is integrated with respect to time, we obtain

$$\int e^{z \sin(\omega t)} dt = I_0(z)t + C + \frac{2}{\omega} \sum_{k=1}^{\infty} (-1)^k \left(\frac{I_{2k-1}(z) \cos((2k-1)\omega t)}{2k-1} + \frac{I_{2k}(z) \sin(2k\omega t)}{2k} \right) \tag{3.13}$$

Since the infinite sum converges to a finite value by the alternating series rule, we can assume that for large ω

$$\int e^{z \sin(\omega t)} dt \approx I_0(z)t \tag{3.14}$$

as the error in this assumption is scaled by ω^{-1}. We can substitute equation 3.14 into equation 3.11; if we first look at the integrating factor

$$\mathcal{I} = e^{\int_0^t g(u)du} \approx \exp\left(e^{(1-\alpha)\eta} I_0((1-\alpha)\Delta E)t + e^{(-\alpha)\eta} I_0(-\alpha\Delta E)t\right) \tag{3.15}$$

which can be written as

$$\mathcal{I} \approx e^{\sigma t}, \tag{3.16}$$

where

$$\sigma = e^{(1-\alpha)\eta} I_0((1-\alpha)\Delta E) + e^{(-\alpha)\eta} I_0(-\alpha\Delta E). \tag{3.17}$$

Consequently, equation 3.11 can be represented as

$$\theta(t) \approx e^{-\sigma t}\left(\theta_0 + e^{(1-\alpha)\eta}\int_0^t e^{\sigma s} e^{(1-\alpha)\Delta E \sin(\omega s)} ds\right), \tag{3.18}$$

the integral can be solved using integration by parts to yield

$$\theta(t) \approx e^{-\sigma t}\left(\theta_0 + e^{(1-\alpha)\eta} I_0((1-\alpha)\Delta E)\left(\frac{e^{\sigma t}}{\sigma} - \frac{1}{\sigma}\right)\right). \tag{3.19}$$

This can be simplified algebraically to the final form of the approximation

$$\theta(t) \approx e^{-\sigma t}(\theta_0 - \theta_\infty) + \theta_\infty, \tag{3.20}$$

where

$$\theta_\infty = \frac{I_0((1-\alpha)\Delta E)}{I_0((1-\alpha)\Delta E) + e^{-\eta} I_0(-\alpha\Delta E)}. \tag{3.21}$$

Finally, equation 3.9 can be expressed in the form

$$i(t) = h(t) - g(t)(e^{-\sigma t}(\theta_0 - \theta_\infty) + \theta_\infty) + O(\omega^{-1}). \tag{3.22}$$

The key result from this derivation comes from equation 3.18; the error in the expression is dependent on the inverse of the dimensionless input frequency ω^{-1}.

Parameter	Symbol (Unit)	Value
Start potential	$\tilde{E}_{start}$ (V)	-0.25
Reverse potential	$\tilde{E}_{reverse}$ (V)	0.25
Reversible potential	$\tilde{E}^0$ (V)	0.1
Rate constant	$\tilde{k}_0$ (s^{-1})	Variable
Uncompensated resistance	$\tilde{R}_u$ (Ω)	0*
Linear double-layer capacitance	$\tilde{C}_{dl}$ (F)	0*
Surface coverage	$\tilde{\theta}_{tot}$ (mol cm^{-2})	1.0E-11
Symmetry factor	α	0.55
Potential frequency	$\tilde{\omega}$ (Hz)	Variable
Potential amplitude	$\Delta\tilde{E}$ (V)	0.25
Area	A (cm^2)	0.03

Table 3.1: Simulation parameters used to generate analytical and numerical simulations as compared in figure 3.2. Parameters marked with an asterisk were not modelled in the analytical approximation, and set to zero for the numerical solution. The values of the parameters were largely taken from Bell et al. [12].

3.1.3 Comparing the analytical and numerical models of PSV

In order to check the validity of the numerical approach, we compare the predicted dimensional current generated using the numerical and analytical solutions, using the simulation parameters in table 3.1. We show the impact of increasing frequency on the analytical and numerical currents in figure 3.1. We see that higher input frequencies lead to increasing agreement between the numerical and analytical solutions, as predicted by the error term discussed in the derivation section. The fact that there is good agreement between the analytical and numerical approaches confirms the accuracy of the numerical approach

In figure 3.2, the relative difference was calculated using the expression

$$\text{Relative difference} = \frac{\text{RMSD}(i_a, i_n)}{\bar{i}_n}, \tag{3.23}$$

where RMSD is the root mean square difference, and i_a and i_n are the analytical and numerical currents respectively. The difference is expressed as a percentage in figure 3.2.
We cannot explicitly calculate the error of the analytical approximation, only the difference

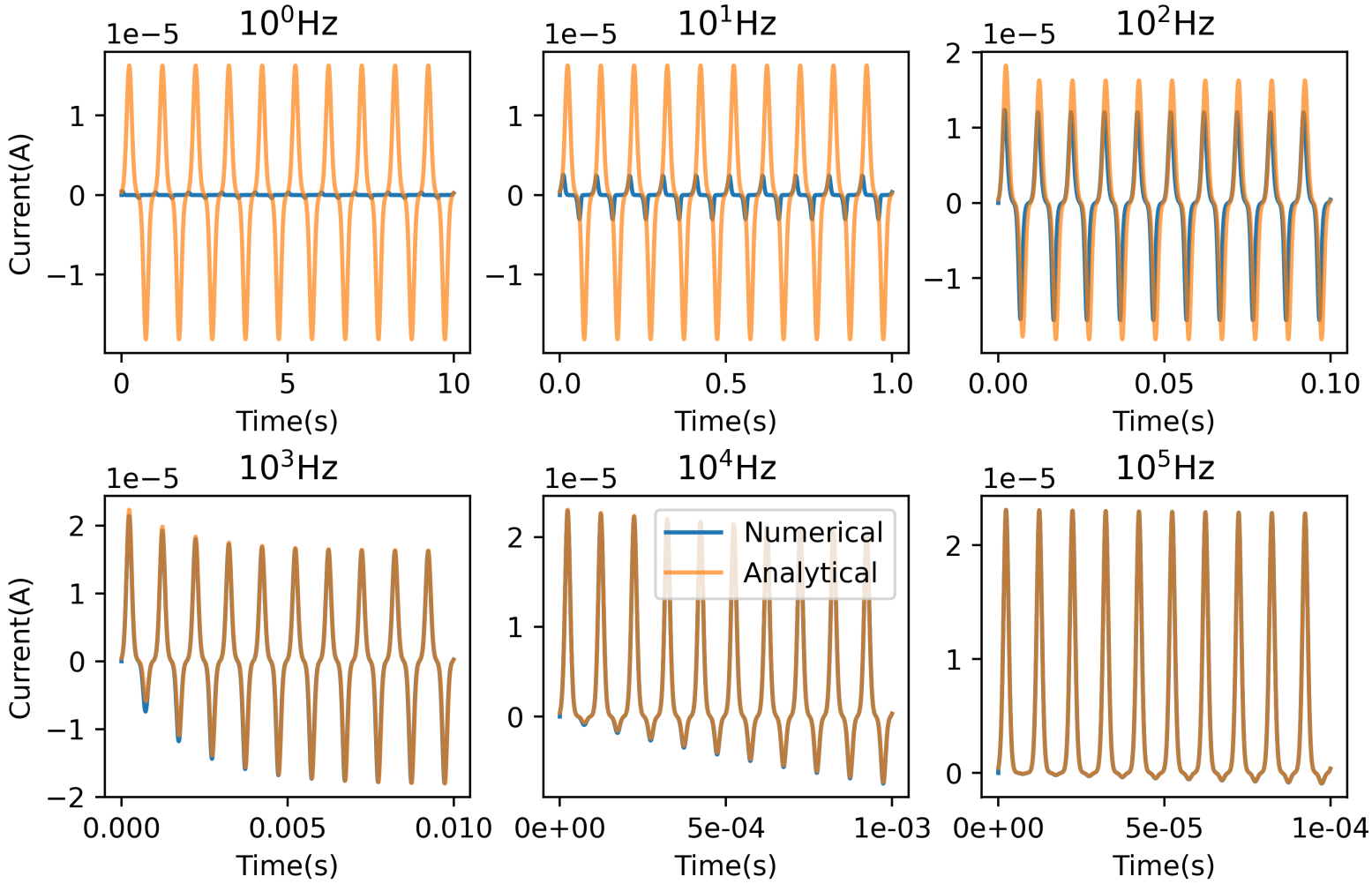


Figure 3.1: Current simulated using analytical and numerical methods against time. The value of the input frequency is shown in the title of each plot, and all other parameters are as in table 3.1, with $\tilde{k}_0$ set to 10 s^{-1}.

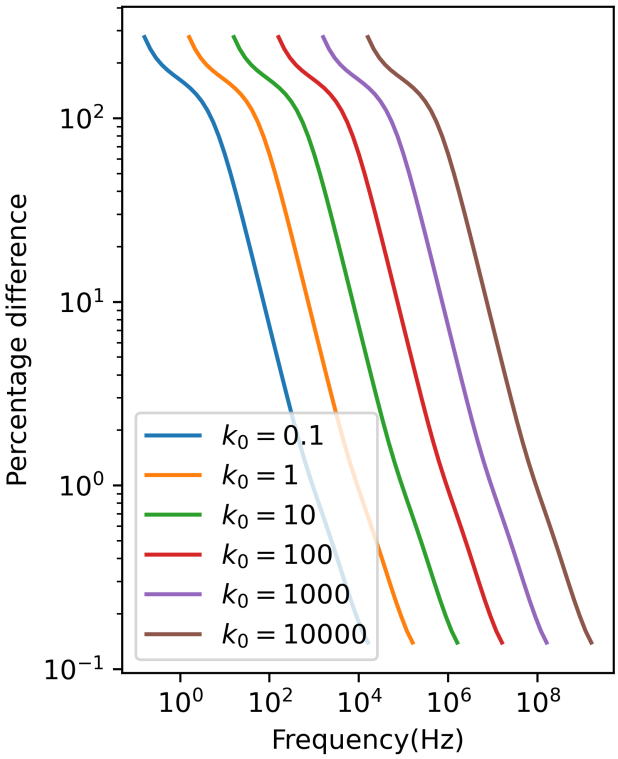


Figure 3.2: Percentage difference between the analytical and numerical currents as function of the kinetic parameter k_0 and the input frequency. The kinetic value $\tilde{k}_0$ in units of s^{-1} is shown in the legend, with the rest of the simulation parameters reported in table 3.1.

between it and the numerical approach. However, as expected, the accuracy (as measured by deviation from the numerical current) of the analytical expression decreases as $\tilde{\omega}$ increases. Because the analytical approximation error is proportional to the inverse of the dimensionless frequency $O(\omega^{-1})$, this can be represented as $O(X\%)$, where X is the ratio $\frac{\tilde{k}_0}{\tilde{\omega}}$. As would be expected, for constant $\tilde{k}_0$, increasing the input frequency by an order of magnitude decreases the difference between the numerical and analytical solutions by approximately an order of magnitude. For the difference between the two approaches to be on the order of 0.1% requires a frequency that is five orders of magnitude larger than $\tilde{k}_0$.

From these comparisons it is apparent that the frequency needs to be very large to obtain an analytical current solution that agrees with the numerical simulations, in some cases larger

than is experimentally realisable. While there are some advantages to solving the forwards problem this way, in particular computational speed, the issue remains that the choice of frequency affects the amount of information contained within the experimental current for a given system. If we use the analytical approach, then we are obligated to use a frequency that minimises the approximation error. In addition, the analytical approximation does not account for the impact of non-linear double layer capacitance and uncompensated resistance, both of which we have found to be key parameters when attempting to fit to experimental PSV data. For the rest of this thesis, all parameters discussed are in dimensional form, and we dispense with the use of $\sim$ to indicate dimensional parameters.

3.2 Exploring the forwards problem

Having satisfied ourselves that our modelling approach for PSV is accurate in the previous section (and with the models for r-FTACV and DCV well-established in the existing literature), we now discuss the implications of the model. It is useful to have an intuition for the effect of each parameter on the voltammetric current, so as to interpret the parameter values returned using the methodology laid out in section 2.4. Consequently, in this section, we graphically present the effect of each model parameter for a surface-confined single-electron redox process, holding the other parameters constant, at the values in table 3.2. The experimental input parameters are taken from a PSV experiment, performed on a glassy carbon electrode modified with a bacterial cytochrome (which is analysed in chapter 5), and the chemical parameters were selected for clarity of exposition, particularly the surface coverage and background current parameters. In a real PFV experiment, where the Faradaic-to-background current ratios are significantly poorer. The constant parameter values used to generate all the simulations are shown in table 3.2.

In figures 3.3 - 3.14 we show simulated total currents arising from DCV, PSV and r-FTACV simulations for a surface-confined species with no dispersion, although figures 3.5, 3.13 and 3.14 do not include DCV, as these parameters are not relevant for a DCV experiment. For PSV and r-FTACV, on the bottom row of each figure we show harmonics 1-7, obtained by Fourier transforming the total current, selecting the appropriate harmonic (by zeroing all other regions of the spectrum) and inverse Fourier transforming the spectrum back into the time domain. One of the weaknesses of PSV is that the effect of the model parameters on the appearance of the PSV harmonics defies easy explanation, in stark comparison to r-FTACV harmonics. Nevertheless,

we include them for completeness. As there is no non-linearity in the DCV current response, and consequently no harmonics in its Fourier spectrum, we have instead shown the effect of the changing parameter on peak-position data, using a trumpet plot, as introduced by Laviron [62]. In a trumpet plot, the peak potential (i.e. the potential at which the oxidative and reductive current peaks can be found) is extracted from multiple DCV experiments conducted at different linear sweep scan rates.

It is interesting to note that PSV seems to be generally more sensitive to changes in Faradaic parameters than the equivalent DCV experiment (for example in figures 3.4, 3.6 and 3.12), despite the potential inputs being largely similar (as mentioned in section 2.2.4.3). This is relatively surprising (albeit in a good direction) and for reasons that are not fully clear, although one plausible explanation is that the input potential of PSV is a smooth function that elicits a non-linear response. As a result, the relative sensitivity possibly relates to our choice of potential input phase and frequency for PSV — for future work, we are interested in obtaining more detailed understanding of how these input parameters can affect the amount of information available about the various parameters of interest. In the following section, we discuss each parameter in turn.

3.2.1 Reversible potential — E^0

From the Nernst equation, we know that the reversible potential E^0 is the potential at which the concentration of reduced and oxidised species are equal at equilibrium, and the Butler-Volmer equations (as in section 2.2.1 and also in equations 3.1 and 3.2) predict that, for a symmetrical reaction, the rates of oxidation and reduction are equal. Consequently, for the single-electron case, it is the potential halfway between reductive and oxidative peaks and changing E^0 alters this position for all three experiments, as can be observed in figure 3.3.

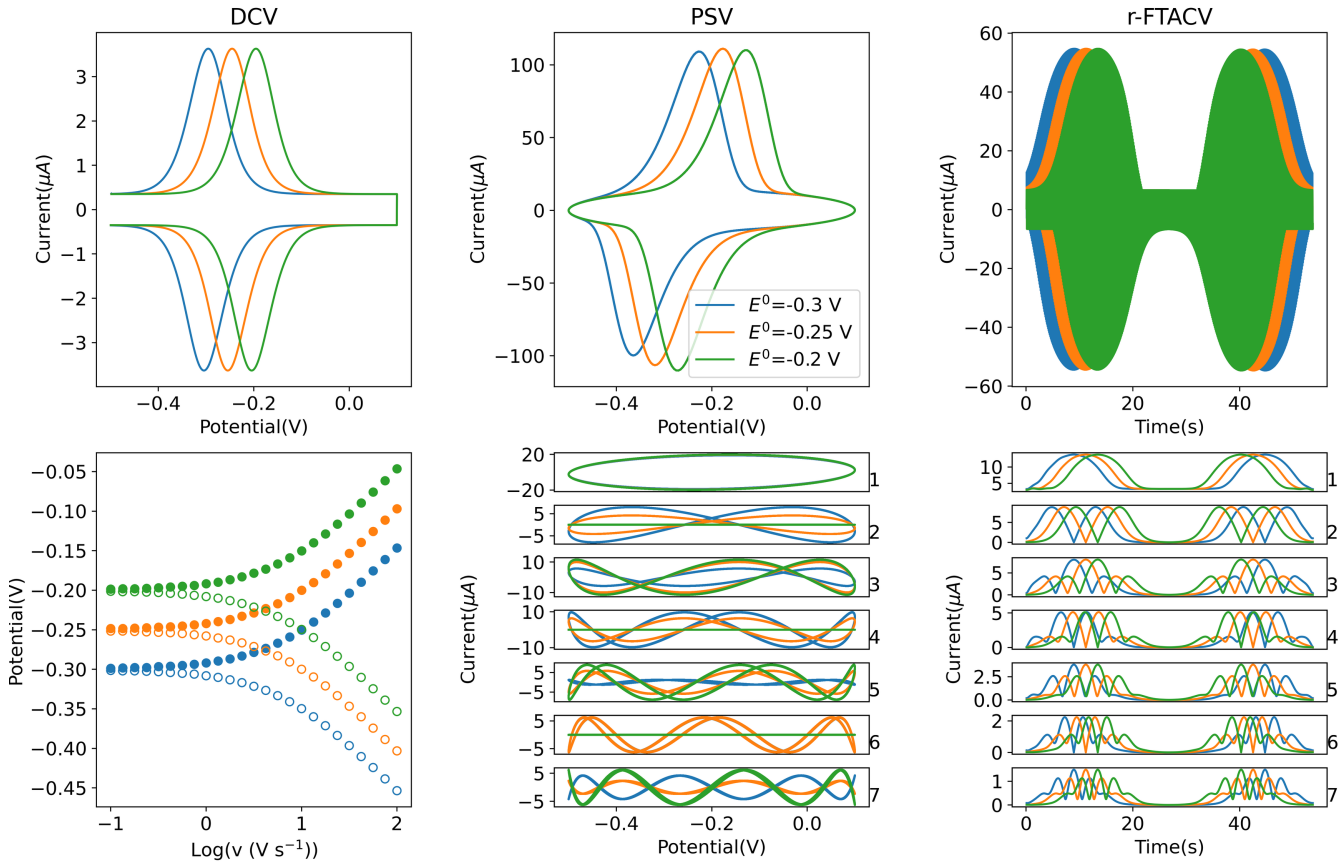


Figure 3.3: Parameter scan for E^0, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the reversible potential E^0 which is as indicated in the legend.

Parameter	Symbol	Value
Start potential	E_{start} (V)	-0.5
Reverse potential	$E_{reverse}$ (V)	0.1
Reversible potential	E^0 (V)	-0.25
Rate constant	k_0 (s^{-1})	100
Uncompensated resistance	R_u (Ω)	0.0
Double-layer capacitance	C_{dl} (F)	1.0E-5
Surface coverage	Γ (mol cm^{-2})	1.0E-10
Symmetry factor	α	0.5
Scan rate	v (Vs^{-1})	0.022/*/1
Potential frequency	ω (Hz)	8.881/8.94/*
Phase	η	$0.0/\frac{3\pi}{2}/*$
Potential amplitude	ΔE (V)	0.15/0.3/*
Sampling rate	(s^{-1})	400.0
Electrode Area	Area (cm^2)	0.07

Table 3.2: Simulation parameters held constant in figures 3.3-3.14. For some experimental parameters, the simulation value is different for each technique, and is written in the following order: r-FTACV/PSV/DCV, with a * indicating that the parameter is not used in the simulation

3.2.2 Rate constant — k_0

The rate constant k_0 alters the shape and position of the Faradaic peak. In more detail, for very fast kinetics (termed "reversible"), further increases in the kinetic parameter do not alter the appearance of the current. The reversible limit can also be thought of as the Nernstian limit, where the concentration of oxidised and reduced species instantaneously equilibrates at the values predicted by the Nernst equation on the timescale of each potential change. In the reversible regime, there is no difference between the potential of the oxidative and reductive Faradaic peaks (as can be observed in figure 3.4 for k_0 =1000s^{-1}). In terms of the r-FTACV harmonics, in the reversible kinetic regime, the number of lobes observed in the peak is the same as the harmonic number. As the kinetic parameter decreases away from the reversible limit, the Faradaic oxidative and reductive peak potentials diverge away from E^0, as can be observed when the current is plotted against potential, as in the PSV and DCV simulations. For the r-FTACV harmonics, the current magnitude decreases, and we observe a merging of the harmonic lobes,

as for the lower values of k_0 in figure 3.4. For all three experiments, decreasing k_0 away from the reversible limit reduces the magnitude of the Faradaic peak. In terms of the effect of the kinetic parameter on the trumpet plot, we can see that the scan-rate at which the peak potentials diverge is dependent on the value of the kinetic parameter.

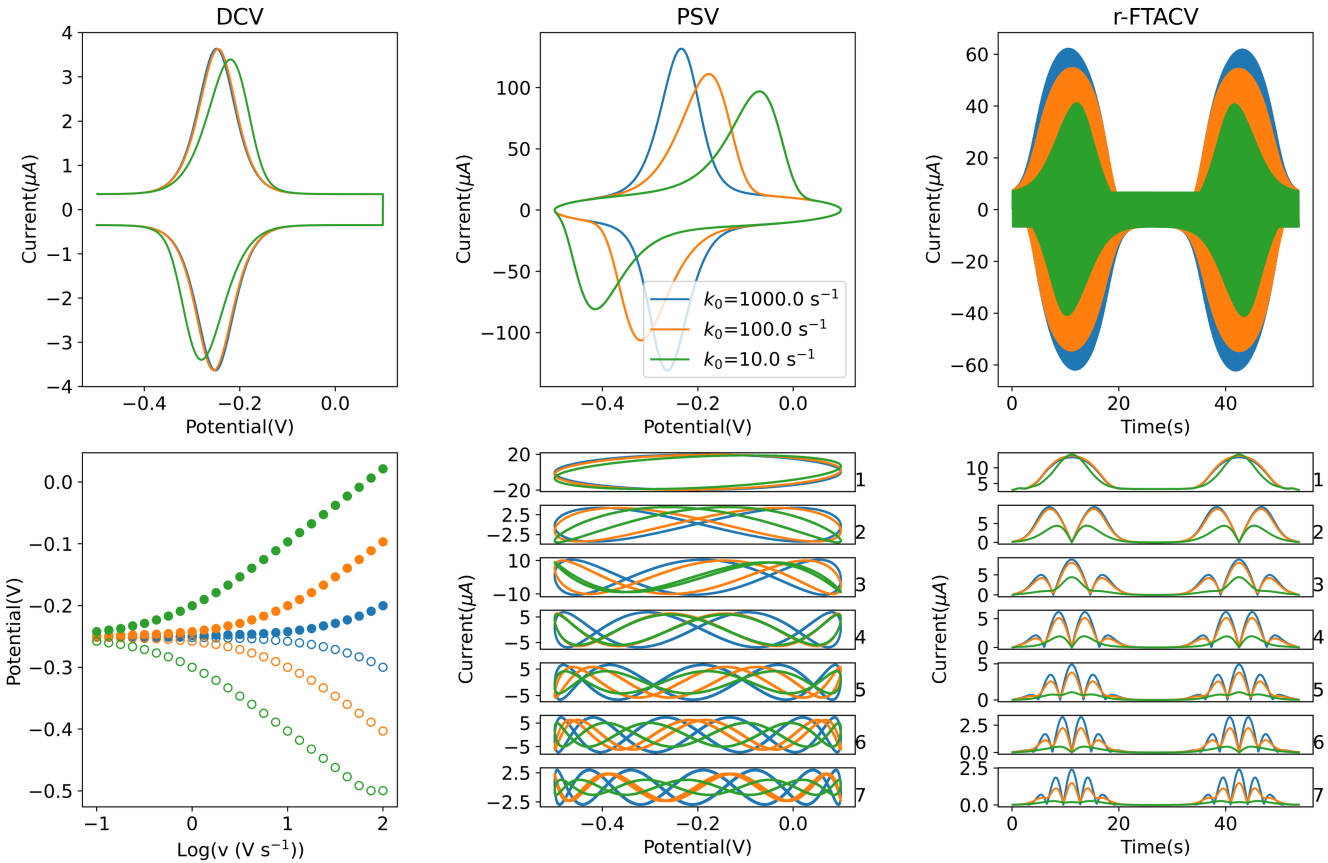


Figure 3.4: Parameter scan for k_0, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the rate constant k_0 which is as indicated in the legend.

3.2.3 Sinusoid frequency — ω

The input frequency ω, relevant only for PSV and r-FTACV, has a similar effect to that of changing the scan rate for the DCV case. For a fixed value of k_0, increasing the input frequency is equivalent to lowering the kinetic parameter with a fixed input frequency, although increasing the frequency does have other effects, such as increasing the magnitude of the total current, as can be observed in figure 3.5. Since the observed kinetic behaviour is a function of the timescale of changes in the potential, increasing this rate naturally alters the observed kinetic behaviour.

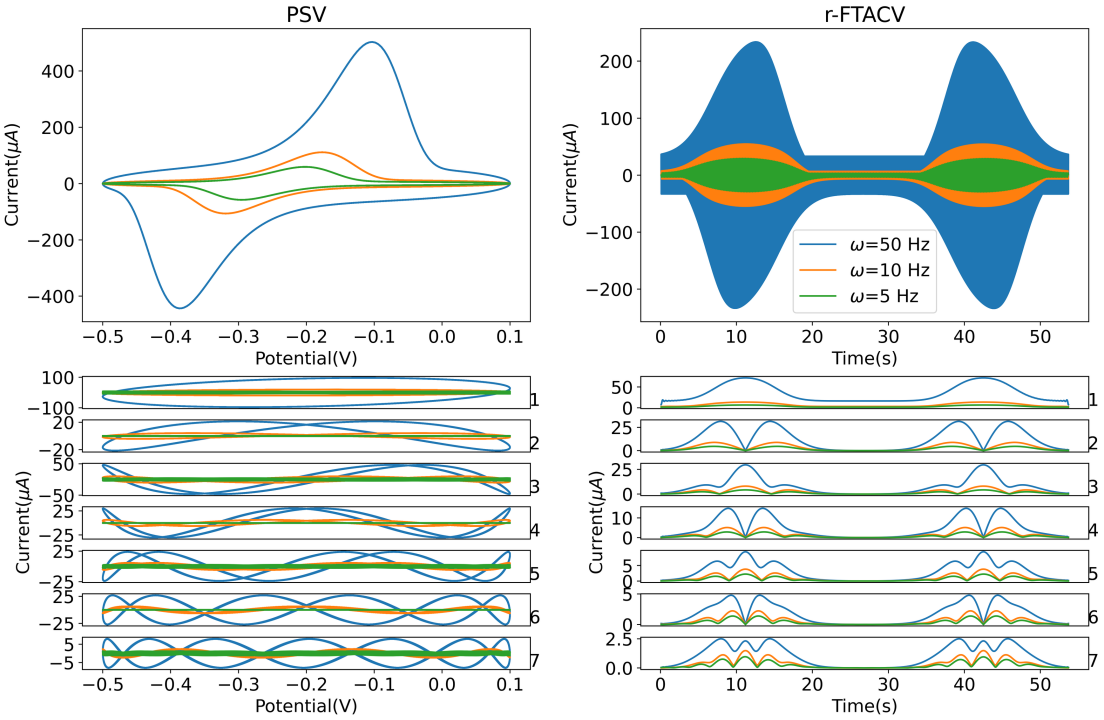


Figure 3.5: Parameter scan for the effect of altering the potential frequency ω on PSV and r-FTACV only, as this parameter is not relevant for DCV

3.2.4 Uncompensated resistance — R_u

The uncompensated resistance affects both capacitive and Faradaic currents through the Ohmic (IR_u) drop effect. The effect of increasing the uncompensated resistance has a similar, but not identical, effect to changing the kinetic parameter. For example, in the trumpet plot, increasing the value of R_u increases the gradient of the peak potentials, rather than the divergence point, as can be seen in figure 3.6. However, a sufficiently high value of R_u can look similar to a different kinetic parameter regime (i.e. a higher k_0 than the true value). Consequently, this can make estimation of the kinetic parameter challenging if the value of R_u is unknown and has to be estimated simultaneously.

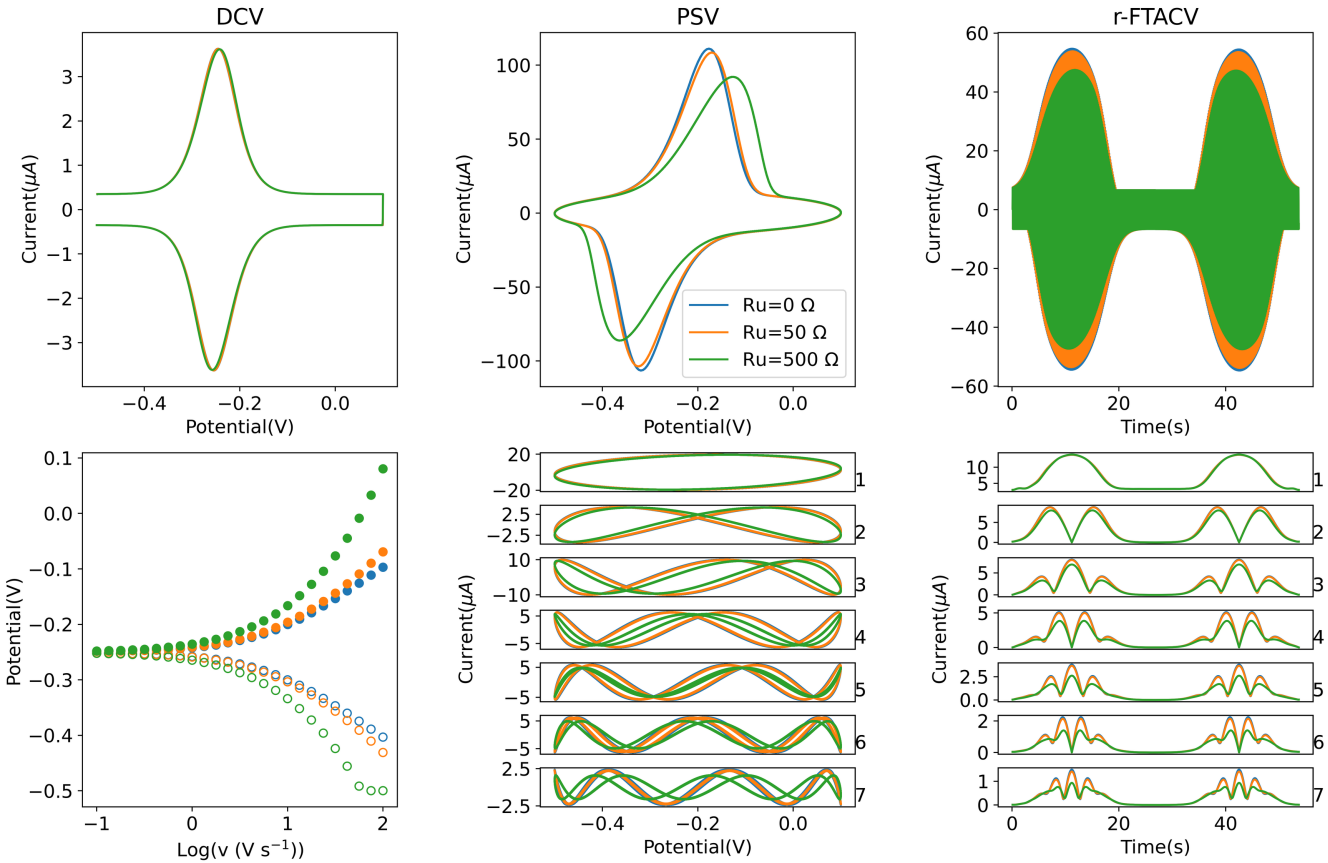


Figure 3.6: Parameter scan for R_u, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the uncompensated resistance R_u which is as indicated in the legend.

3.2.5 Capacitance — C_{dl} and C_{dlEX}

Linear capacitance C_{dl} linearly increases the magnitude of the total current, independent of the potential, as shown in figure 3.7. Additionally, we see that changing the linear capacitance has an impact only on the first harmonic of the PSV and r-FTACV timeseries. Higher order C_{dlEX} terms, as shown in figures 3.7-3.10 introduce capacitance effects as a function of the applied potential. We can also see, through inspection of the PSV and r-FTACV harmonics in figures 3.7 - 3.10, that each term affects the harmonics at X+1 and below.

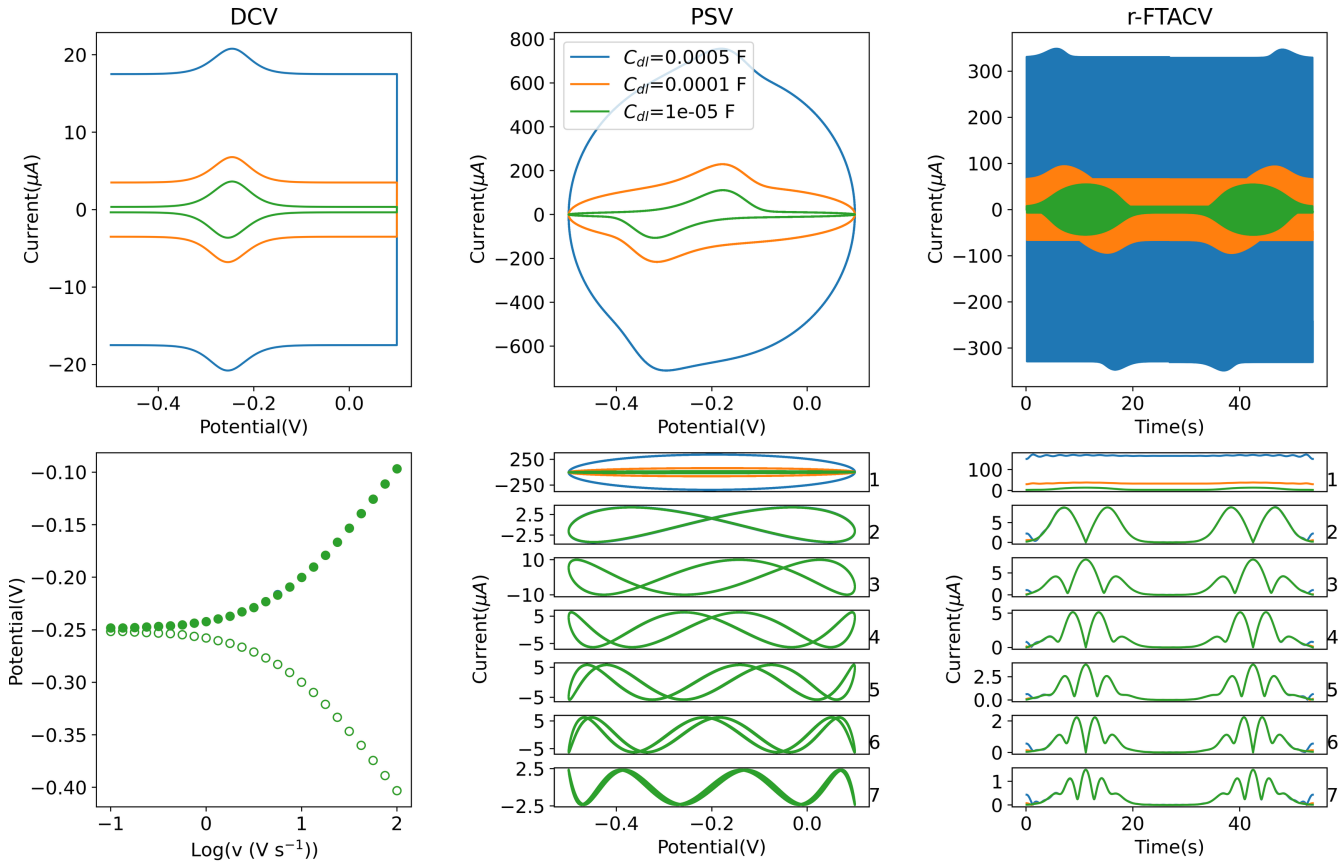


Figure 3.7: Parameter scan for C_{dl}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the linear double-layer capacitance C_{dl} which is as indicated in the legend.

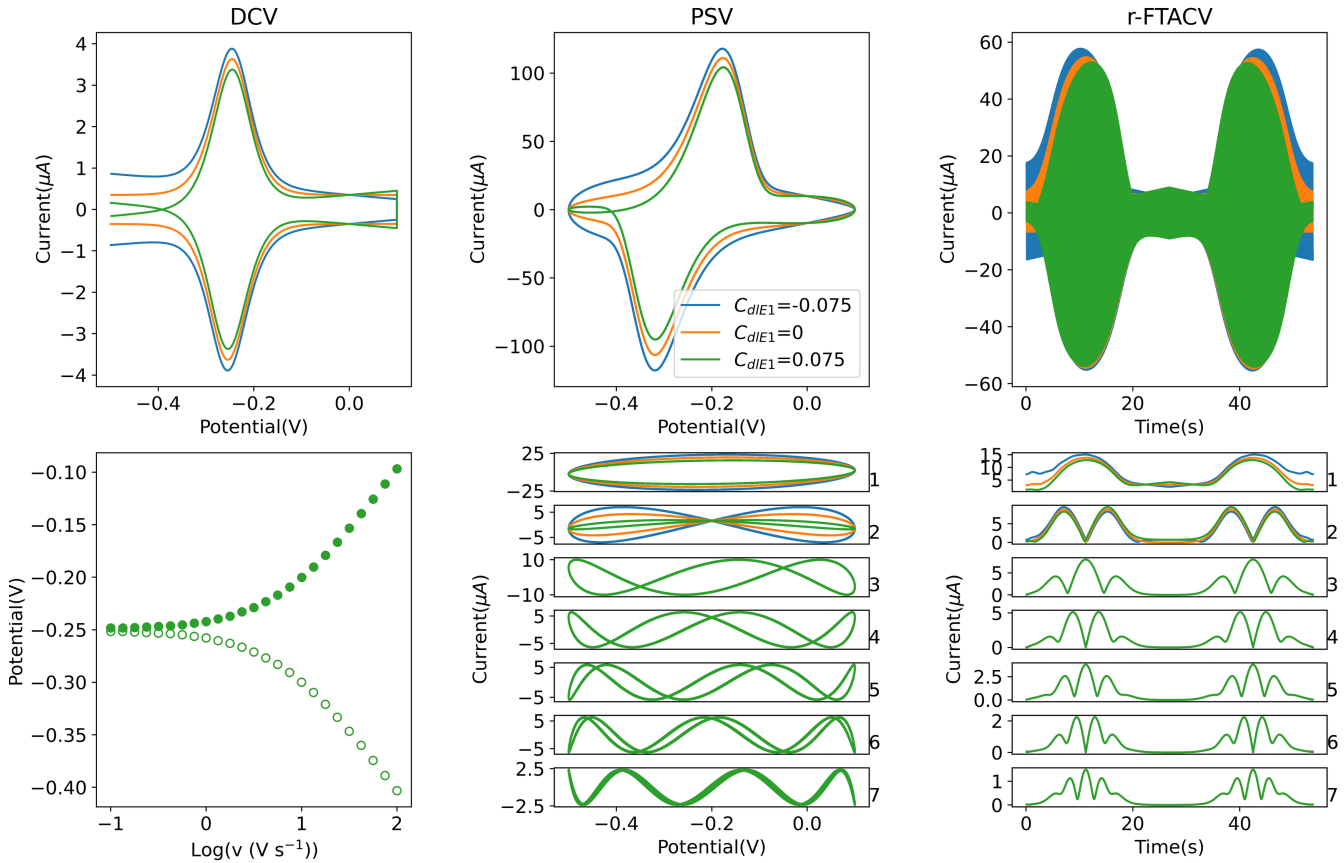


Figure 3.8: Parameter scan for C_{dlE1}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the higher-order capacitance term C_{dlE1} which is as indicated in the legend.

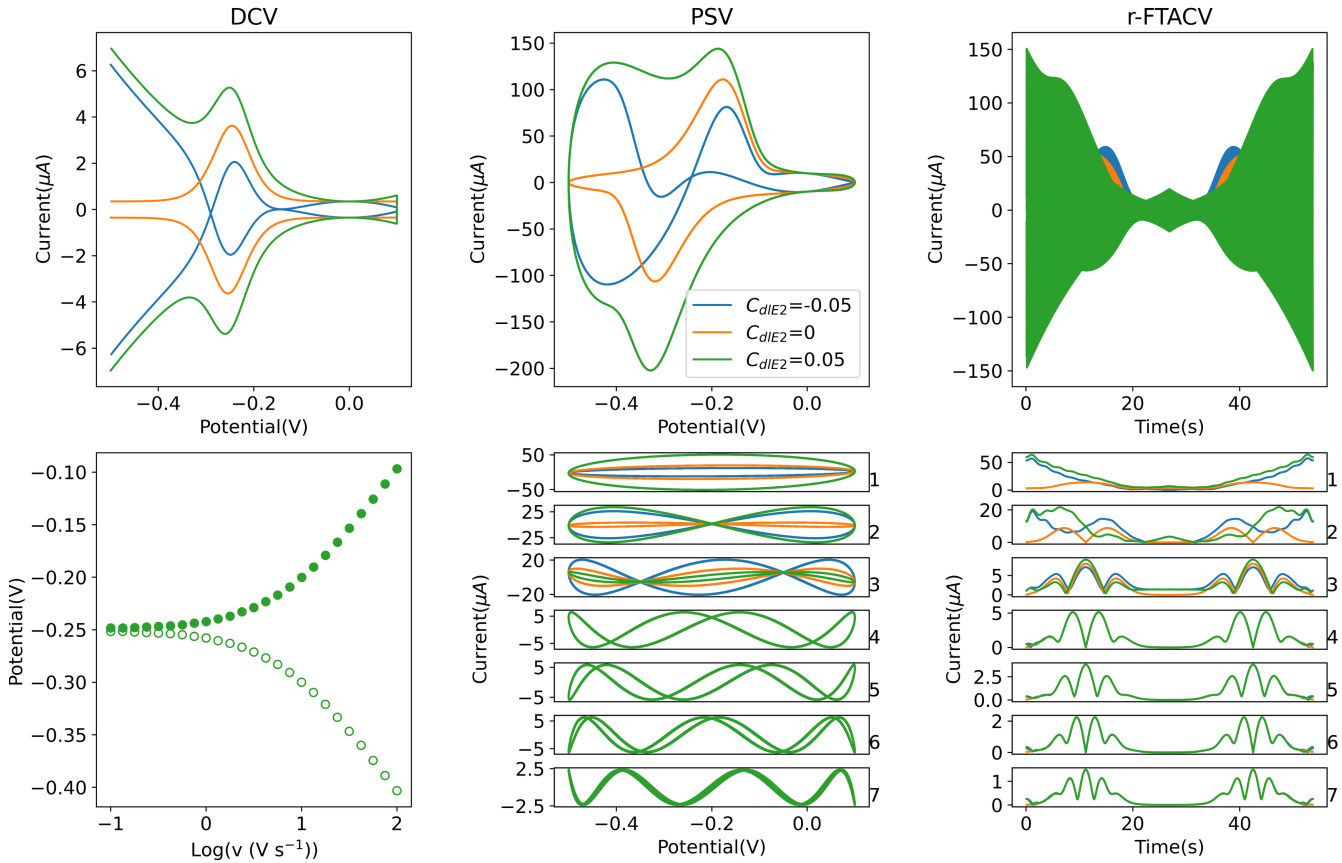


Figure 3.9: Parameter scan for C_{dlE2}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the higher-order capacitance term C_{dlE2} which is as indicated in the legend.

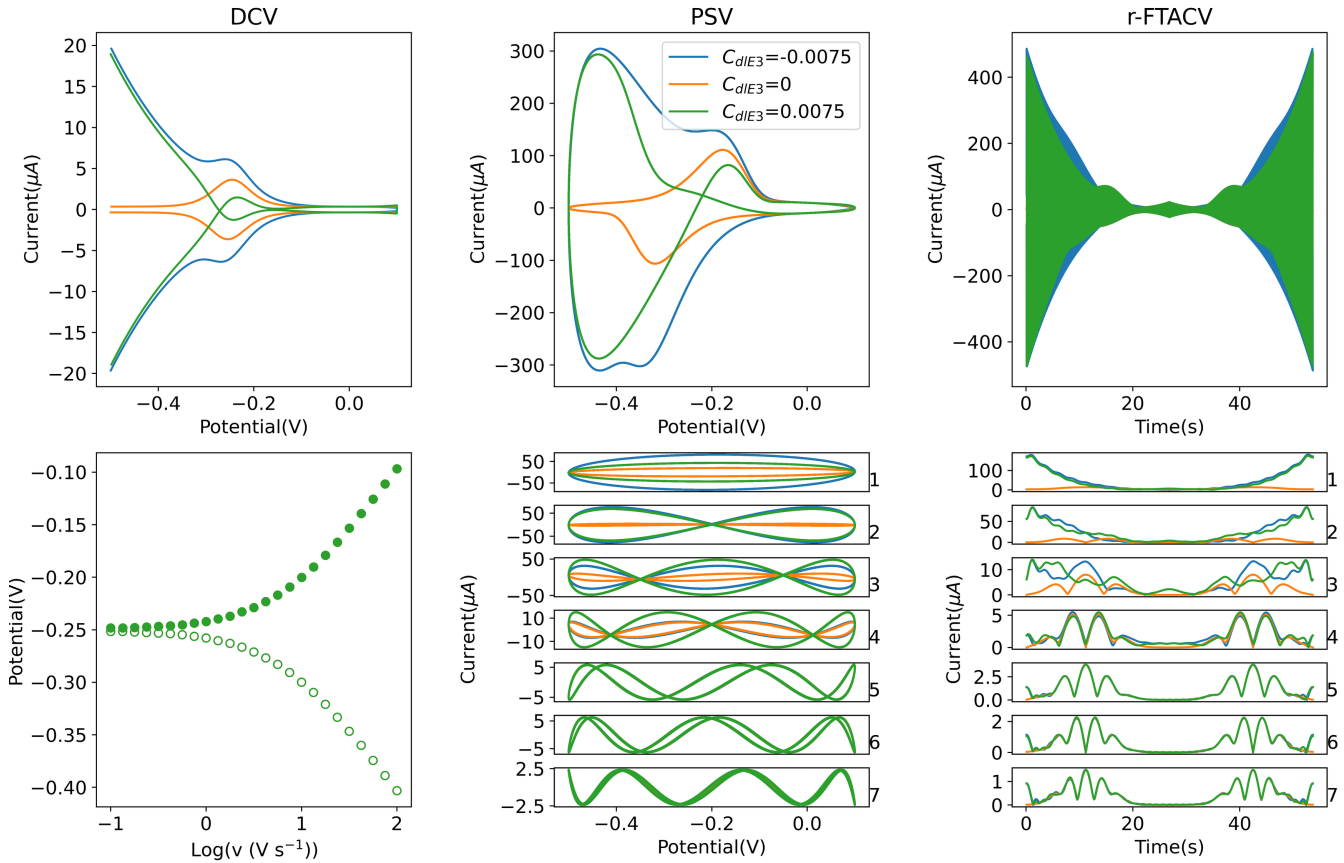


Figure 3.10: Parameter scan for C_{dlE3}, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the higher-order capacitance term C_{dlE3} which is as indicated in the legend.

3.2.6 Surface coverage — Γ

In figure 3.11, we see that the surface coverage Γ acts as a simple scaling factor for the magnitude of the Faradaic current, as a greater quantity of protein requires a greater quantity of electrons to be reduced, and vice versa.

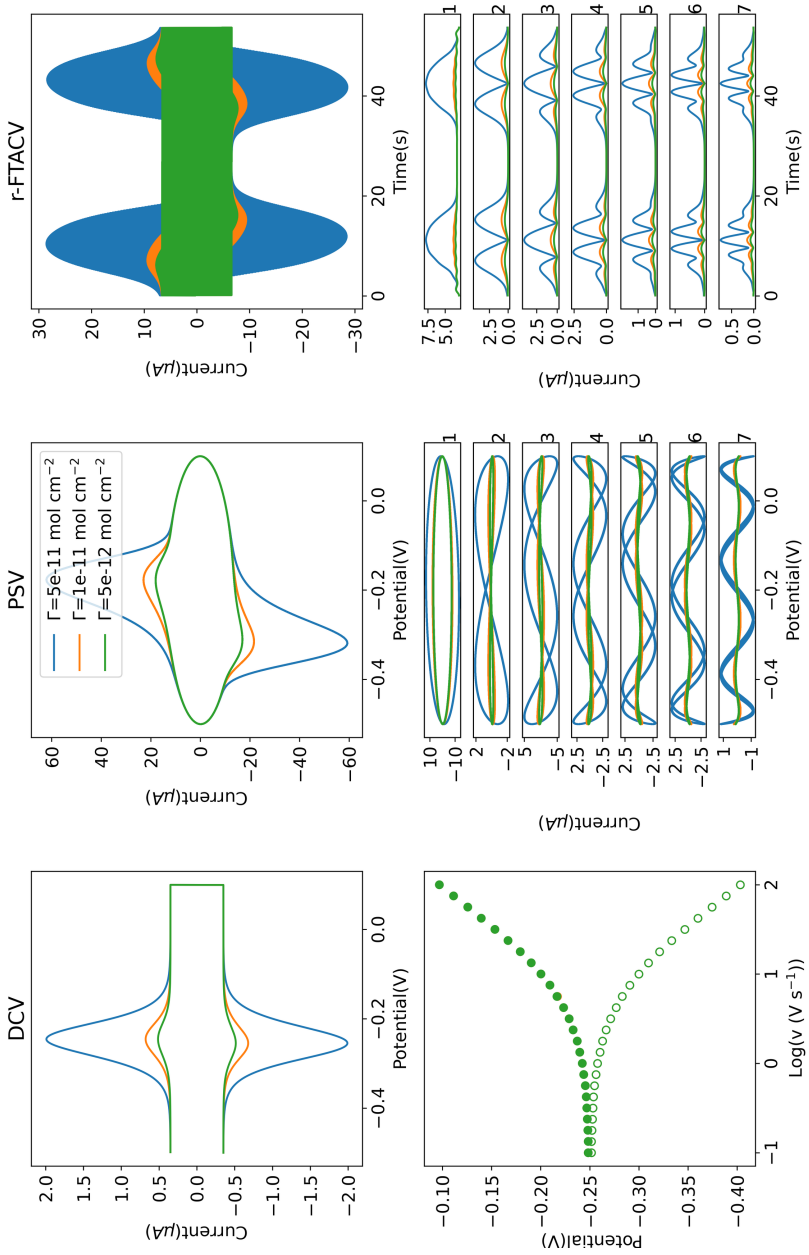


Figure 3.11: Parameter scan for Γ, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the surface coverage Γ which is as indicated in the legend.

3.2.7 Symmetry factor — α

As stated in section 2.2, α is the measure of the sensitivity of the transition state to the potential. As such, the ratio of the magnitude of the oxidative and reductive Faradaic peaks is controlled by α — when α is 0.5, the peaks are the same size. This ca be seen in the DCV and PSV total currents in figure 3.12. In addition, it has an effect on the symmetry of the r-FTACV harmonic lobes — a value of 0.5 means that the lobes are symmetrical, whereas other values increase the magnitude of one half relative to the other.

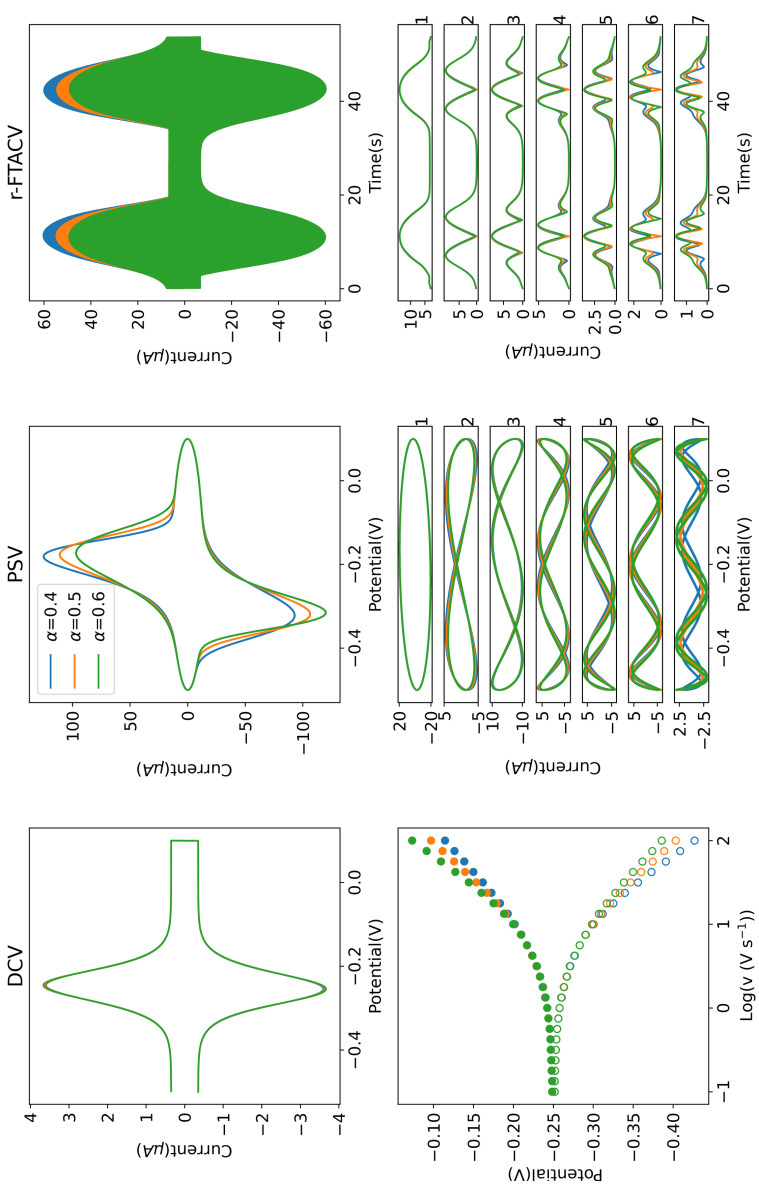


Figure 3.12: Parameter scan for α, for DCV, PSV and r-FTACV. Simulation parameters are as in table 3.2, except the symmetry factor α which is as indicated in the legend.

3.2.8 Phases — η and C_{dl} η

The phase of the applied potential appears to have a very strong effect on the PSV current, and no effect on the absolute r-FTACV harmonics — however, this is somewhat misleading because phase shifts will be more apparent when the current is plotted against potential as with PSV. In addition, taking the absolute value of the harmonics (as we do for r-FTACV) removes the phase information. Consequently, in figure 3.13 and 3.14, for r-FTACV we plot both the absolute harmonics, and the imaginary component of the harmonics. Capacitance phase, which is a parameter introduced in response to challenges with fitting experimental PSV data, as discussed in 4.3.2.2, alters the angle of the ellipse of the background current, when the current is plotted against potential. To model this, equation 2.22 is modified such that

$$\begin{aligned} I = C_{dl}\left(\frac{dE_{c_{app}}}{dt} - R_u\frac{dI}{dt}\right) \\ +\Gamma\Big((1-\theta)k_0\exp\big((1-\alpha)(E_{f_{app}} - IR_u - E^0)\big) \\ -(\theta)k_0\exp\big(-\alpha(E_{f_{app}} - IR_u - E^0)\big)\Big), \end{aligned} \tag{3.24}$$

such that

$$E_{f_{app}}(t) = E_{half} + \Delta E\sin(\omega t + \eta), \tag{3.25}$$

and

$$E_{c_{app}}(t) = E_{half} + \Delta E\sin(\omega t + C_{dl}\eta). \tag{3.26}$$

There is also an effect on the fundamental harmonic, but this is because we are only modelling linear capacitance in figure 3.14; if the values of C_{dlE1-3} were not zero, we would also observe changes in harmonics 2-4 with changing values of C_{dl} η. A further discussion of phase in the context of the theory of electrochemical impedance spectroscopy is provided in appendix D

3.3 The impact of dispersion — a spotter's guide

Content in this section has been re-purposed from the following reference: Henry Lloyd-Laney, Martin Robinson, Alan Bond, Alison Parkin, and David Gavaghan, 2021. A spotter's guide to dispersion in non-catalytic surface-confined voltammetry experiments. ***Journal of Electroanalytical Chemistry****, p.115204*

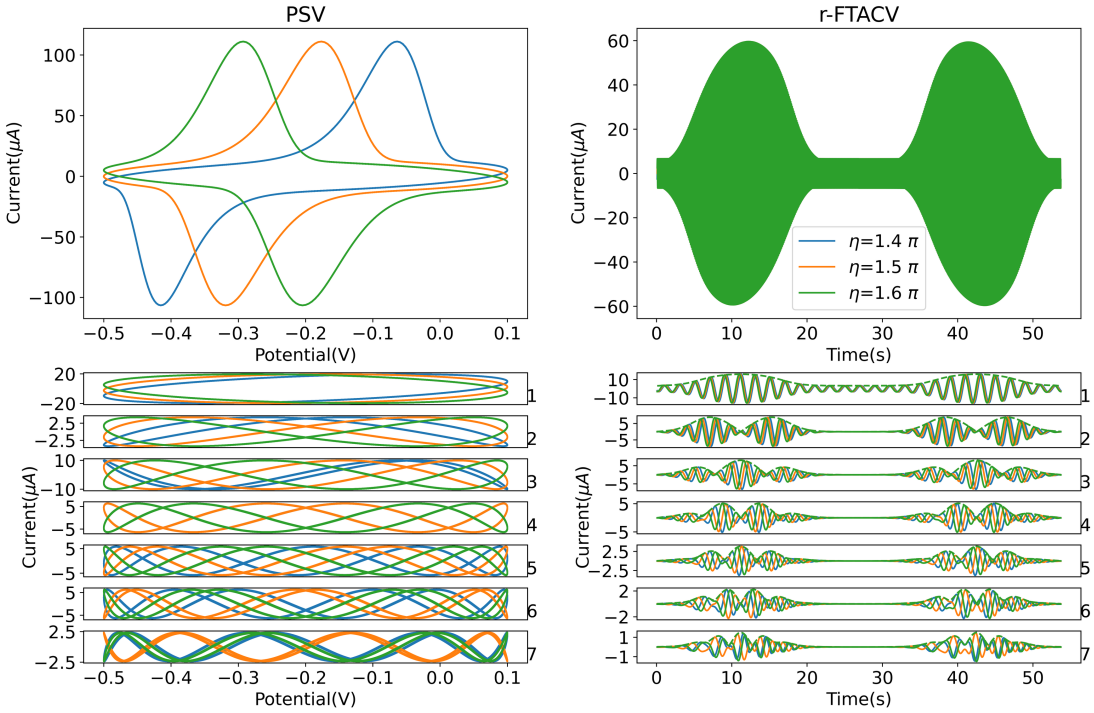


Figure 3.13: Parameter scan for η, for PSV and r-FTACV. Simulation parameters are as in table 3.2, except the Faradaic phase η which is as indicated in the legend. The r-FTACV plots incorporate both the absolute harmonics as a dotted line, and the imaginary harmonics, as the latter shows the impact of the phase parameter.

In the previous section, we have shown and described the effect of the various model parameters that we use to model the current response of a surface-confined voltammetry experiment. However, as alluded to in section 2.2.7, there are further experimental complications that necessitate modifications to this model. When fitting experimental data, as shall be presented in subsequent chapters, we have found that it is necessary to explicitly model dispersion.

Dispersion corresponds to an underlying distribution of certain parameter values; this is distinct from an inferred parameter distribution as recovered, for example, by the Bayesian methods introduced in section 2.4.2, which represents uncertainty about point values of parameters. Dispersion has been acknowledged as a complicating factor in the analysis of surface-confined voltammetry experiments since the 1990s, and has been demonstrated and quantified for a range of systems with multiple different methods [63, 64, 65]. These studies have focused on dispersion in two key parameters, the rate constant for electron transfer (k_0) and the reversible potential (E^0) — we refer to these two forms as kinetic and thermodynamic dispersion, respectively. The

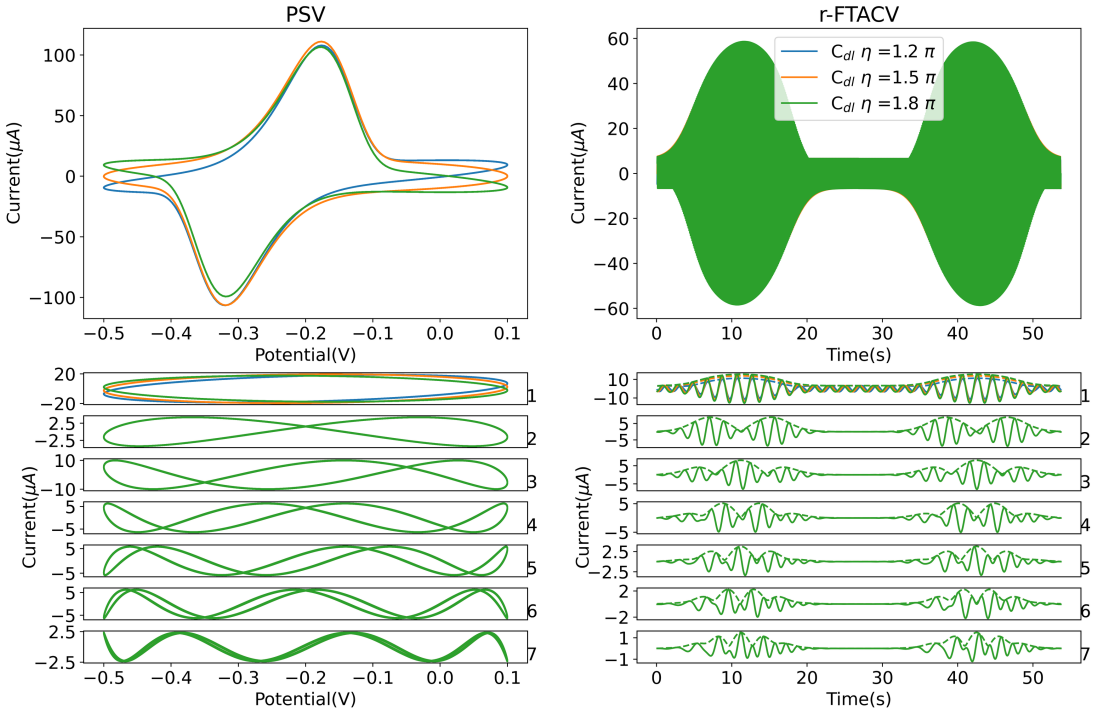


Figure 3.14: Parameter scan for $C_{dl}\eta$, for PSV and r-FTACV. Simulation parameters are as in table 3.2, except the capacitance phase $C_{dl}\eta$ which is as indicated in the legend.The r-FTACV plots incorporate both the absolute harmonics as a dotted line, and the imaginary harmonics.

aim of this section is to provide a guide on the impact that dispersion can have on the current output of a voltammetric experiment. It should be noted that, for solution phase voltammetry, which we do not analyse, it is not possible to have a distribution in the E^0 parameter [16].

For a two electron catalytic oxidation reaction, Léger et al. [52] showed that the presence of dispersion in k_0 could be detected by a change in the relationship between overpotential and observed current, in a study of the H_2-oxidising ability of a [NiFe]-hydrogenase. Since this initial study, Léger and co-workers, have shown that incorporating kinetic dispersion is an essential element of the modelling process for PFV experiments with catalytic processes [50, 66]. We have therefore included an example of a simple surface-confined catalytic mechanism in our considerations.

With regards to a surface-confined one-electron reaction, Patil et al. [67] detected dispersion in the reversible, non-catalytic single electron-transfer chemistry of the protein azurin through

the addition of a Forster resonance donor that fluoresces when reduced, allowing for quantification of the distribution of reversible potentials (as defined by the Nernst equation, as in equation 2.4) [67]. This technique was extended in work by Salvedera et al. by recording fluorescence movies, such that the reversible potentials and rate constants of distinct parts of the electrode could be determined [68]. The resulting appearance of the distribution was used by Morris [37] to suggest the true distribution of kinetic and thermodynamic parameters that generated the observations [68] were log-normal and normal respectively[1]. This analysis has informed our choice of parameter distribution, but it should be noted that other approaches to modelling kinetic dispersion choose a different distribution, such as the log-uniform distribution, as in the work by Léger et al [52], the modelling of which is explored in section 3.3.2.3.

The assumption that the redox-reactivity of all the subject molecules in the experiment can be described by the same reaction parameter values (i.e. no dispersion) is just one of a series of assumptions introduced to simplify the simulation of surface-confined voltammetric experiments. As laid out in section 2.2.7, additional assumptions include the choice of isotherm [69] and the theory used to describe electron transfer [70]. In terms of isotherm, the Langmuir isotherm is the simplest, assuming non-interaction between individual chemical moieties, while other isotherms, such as the Freundlich, Temkin and Frumkin isotherms include terms for the interaction of these species, along with different descriptions for their adsorption behaviours. The rate of electron transfer can be described by Butler-Volmer theory [19], Marcus theory [71] or other quantum approaches [72]. While we only address the presence of dispersion in this section, it should be noted that all three assumptions have areas of conceptual overlap; for example, the Temkin isotherm assumes a distribution of adsorption states, along with molecule self-interaction, which could lead to dispersion-like behaviours. Consequently, the experimental artefacts we ascribe to dispersion are likely the result of a highly complex surface chemistry. We have focused on dispersion in this case as it has received more extensive attention in the theoretical literature [37], with most models of voltammetry experiments assuming a Langmuir isotherm and Butler-Volmer kinetics, although there are exceptions [72].

When analysing surface-confined voltammetric data, obtaining the solution to the forward prob-

[1] A log-normal distribution of kinetic values can also be observed when attempting to solve the inverse problem and a logarithmic transform has been applied when non-dimensionalising the kinetic parameter. Normally distributed uncertainty about the dimensionless kinetic value would result in log-normally distributed uncertainty about the dimensional kinetic value. It should be noted that this effect is distinct from an underlying spread of kinetic values, and could be observed even if the true distribution was a delta function.

lem when dispersion is incorporated is significantly more computationally taxing. In addition, modelling dispersion adds more degrees of freedom to the inverse problem due to the need to parameterise a probability distribution for each dispersed parameter, resulting in a greater frequency of spurious results, due to the optimisation algorithm becoming trapped in local minima. Consequently, determining which form of dispersion is present (if at all), and the likely order of magnitude for the required parameters, is an essential step in the process of solving the inverse problem in reasonable time. In this section we demonstrate the effect that dispersion of key parameters has on the current response in a range of surface-confined voltammetry experiments, and to act as a reference for when the nature of the dispersion is ambiguous, which will be pertinent to our discussion of parameter inference for experimental data in subsequent chapters. The values of the dispersion parameters that we describe here are informed by both observations of experimental data, and a desire for clarity of demonstration. The interested reader can explore the phenomenon further using our online simulation package. This section builds on the basic theoretical work described previously [37], extending the analysis to PSV and DCV, and including sensitivity information about the various dispersion parameters.

3.3.1 Modelling approach

We assume that any parameter which is affected by dispersion no longer has a single fixed value but instead is drawn from a probability distribution which captures the degree of the dispersion. If we have n_p dispersed parameters in the model, then we assume that in any particular experiment, those n_p parameters are drawn from a joint probability $\phi(p_1, p_2, ..., p_n)$. The total current recorded given the distribution is the expectation $E[I(t, (p_1, ..., p_n))]$, which we refer to as I_{disp} [37]

$$I_{disp} = \int_{A_n} ... \int_{A_1} I(t, (p_1, ..., p_n))\phi(p_1, p_2, ..., p_n)dp_1...dp_n, \tag{3.27}$$

where $I(t, (p_1, ..., p_n))$ is the simulated current as a function of time and the n_p dispersed parameters $p_1, p_2, ..., p_n$. The range of the i^{th} integral is represented as A_i. This integral cannot be calculated analytically, and so we approximate it using numerical quadrature. In numerical quadrature the function to be integrated is evaluated at points over the range of the integral, and these evaluations are weighted appropriately and summed. This is represented in figure 3.15,

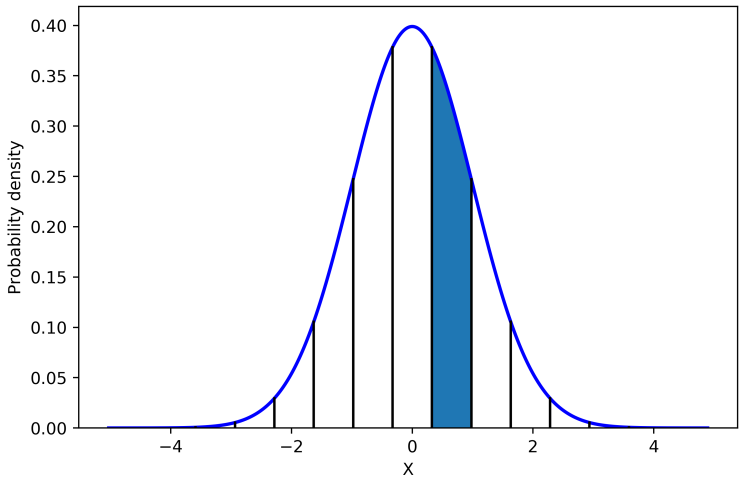


Figure 3.15: A representation of numerical integration of a Gaussian distribution

where the distribution has been partitioned into b equally-spaced "bins". The sum of the area of all bins (such as the one filled in blue) is an approximation to the value of the integral. In the case of normal distribution we use Gauss-Hermite quadrature, and for the log-normal and log-uniform distributions we use the midpoint rule. For both cases, a dispersed simulation takes b^{n_p} times longer than a non-dispersed simulation.

3.3.2 Results and Discussion

To show the effect of each dispersion parameter, we perform parameter scans where all parameters are held constant except the one of interest. Although the effect observed from this process may not be exclusive to the parameter in question (as other parameters have the same effect), and there may be parameter regimes where the observed effect is less apparent, this process is the easiest way to gain intuition about the effects of the various parameters. Intuition about the approximate order of magnitude for the dispersion parameters allows the experimentalist to constrain parameter space (for either heuristic or automatic inference approaches). The more parameter space is constrained, the easier it is to find the global optimum, as long as the constraint process has not, in fact, excluded this optimum. In figures 3.16-3.21, we show the results of holding all parameters constant (as reported in table 3.3) except the parameter named in the upper left hand corner of each plot, the value of which is indicated. We have chosen this range of parameters as a compromise between clear visual separation between different parameter val-

ues, and chemical realism. For this reason, the parameters that contribute to the background current (the capacitance and uncompensated resistance) are in the lower range of what would be observed in a typical voltammetry experiment, and the surface coverage is in the higher range, as in section 3.2. We have chosen to show two kinetic regimes; an irreversible, low k_0 regime, and a quasi-reversible regime where the k_0 values were chosen as a compromise between providing the "ideal" r-FTACV harmonic appearance, and sensitivity to these parameters (all kinetic distributions where the majority of the range is approaching or exceeding the reversible limit will look identical). With regards to the choice of distribution, for the reversible potential E^0, we use a skewed Gaussian distribution (where a skew of zero is a regular Gaussian), and for the rate constant k_0 we use a log-normal distribution. The choice of a Gaussian and log-normal distributions are informed by previous work [37, 68]. The skew parameter was included as observations of r-FTACV data in subsequent chapters implied the presence of an asymmetric distribution of E^0 values, which for normally distributed values indicates the presence of skew. Although other parameters are suspected of being dispersed, such as the symmetry factor α and the uncompensated resistance, we neglect them here as they are secondary to the main parameters of interest. For each subgroup of plots, the input parameter distribution is shown on the top, and the resultant current-potential output is shown below. In each sensitivity figure, each current plot shares a common x-axis, with the values and units shown in the bottom figure of each column. The x-axis for each distribution is shown on the top of each distribution plot.

For the r-FTACV experiment, in figure 3.16 we show harmonics 3, 5 and 7 plotted against time instead of a current-potential plot, as the harmonics are more sensitive to dispersive effects, and are easily interpretable. In terms of the thermodynamic distribution parameters, the value of the mean of the distribution E^0_μ affects the position of the harmonic relative to the potential, as for E^0 in section 3.2. The standard deviation of the distribution E^0_σ affects both the width and magnitude of the harmonic, with higher harmonics being affected more strongly. The skewness of the distribution E^0_κ has a slightly more complex effect — the skewed distributions are less broad than the unskewed distribution (i.e. $E^0_\kappa = 0$), which accounts for why the magnitude of each skewed harmonic is larger than for the unskewed harmonic. In addition, the presence of skew means that the position of the harmonic within the timeseries is a function of the order of the harmonic, and that higher-order harmonics have an asymmetric shape relative to an unskewed distribution.

The effect of the parameters for the log-normal k_0 distribution on the r-FTACV simulation output in figure 3.16 are less immediately obvious. Increasing the scale parameter (which is the mean of the natural logarithm of k_0, i.e. $\log(k_0)\mu$ largely affects the resolution of the individual harmonics, with the higher harmonics being more sensitive to this effect; this is especially apparent in the 7^{th} harmonic, where the peaks of the $\log(k_0)\mu = 50$ plot are less well-resolved than the $\log(k_0)\mu = 100$ and $\log(k_0)\mu = 500$ plots. As expected, this is similar to the effect of altering the value of k_0 as in figure 3.4. The $\log(k_0)\sigma$ parameter, which also affects the width of the distribution has the effect of reducing the magnitude of the harmonics, where the higher harmonics are also more strongly affected. This is distinct from the E^0_σ parameter, which reduces *and* broadens the harmonics.

Parameter	Symbol	Value
Start potential	E_{start} (V)	-0.5
Reverse potential	$E_{reverse}$ (V)	0.1
Reversible potential	E^0 (V)	-0.25
Rate constant	k_0 (s^{-1})	100/1
Uncompensated resistance	R_u (Ω)	0.0
Double-layer capacitance	C_{dl} (F)	1.0E-5
Surface coverage	Γ (mol cm^{-2})	1.0E-10
Symmetry factor	α	0.5
Scan rate	v (V s^{-1})	0.022/*/1
Potential frequency	ω (Hz)	8.881/8.94/*
Phase	η (rads)	0.0/$\frac{3\pi}{2}$/*
Potential amplitude	ΔE (V)	0.15/0.3/*
Sampling rate	(s^{-1})	400.0
Electrode Area	Area (cm^2)	0.07
Reversible potential mean	E^0_μ (V)	-0.25
Reversible potential standard deviation	E^0_σ (V)	0.05
Reversible potential skew	E^0_κ	0.0
Rate constant shape	$\log(k_0(s^{-1}))\sigma$	0.5
Rate constant scale	$\log(k_0(s^{-1}))\mu$	100

Table 3.3: Simulation parameters used to generate plots in figures 3.16-3.21. For some experimental parameters, the simulation value is different for each technique, and is written in the following order: r-FTACV/PSV/DCV, with a * indicating that the parameter is not used in the simulation

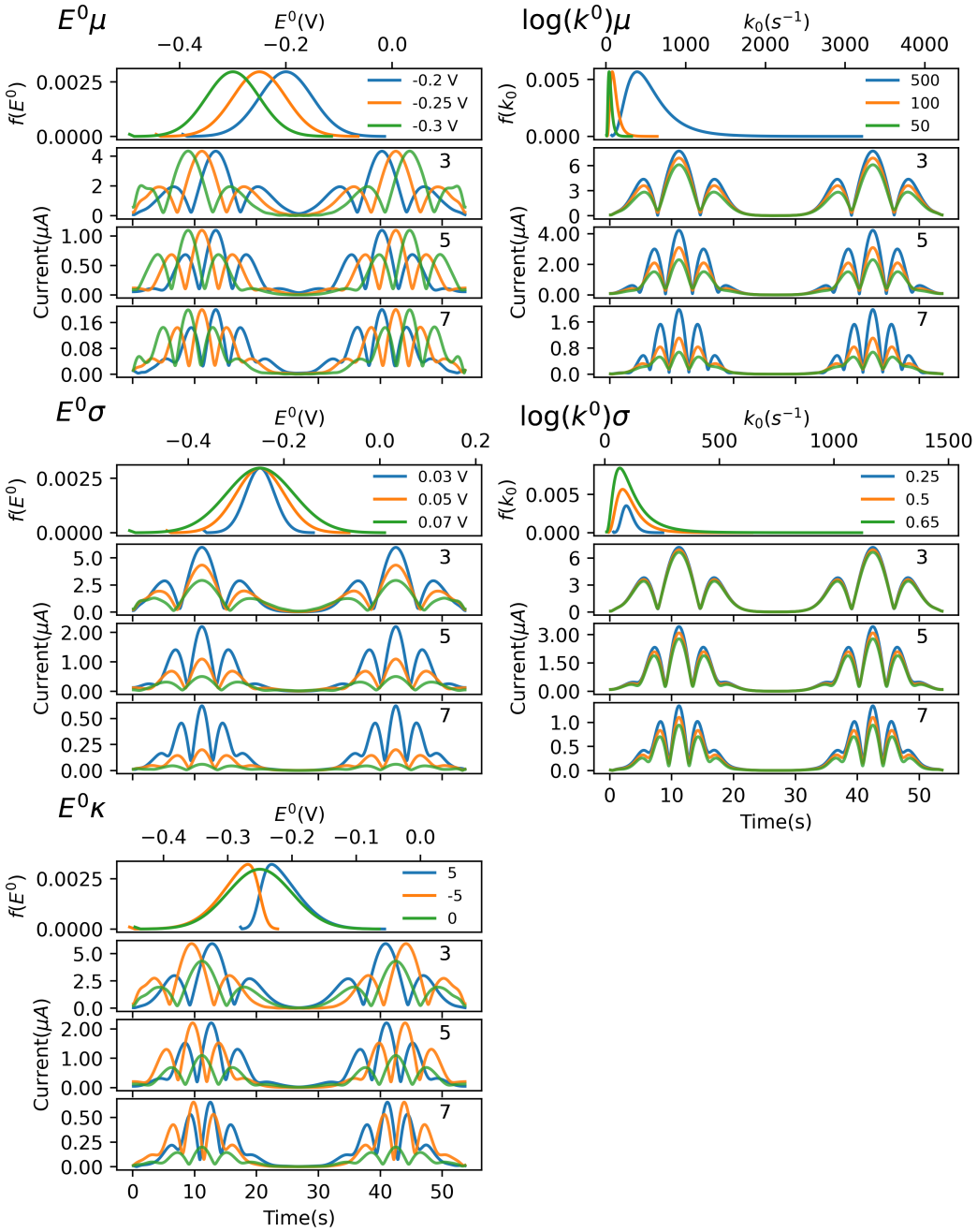


Figure 3.16: Plots of simulated r-FTACV harmonics vs time , with dispersion parameters for E^0 — mean, standard deviation and skew $(E^0_\mu, E^0_\sigma, E^0_\kappa)$ and k_0 — shape and scale $(\log(k_0)\mu, \log(k_0)\sigma)$ where the parameter being varied is in the top left hand corner. The resulting distribution, where $f(\text{parameter})$ is the assumed distribution of the parameter, is shown at the top of the plot, and the 3^{rd}, 5^{th} and 7^{th} harmonics shown. The colour of the harmonic corresponds to the colour of the distribution. Simulation parameters are found in table 3.3. For all cases, the kinetics of the reaction are in the quasi-reversible to reversible regimes.

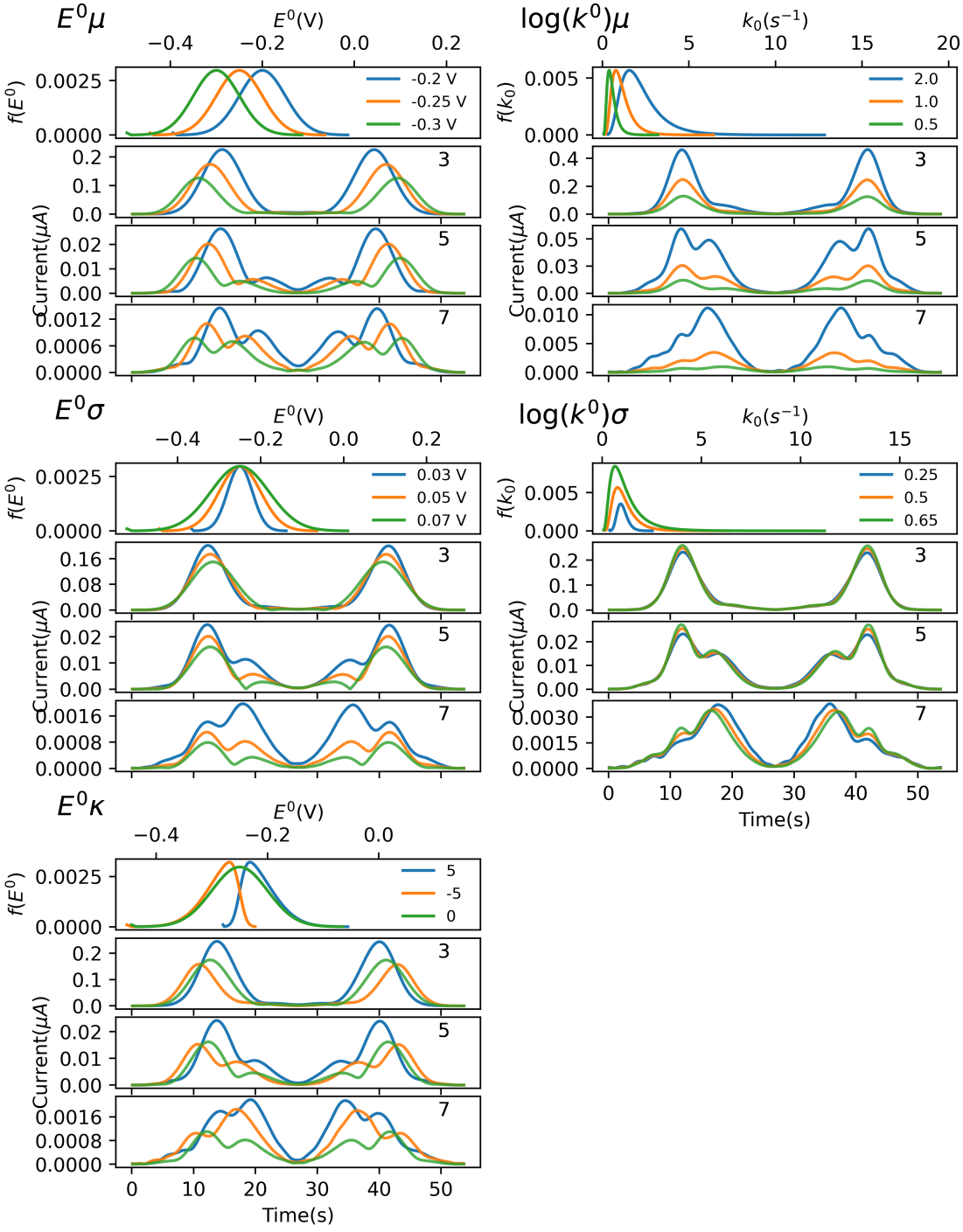


Figure 3.17: Plots of simulated r-FTACV harmonics vs time, where the parameter distributions under investigation are plotted in the top panel of each plot subgroup. The remaining simulation parameters can be found in table 3.3. These simulations were generated using irreversible kinetic values.

For the irreversible r-FTACV case in figure 3.17, we lose definition so that the individual harmonic peaks (as observed in the harmonics in, for example, figure 3.23) blend into one another, as noted in section 3.2 in figure 3.4 for non-dispersed kinetics. It should be noted that for the r-FTACV harmonics, the magnitude of the signal harmonic appears to be dependent on the proximity to the middle of the time-series. This is because we have used the Hann window, which suppresses the edges of the time-series, to improve signal-to-noise — this is weaker in the higher harmonics in an irreversible dispersed system. This is an artefact of the filtering process, and does not represent an underlying chemical property. For figures 3.17, 3.19 and 3.21, all parameters are as in table 3.3, except for k_0 and $\log(k_0)\mu$ which were set to 1s^{-1} as appropriate.

For the PSV case (figure 3.18 and 3.19), we assess the effect of each parameter on the current-potential plots, instead of the harmonics. This is because the harmonics of the PSV timeseries are not as amenable to immediate visual interpretation as the harmonics of the r-FTACV experiment, as noted in section 3.2. The effects of the various thermodynamic and kinetic dispersion parameters are similar for PSV and DCV (where the latter case is shown in figure 3.20 and 3.21), which is not unexpected. The reason for this similarity is that the PSV potential input is highly similar to multiple cycles of the DCV potential input, except without a discontinuity at the switching potential, and the fact that the PSV input elicits a non-linear current response. With this similarity in mind, we note that the effect of each parameter controlling the shape of the E^0 distribution on both the PSV and DCV current-potential outputs is as described for the r-FTACV harmonics, namely E^0_μ alters the position of the reductive and oxidative peaks, E^0_σ broadens and reduces the magnitude of these peaks, and E^0_κ affecting the position and amplitude of the peak. For the PSV plots, the $\log(k_0)\mu$ parameter also exerts a similar effect to the one described for the r-FTACV harmonics, i.e. similar to the effect of k_0 as a point value. The exact nature of this effect depends upon which regime (irreversible, quasi-reversible or reversible) the dimensionless k_0 value is in. In more detail, for a current-potential PSV plot of a completely reversible system, the oxidative and reductive peaks will be stacked on top of one another. As the kinetics becomes progressively slower, the relative peak position will start to diverge, along with peak broadening and a reduction in maximum amplitude. The parameter $\log(k_0)\sigma$ affects the gradient of the current returning from a Faradaic peak to background current. As the effects of the kinetic parameters are similar for both the DCV and PSV timeseries, for the DCV case we also show trumpet plots, as previously. The $\log(k_0)\sigma$ parameter slightly reduces the gradient of the oxidative and reductive curves. The $\log(k_0)\mu$ parameter alters the scan rate at which the

locations for the oxidative and reductive peaks start to diverge.

The descriptions of all of the effects of these parameters are caveated with the fact that varying other parameters used in the simulation of the current response can have similar effects. Obtaining an understanding of the effects and interactions of the various parameters requires iteratively comparing multiple simulations. As this is not feasible in the static format of a thesis chapter, we have developed an online interactive application for the 1-electron redox case. This can be found here, and allows the user to simulate the three experiments described in this section, with or without dispersion, and compare multiple different simulations.

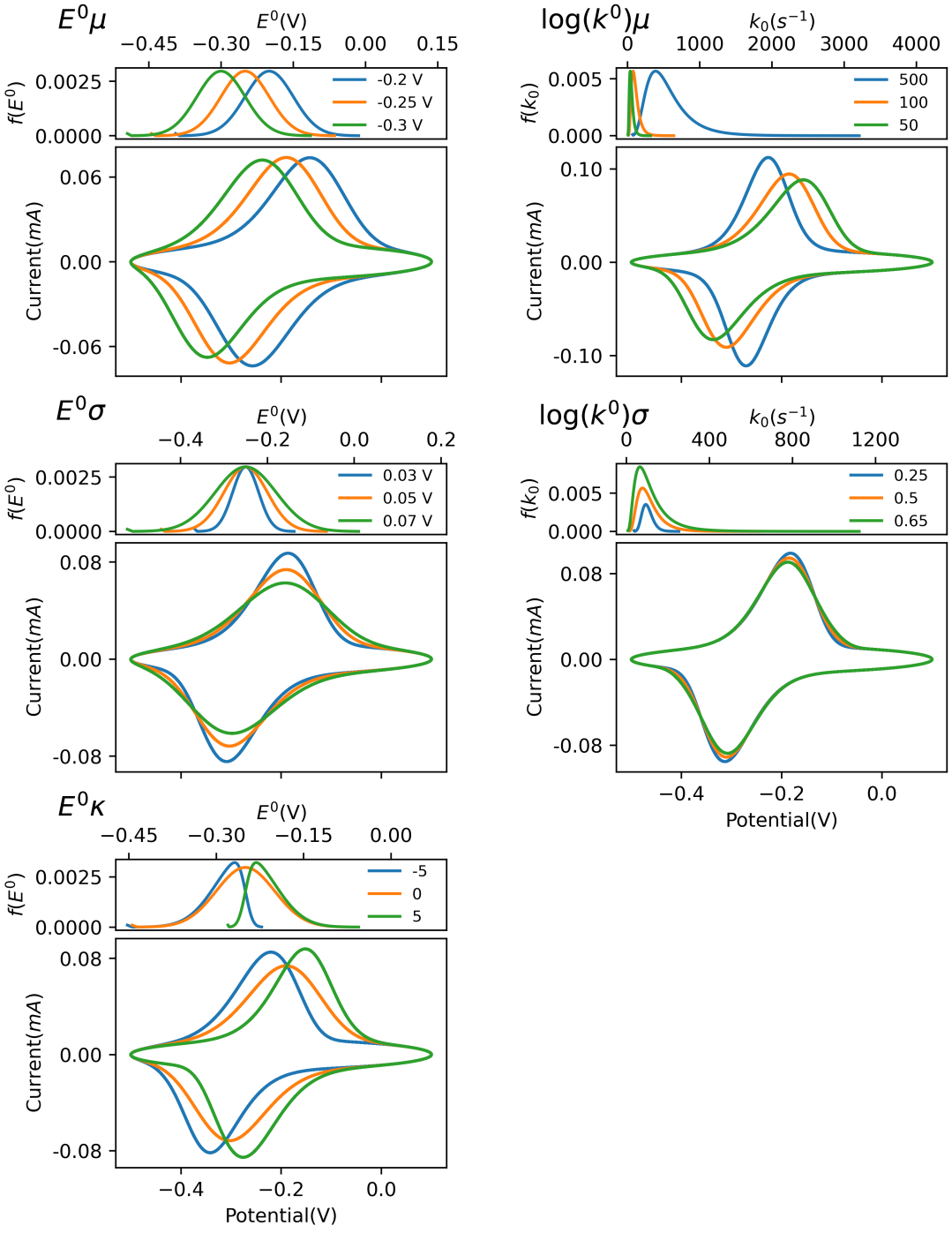


Figure 3.18: Parameter scans for a PSV current-potential plot, where the parameter distributions under investigation are plotted in the top panel of each plot subgroup. The remaining simulation parameters can be found in table 3.3.

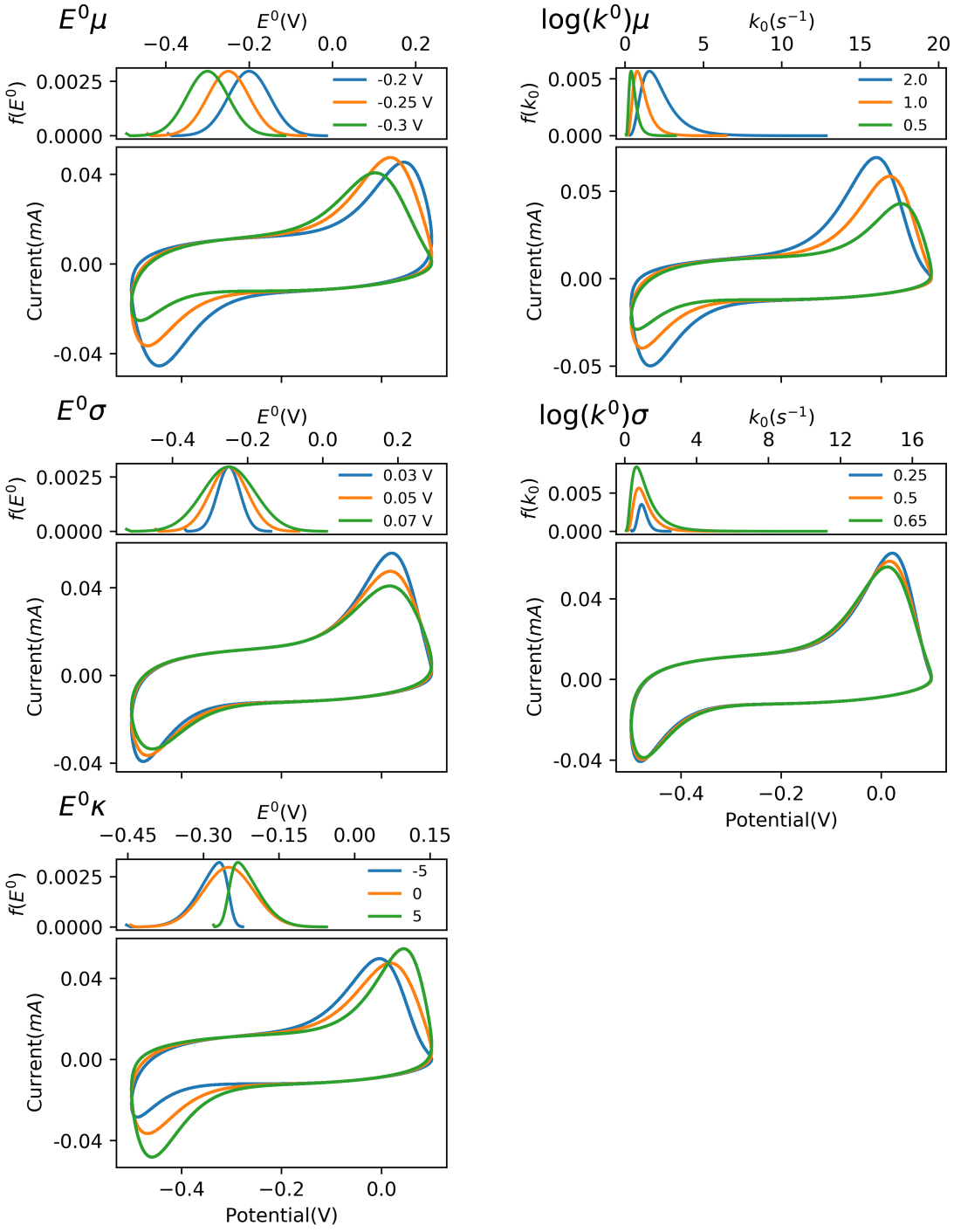


Figure 3.19: Parameter scans for a PSV current-potential plot, where the parameter distributions under investigation are plotted in the top panel of each plot subgroup. These simulations were generated using irreversible kinetic values.

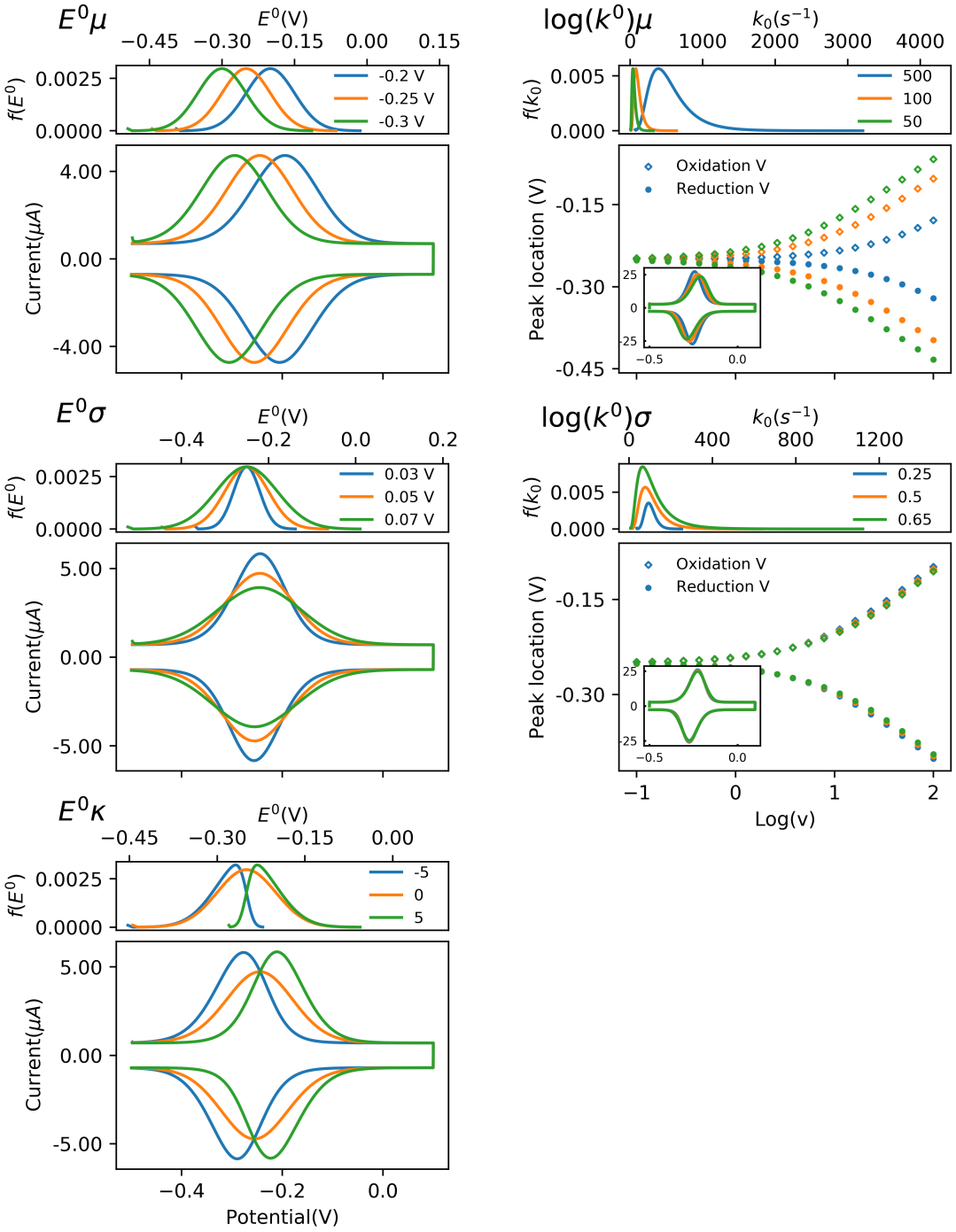


Figure 3.20: Parameter scans for a DCV current-potential plot (left), and trumpet plots (right) which plot the potential-values of the maximum and minimum current peaks against the base 10 logarithm of the scan rate. The inset plot in the right hand column is a current(μA)-potential(V) plot of a DCV simulation using a scan rate from the middle of the range defined in the x-axis of the containing plot. The remaining simulation parameters can be found in table 3.3.

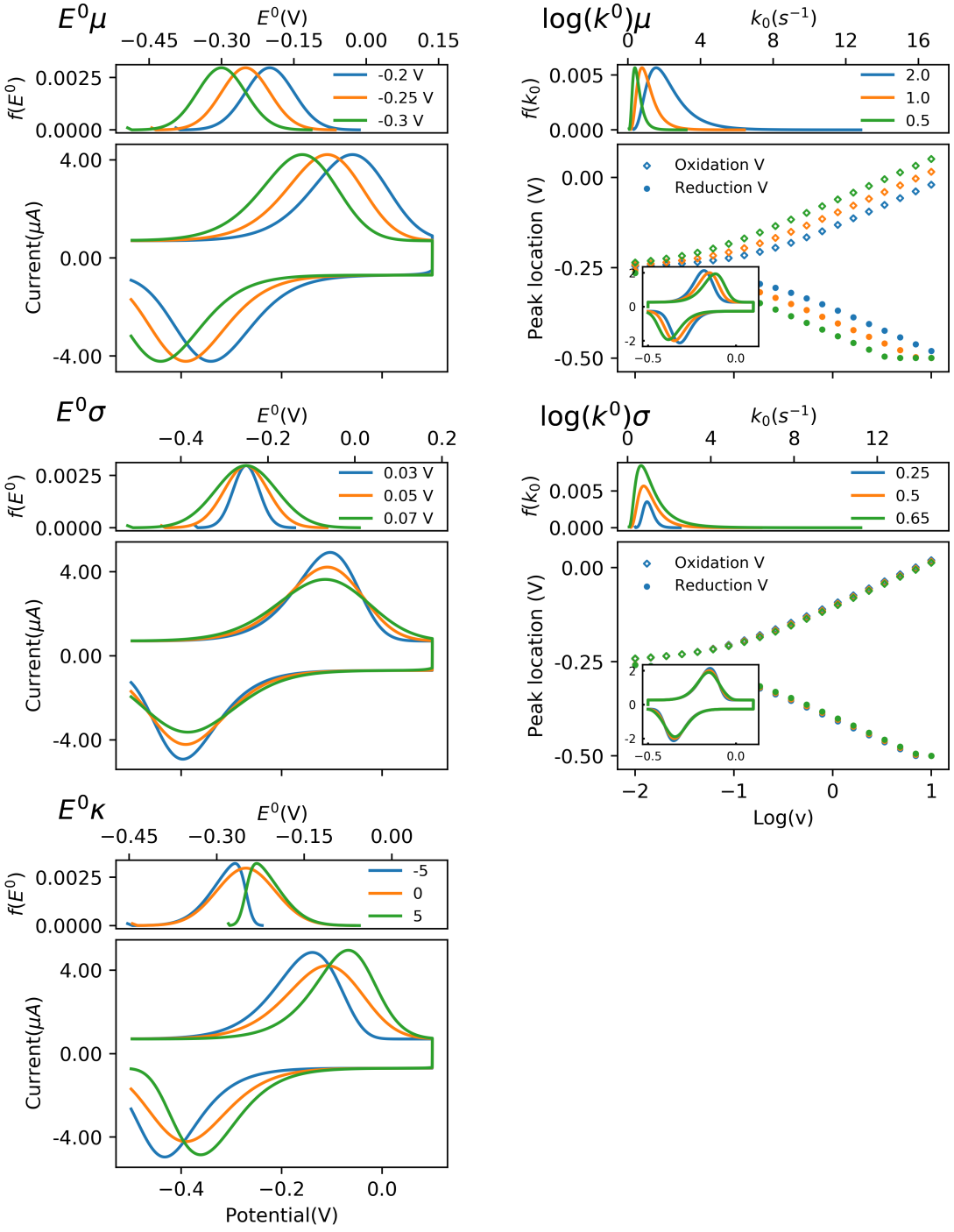


Figure 3.21: Parameter scans for a DCV current-potential plot (left), and trumpet plots (right) which plot the potential-values of the maximum and minimum current peaks against the base 10 logarithm of the scan rate. The inset plot in the right hand column is a current(μA)-potential(V) plot of a DCV simulation using a scan rate from the middle of the range defined in the x-axis of the containing plot. These simulations were generated using irreversible kinetic values.

3.3.2.1 Distinguishing different forms of dispersion

Dispersion is predominantly relevant when solving the inverse problem, that is when fitting experimental data. Naturally, ascertaining whether or not dispersion is actually present is an essential first step in this process, as the spurious inclusion of dispersive effects increases the chances of error in fitting. The risk is that a "good fit" could be achieved with parameters that do not reflect the underlying chemical reality, and that a proposed solution only represents a local, rather than the global, minimum in parameter space. With this in mind, in figure 3.22, we show the effects of dispersion for three cases, E^0, k_0 and a combination of the two, as shown in the legend, relative to the case with no dispersion in either parameter, which is marked as "None".

Distinguishing between purely thermodynamic and purely kinetic dispersion is relatively simple, at least in the r-FTACV case, as, while both result in a reduction in the expected magnitude of the harmonic, only E^0 broadens it. The difficulty lies in resolving a combination of E^0 and k_0 dispersion from pure E^0 dispersion, as the latter dominates. This is also complicated by the fact that the reduction in harmonic magnitude is also an effect of other parameters, such as the uncompensated resistance. As before, for the simple reaction model under investigation, both the PSV and DCV cases are less sensitive to the presence of dispersion and more challenging to interpret. For the trumpet plot case, there is a slight difference in gradient between the kinetically-dispersed, thermodynamically dispersed and non-dispersed cases. In figure 3.23, we show that when k_0 is highly reversible, the presence of k_0 dispersion is identical to the non-dispersed case, and in figure 3.24 we show that the effect of k_0 is greatest relative to the non-dispersed case in the irreversible and quasi-reversible regimes. This kinetic dispersion effect is (relatively) more pronounced in figure 3.24, for the r-FTACV and PSV cases, which is to be expected as the current is more sensitive to changes in the kinetic parameter in these regimes. Interestingly, in this parameter regime, the trumpet plot shapes are not particularly sensitive to the change in the kinetic values (figure 3.24, bottom row). This is because the timeseries is more sensitive to changes in k_0 in the irreversible regime, and consequently dispersion has a greater effect. Consequently increasing either the scan rate (for DCV) or the input frequency (for r-FTACV and PSV) will make the results of the experiment more sensitive to kinetic dispersion.

As we show in figure 3.25, it is possible to tune the potential frequency to maximise the difference between dispersed and non-dispersed cases. This allows us to maximise the chances

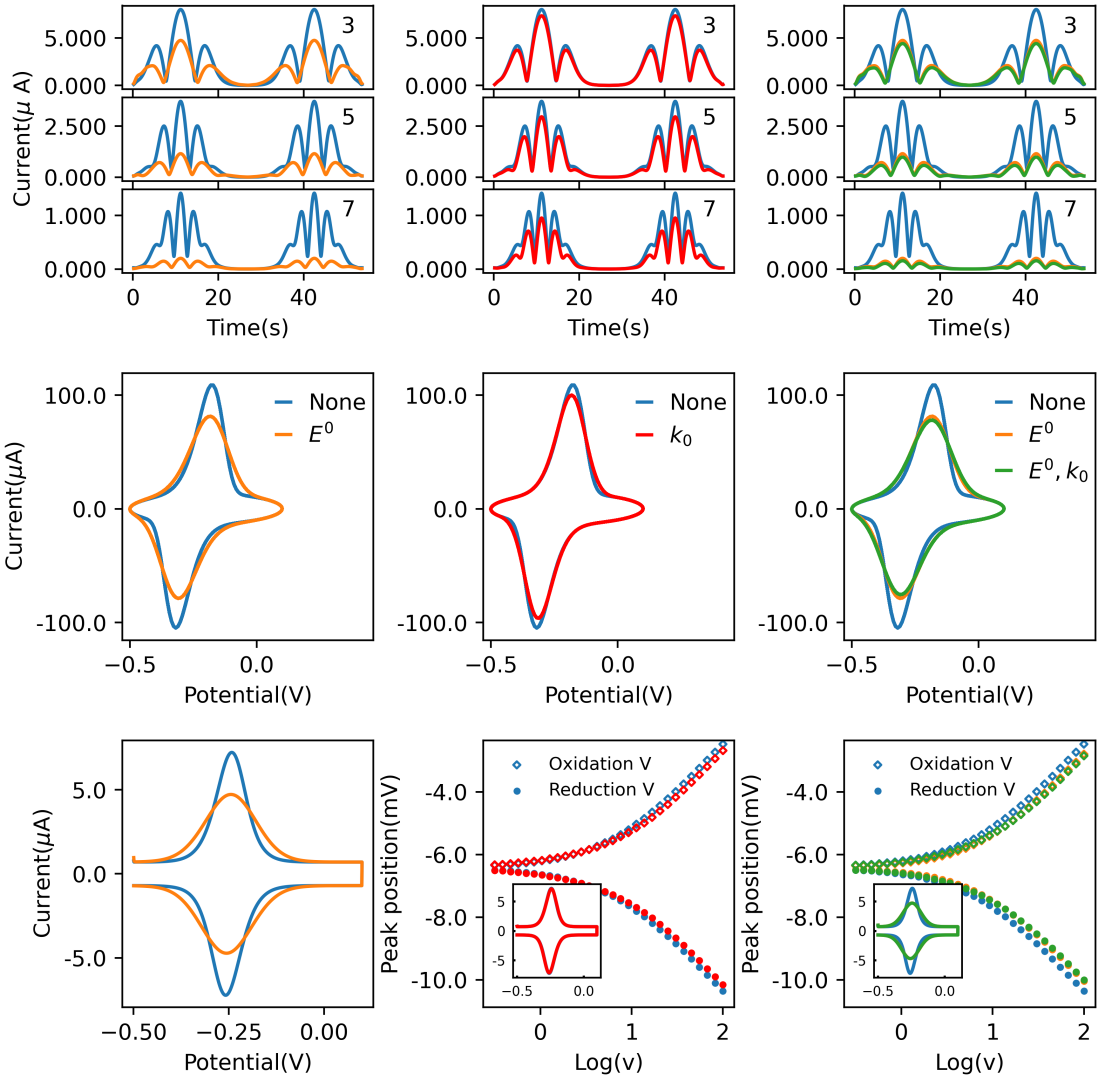


Figure 3.22: Comparison of different forms of dispersion for r-FTACV, PSV and DCV. Each column represents a different form of dispersion, compared against the non-dispersed case (labelled as "None"). From left to right, the dispersed parameters are E^0, k_0 and a combination of the two, as listed in the legend of the middle row. All parameters are as in table 3.3. Harmonics were generated without the usage of the Hann window.

that the two cases can be distinguished. This makes most sense for PSV, as the input frequency does not make a difference to simulation times. The plots in figure 3.25 show the root mean square difference between PSV current simulated with and without kinetic dispersion against the logarithm of the input frequency, for various values of $\log(k_0)\mu$ and $\log(k_0)\sigma$. The largest (and therefore most apparent) deviations are a function of frequency for varying values of $\log(k_0)\mu$.

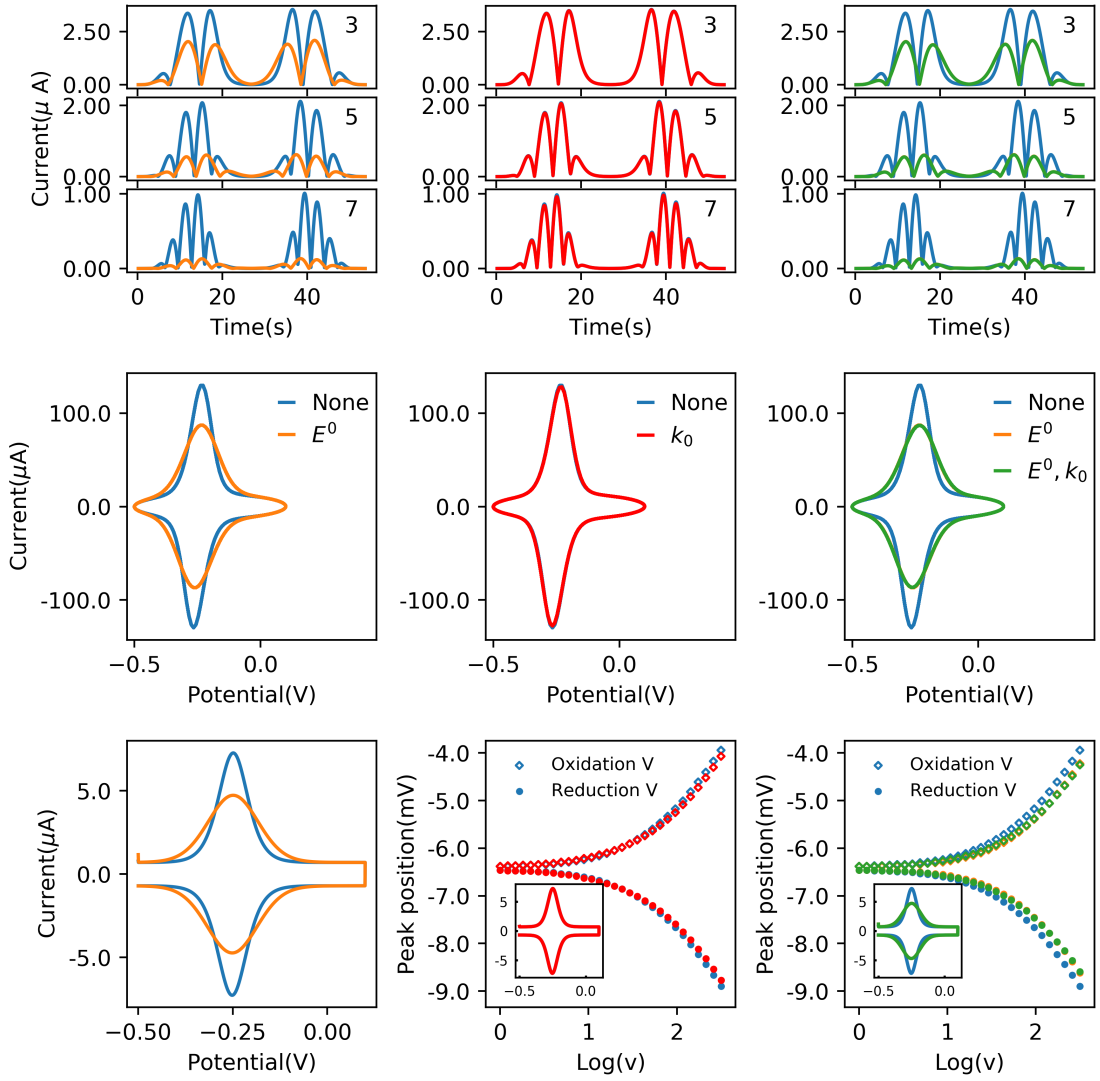


Figure 3.23: Comparison of different forms of dispersion for r-FTACV, PSV and DCV. Each column represents a different form of dispersion, compared against the non-dispersed case (labelled as "None"). From left to right, the dispersed parameters are E^0, k_0 and a combination of the two, as listed in the legend of the middle row. For this plot, the values for k_0 in the non-dispersed case, and $\log(k_0)\mu$ in the dispersed case have been set to 1000 s^{-1}, i.e. totally reversible at the input frequency in table 3.3.

There is a different relationship between the observed difference and the input frequency when tracking changes in the $\log(k_0)\sigma$ parameter, but it is apparent that there is only a weak relationship between the frequency and the largest difference. However, when we discuss how to distinguish kinetic and thermodynamic dispersion, it should be noted that a distinction between

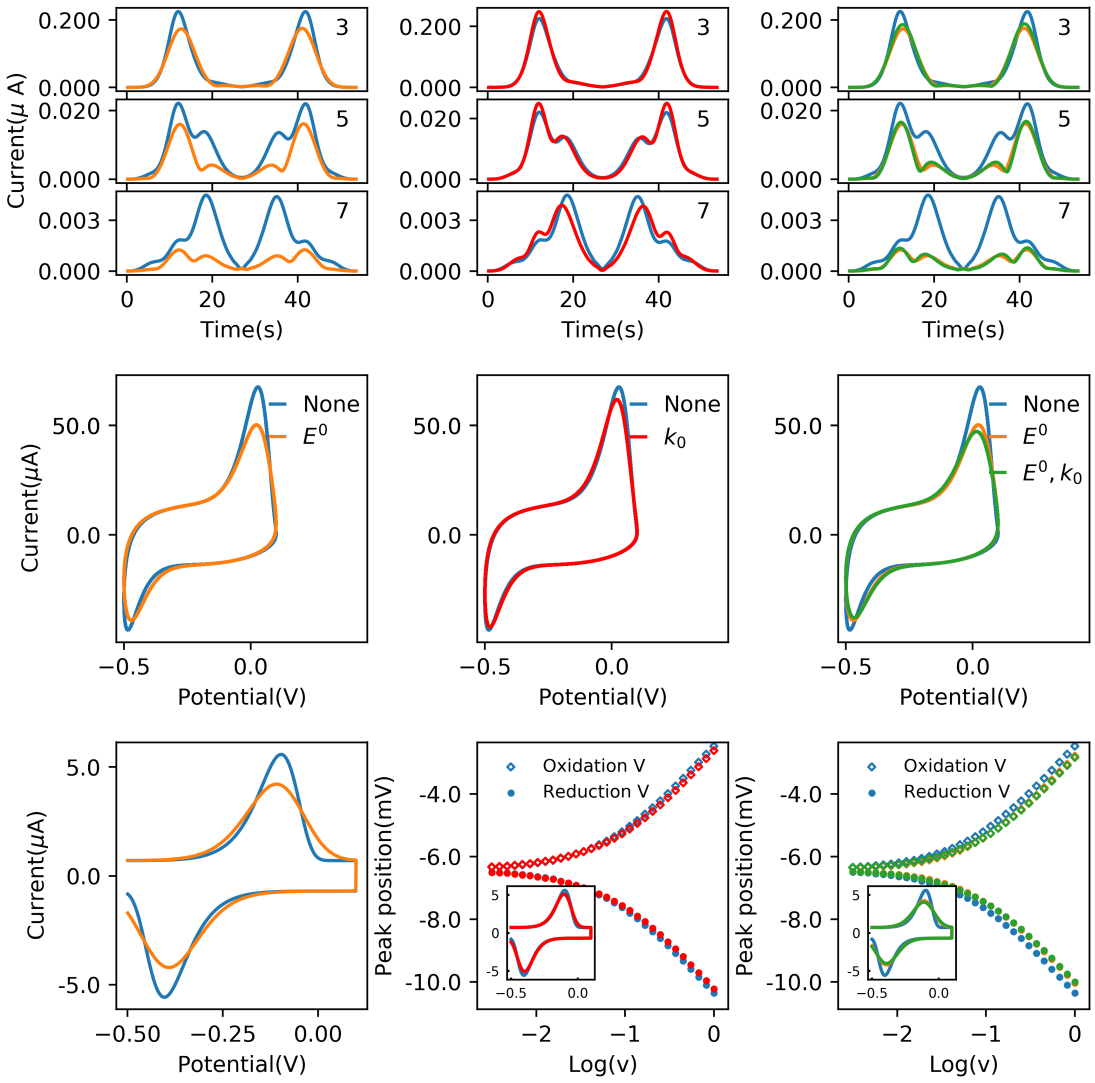


Figure 3.24: Comparison of different forms of dispersion for r-FTACV, PSV and DCV in the quasi-irreversible kinetic regime. Each column represents a different form of dispersion, compared against the non-dispersed case (labelled as "None"). From left to right, the dispersed parameters are E^0, k_0 and a combination of the two, as listed in the legend of the middle row. k_0 and $\log(k_0)\mu$ are both set to 1 s^{-1} as appropriate

these two effects is somewhat arbitrary. Marcus theory predicts that the thermodynamics of an electron transfer reaction will impact the rate constant [73], and consequently we should expect that presence of dispersion in one parameter will then imply dispersion in the other parameter. The separation is only a valid approximation when one of the dispersed parameters can be assumed to be a point value with reasonable accuracy, as is the case with low values of the E^0_σ

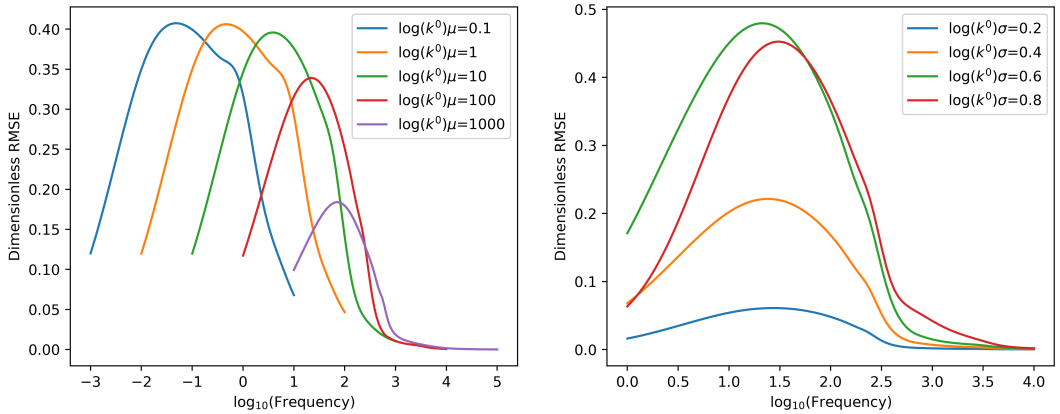


Figure 3.25: The dimensionless root mean squared difference observed between simulated PSV current in the kinetically dispersed and non-dispersed cases, as a function of varying frequency, for varying values of the shape and scale parameters of the log-normal distribution of k_0, as shown in the legend.

parameter, or a distribution of k_0 values which are largely reversible.

Given this analysis, we judged that the best way to distinguish different forms of dispersion was to compare experimental data against non-dispersed simulations, and select which dispersion parameter it is necessary to include in the model based on the observed discrepancy. Furthermore, we have demonstrated that the best medium for this comparison is the r-FTACV harmonics. If it is possible to obtain a good fit between the lower-order r-FTACV harmonics, but the simulated magnitude for the higher harmonics is too large, and the experimental harmonics are broader than predicted by the simulation, then it is likely that E^0 dispersion is present. If it is the case that the discrepancy is only in the magnitude, then the dispersion is likely kinetic, as long as the experimentalist is confident that the uncompensated resistance is not causing this effect.

This method requires non-dispersed simulations that are a reasonably good fit to existing experimental data. We generate these simulations using our existing workflows for automatic fitting processes, as laid out in section 2.4. Because the simulation of a non-dispersed current is b^n times faster than simulating dispersed current (i.e. the number of bins or nodes b to the power of the number of dispersed parameters n), the time impact of the additional non-dispersed fitting attempts is relatively minimal. These non-dispersed simulations could be generated without automatic parameter fitting, but this comes with all the attendant issues with such a manual

approach to fitting, as discussed in the commentary by Bond [16].

As we shall demonstrate in section 4.3.3, we have found that when solving the inverse problem for single-electron systems of this type there is a clear contribution from thermodynamic dispersion, but that the inclusion or absence of kinetic dispersion does not improve the fits to the data. This is not to say that kinetic dispersion is not taking place, rather that in the experimental systems we have investigated, our analysis of voltammetric data has not required kinetic dispersion to obtain acceptable fits to the data. This assertion is supported by our observations in this section, specifically the limited effect that kinetic dispersion has on the appearance of the experimental current. It should be noted that our inability to detect kinetic dispersion is in contrast to other established literature [50, 66]. As described in the introduction, in Léger's work on analysing catalytic current in the steady state [52], analysis of the voltammetric data heavily emphasises kinetic dispersion. A possible reason for this discrepancy is the difference in complexity of mechanism analysed; in the following section we show that, under certain parameter regimes, the effect of kinetic dispersion is more apparent when the current is simulated with a model that incorporates catalytic chemistry.

3.3.2.2 Catalytic processes

First, we show the impact of changing to the redox reaction model to include catalytic processes. We model this by including more terms in the calculation of $\frac{d\theta}{dt}$, written here in dimensionless form

$$\frac{d\theta}{dt} = k_{ox}(t)(1-\theta) - k_{red}(t)(\theta) + k_2(1-\theta) - k_{-2}(\theta), \tag{3.28}$$

where k_2 and k_{-2} are dimensionless first order rate constants. This is an example of a so-called "EC" model, described in work by Fourmond [50, 66], and represented graphically in figure 3.26(ii). In figure 3.27 we show that the effect of k_0 dispersion in this system is more noticeable relative to the non-dispersed case than in figures 3.22-3.24 (derived from a non-catalytic reversible electron-transfer reaction model), even though the majority of the simulation parameters are identical.

3.3.2.3 Distribution choice

The choice of distribution function will also affect the simulated current. We have used the log-normal distribution for the kinetic parameter but this is not the only distribution that has

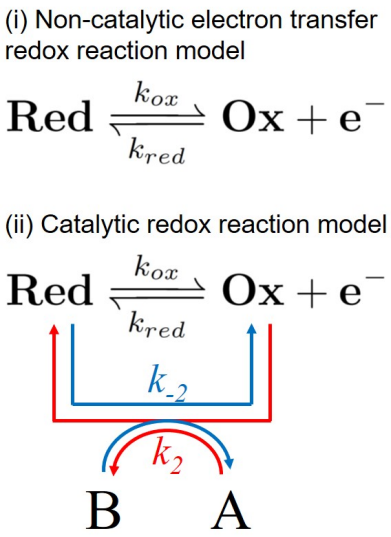


Figure 3.26: (i) Single-electron redox process, as modelled throughout this thesis. (ii) EC model, incorporating Faradaic and catalytic processes, as modelled in equation 3.28.

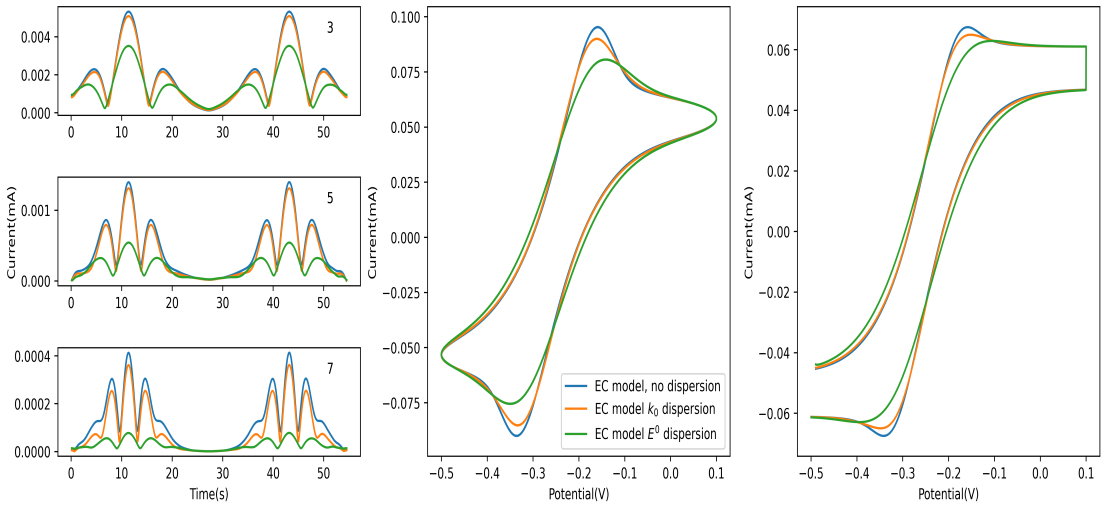


Figure 3.27: Simulated current for a r-FTACV, PSV and DCV experiment from left to right, simulated according to an EC model. The forward and backward first-order catalytic rates constants (k_2 and k_{-2}) were set at 80 s^{-1}, and all other parameters as in table 3.3.

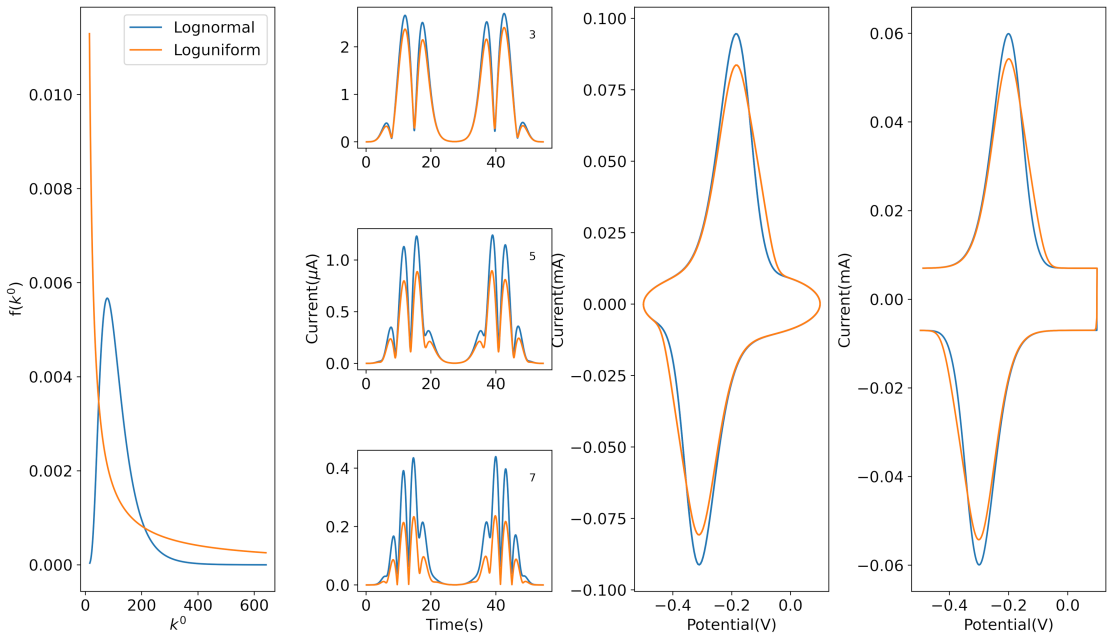


Figure 3.28: Comparison between r-FTACV, PSV and DCV experiments simulated using a log-normal and a log-uniform distribution, with the two distributions shown on the left-most figure. Parameters are as in table 3.3. For the log-uniform distribution, the upper and lower values chosen were the upper and lower values used to approximate the log-normal distribution.

been used to model kinetic dispersion, with the log-uniform distribution being used by Léger et al. [52]. We show the difference between a kinetically dispersed current simulated assuming a log-normal and log-uniform distribution respectively in figure 3.28. Since the log-uniform k_0 distribution has higher density for lower values of k_0 than the corresponding log-normal k_0 distribution, consequently current simulated using the former distribution has a more irreversible character than the latter.

3.3.2.4 Experimental example

After this section on how to spot different forms of dispersion, we briefly conclude with a demonstration of determining the presence of dispersion from experimental data. When fitting PSV data, we found we could fit the data without including dispersion, but when comparing the resultant r-FTACV simulation to the data, we discovered a discrepancy between experimental and simulated harmonics that could only be explained by the presence of thermodynamic dispersion.

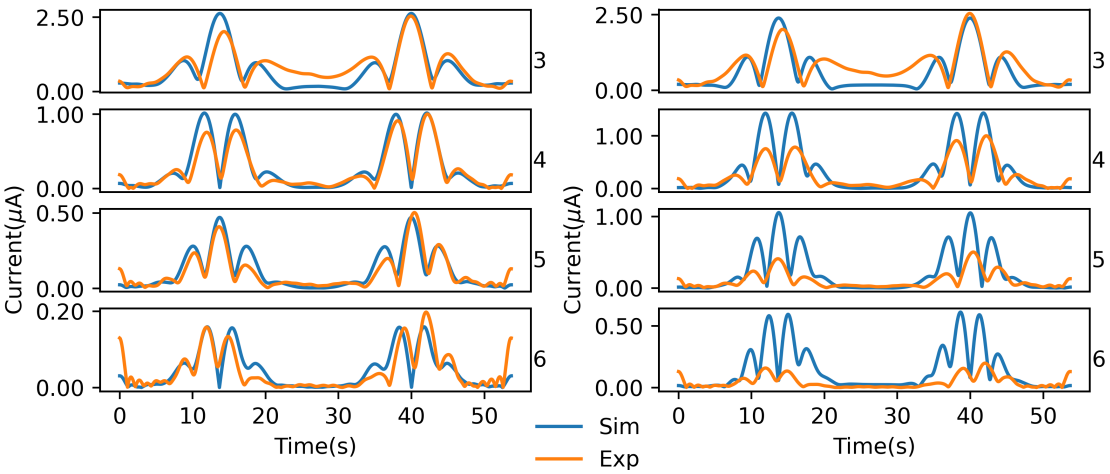


Figure 3.29: Simulated and experimental harmonics 3-6 from r-FTACV current. The left hand side was simulated using dispersion (using the r-FTACV parameters described in chapter 4). On the left, simulations were generated by setting the $E^0\mu$ value as the E^0 value, and reducing the value of the surface concentration Γ by half.

We show this in figure 3.29; on the left is a fit to the r-FTACV harmonics generated using thermodynamic dispersion. On the right, the simulation is generated without dispersion. For the non-dispersed case, although there is a reasonably good fit to the third harmonic, the amplitudes of experimental higher harmonics decrease more rapidly than is predicted by the simulation, and in addition, the non-dispersed harmonic peaks are insufficiently broad.

3.4 Inference on synthetic noisy data

Having explored the effect of every parameter in the model laid out in section 2.2, we turn to assessing the inference procedure laid out in section 2.4. In this section, we infer parameters from synthetic noisy surface-confined single-electron voltammetric data for the three "core" techniques; DCV, PSV and r-FTACV, and show that we can infer the correct parameters using our approach.

3.4.1 Noise

When creating artificially noisy voltammograms I_{noisy}, our procedure for the addition of i.i.d. noise is simply the addition of a random variable $\epsilon(t)$ to the simulated current I_{sim}

$$I_{noisy}(t) = I_{sim}(t) + \epsilon(t), \quad \epsilon(t) \sim \mathcal{N}(0, (\max(I_{sim})p)^2), \tag{3.29}$$

such that $\mathcal{N}(0, (\max(I_{sim})p)^2)$ indicates a normally-distributed random variable with 0 mean, and a standard deviation proportional to $\max(I_{sim})$, and p is a scaling factor in the range of 0.005-0.05, or 0.5-5% of the maximum. We have found that a noise level of 0.05% of the maximum current is usually a good approximation to the noise observed in actual experimental voltammetric data.

As we are comparing our own simulations against themselves, this effort does not provide much indication about how successful we will be in inferring parameters from experimental current, and in subsequent chapters we will show the many model modifications required to do so. What it does show is that our inference procedures are working correctly, i.e. if we can perfectly describe the behaviour of a system exhibiting i.i.d. noise then we can accurately recover parameters that describe I_{sim}. Secondly, using MCMC, we can compare the relative widths of the inferred distributions to assess the relative amount of recoverable information for each technique.

3.4.2 Total current fitting

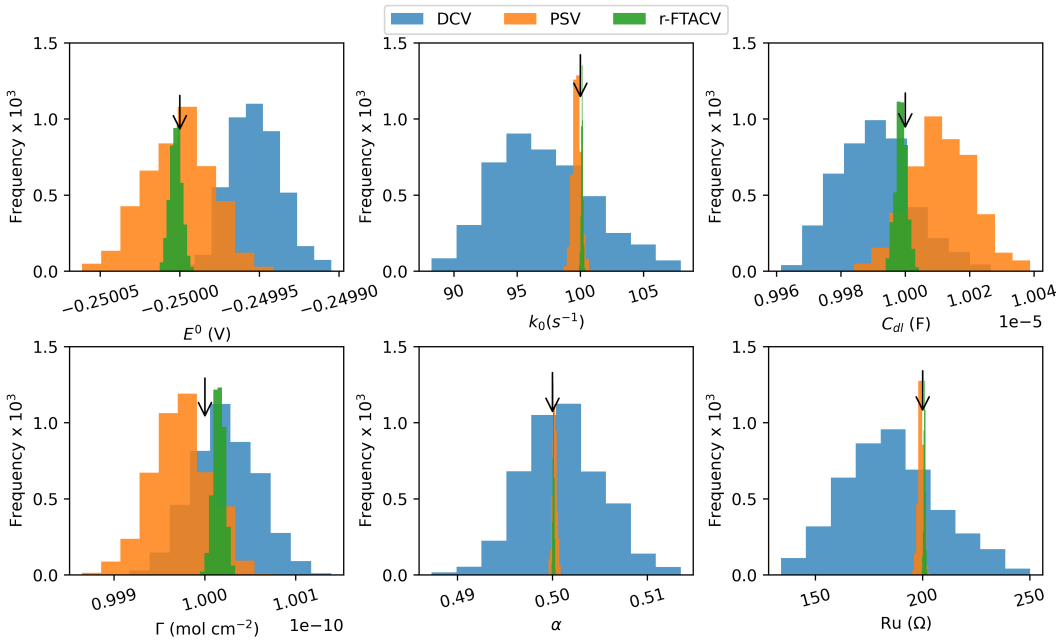


Figure 3.30: Parameter distributions inferred from synthetic noisy data for DCV, PSV and r-FTACV, where the magnitude of the noise added was 0.5% of the maximum current value. The true parameter value is indicated by an arrow on each histogram.

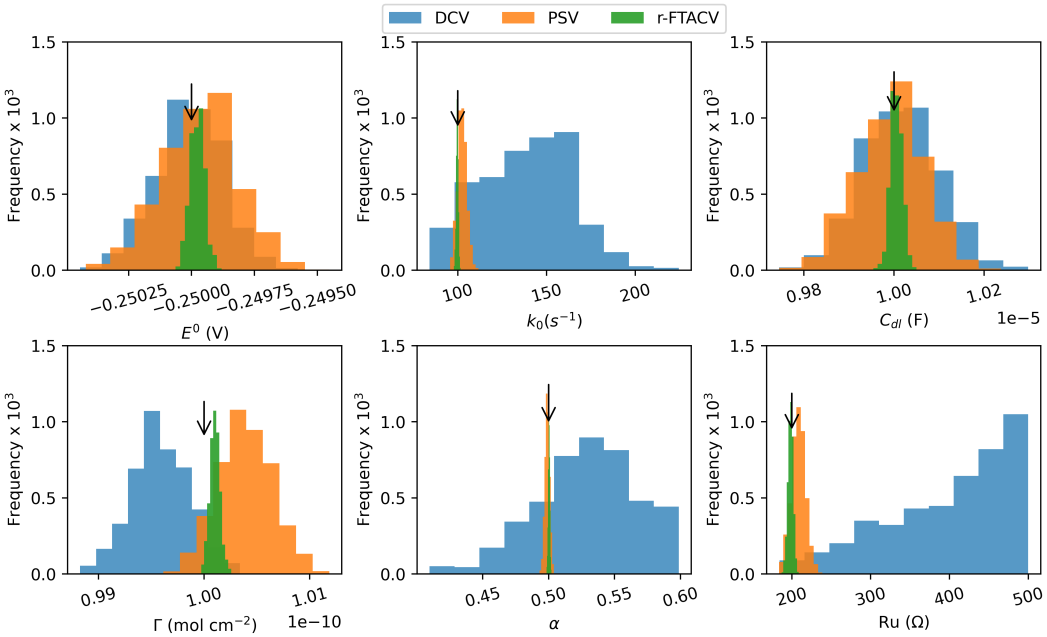


Figure 3.31: Parameter distributions inferred from synthetic noisy data for DCV, PSV and r-FTACV, where the magnitude of the noise added was 4% of the maximum current value. The true parameter value is indicated by an arrow on each histogram.

For all three experiments, we simulated total current using the parameters found in table 3.2, except for the value of R_u, which was set to 200 Ω, for reasons outlined below. Noise was added as in equation 3.29, and we fitted the resulting noisy current using MCMC. Three chains were initialised from the correct parameter values, run independently for 10000 samples, and inspected for convergence. The first 8500 samples for each chain were discarded as burn-in; this value was chosen after inspecting the traces of every chain and determining the minimum sample number after which every fitting attempt (across all techniques and noise percentages) converged. The resulting samples were pooled and plotted as histograms, for the 0.5% and 4% noise cases in figures 3.30 and 3.31 respectively, where the true parameter value used in the simulation of the total current prior to adding noise is plotted as an arrow. The standard deviations, represented as a percentage of the true parameter value, are presented in table 3.4 for every noise percentage. Comparing the three techniques, the parameter distributions inferred using r-FTACV are always more confident than for both PSV and DCV. Directly comparing r-FTACV and PSV, the inferred standard deviations are usually 2-5 times larger for the PSV case. When directly comparing PSV and DCV, the standard deviations are similar for some parameters (E^0, C_{dl} and Γ), but much larger for the DCV case in others (k_0, α and R_u), such that the standard deviation inferred for

R_u is almost 20 times larger than for PSV. This is because DCV struggles in relatively high resistance regimes (which is why R_u was set to 200Ω, for demonstration of this effect). In fact, the distributions of k_0 and R_u inferred by DCV become very broad with a noise value of 2%, losing their Gaussian appearance, as can be observed in figure 3.31, and the inferred standard deviations no longer increase monotonically with increasing noise, as they do for all parameters for the PSV and r-FTACV cases. It should be noted that the parameter regime indicated in table 3.2 is favourable relative to experimental systems, with a high surface coverage, a low value for the double layer capacitance and no non-linear capacitance effects. PSV and r-FTACV both perform well under these conditions, and the very broad distributions from DCV provide a rationale for use of the other two techniques.

Parameter	Experiment	Noise %				
		0.5%	1%	2%	3%	4%
E^0	DCV	6.6e-03%	1.4e-02%	2.4e-02%	3.9e-02%	5.3e-02%
	PSV	8.4e-03%	1.6e-02%	3.1e-02%	5.0e-02%	5.9e-02%
	r-FTACV	1.5e-03%	3.1e-03%	6.4e-03%	8.9e-03%	1.2e-02%
k_0	DCV	3.84%	8.89%	27.39%	15.75%	25.73%
	PSV	0.33%	0.64%	1.33%	2.14%	2.58%
	r-FTACV	7.5e-02%	0.15%	0.32%	0.48%	0.6%
C_{dl}	DCV	0.12%	0.26%	0.47%	0.66%	0.85%
	PSV	9.7e-02%	0.18%	0.34%	0.56%	0.74%
	r-FTACV	1.8e-02%	3.8e-02%	6.7e-02%	0.11%	0.13%
Γ	DCV	3.6e-02%	7.8e-02%	0.14%	0.23%	0.27%
	PSV	3.2e-02%	5.9e-02%	0.12%	0.18%	0.26%
	r-FTACV	6.2e-03%	1.2e-02%	2.4e-02%	3.5e-02%	4.8e-02%
α	DCV	0.82%	1.85%	4.81%	4.63%	7.59%
	PSV	4.7e-02%	9.4e-02%	0.18%	0.28%	0.36%
	r-FTACV	1.5e-02%	2.9e-02%	5.6e-02%	9.2e-02%	0.12%
R_u	DCV	11.23%	21.52%	37.92%	44.66%	39.41%
	PSV	0.6%	1.14%	2.33%	3.79%	4.53%
	r-FTACV	0.2%	0.41%	0.85%	1.28%	1.64%

Table 3.4: Standard deviations for inferred parameter distributions (such as the ones plotted in figures 3.30 and 3.31 for the 0.5% and 4% noise cases respectively), represented as a percentage of the true parameter value, which is recorded in table 3.2, except R_u which was set to 200 Ω

3.4.3 Fourier domain fitting

We also performed inference on the Fourier spectrum of synthetic noisy currents, for PSV and r-FTACV. We would expect the majority of the observations obtained in the previous section to also hold true for the frequency domain case. Therefore, we instead used MCMC methods to investigate the effects of filtering the Fourier spectrum. As detailed in section 2.2.5.5, the higher harmonics of PSV and r-FTACV report exclusively on the Faradaic process, and so the lower harmonics can be discarded to remove the impact of large and complex background currents [9]. However, because these lower harmonics also contain information about the Faradaic process, we

are effectively throwing information away, which may have an effect on the inferred distributions. We can use MCMC to investigate the impact of this filtering, by comparing the widths of the inferred distributions obtained by fitting to the filtered and un-filtered Fourier spectrum. In figures 3.32 and 3.33, we see that the only parameter affected by discarding harmonics 1-3 in the fitting process is Γ, with the widths of the inferred distributions for the other parameters essentially unchanged. We did not fit C_{dl} in either case as there is minimal information present on that parameter in the high harmonics for this parameter regime. However, we again note that this parameter regime (in table 3.2) is a highly favourable one, and in later chapters we will shown that for experimental systems, the discarding of the lower harmonics has less benign consequences.

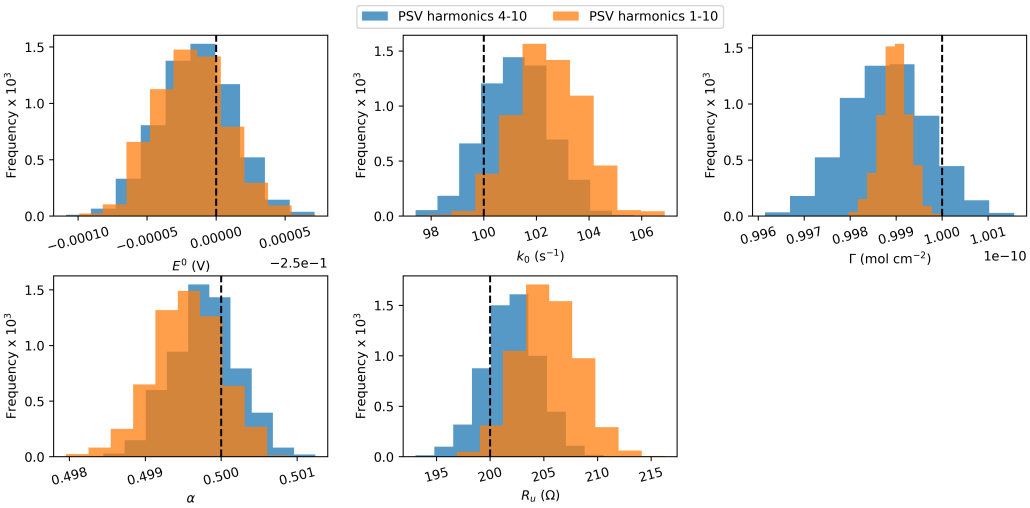


Figure 3.32: Parameter distributions inferred from fitting to PSV current harmonics 1-10 and 4-10 with 0.5% noise added to the total PSV current, where the true parameter values are shown as black dotted lines.

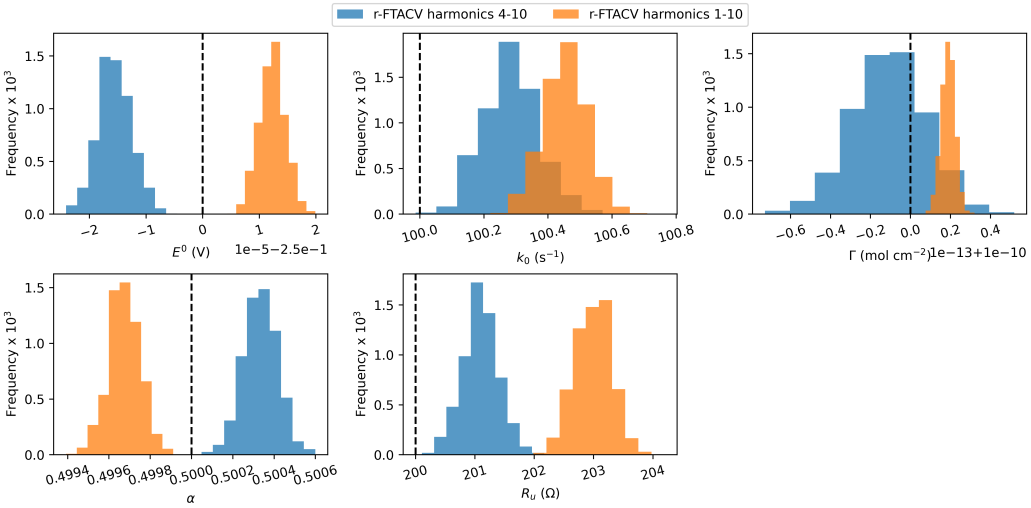


Figure 3.33: Parameter distributions inferred by fitting to synthetic r-FTACV current harmonics 1-10 and 4-10, with 0.5% noise added to the total r-FTACV current, where the true parameter values are shown as black dotted lines.

3.5 Conclusions

In this chapter, we have discussed many aspects of the modelling and inferential process detailed in chapter 2. Having validated our numerical approach for simulating PSV by comparison to an alternative analytical approximation method, we explored the forwards problem, characterising the impacts of various key model parameters, and have also investigated the impact that dispersion has on the appearance of voltammetric current. We have demonstrated that we can successfully obtain accurate parameter estimates from synthetic noisy data, and in the subsequent chapters, we move on to analysis of real data, informed by the results obtained in this chapter. In the next chapter, we start by inferring parameters from current generated by a simple chemical system — ferrocene.

Chapter 4

Ferrocene

Content in this section has been re-purposed from the following reference: Henry Lloyd-Laney, Nicholas Yates, Martin Robinson, Alice Hewson, Jack Firth, Darrell Elton, Jie Zhang, Alan Bond, Alison Parkin, and David Gavaghan, 2021. Using purely sinusoidal voltammetry for rapid inference of surface-confined electrochemical reaction parameters. ***Analytical Chemistry****, 93(4), pp.2062-2071. All experimental data were collected by Nicholas Yates and Alice Hewson at the University of York. All parameter inference and analysis was undertaken by Henry Lloyd-Laney, except the DCV surface coverage estimation shown in figure 4.9, which was performed by Nicholas Yates.*

4.1 Introduction

Having demonstrated in the previous chapter that it is possible to infer correct parameters from synthetic noisy PSV data, in this chapter we repeat the process for experimental data. In order to characterise the behaviour of PSV in an experimental system, we chose an extensively studied electrochemical species, such that any challenges in analysing the voltammetric current would be related to PSV rather than the analyte. Our aim was to demonstrate the advantages of the reduced simulation time for PSV (relative to r-FTACV), and to assess whether the parameters inferred from PSV were comparable to those obtained from the more information rich r-FTACV.

For our model system we chose ferrocene, which is a single iron atom co-ordinated by two five-carbon cyclopentadienyl rings (shown in surface-linked form in the bottom right of figure

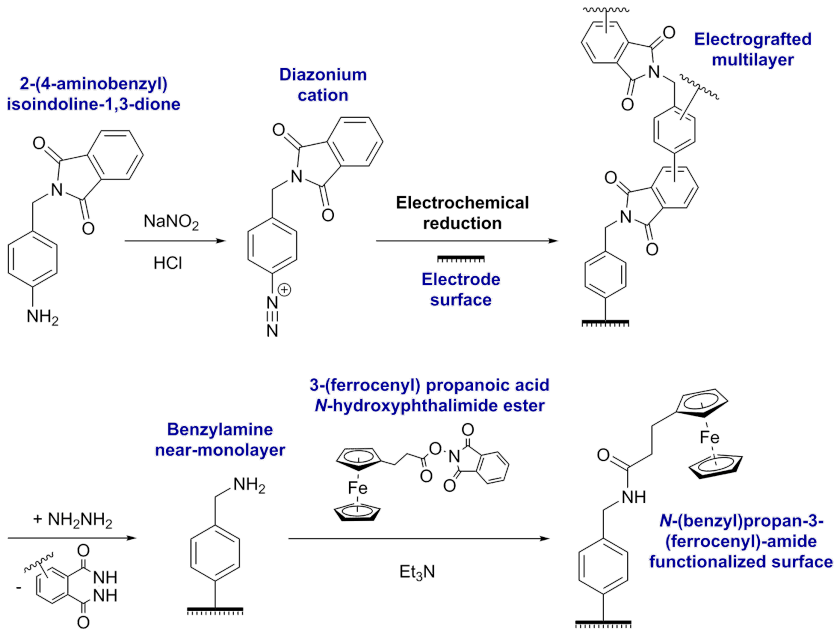


Figure 4.1: Cartoon representing the diazonium functionalisation procedure used to generate a ferrocene-coated glassy carbon electrode surface.

4.1), the electrochemistry of which is extremely well understood. The rate of the redox reaction that interconverts the Fe atom between its 2+ and 3+ oxidation state is rapid, which can be explained by the fact that the length of the Fe-C bond does not change significantly with the addition or removal of a single electron, which inhabits a molecular orbital that does not significantly participate in bonding [74]. The robustness of the ferrocene redox process is so well established that IUPAC has designated it as a "redox reference standard" for electrochemistry in non-aqueous solvents [75]. As the ultimate aim of this thesis is to use PSV to study the electron-transfer reactions of proteins, we performed the characterisation on ferrocene that had been chemically attached to the surface of the electrode, as proteins we study in later chapters will also be surface-linked. A cartoon representing the chemical steps of this process is shown in figure 4.1. This introduced significant challenges related to dispersion and non-linear capacitance that we have outlined in section 3.3 and 2.2.2 respectively.

4.2 Experimental methods

A glassy carbon electrode modified with the ferrocene diazonium derivative (N-(benzyl)-propan-3-(ferrocenyl)-amide, shown in the bottom right of figure 4.1) was subjected to ten PSV experi-

ments followed by ten separate r-FTACV experiments. The PSV experiments used a sinusoidal waveform with a frequency of 8.94 Hz and an amplitude of 300 mV over a potential range of -40 to +560 mV vs Ref of +0.171 mV vs SHE, for a time period of 26.8 s, equivalent to 240 sine-wave oscillations. For all PSV inference attempts, we compared simulations to the total current response. The PSV ΔE value was selected such that the PSV input would encompass the same potential range as a complete DCV/r-FTACV linear sweep, which in itself is designed to encompass all potentials at which Faradaic processes occur. If this range is smaller then information about the Faradaic process is lost. Conversely, including potentials outside of this Faradaic window would only yield additional background current. For r-FTACV, we compared simulations to a portion of the Fourier transform of the total current, corresponding to harmonics 2-6, using equation 2.55. When comparing experimental and simulated PSV currents using equation 2.54 in chapter 2, the initial transient current (corresponding to the first oscillation of the potential input) was discarded, and for PSV data-simulation comparisons, the current was truncated to the times before the 30th oscillation. These manipulations represent a compromise between reducing the number of points for simulation speed, and to ensure that the results of parameter inference attempts were consistent — we found that truncating the data to under 20 oscillations resulted in a loss of convergence between Markov chains when analysing the data using MCMC. By comparison, the AC component of the r-FTACV experiment requires approximately 475 oscillations for complete simulation. Although there was scope for assessing the impact of the experimental input frequency ω on the fidelity of the inference process, this work was not carried out. This was predominantly due to time constraints, and a need to work on more complex biological systems to fulfill the requirement of the DPhil course, instead of continuing to characterise the ferrocene inference process. Additionally, it was observed that the Faradaic signal appeared to decrease in current collected over 10 Hz. This is contrary to what would be predicted, as for example in figure 3.5, and is an effect that requires further investigation.

4.3 Results and discussion

The results from the parameter inference process are shown in figures 4.2-4.6 with each figure showing both experimental and best-fit simulated current in several forms: the total current, the real and imaginary components of the Fourier spectrum up to harmonic 12, and harmonics

one through 7 inverse Fourier transformed back into the time domain. Because of the close similarity of the simulated and experimental currents, the experimental (orange) line has been made partially transparent, and as such regions of very good agreement are brown. The inferred parameters for all ten experiments can be found in table 4.1, and were obtained for each experimental current by running the CMAES algorithm 10 times, with random initial parameter values, with chemically informed upper and lower bounds for each parameter, which are reported in table 4.1, and noting the best scoring parameter vectors. The values of the upper and lower bounds were largely drawn from observation of the data. The kinetic parameter k_0 has a lower bound of 50, which neglects values in the irreversible and quasi-reversible regimes. This was done as there was a local minimum in parameter space observed when fitting the ferrocene PSV data. However, simulating the parameter vector associated with this minimum with an r-FTACV input, and comparing to the r-FTACV data showed that this fit was an artefact, as the simulation did not match the r-FTACV experiment data at all. The value for $\alpha\mu$ was selected as values of α outside the range of 0.35 and 0.65 was determined to be chemically implausible.

Parameter	Symbol	PSV 1	PSV 2	PSV 3	PSV 4	PSV 5	PSV 6
Midpoint potential mean	$E^0\mu$ (V vs. Ref)	0.247	0.244	0.244	0.244	0.242	0.241
Midpoint potential standard deviation	$E^0\sigma$ (V)	0.028	0.047	0.045	0.046	0.047	0.048
Rate constant	k_0 (s^{-1})	104.61	144.066	142.214	142.928	140.303	139.955
Uncompensated resistance	R_u (Ω)	732.021	517.171	539.316	540.249	516.33	495.788
Linear double-layer capacitance	C_{dl} (F)	7.689E-5	7.681E-5	7.636E-5	7.609E-5	7.609E-5	7.613E-5
1^{st} order C_{dl}	C_{dlE1}	3.207E-3	3.294E-3	3.137E-3	2.976E-3	2.738E-3	2.541E-3
2^{nd} order C_{dl}	C_{dlE2}	-3.993E-4	-4.402E-4	-4.265E-4	-4.205E-4	-4.219E-4	-4.226E-4
Surface coverage	Γ (mol cm^{-2})	7.637E-11	7.335E-11	7.330E-11	7.308E-11	7.288E-11	7.271E-11
Potential frequency	ω (Hz)	8.941	8.941	8.941	8.941	8.941	8.941
C_{dl} phase	C_{dl} η	4.392	4.335	4.342	4.342	4.335	4.329
Phase	η	5.082	4.956	4.969	4.968	4.958	4.95
Symmetry factor mean	$\alpha\mu$	0.588	0.614	0.61	0.609	0.604	0.603
Symmetry factor standard deviation	$\alpha\sigma$	0.163	0.188	0.188	0.189	0.19	0.192
Error	RMSE (μA)	0.997	0.822	0.837	0.828	0.782	0.753

Parameter	Symbol	PSV 7	PSV 8	PSV 9	PSV 10	r-FTACV1	
Midpoint potential mean	$E^0\mu$ (V vs. Ref)	0.241	0.241	0.24	0.241	0.229	
Midpoint potential standard deviation	$E^0\sigma$ (V)	0.048	0.048	0.048	0.048	0.046	
Rate constant	k_0 (s^{-1})	141.737	136.411	133.439	136.212	123.33	
Uncompensated resistance	R_u (Ω)	491.333	493.475	488.66	499.132	873.541	
Linear double-layer capacitance	C_{dl} (F)	7.602E-5	7.582E-5	7.571E-5	7.537E-5	3.341E-5	
1^{st} order C_{dl}	C_{dlE1}	2.475E-3	2.391E-3	2.221E-3	2.277E-3	0.058	
2^{nd} order C_{dl}	C_{dlE2}	-4.228E-4	-4.195E-4	-4.153E-4	-4.106E-4	-2.122E-3	
Surface coverage	Γ (mol cm^{-2})	7.241E-11	7.220E-11	7.205E-11	7.166E-11	7.178E-11	
Potential frequency	ω (Hz)	8.94	8.941	8.941	8.941	8.885	
C_{dl} phase	C_{dl} η	4.328	4.328	4.326	4.328	0.00[†]	
Phase	η	4.947	4.95	4.949	4.951	0.0 [†]	
Symmetry factor mean	$\alpha\mu$	0.603	0.599	0.596	0.599	0.438	
Symmetry factor standard deviation	$\alpha\sigma$	0.192	0.191	0.191	0.191	0.155	
Error	RMSE (μA)	0.745	0.742	0.722	0.747	0.337*	

Table 4.1: Inferred CMAES parameters for PSV experiments 1-10 and the first r-FTACV experiment. [†] The phase and the capacitance phase were both held at 0 during the fitting process. *The error reported for the r-FTACV experiment is the mean error across the fitted harmonics (i.e. 2-6), and so is not directly comparable to the error for the PSV experiments, which is the error for the total timeseries.

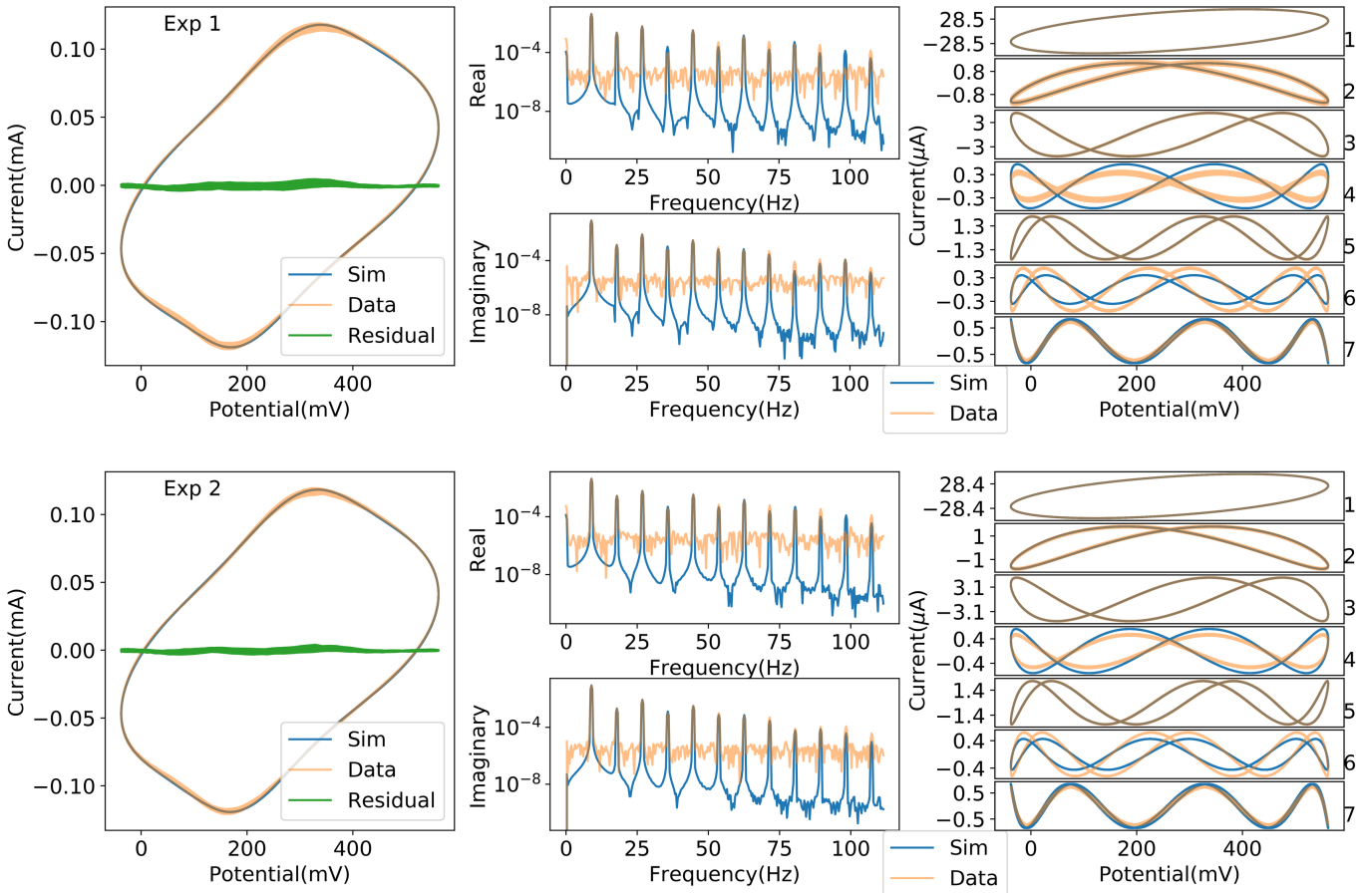


Figure 4.2: Simulation vs experimental data for experiments 1 and 2, with parameter values as reported in table 4.1. The data and simulation are plotted as a current vs. time plot, semi-logarithmic plots of the real and imaginary parts of the absolute Fourier spectrum, and the 1-7th harmonics.

Parameter	Lower bound	Upper bound
$E^0\mu$ (V)	0.2	0.3
$E^0\sigma$ (V)	1e-5	0.1
k_0 (s^{-1})	50	1000
R_u (Ω)	0	1000
C_{dl} (F)	0	1e-4
C_{dlE1}	-0.15	0.15
C_{dlE2}	-0.01	0.01
Γ (mol cm^{-2})	1e-11	1e-9
ω (Hz)	8.49	9.39
Phase	π	2π
C_{dl} phase	π	2π
$\alpha\mu$	0.35	0.65
$\alpha\sigma$	1e-3	0.3

Table 4.2: Bounds for the values drawn by the CMAES algorithm.

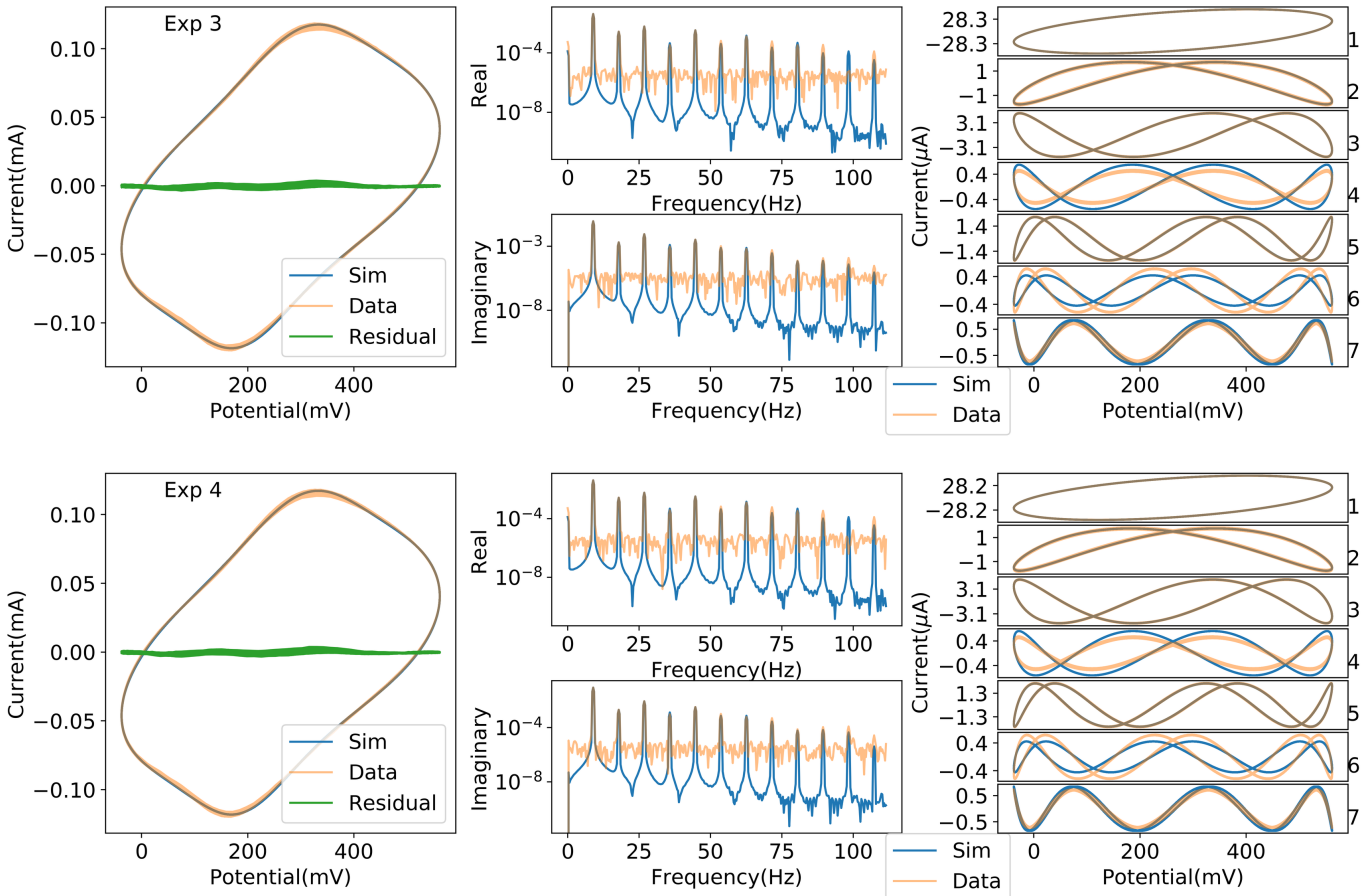


Figure 4.3: Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 3 and 4, using parameters as reported in table 4.1

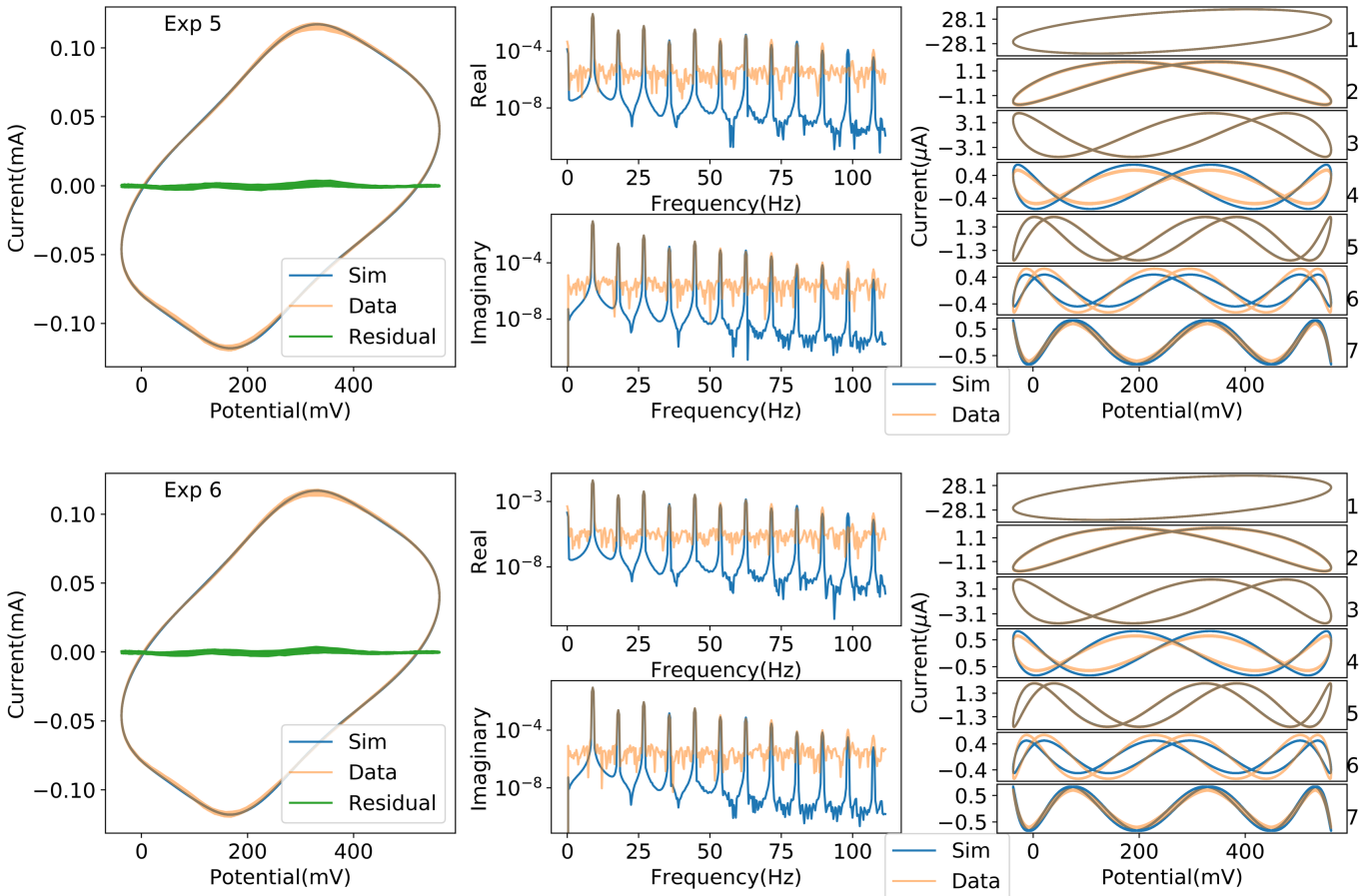


Figure 4.4: Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 5 and 6, using parameters as reported in table 4.1

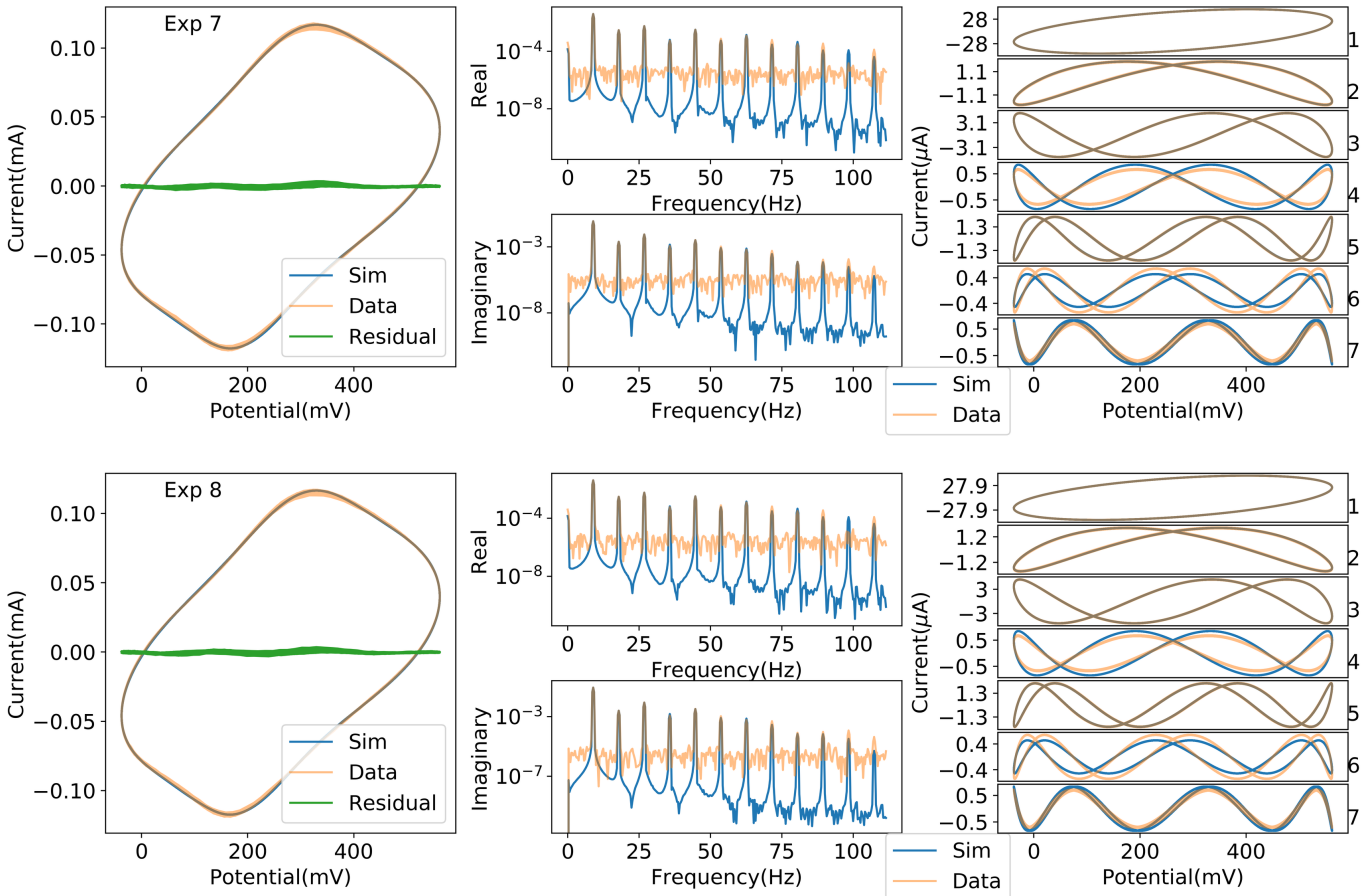


Figure 4.5: Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 7 and 8, using parameters as reported in table 4.1

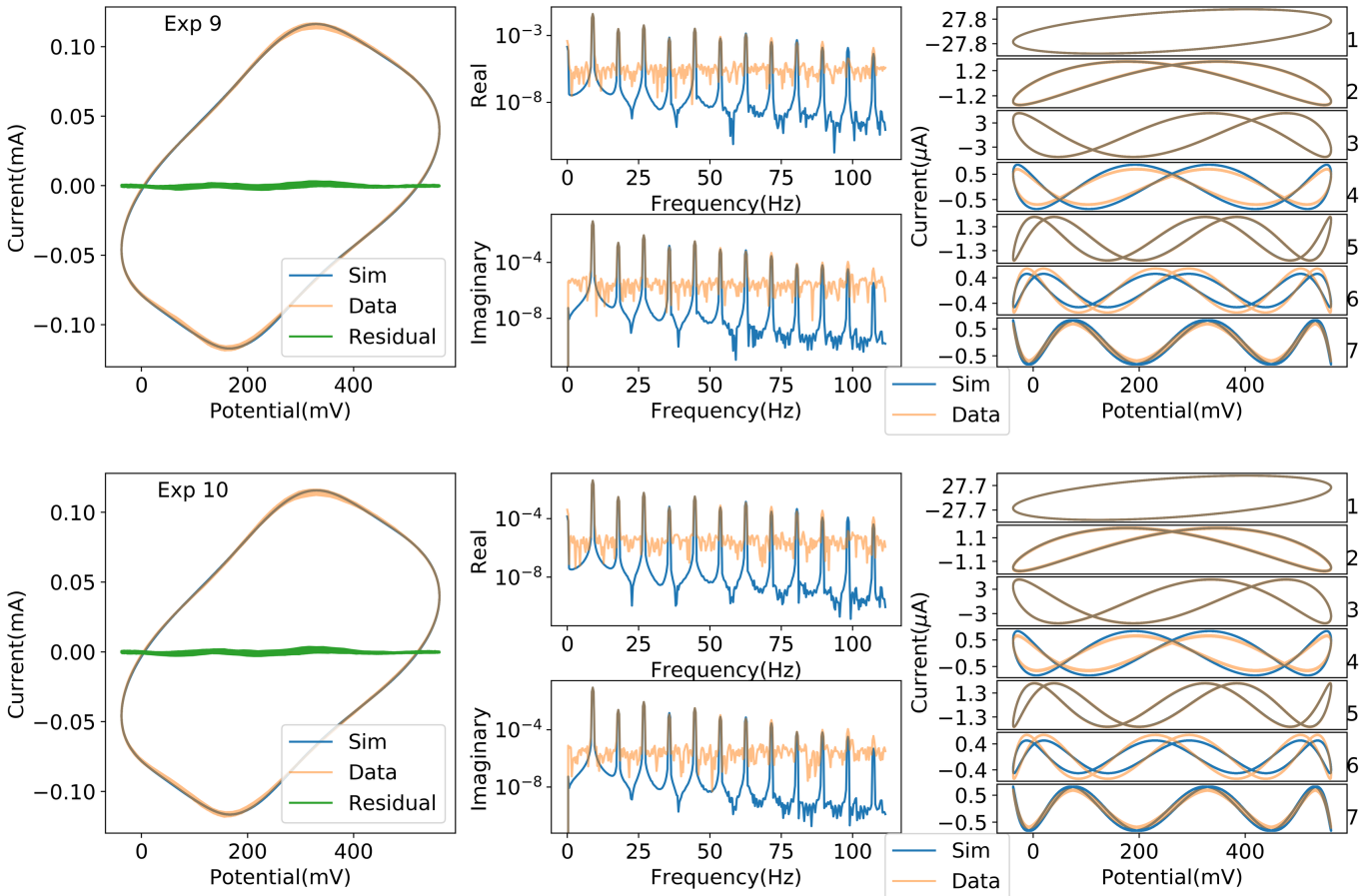


Figure 4.6: Three forms of PSV simulated and experimental current, as in figure 4.2, for experiments 9 and 10, using parameters as reported in table 4.1

4.3.1 Inferring parameters from experimental data

The initial parameter values for each run of the CMAES algorithm were obtained by randomly sampling from within the boundaries, as defined in table 4.2, for each value. We believe that these best-fit values, presented in table 4.1 represent the global minimum (at least as conditional on the parameter boundaries), as the 10 sets of random initialisations will have explored a large portion of parameter space. On average, the solution vectors presented in table 4.1 were returned by the CMAES algorithm 4-6 times. Standard deviations, as determined by MCMC, can be found in table 4.5, which we will discuss in more detail below.

From figures 4.2-4.6, we see that the simulations consistently provide a good fit to the data in both the time and frequency domains, albeit with a slightly worse fit in the even harmonics, especially the fourth harmonic, where the simulation consistently overestimates the magnitude relative to the data. We believe that this is because the even harmonics contribute very little to the total signal (note the magnitude of the peaks in the Fourier spectrum, and the size of the current response in the harmonic plots), and are therefore difficult to fit when using the likelihood function in equation 2.54, in chapter 2. It should be noted that the inaccuracies in the fit (such as overstating the magnitude of the 4th harmonic) are consistent across all ten fitting runs. The fact that the error in the fit relative to the data is consistent implies that this error is systematic, and that we are not correctly modelling every process that contributes to the total observed current. This observation is not surprising, as we are only approximating the current arising from capacitance, and are assuming several features about the redox process generating the Faradaic current (i.e. the Langmuir isotherm, and Butler-Volmer electrode kinetics). While we may have to revisit these assumptions for more complex systems, we do not believe these errors represent a fundamental mismatch with the chemical processes occurring during the PSV experiment, and furthermore they do not affect in any material way the results and conclusions presented in this chapter. We have not combined these results and reported the average, as is common in electrochemical literature, as the parameter variation between experiments in table 4.1 is larger than the calculated standard deviations in table 4.5. Pooling the results would obscure the experimental trends in inferred parameter values observed, which we discuss below.

4.3.2 Model modifications

In order to obtain the fits shown in figures 4.2-4.6, two major inclusions to the model were required; dispersion (as encountered in section 3.3) and the inclusion of a capacitance phase parameter. We will discuss each of these in turn.

4.3.2.1 Dispersion

Normally-distributed dispersion was ultimately incorporated for two parameters, E^0 and α. E^0 was selected as the width observed in the harmonics of the r-FTACV data was much better described by incorporating a distribution of E^0 values. α was also selected for normally-distributed dispersion because we observed that fitting α as a point value resulted in the CMAES algorithm repeatedly returning the upper bound for α as the inferred value, regardless of where this boundary was set. This is probably attributable to the fact that k_0 values are observed to approach the reversible limit and accordingly the time series is insensitive to α. The alternative approach would be to fix the value of α at 0.5 which has the effect of increasing the inferred kinetic value to around 270s^{-1}, but the value of the other parameters remain close to those reported in table 4.1. Including dispersion in the kinetic parameter did not improve the inferred fit, a phenomenon which we discuss below. In addition, the presence of dispersion in the Faradaic parameters suggests that dispersion may also be important in the simulation of current arising from double layer capacitance, although for reasons of computational feasibility, we did not explore this possibility.

In figure 4.7 we show the predicted distributions for parameters where we included dispersion. There are two features in these inferred distributions that we did not expect. Firstly, the distribution of E^0 values for experiment 1 is much narrower than for the other experiments, which we believe is due to a significant change in the composition of the ferrocene layer of the electrode, between PSV experiment 1 and PSV experiments 2-10, as we will discuss below. Secondly, the inferred α parameter distributions encompass all possible values of α (i.e. 0-1), indicating that the simulated current is largely insensitive to α in this parameter regime. As noted in 3.3, the inclusion of dispersion increases the simulation time by b^n, where b is the number of bins used to approximate the distribution and n is the number of dispersed parameters. In this work, we used 5 bins for both parameters, meaning that the simulation time was 25 times slower. This increased the average inverse problem solution time to approximately 45 minutes for PSV using CMAES, whereas for r-FTACV this process took approximately 8 hours. The utility of the in-

creased speed for PSV is that it allows for the assessment of many more modelling options (such as which parameters to include dispersion for) relative to r-FTACV in the same time-frame.

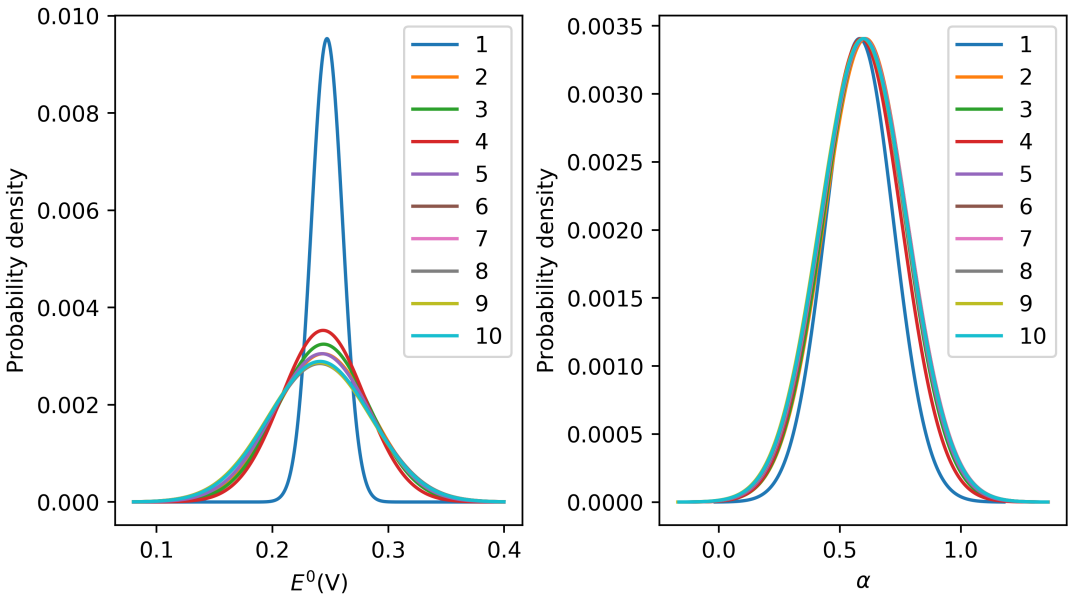


Figure 4.7: Inferred normal distributions for E^0 (left) and α (right) for each experiment, as determined by the values obtained by the CMAES fitting process (shown in table 4.1)

4.3.2.2 Capacitance phase

Incorporation of a capacitance phase as a parameter distinct from the Faradaic phase, as described in section 3.2.8, was essential to achieve good agreement between simulation and experiment. A difference between the two phases of about 35 degrees is consistently observed with the ferrocene process. In order to investigate whether the inclusion of this parameter was justified, we collected data from a blank (unmodified) glassy-carbon electrode using PSV. Because there is no redox-active molecule under investigation, the resulting current will be purely capacitive. When attempting to fit this data, using the parameters $C_{dl}, C_{dlE1}, C_{dlE2}, R_u$ and then either including the C_{dl} phase parameter or setting it at the experimentally set value of $\frac{3\pi}{2}$, we demonstrate that achieving a good fit to the blank data requires the inclusion of the phase, as demonstrated in figure 4.8. This confirms that explicitly including the capacitive phase angle is essential to obtaining a good fit. An ideal capacitor is phase-shifted by 90 degrees relative to an input sinusoid (i.e. it has the form of a cosine wave). The Faradaic current arising from a reversible process

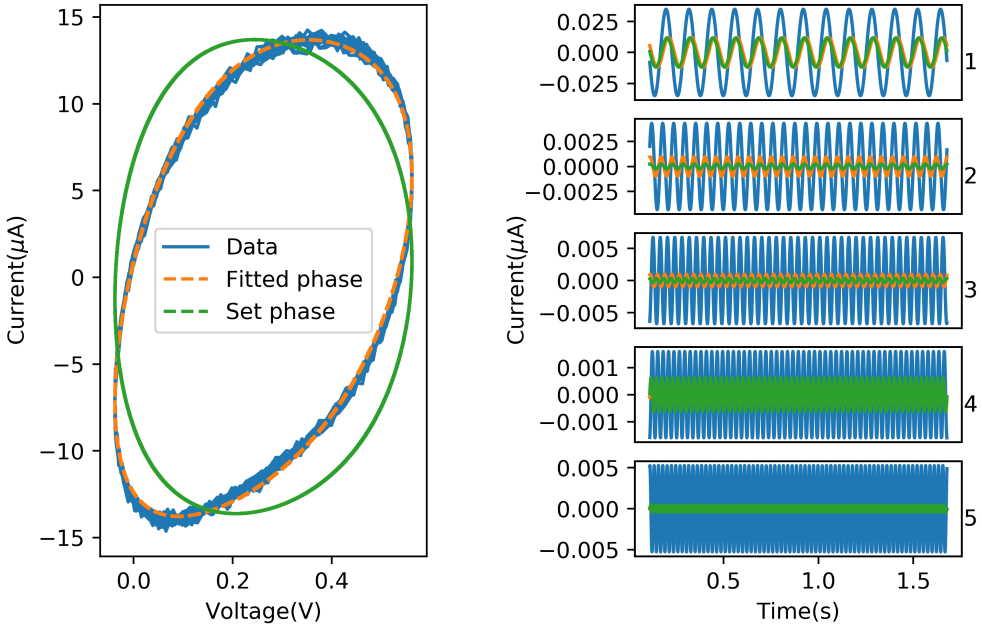


Figure 4.8: Blank electrode data (blue) overlaid with fitted capacitive current, when the phase was fitted explicitly as a parameter (orange), and when the phase was fixed at the experimentally set phase of $\frac{3\pi}{2}$ (green). Harmonics 1-5 of the three current-voltage plots are shown on the right.

is predicted to be phase shifted by about 45 degrees relative to the input sinusoid. However, these phase angles will be affected both by the fact that the Faradaic process is not completely reversible, and that the capacitance is non-ideal. Figure 4.8 also demonstrates that, as expected, a purely capacitive PSV current has negligible current response above the second harmonic ($\sim$3 orders of magnitude smaller than observed in e.g. figure 4.2), as with r-FTACV.

4.3.3 Discussion of values inferred by CMAES

In the parameter inference exercise based on fitting the total PSV current, a global best fit of 13 parameters was achieved, as listed in table 4.1. On the basis of results summarised in the table, experiment 1 appears to be an outlier with significantly higher error estimates than found with the other experiments. This implies that a structural, orientation or other change of the surface confined ferrocene accompanies the initial oxidation-reduction cycling of potential that

Parameter	Value
Anodic peak potential (V vs Ref)	0.291
Cathodic peak potential (V vs Ref)	0.23
Number of electrons n_s	1
Apparent number of electrons n_{app}	0.525
Averaged Γ (mol cm^{-2})	7.45E-11

Table 4.3: Parameters fit to background-subtracted current, using equation 4.1, where the resulting simulation is shown in purple in figure 4.9.

occurs in experiment 1. In experiments 2 to 10 the surface coverage decreases marginally with each experiment, which can be attributed to a loss of a small quantity of ferrocene from the electrode surface that accompanies each experiment. "Film-loss" is a common feature of surface confined voltammetry [19, 76], and the value of the surface concentration is corroborated to within picomolar values by analysis of the DCV current. This coverage validation was obtained by analysis of DCV data obtained from the same electrode film as the PSV data shown above. The data, and the resulting background-subtracted current is shown in figure 4.9. Simulations for the background-subtracted current were obtained assuming a Nernstian process,

$$i = \frac{n_s n_{\text{app}} F^2 v A \Gamma}{RT} \frac{\exp\left\{n_{\text{app}} F \left(E - E^0\right)/RT\right\}}{\left(1 + \exp\left\{n_{\text{app}} F \left(E - E^0\right)/RT\right\}\right)^2}, \tag{4.1}$$

using the parameters defined in table 4.3, as used in [39].

In addition to the gradual decline in the value of the surface coverage from experiment 1-10, there are other trends in the inferred parameters, particularly in values for R_u and k_0, with results that differ by approximately $\sim$4 standard deviations (as calculated using MCMC, shown in table 4.5). This is presumably in response to changes in the composition of the film, and takes this form because of correlations between the parameters and observable in figure 4.12. Significantly, the predicted distribution for α plotted in figure 4.7 is extremely broad, relative to the commonly "expected" α value in the range 0.4-0.6. This represents the low impact α has on PSV data when the electrode kinetics lie near to the reversible limit. The relatively wide spread of uncompensated resistance values also reflects the insensitivity of data to this parameter in highly conducting aqueous electrolyte media. Finally, the consistent separation between the η and C_{dl} η parameters is interesting, for a number of reasons. In the ideal case as described in appendix D, there should be a 90 degree phase shift between the capacitive and Faradaic

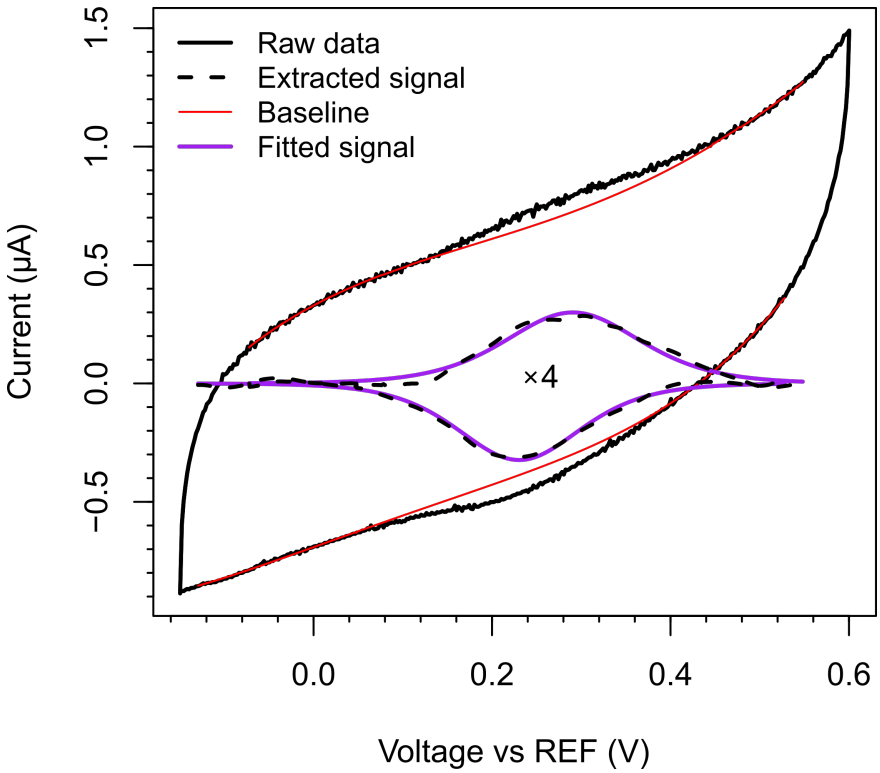


Figure 4.9: DCV current-voltage plot, showing the total DCV current (black), the background subtracted current (dashed), the polynomials used to obtain the background-subtracted current, and the fitted waveshape.

currents (which would appear as 0 degrees in the table, as a 90 degree phase shift is introduced by the cosine function). If the double-layer capacitance is better modelled as a constant phase element, then the phase shift of the capacitance could be anything between 0 and 90 degrees. Additionally, the Faradaic phase shift is not as expected — it would be approximately 4.71 if the Faradaic process could be modelled as a pure resistor. However, even if we use modified EIS components to contextualise the phase shift of the two processes, this still would not capture the fact that the separation between the two phases is correlated, as can be observed in figure 4.12. This represents a further departure from the impedance model of the redox process in appendix D, which assumes independence. This correlated separation is presumably representative of an un-modelled physical process, and as such is a target for potential future research.

It is a useful control exercise to use the parameters presented in table 4.1, based on analysis of PSV data, to predict the outcome of an equivalent r-FTACV experiment. With r-FTACV,

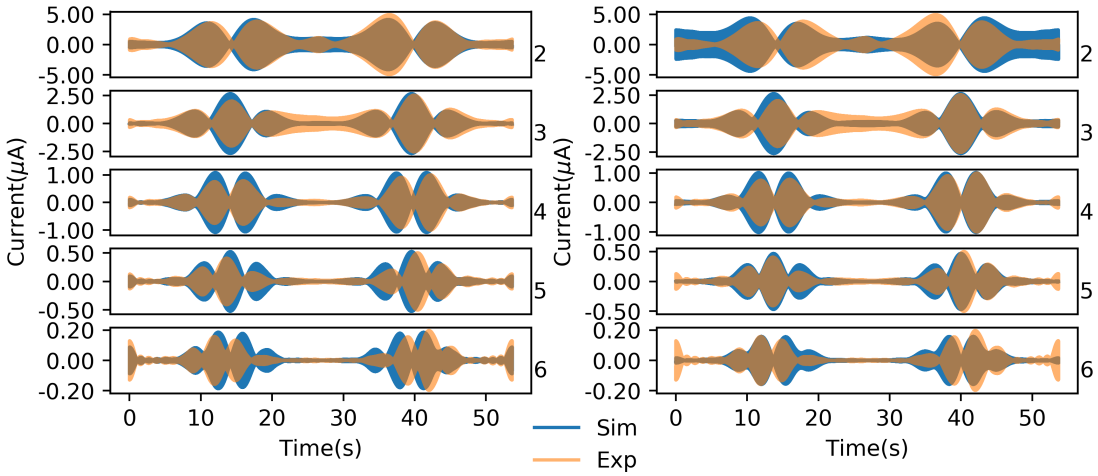


Figure 4.10: Comparison of harmonics 2-6 for r-FTACV experimental data (orange) with the output of simulations (blue). The DC ramp was applied over the potential range of -0.18 to +0.62 V vs Ref at a scan rate of 29.8 mV s^{-1}. The superimposed sine wave had an amplitude of 150 mV and frequency of 8.94 Hz. The simulation parameters used for the left hand side plots were as for experiment 10 in table 4.1, except for ω and phase, which were set at 8.885Hz and 0 rads respectively, as obtained by fitting these parameters for r-FTACV data. The parameter values used to simulate the harmonics on the right hand side are provided in table 4.1. and were inferred by CMAES fitting of the 4-6th harmonics of the experimental data (shown in orange).

the effect of capacitance can be minimised by focusing on the second and higher order harmonics which are essentially devoid of this non-Faradaic contribution [34]. Figure 4.10 provides a comparison of the second to sixth harmonics derived from an r-FTACV experiment with the simulation output generated using the parameters inferred by the CMAES algorithm to describe PSV experiment 10. PSV experiment 10 was chosen for this exercise as it was performed immediately prior to undertaking the r-FTACV experiment. The fit to the r-FTACV data is generally excellent. However, simulated and experimental plots diverge slightly. In the experimental data, the potential of the centre-point of the harmonic [9] (E_{cp}) is slightly dependent on the harmonic number, whilst the simulated E_{cp} is uniform across harmonics. It is possible to more closely replicate the experimental effect within the simulated r-FTACV harmonics by using a distribution of E^0 values that is asymmetric, (not a "true" normal distribution). However, this experimental artefact could also be a consequence of non-idealities within the surface-confined moiety, particularly with regards to the assumption of a Langmuir isotherm, and monolayer coverage. This form of disagreement is detected in the r-FTACV experiment because the data returned by the purely Faradaic higher order harmonics are in a format that is conducive to detecting non-ideality by visual inspection. Departures in modelling surface confined DC cyclic voltammetry based on a

Butler-Volmer model of electron transfer and a Langmuir isotherm are frequently encountered in metalloenzyme studies [48, 50]. However, these subtle nuances do not significantly challenge the fidelity of the parameters recovered from the PSV experiment.

The fact that the Faradaic parameters inferred by CMAES analysis of the PSV experiments also provide a good fit to r-FTACV data indicates that the parameters inferred for the surface-linked ferrocene are all likely to be physically and chemically realistic. There is also acceptably good agreement between the parameters inferred by fitting to the 2-6th harmonics of the r-FTACV experiment and those inferred from the PSV experiment 10 as shown in figure 4.10 and the fitted parameter estimates are included in table 4.1. The time taken to estimate parameters for r-FTACV data was significantly greater, such that the process had to be run overnight. It should also be noted that, for the r-FTACV fits, as the effect of capacitance is confined to the lower harmonics and these harmonics are not included in the objective function, the estimates of the parameters C_{dl}, C_{dlE1} and C_{dlE2} are unlikely to be realistic.

Analysis of the r-FTACV data, in addition to supporting the fidelity of the computation-

Parameter	PSV experiment 10
E^0 (V)	0.241
$\log(k_0(s^{-1}))\sigma$	0.916
$\log(k_0(s^{-1}))\mu$	54.577
R_u (Ω)	544.041
C_{dl} (F)	7.588E-5
C_{dlE1}	2.410E-3
C_{dlE2}	-4.219E-4
Γ ($mol\ cm^{-2}$)	7.062E-11
ω (Hz)	8.941
C_{dl} phase	5.088
Phase	4.359
$\alpha\mu$	0.547
$\alpha\sigma$	0.072

Table 4.4: Parameters inferred by the CMAES algorithm for PSV experiment 10, incorporating kinetic dispersion (where the kinetic parameter was distributed log-normally).

ally much less intensive PSV approach, also reveal why the electrode kinetic parameter only needed to be fitted as a point value and not with dispersion. With the FTACV approach, the broadness of the higher order harmonics in the experimental r-FTACV data is well-mimicked when the modelling included thermodynamic dispersion. In figure 4.11 we show the results of using the CMAES algorithm to fit the data by incorporating kinetic rather than thermodynamic dispersion, where the inferred parameters used to simulate the PSV experiment are in table 4.4. Although the fit to the sinusoidal harmonics is about as good as the fits observed when including thermodynamic rather than kinetic dispersion (i.e. the fits observed in figures 4.2-4.6), the fits to r-FTACV experiment 1 harmonics are not as good. In particular, the simulated harmonics have a larger magnitude and are narrower than the experimental harmonics, especially in harmonics 4-6. In addition, the peak separation in the simulated harmonics is not as defined as in the experimental harmonics. With regards to the first two observations (i.e. the magnitude and width of the higher harmonics), these are both features that are altered by the inclusion of thermodynamic dispersion, as discussed in section 3.3, as this effect broadens and reduces the size of the harmonics, and the effect is more pronounced at higher frequencies. With regards to the resolution, this is a result of the distribution of kinetic values being mostly lower than the point value at which the harmonic peaks fully separate, at approximately 120 s^{-1}.

Interestingly, metalloprotein and metalloenzyme film voltammetry analysis has often required kinetic dispersion [48, 50]. The redox centres in metalloproteins and enzymes are buried within a very large structure and are attached to the electrode surface with different orientations, and hence distances from the electrode surface. In comparison, there should be less variation in the diazonium tethered ferrocenes with respect to orientation and distance from the GC surface. Furthermore, the k_0 value at ~ 145 s^{-1} approaches the reversible limit at the frequency of the experiments undertaken so that minor differences do not affect the response as significantly as would occur for slower rates of electron transfer.

4.3.4 Bayesian inference

A measure of the confidence in parameter values for the PSV experiments, as inferred by the CMAES algorithm, excluding $\alpha\mu$, was obtained by use of MCMC, as detailed in 2.4.2, which was run using three chains for 100000 iterations, with the CMAES parameters used as the initial

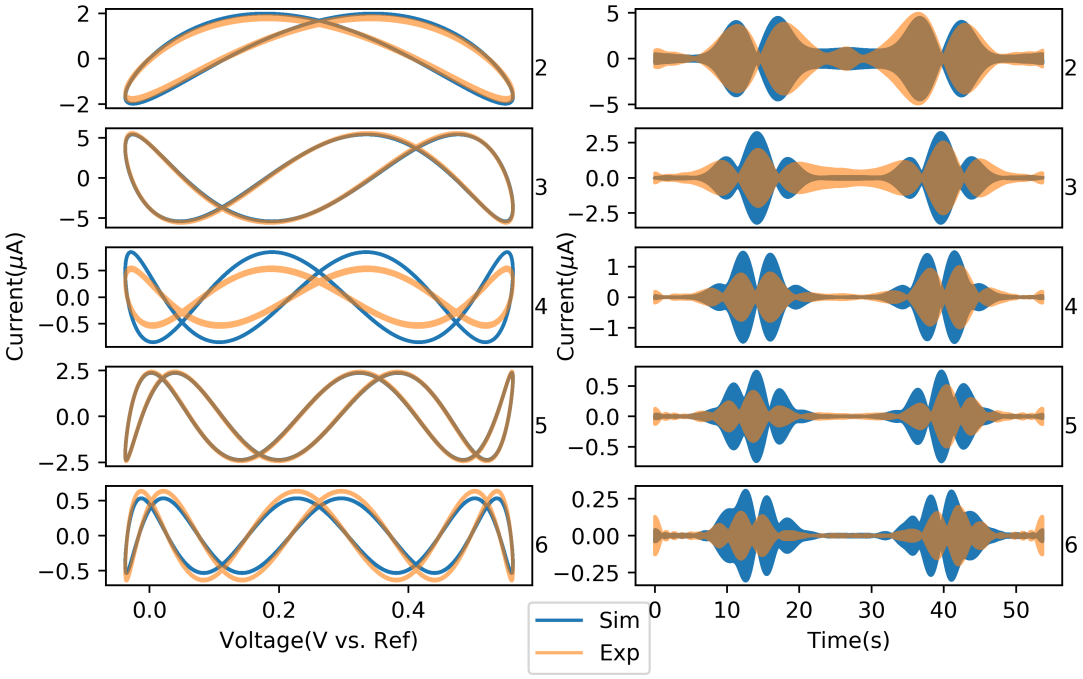


Figure 4.11: Left — current-potential harmonics from PSV experiment 10, plotted against PSV harmonics simulated using the parameters in table 4.4. Right — current-time harmonics from r-FTACV experiment 1 plotted against simulated r-FTACV harmonics simulated using the same parameters (except for the input frequency and phase, which were set to 8.885Hz and 0 rads respectively)

values. The means and standard deviations of the distributions for each parameter for each experiment as inferred by the MCMC algorithm are reported in table 4.5. The results are highly consistent with the parameters inferred by the CMAES algorithm, with the difference between the two values being significantly less than 1 standard deviation (as calculated using MCMC). Although this is expected, as we initialised the MCMC algorithm using these CMAES values, it is nevertheless a confirmation that the inferred parameters represent a good fit to the data. The confidence in the inferred parameters (represented by the standard deviation) is also high, where the average coefficient of variation for the parameters across experiments is 2.1%. The fact that the trends within the parameters discussed above are reproduced here (and that the difference between experiments for individual parameters is usually larger than 1 standard deviation) implies that these trends in parameter values are not due to noise.

Parameter	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
$E^0\mu$ (V vs. Ref)	0.247 (6.730E-5)	0.244 (6.064E-5)	0.244 (6.087E-5)	0.244 (6.061E-5)	0.242 (5.516E-5)
$E^0\sigma$ (V)	0.028 (1.242E-3)	0.047 (4.141E-4)	0.045 (4.551E-4)	0.046 (4.717E-4)	0.047 (3.985E-4)
k_0 (s^{-1})	104.981 (1.594)	144.167 (1.806)	142.427 (1.98)	142.783 (1.973)	140.441 (1.796)
R_u (Ω)	732.531 (14.544)	517.468 (6.193)	540.034 (6.805)	540.692 (7.003)	516.687 (5.908)
C_{dl} (F)	7.688E-5 (5.456E-8)	7.681E-5 (3.248E-8)	7.636E-5 (3.347E-8)	7.609E-5 (3.465E-8)	7.609E-5 (2.930E-8)
C_{dlE1}	3.215E-3 (1.852E-4)	3.297E-3 (1.171E-4)	3.140E-3 (1.225E-4)	2.973E-3 (1.293E-4)	2.740E-3 (1.123E-4)
C_{dlE2}	-3.995E-4 (8.815E-6)	-4.403E-4 (5.692E-6)	-4.267E-4 (5.963E-6)	-4.204E-4 (6.266E-6)	-4.219E-4 (5.469E-6)
Γ (mol cm^{-2})	7.637E-11 (3.982E-13)	7.335E-11 (1.699E-13)	7.331E-11 (1.788E-13)	7.309E-11 (1.866E-13)	7.289E-11 (1.619E-13)
ω (Hz)	8.941 (1.488E-5)	8.941 (1.251E-5)	8.941 (1.272E-5)	8.941 (1.293E-5)	8.941 (1.220E-5)
C_{dl} η	4.392 (4.164E-3)	4.335 (1.780E-3)	4.342 (1.952E-3)	4.342 (1.993E-3)	4.335 (1.680E-3)
η	5.082 (5.836E-3)	4.956 (3.063E-3)	4.969 (3.222E-3)	4.968 (3.387E-3)	4.958 (2.957E-3)
$\alpha\sigma$	0.163 (3.700E-3)	0.187 (3.451E-3)	0.188 (3.481E-3)	0.188 (3.425E-3)	0.19 (3.171E-3)
σ	0.165 (7.871E-4)	0.136 (6.614E-4)	0.139 (6.762E-4)	0.137 (6.513E-4)	0.13 (6.236E-4)
Parameter	Exp 6	Exp 7	Exp 8	Exp 9	Exp 10
$E^0\mu$ (V vs. Ref)	0.241 (5.560E-5)	0.241 (5.554E-5)	0.241 (5.416E-5)	0.24 (5.457E-5)	0.241 (5.628E-5)
$E^0\sigma$ (V)	0.048 (3.616E-4)	0.048 (3.439E-4)	0.048 (3.500E-4)	0.048 (3.288E-4)	0.048 (3.536E-4)
k_0 (s^{-1})	140.034 (1.807)	141.825 (1.841)	136.384 (1.775)	133.434 (1.767)	136.206 (1.824)
R_u (Ω)	495.782 (5.351)	491.194 (5.087)	493.499 (5.185)	488.835 (4.858)	498.912 (5.316)
C_{dl} (F)	7.613E-5 (2.869E-8)	7.601E-5 (2.734E-8)	7.581E-5 (2.768E-8)	7.571E-5 (2.648E-8)	7.537E-5 (2.744E-8)
C_{dlE1}	2.534E-3 (1.016E-4)	2.478E-3 (1.038E-4)	2.392E-3 (1.004E-4)	2.221E-3 (9.558E-5)	2.277E-3 (1.025E-4)
C_{dlE2}	-4.222E-4 (4.964E-6)	-4.228E-4 (5.113E-6)	-4.196E-4 (4.934E-6)	-4.153E-4 (4.716E-6)	-4.106E-4 (4.996E-6)
Γ (mol cm^{-2})	7.271E-11 (1.474E-13)	7.241E-11 (1.421E-13)	7.220E-11 (1.416E-13)	7.205E-11 (1.349E-13)	7.167E-11 (1.460E-13)
ω (Hz)	8.941 (1.198E-5)	8.94 (1.113E-5)	8.941 (1.143E-5)	8.941 (1.092E-5)	8.941 (1.143E-5)
C_{dl} η	4.329 (1.526E-3)	4.328 (1.448E-3)	4.328 (1.470E-3)	4.326 (1.375E-3)	4.328 (1.487E-3)
η	4.95 (2.795E-3)	4.947 (2.655E-3)	4.95 (2.705E-3)	4.949 (2.567E-3)	4.951 (2.776E-3)
$\alpha\sigma$	0.192 (3.058E-3)	0.192 (3.045E-3)	0.191 (2.890E-3)	0.191 (2.923E-3)	0.191 (3.098E-3)
σ	0.125 (5.971E-4)	0.123 (5.891E-4)	0.123 (5.868E-4)	0.12 (5.645E-4)	0.124 (5.855E-4)

Table 4.5: Mean and standard deviation (in brackets) for each distribution inferred during an MCMC run with 100000 samples. The first 50000 samples were excluded as burn-in.

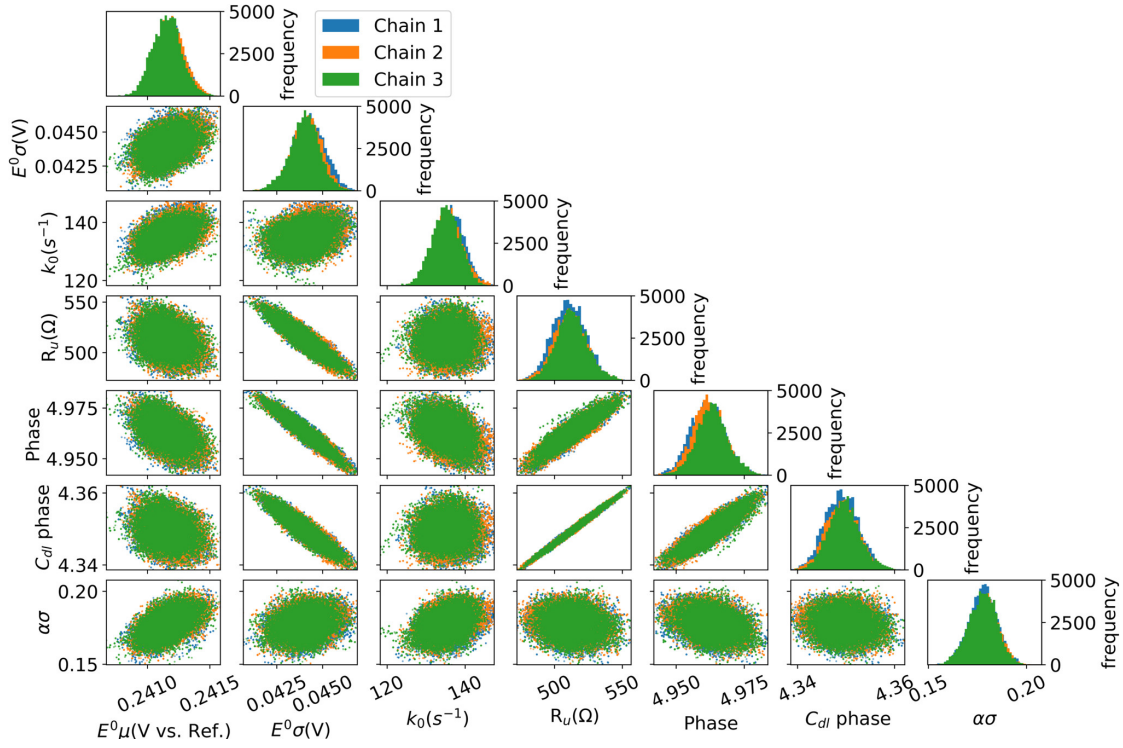


Figure 4.12: 2D scatter plots for individual parameters used to infer parameter distributions from PSV experiment 10, which is plotted in figure 4.6 on the bottom row. The diagonal represents the 1D histograms for each parameter. The MCMC used to generate these chains were run for 100000 samples, with the first 50000 discarded as burn-in.

Two-dimensional histograms of the converged chains are provided in figure 4.12, where all three chains exhibit similar levels of pairwise correlation between individual parameters. Uncorrelated chains appear as a circle, and correlated chains appear as an ellipse; the narrower the ellipse the stronger the correlation. Correlation in this case indicates that the two parameters are compensating for each other. There is a strong positive correlation (i.e. an ellipse angled to the right) between the two phases, which are both also correlated with the resistance. The correlation of the phases is understood in terms of the need to maintain a constant distance between them. In addition, the fact that the uncompensated resistance is negatively correlated with the standard deviation of E^0 may be a consequence that both exert broadening effects on the PSV, as shown in the sensitivity analysis in section 3.2. With regards to the correlation of resistance and phase, in electrical impedance, the phase of the impedance is defined by the relative sizes of the (real) resistive and (imaginary) reactance components, and so we suggest that the observed correlation is a result of this relationship. This would also explain the negative correlation between phase

and thermodynamic standard deviation, through second-order correlations.

When attempting to infer distributions for every parameter using MCMC we found that the inclusion of the $\alpha\mu$ parameter during the fitting process did not result in convergence of multiple chains to a single distribution, although the means were similar. This is illustrated in figure 4.13, which is a scatter plot where the chains for individual parameters are plotted together. It was ultimately determined that, while the chains of each Markov process had similar means, they often exhibited entirely different correlations (as shown in figure 4.13), which implies that "true" distribution is multi-modal within parameter space. The overlapping densities provide an explanation as to why the CMAES algorithm does not appear to be affected by this apparent multi-modality, as the centre of each distribution does overlap, over an area which encompasses the inferred CMAES parameters, which presumably represents the minimum in parameter space. It is not a difference between a least-squares minimisation problem and log-likelihood maximisation problem, as we inferred nearly identical parameters using both approaches for the score function for the CMAES algorithm. We see that individual chains have different correlations, which indicates that they have not converged to the same distribution. When we dispensed with the $\alpha\mu$ parameter, the chains converged to similar sampling distributions.

This whole process again reinforced the benefits of the faster simulation times for PSV — a 100000 sample MCMC run with 2 dispersed parameters (requiring 7,500,000 simulations of the forward problem) required approximately 7 hours. Using r-FTACV, we estimated that repeating this process would require 4 days of computing time. Consequently, determining the impact of the $\alpha\mu$ parameter discussed above would have been on the order of weeks rather than days.

When not fitting the $\alpha\mu$ parameter during the MCMC fitting process, we are assuming total certainty about that parameter, which is not realistic. This may have a knock-on effect of spuriously lowering the uncertainty of the other inferred parameter distributions, as the parameter space that the algorithm has to explore has been reduced by a dimension. However, in this case, it does not appear to affect the width of the inferred distribution. In figure 4.14, we can see that removing the $\alpha\mu$ parameter from the fitting process does not significantly alter the inferred distributions of the other parameters. The MCMC results also establish the confidence in the mode as the inferred distributions in table 4.5 are almost exactly the same as those inferred by the CMAES fitting process in table 4.1.

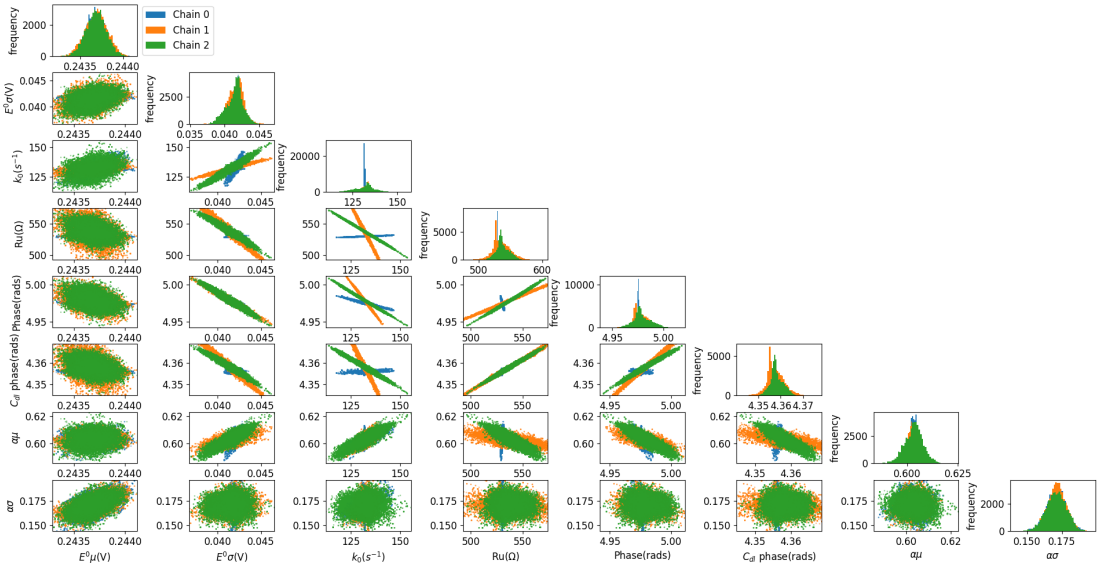


Figure 4.13: 2D scatter plots showing individual Markov chains for each parameter plotted against each other, with their individual histograms being shown on the diagonal. Chains were run for 100000 iterations and the first 50000 of these samples were discarded as burn-in.

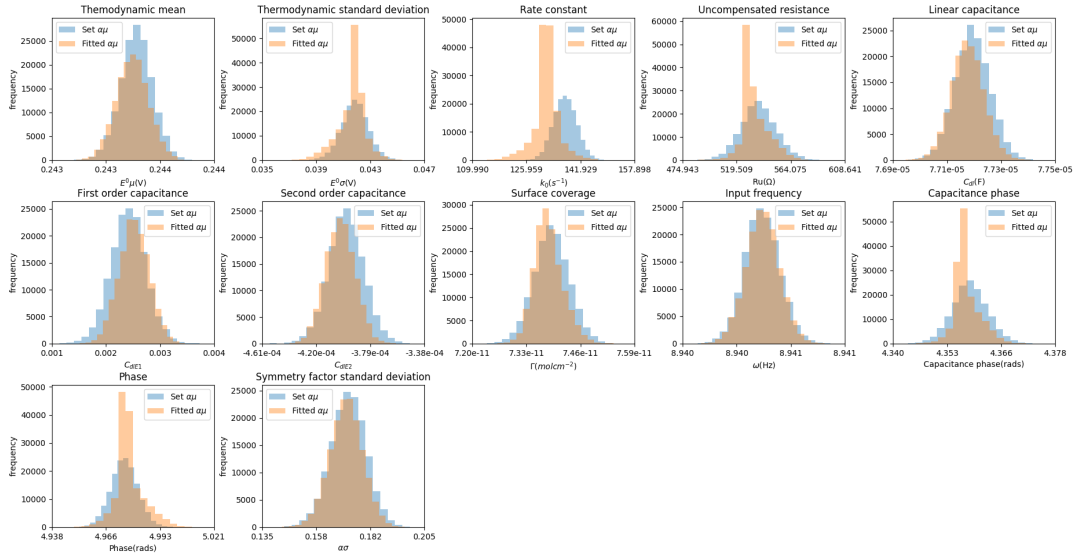


Figure 4.14: Inferred distributions of parameters when $\alpha\mu$ is fitted as a parameter vs when it is fixed at the value inferred by the CMAES.

4.4 Conclusions

The results detailed in this chapter demonstrate that the computing time required to use Bayesian inference to quantify the 13 parameters associated with modelling of the PSV of a surface-linked diazonium based ferrocene derivative is substantially less than the dual time domain technique, r-FTACV. Although (as noted in section 3.2) it is harder to assess the effect of each parameter than for r-FTACV data, parameters that accurately describe the PSV experiment are still accessible, and verifiable by comparison to r-FTACV. The fidelity of the PSV derived parameters was demonstrated by showing that their use in simulations of the r-FTACV experiment provided a very good fit to the data obtained with the same ferrocene modified electrode. Having satisfied ourselves that it is possible to make consistent inferences about an electro-active surface-confined system using PSV, in the following chapter we again use PSV, r-FTACV and DCV to interrogate a surface-confined protein undergoing a single-electron redox reaction.

Chapter 5

Cytochromes

Content in this section has been repurposed from the following reference: Henry Lloyd-Laney, Nicholas Yates, Martin Robinson, Alice Hewson, Alan Bond, Alison Parkin, and David Gavaghan, 2021. A multiple-technique approach to inferring bio-electrochemical reaction parameters ***Chemrxiv****. We are currently in the final stages of preparation before submission to the Journal of the American Chemical Society. All experimental data were collected by Nicholas Yates and Alice Hewson at the University of York. All parameter inference and analysis was undertaken by Henry Lloyd-Laney.*

5.1 Introduction

In the previous chapter, we showed that we could infer Faradaic and capacitive parameters for a simple chemical system using PSV, and that these parameters were also a good fit to r-FTACV data collected from the same electrode. Although there were challenges related to dispersion and non-linear capacitance, the ferrocene voltammetry data had excellent Faradaic-to-background current ratios, allowing us to fit PSV total current. In this chapter, we analyse PSV, r-FTACV and DCV experiments interrogating a simple protein, a type-c cytochrome from *Cellvibrio japonicus*, which undergoes a reversible single-electron redox process. This is a small domain of Cbp2D, a probable activating partner for the lytic polysaccharide monooxygenase (LPMO) produced by *C. japonicus*. This LPMO catalyses the oxidative cleavage of glycosidic bonds in cellulose and such enzymes facilitate industrial biofuel production from biomass [77]. It has recently been shown that *Cj*X183, a type-c cytochrome domain, can transfer electrons to LPMOs via the re-

versible redox reactivity of the heme centre [78], with these authors making use of direct DCV measurements of the Fe3+/2+ redox chemistry of *Cj*X183 can be obtained via adsorption of purified *Cj*X183 onto a pyrolytic graphite edge (PGE) electrode.

*Cj*X183 was selected for the characterisation of PFV PSV parameter inference because it gives comparatively strong signals when adsorbed to a PGE electrode. However, this parameter inference approach was much more challenging than for ferrocene, because the Faradaic-to-background current ratio is smaller. Additionally, as we have demonstrated theoretically in chapter 3 and experimentally in chapter 4, several of the parameters that we fit have similar impacts on the voltammetric current, which results in parameter compensation effects. The end result is that we are typically presented with an inverse problem that has multiple local minima within parameter space. We resolve these issues firstly by using insights obtainable from the three voltammetric experiments to appropriately bound parameter space, and consequently exclude many of the spurious minima, and secondly by using Bayesian inference to analyse and interpret the degree of parameter compensation. We ultimately show that we can infer biological electron-transfer parameters quickly and consistently using a multiple-experiment approach.

5.2 Experimental methods

The expression and purification of *Cj*X183 was based on protocols reported by the Hemsworth group in Leeds [78], with several alterations. *Escherichia coli* BL21(DE3) competent cells were co-transformed with the two plasmids, pCW-*Cj*X183 and pEC86, simultaneously. Prior to inducing with IPTG the cultures were cooled to 16 °C (not 20 °C) and IPTG was added to a final concentration of 1 mM (not 0.4 mM). Post cell harvest, the cells were resuspended in 50 mL of cold lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 10% v/v glycerol). A protease inhibitor tablet (Thermo Scientific) and 20 μL of Benzonase (2.5 U μL^{-1}) were added to the resuspended solution. The solution was placed on ice and sonicated in 30 second bursts with the same amount of cooling time in between until the cells were lysed. The lysed cell solution was centrifuged (30750 g, 4 °C, 30 min) to isolate the supernatant and imidazole added such that that imidazole concentration reached 30 mM. The protein was purified per the established protocol except with buffer A as pH 8.0, 20 mM Tris, 200 mM NaCl, and 30 mM imidazole, and buffer B which is the same except with 300 mM imidazole.

A custom-built electrochemical cell was set-up exactly as previously used for studies of wild type *Cj*X183 [78], inside an anaerobic chamber (< 40 ppm oxygen). A 10 μL aliquot of protein was pipetted onto the working electrode tip and left to adsorb until a film has formed. The cell was maintained at 5 °C (realistically between 4 and 6 °C) and all measurements were performed in pH 7, 50 mM sodium phosphate, 150 mM NaCl buffer.

Three different voltammetric methods were carried out, DCV, r-FTACV and PSV. The DCV experiments were carried out initially after protein absorption and then after r-FTACV and PSV experiments. The experiments were conducted using an Ivium potentiostat and IviumSoft control software. The voltage was cycled between -0.39 V to 0.30 V (vs the reference electrode, which is +0.239 V vs. SHE) for 4 scans with a 5 second equilibration at the start potential, using a scan rate of 30 mV s^{-1} and a potential step of 5 mV. The same method was carried out for the trumpet plot analysis with a scan rate varying from 10 to 30000 mV s^{-1}.

Three successful repeats of r-FTACV were recorded for each protein film. The r-FTACV experiments were conducted using a custom potentiostat developed by collaborators at Monash University and the instrument was controlled using custom software. Each r-FTACV experiment commenced with a 5 second pre-treatment at -345 mV (vs the reference electrode) and an r-FTACV potential input between -345 mV and 255 mV was applied with a scan rate of 22.35 mV s^{-1}, as well as a sinusoidal oscillation with a frequency of 9 Hz and amplitude of 150 mV.

Ten successful repeats of PSV were recorded for each protein film. The PSV experiments were conducted using the same potentiostat as for r-FTACV and the instrument was controlled using custom software. Each PSV experiment commenced with a 5 second pre-treatment at -45 mV (vs the reference electrode). The voltage was cycled between -344 mV to 254 mV with a frequency of approximately 9 Hz and a phase of 270°for 26.8 seconds (equivalent to 268 oscillations).

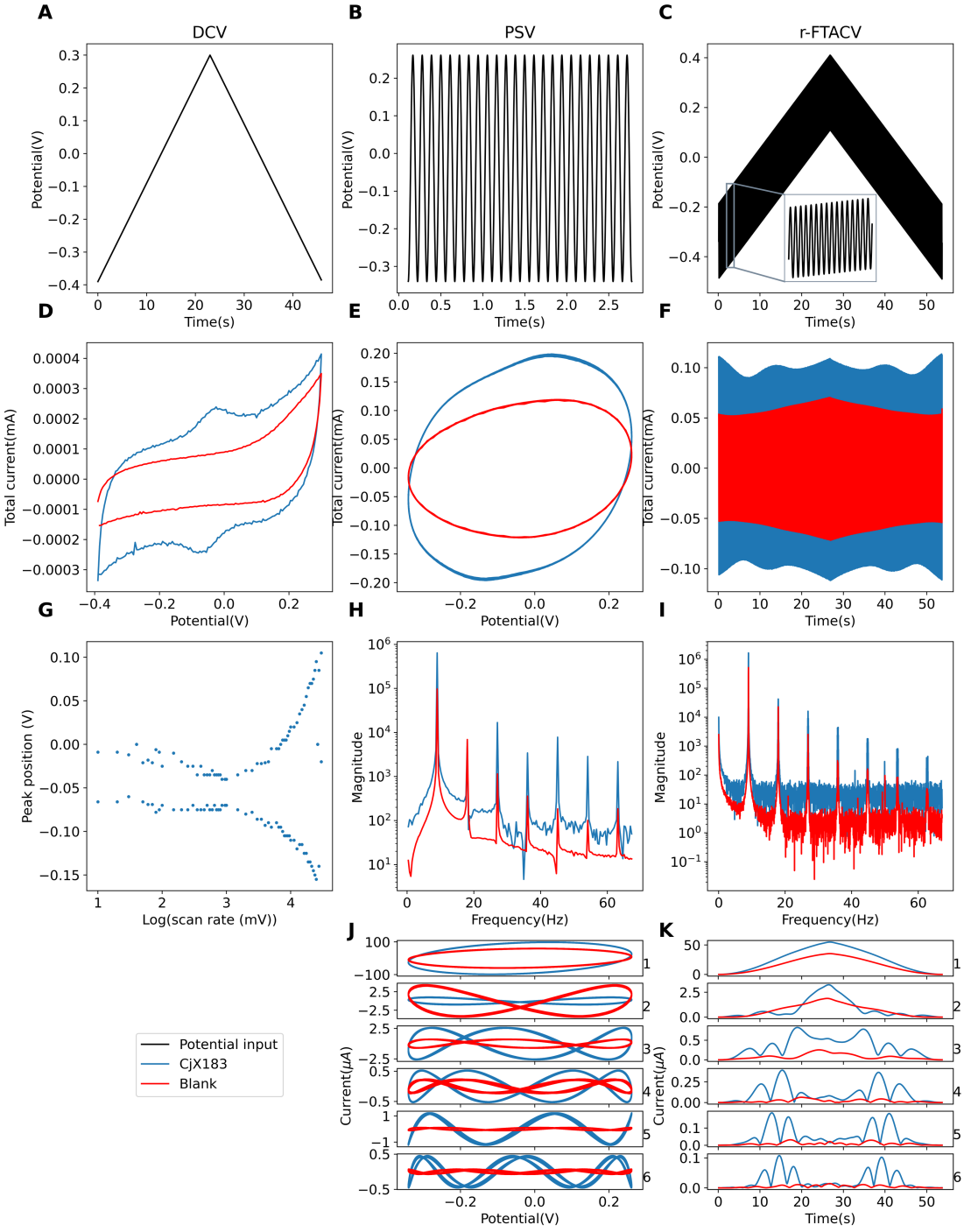


Figure 5.1: DCV (left), PSV (centre) and r-FTACV (right) experimental data. A-C: Potential inputs for the three experiments. D-F: Total current response of *Cj*X183 (blue) and a bare PGE electrode (red) to the three experimental inputs. G: DCV peak position data from *Cj*X183. H and I: magnitude of the Fourier spectrum corresponding to harmonics 1-10 for *Cj*X183 (blue) and a bare electrode (red) for PSV and r-FTACV. J and K: PSV and r-FTACV harmonics 1-6 from *Cj*X183 (blue) and a bare electrode (red)

5.3 Results and discussion

5.3.1 Preliminary investigations

Examples of the experimental data generated by DCV, PSV and r-FTACV voltammetric experiments interrogating *Cj*X183 adsorbed to the surface of a graphite electrode are shown in figure 5.1, with figures 5.1(A-C) showing the potential input and the subsequent figures showing various forms of the current for the three experiments. As can be seen in figures 5.1(D-F), a comparison of experiments conducted in the presence and absence of protein ("blank" and "*Cj*X183") shows that the contribution of the Faradaic current to the total current response is significantly stronger in the DCV and r-FTACV total currents than in the PSV total current. Within the PSV current, to the eye, the signal attributed to the protein (the two "bumps" at $\sim \pm$ 0.18V) is almost totally overwhelmed by background current. One method of voltammetric analysis in the presence of large background currents is background subtraction, where the current from an unmodified electrode (i.e. purely capacitive current) is subtracted from the current of a modified electrode, to obtain a purely Faradaic current response, hypothetically free of non-Faradaic background contributions. However, as shown in figure 5.1, processing the *Cj*X183 experimental data in this way would not yield purely Faradaic current: the presence of protein on the electrode clearly alters the observed background current, meaning the background subtraction would only serve to bias the results.

In chapter 4, where we inferred parameters from a voltammetric PSV experiment interrogating ferrocene attached to a glassy carbon electrode, the Faradaic-to-background current ratio was significantly higher, and consequently we could consistently obtain accurate parameters by fitting to the total PSV current. However, when fitting to the total current generated by *Cj*X183, it was determined that it was possible to obtain multiple "plausible" fits with highly dissimilar parameter values (as recorded in table 5.1), with none of the proposed simulations providing a good fit to either the higher PSV or r-FTACV harmonics. Coupled with the fact that two markedly different sets of Faradaic parameters generate similar total current fits indicate that the inferred Faradaic parameters have little effect on the goodness-of-fit of the two simulations. As can be seen in figure 5.2, the best-fit total-current simulations are primarily a good fit to the first harmonic, which, as can be seen in figure 5.1, is two orders of magnitude larger than any of the higher harmonics. This is a demonstration that, within the total current of the protein-film PSV experiment, the relatively small amount of Faradaic current arising from the biological elec-

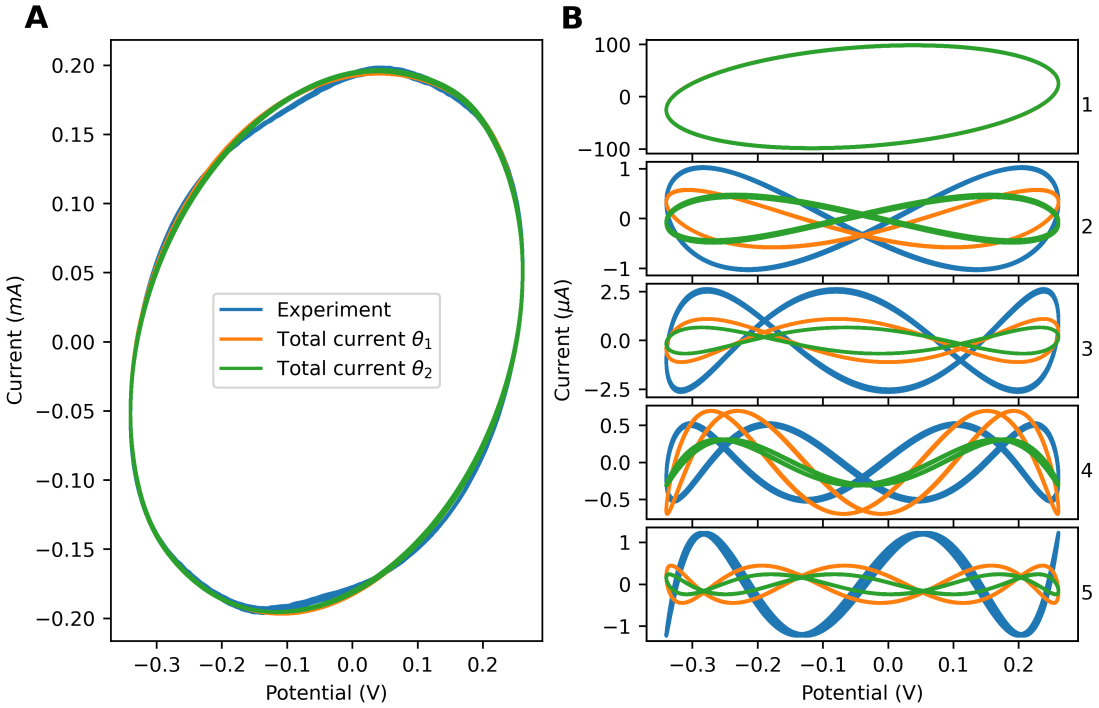


Figure 5.2: A: PSV total current, and two simulations with highly differing parameter values obtained by fitting to PSV total current (referred to as ϕ_1 and ϕ_2, the parameter values of which can be found in table 5.1). When observing the fit in harmonics 1-6 (B), we see that the relatively good fit between simulation and experiment in the time-domain is not replicated in the higher harmonics.

tron transfer process is swamped by the background current contributions, as was not the case for surface-linked ferrocene. Additionally, background currents arising from a PGE electrode are generally more complex than for a glassy carbon electrode, because of the roughness of the surface.

Given that fitting the total PSV current in the time domain did not provide a satisfactory result, we turned to filtering approaches. By inspection of the harmonics for the "blank" (protein-free) experiment in figure 5.1, we judged that PSV harmonics 0-3 contained most of the capacitive current contribution, and filtered this out by band selecting harmonics 4 and above from the Fourier spectrum of the *Cj*X183 PSV current. We note that this filtered data contains little information about capacitance — a detail we will return to in section 5.3.3.2.

It was also determined that we should include the effects of thermodynamic dispersion in the

Parameter	Symbol	ϕ_1	ϕ_2
Midpoint potential	E^0 (V vs. ref.)	-6.945E-2	-4.147E-2
Rate constant	k_0 (s^{-1})	54.1	2651.3
Uncompensated resistance	R_u (Ω)	1344.4	1188.13
Linear double-layer capacitance	C_{dl} (F)	4.00E-4	4.45E-4
1^{st} order C_{dl}	C_{dlE1}	0.001	-0.017
2^{nd} order C_{dl}	C_{dlE2}	-0.013	-0.002
3^{rd} order C_{dl}	C_{dlE3}	-0.002	5.98E-4
Surface coverage	Γ (mol cm^{-2})	9.0E-11	5.45E-11
Potential frequency	ω (Hz)	9.015	9.016
C_{dl} η	C_{dl} η	5.61	5.65
Phase	Phase	5.15	4.97
Symmetry factor	α	0.5	0.502

Table 5.1: Simulation parameters for the two fits shown in figure 5.2 obtained by fitting to total PSV current.

model, from inspection of the r-FTACV experimental harmonics, as described in section 3.3. This can be achieved by comparing the r-FTACV harmonics against a non-dispersed simulation. Such a simulation can be obtained by fitting to PSV data, and then using the best-fit parameters to obtain an r-FTACV simulation — if the higher r-FTACV experimental harmonics are of a lower magnitude and are wider than the non-dispersed simulation, then this is an indication that thermodynamic dispersion is present. The presence of thermodynamic dispersion also implies the presence of kinetic dispersion, but as detailed in section 3.3, in the close-to-reversible kinetic regime, kinetic dispersion does not have a significant effect on the Faradaic current, and so we did not model it for reasons of computational efficiency.

After an initial parameter inference attempt, fitting to both real and imaginary harmonics 4-10 of the *Cj*X183 current, we determined that it was possible to obtain parameter vectors that provided a reasonably good fit to these PSV harmonics. However, when compared to r-FTACV harmonics data as shown in figure 5.3, the fit is poor — with a mismatch between the predicted and experimental harmonic magnitudes, and implausible waveshapes — the parameters used to generate the simulation can be found in table 5.2 under "low k_0". The poor fit is particu-

Parameter	Symbol	Low k_0	High C_{dl}
Midpoint potential mean	E^0_μ (V vs. ref.)	-7.000E-2	-7.900E-2
Midpoint potential standard deviation	$E^0\sigma$ (V)	0.055	0.021
Rate constant	k_0 (s^{-1})	20.42	63.02
Uncompensated resistance	R_u (Ω)	361.219	112.807
Linear double-layer capacitance	C_{dl} (F)	5.089E-4	7.989E-4
1^{st} order C_{dl}	C_{dlE1}	-5.507E-2	-8.716E-3
2^{nd} order C_{dl}	C_{dlE2}	2.265E-3	1.265E-3
3^{rd} order C_{dl}	C_{dlE3}	0	-2.602E-5
Surface coverage	Γ (mol cm^{-2})	7.263E-11	2.968E-11
Potential frequency	ω (Hz)	9.015	9.015
C_{dl} η	C_{dl} η	1.571	5.495
Phase	Phase	5.247	5.193
Symmetry factor	α	0.484	0.6

Table 5.2: Simulation parameters for simulations that were a good fit to higher PSV harmonics whilst being physically unrealistic, as in figures 5.3 and 5.5, for the "low k_0" and "high C_{dl}" cases respectively.

larly apparent in r-FTACV harmonic 4, where one single merged set of peaks is visible in the simulation while two separately resolved sets of peaks are observed in the experimental data. Based on knowledge of how k_0 affects r-FTACV harmonic waveshape (detailed in section 3.2), this type of deviation between simulation and experimental r-FTACV data is attributable to the simulation being generated by a parameter set in which the electron-transfer rate k_0 is too small. We concluded that exploring a very broad region of parameter space during the modelling of PF-PSV high harmonic data is not an effective way to obtain realistic electrochemical reaction parameters. This is a direct illustration of the phenomenon of parameter compensation. In these circumstances, it is often the case that the optimisation algorithm returns a parameter set corresponding to a local (rather than the global) minimum in parameter space; in the example shown in figure 5.3 this is driven by the kinetic parameter.

The next section therefore details how we can avoid this problem by using prior information and information from multiple experimental techniques to constrain the search space and return a consistent and reproducible global optimum.

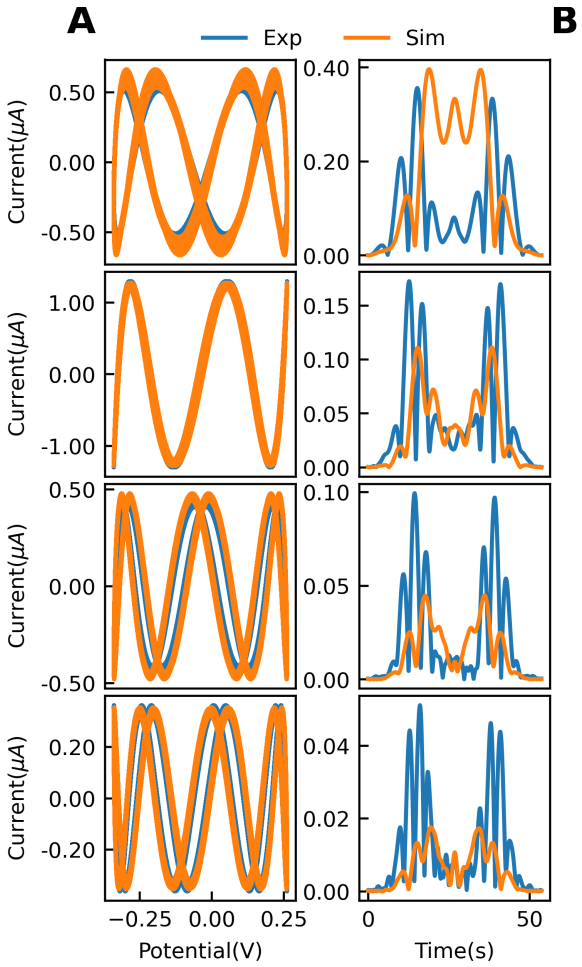


Figure 5.3: An example of how a relatively good agreement between PSV experimental (blue) and simulated (orange) harmonics 4-7 (A) does not translate well to an r-FTACV harmonic comparison (B), driven by a low inferred k_0 value.

5.3.2 Constraining the search space

From the preliminary investigation above, it became apparent that an important part of obtaining a robust fit is to constrain the parameter space in which the minimisation algorithm searches. We constrain parameter space to prevent physically implausible parameter values, aiming to exclude the areas of space that resulted in good fits only to the filtered portion of the current from one technique, but failed to generate a good-fit to data from other techniques. This is a process that must be undertaken with great caution; it carries the obvious danger of excluding the region of parameter space in which the "true" parameter values reside. The bound we used for each parameter is shown in table 5.3, and in figure 5.7, we summarise this process graphically. We

will now detail how we obtained the bounds for each parameter in turn.

5.3.2.1 Midpoint potential mean — E_μ^0

The boundaries applied to the midpoint potential mean E_μ^0 are selected from inspection of the potential position of the harmonic peaks from r-FTACV experiments. This can be achieved by plotting the individual r-FTACV harmonics against the linear ramp portion of the r-FTACV potential input and noting the peak potential at approximately -0.065 V; this is shown in figure 5.4, and the bounds were selected to be within 60 mV of this value. If there was no thermodynamic dispersion present in the system, then the E^0 parameter would be bounded in the same way.

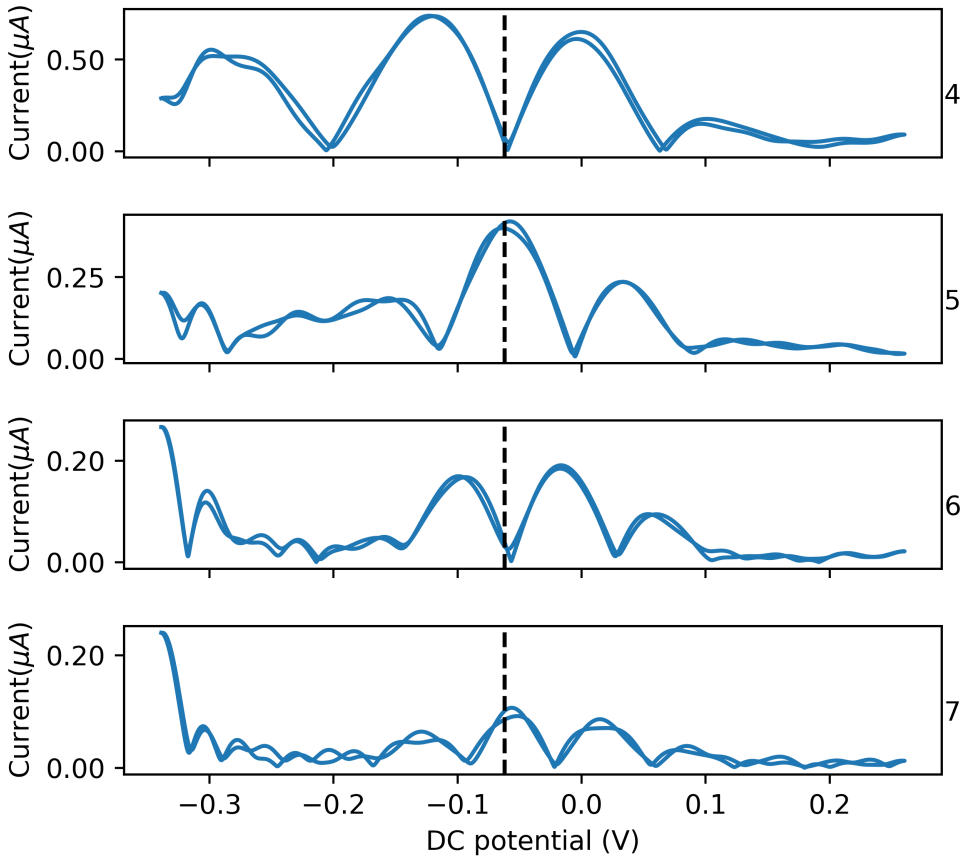


Figure 5.4: R-FTACV harmonics, as shown in figure 5.1, plotted against the linear portion of the r-FTACV potential input. The approximate mid-point of the harmonics can be read off the x-axis (in this case, we have drawn a line at -0.065V), and used to define E_μ^0 boundaries

5.3.2.2 Midpoint potential standard deviation — E^0_σ

For the midpoint potential standard deviation, the lower bound at 1e-4 V is the value at which there is no noticeable difference between currents simulated in the presence or absence of thermodynamic dispersion. The maximum value of 0.06V is the result of a modelling choice — we approximate a distribution of E^0 values using a defined number of bins. At values of the standard deviation above this upper bound, we observe aliasing in the simulated current. If we had a meaningful suspicion of a larger standard deviation (through observation of the width and magnitude of the experimental r-FTACV harmonics, as detailed in section 3.3) then we would have increased the number of bins so we could accurately model these higher values. Additionally, we suggest that values of E^0_σ that predict a distribution of E^0 values that is greater than 50% of the experimental potential window is chemically implausible.

5.3.2.3 Rate constant — k_0

Our initial method for determining the appropriate bounding for the kinetic parameter was based on analysis of the waveshape of the r-FTACV harmonics. The "classical" appearance of r-FTACV harmonics is the one usually presented when introducing r-FTACV [34] ; the harmonics are symmetrical about E^0, each lobe is clearly identifiable, and the number of lobes matches the harmonic number. The classical appearance is observed when the redox process is approaching the reversible limit (where the redox system is at equilibrium on the timescale of the potential change) [10], which in a 10 Hz measurement requires a k_0 of at least 500s^{-1}. The further the rate-constant decreases away from this limit, the less well-defined the individual lobes of the higher harmonics become. However, evaluating the range of likely kinetic values via the visual inspection of r-FTACV harmonics requires extensive familiarity with the appearance of these signals under various conditions, and consequently our suggestion for determining the approximate kinetic regime is instead to use analysis of DCV data in the form of a trumpet plot, as pioneered by Laviron [62]. This approach uses the fact that at higher experimental scan rates the oxidative/reductive peak potentials, measured for a non-catalytic electron-transfer process, diverge in a manner that is characteristic of the electron-transfer kinetics. In figure 5.1(G) we show trumpet plot data obtained for *Cj*X183. To generate the trumpet plots, the current, obtained from DCV experiments performed at different scan rates, was background-subtracted, as shown in figure 5.16, and the position within the potential window of the largest current magnitude for the reductive and oxidative peaks was extracted (from here onwards referred to as "peak posi-

tion"). To obtain peak-potentials as a function of chemical parameters, purely Faradaic DCV current (i.e. with the value of C_{dl} set to zero) was simulated at every experimental scan rate and the peak positions extracted. Because we were only interested in peak position, the only relevant simulation parameters were E^0 (modelled as a point value, as dispersion does not affect peak potential, as shown in 3.3), k_0 and a separation parameter to account for the non-zero difference in oxidative and reductive peak position observed even at very slow scan rates. This is discussed in more detail in section 5.3.5.

Using MCMC, we then determined the distribution of the most likely values of these parameters given the experimental data, where the resultant best-fit is shown in figure 5.18. From this effort, the most likely k_0 value that described the trumpet plot data was determined to be 109 s^{-1}, which can be used to inform the kinetic boundary values, by suggesting the order of magnitude. We ultimately settled on a lower bound of 50 s^{-1} and an upper bound of 500 s^{-1}. The lower bound was determined by observation of r-FTACV waveshapes in forward problem simulations (i.e. the point at which harmonic lobes started to merge, which is not observed in the r-FTACV harmonics in figure 5.1), and the upper bound by the knowledge that this was the approximate point of the reversible limit at 10 Hz. As before, if after multiple fitting runs the k_0 value returned is at the boundary value, then that boundary can be raised or lowered accordingly. If the kinetics are truly reversible at the experimental frequency, as previously [10], then it is likely that the kinetic value returned will be the value of the upper bound, regardless of where it is set.

5.3.2.4 Surface coverage — Γ

With regards to the surface coverage parameter, we estimated this value by finding the area (or charge) under the reductive and oxidative curves from a single scan-rate background-subtracted DC voltammogram, as in [6]. Although the value of the estimate is sensitive to the method of background subtraction, as we show in figure 5.17, the inferred value was consistently within 1.5 to 3.5e-11 mol cm^{-2}, and so we set the surface coverage bounds accordingly.

5.3.2.5 Linear double layer capacitance — C_{dl}

In terms of bounding C_{dl}, we must first discuss our modelling approach for capacitance. Our solution was ultimately to vary both capacitive and Faradaic parameters while fitting to harmonics 4-10. This is in comparison to previous work on biological parameter inference using r-FTACV,

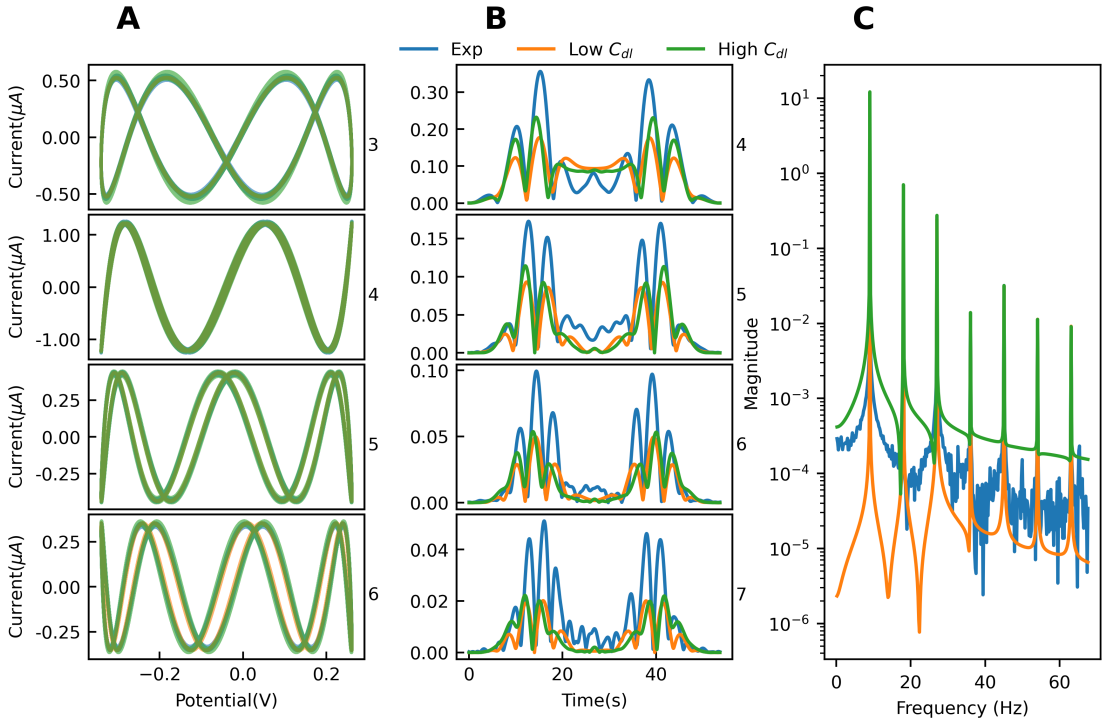


Figure 5.5: Comparison of two plausible parameter vectors generated by fitting to PSV harmonics 4 and above. In A, we see the very good agreement between simulated and experimental PSV harmonics 4-7 for both parameter vectors. Neither vector is clearly better than the other when compared against r-FTACV experimental harmonics in B. The "high C_{dl}" parameter was discounted on the evidence figure C, as there is a significant, unrealistic quantity of spectral bleed from the fundamental harmonic of the "high C_{dl}" simulation.

which uses a "two-step" approach to fitting, first finding best-fit capacitance parameters, by fitting to the total current with zero Faradaic current, then finding the Faradaic parameters which give the best fit to the high harmonic data [5], with the capacitance parameters set to the values inferred in the previous step.

However, for analysis of the Cj1X183 PSV data, we did not find the two-step process satisfactory, because the parameter values were not consistent between different experiments, in particular for the k_0 and E^0 parameters, which we would expect to be consistent as they report on the fundamental electrochemical properties of the protein; we discuss this further in the section 5.3.3.2. Additionally, the value of the surface coverage was consistently overestimated relative to the values inferred from DCV analysis for all three fitted PSV experiments, shown in figure 5.10. However, the choice to vary both capacitive and Faradaic parameters when fitting to harmonics

4-10 introduced additional difficulties. In figure 5.5 we show two simulations that provided good fits to PSV harmonics and r-FTACV harmonics 4-7. Most of the inferred parameters for the two simulations (in table 5.5) are reasonably similar, except the inferred C_{dl} parameter, which differs by an entire order of magnitude — ∼1e-5 F in the "low C_{dl}" case, and ∼1e-4 F in the "high C_{dl}" case. The reason for this discrepancy can be seen when we plot the absolute Fourier transform, as on the right-hand plot in figure 5.5. The magnitude of the first harmonic is so large in the high C_{dl} case that it raises the baseline amplitude of all the higher harmonics, up to number 10. We can see that this effect (also referred to as "spectral bleeding") is somewhat present in the experimental data, but not on the same order of magnitude as for the high C_{dl} simulation. This is a dilemma introduced by filtering; inclusion of the low harmonics means that the solver only fits to that part of the frequency domain (such as is observed in figure 5.2). However, when they are excluded, without appropriate bounds it is possible to explore regions of unrealistic parameter space in order to obtain a better fit to the higher harmonics. Consequently, we set the upper bound for capacitance at 1e-5 F, a level at which we could be certain (through inspection of the Fourier spectrum) that spectral bleeding effects from the fundamental harmonic were not biasing the inferred parameters in the manner of the high C_{dl} case.

These observations contributed to our decision to vary the capacitance parameters while only fitting to the higher harmonics — because some level of spectral bleed is observed in the actual data, the degree of capacitance does have some effect on the higher harmonics, and explicitly varying both the Faradaic and capacitive parameters allows us to capture this effect in our modelling. Additionally, as we shall see below, the MCMC effort has shown us that there was some parameter compensation between capacitive and Faradaic parameters.

5.3.2.6 Nonlinear double layer capacitance — C_{dlEX}

Although similar considerations with regards to spectral bleeding could be made for setting the bounds for the higher order C_{dlEX} terms, we used the ones provided in table 5.3 for reasons of numerical stability, as certain parameter C_{dlEX} combinations outside of these bounds caused our forwards-problem solving programs to crash. These values were selected to ensure that we never observed solver errors, and consequently will be dependent on the local implementation of the solving program.

5.3.2.7 Uncompensated resistance — R_u

With regards to the uncompensated resistance, we found that when we simulated PSV high harmonic data, the "best-fit" values returned were always between the upper or lower bounds initially set for R_u in table 5.3, which were informed by the previous parameter inference effort on surface-linked ferrocene in chapter 4. This is a valid approach because in both the ferrocene study and the *Cj*X183 work, an aqueous solvent containing a total of 200 mM electrolyte salts was used within a 3-electrode setup comprising a carbon working, platinum counter and commercial reference electrodes. We will further discuss the impact of our capacitive and resistive modelling choices in a later section.

5.3.2.8 Potential frequency and phase — ω, η and $C_{dl}\eta$

The frequency of the potential-time sinewave (ω) was experimentally defined at 9.015 Hz in PSV experiments, and so the bounds were set to be close ($\pm$ 5%) of this value. The phase of the input potential sinusoid was set (at $\frac{3\pi}{2}$), we know from section 4.3.2.2 that the inferred phase for the current can differ from the experimentally set value of the potential, in a manner in keeping with the theory of electrochemical impedance. In order to assess the likely magnitude of this phase shift, we determined the phase shift between the current and potential waves. For a pure capacitor, the phase shift observed relative to a purely sinusoidal potential should be π radians, but we determined (as can be seen in figure 5.6) that this shift was approximately $\frac{5\pi}{6}$ radians, and set the phase boundaries for both the capacitance and Faradaic phase accordingly.

5.3.2.9 Symmetry factor — α

Finally, with regards to α, we considered values outside of the 0.4-0.6 range to be chemically implausible, and consequently excluded them.

5.3.2.10 Conclusions on constraining the search space

As can be gleaned from figure 5.7, the process of bounding parameter space required many fitting runs to gain a deeper understanding of the inference problem. This is a situation in which the speed advantage of PSV becomes relevant. A single fitting run, returning a best-fit parameter set for PSV took approximately 45 minutes, while the same attempt with r-FTACV took several

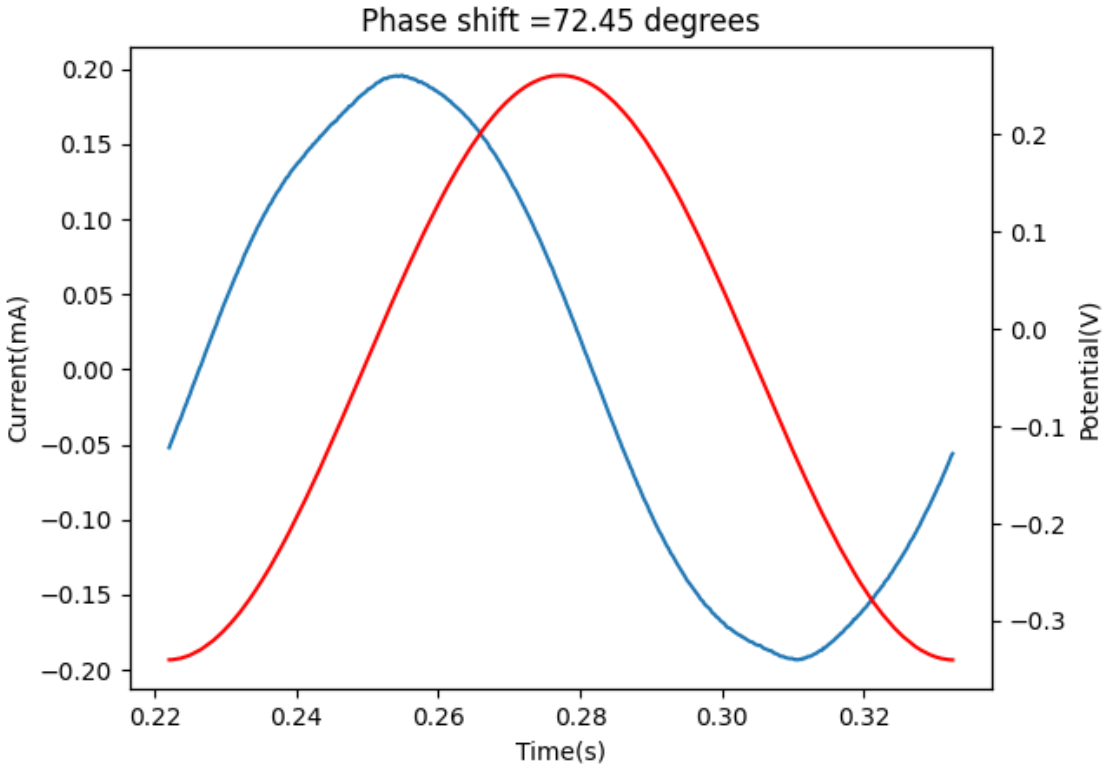


Figure 5.6: Single oscillation of PSV current and potential on the same time axis. The phase difference can be determined by taking the time-difference between the peaks of the two waves. As the capacitive current is proportional to the derivative of the potential, an ideal current will be shifted by 90 degrees.

hours. Therefore, using PSV analysis, multiple model combinations can be tried in the time it takes to undertake one r-FTACV fitting attempt. As we repeat the fitting process ten times to ensure a high search coverage of parameter space, r-FTACV fitting timescales quickly become untenable.

5.3.3 Inferring parameters from multiple films

Using the bounds shown in table 5.3, we were able to infer parameters from three PSV experiments interrogating different protein preparations of *Cj*X183; for each preparation we obtained both PSV and r-FTACV measurements. The resulting best fits are shown in figure 5.8, with the inferred parameters shown in table 5.3, where each column shows data from a different electrode film. In figures 5.8(A-C), we show PSV harmonics 4-10 for both the experimental and simulated

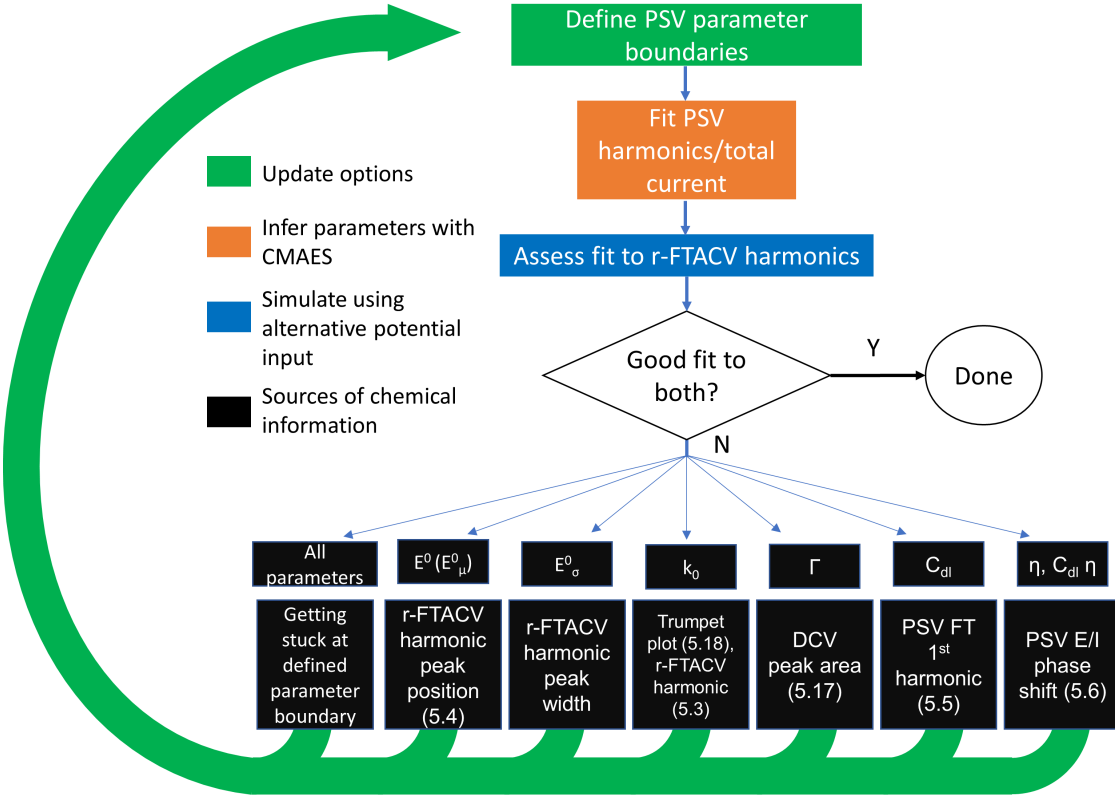


Figure 5.7: Flowchart representation of the fitting methodology proposed in this chapter. For certain parameters we have presented graphically how the boundary values can be obtained; these figures are written in brackets as appropriate.

currents, with the simulations using the parameters in table 5.3. In figures 5.8(D-F), r-FTACV simulations also use the input simulation parameters written in table 5.3, except for the input frequency and phase. For figures 5.8(G-I), we performed a limited optimisation on the absolute Fourier spectrum corresponding to harmonics 4-10, optimising the value of parameters thought likely to change as a result of so-called "film-loss" effects, which change as a function of CjX183 molecules coming off the electrode over time, specifically the E^0_μ, E^0_σ and Γ parameters. In addition, we fitted the sinusoidal frequency ω, which is different for r-FTACV and PSV experiments, and set the phase to zero. The altered values are written in brackets in table 5.3. We do this to show that only a relatively small modification in the PSV parameters is required to go from an excellent fit to the PSV harmonics to an excellent fit to the r-FTACV harmonics. Because the r-FTACV inference approach requires searching in four-dimensional parameter space, the time for a single fitting run is shorter than it would be for the full 13-dimensions searched in the PSV case. In practice, this was 2-3 hours for an r-FTACV harmonic data fitting attempt, as opposed

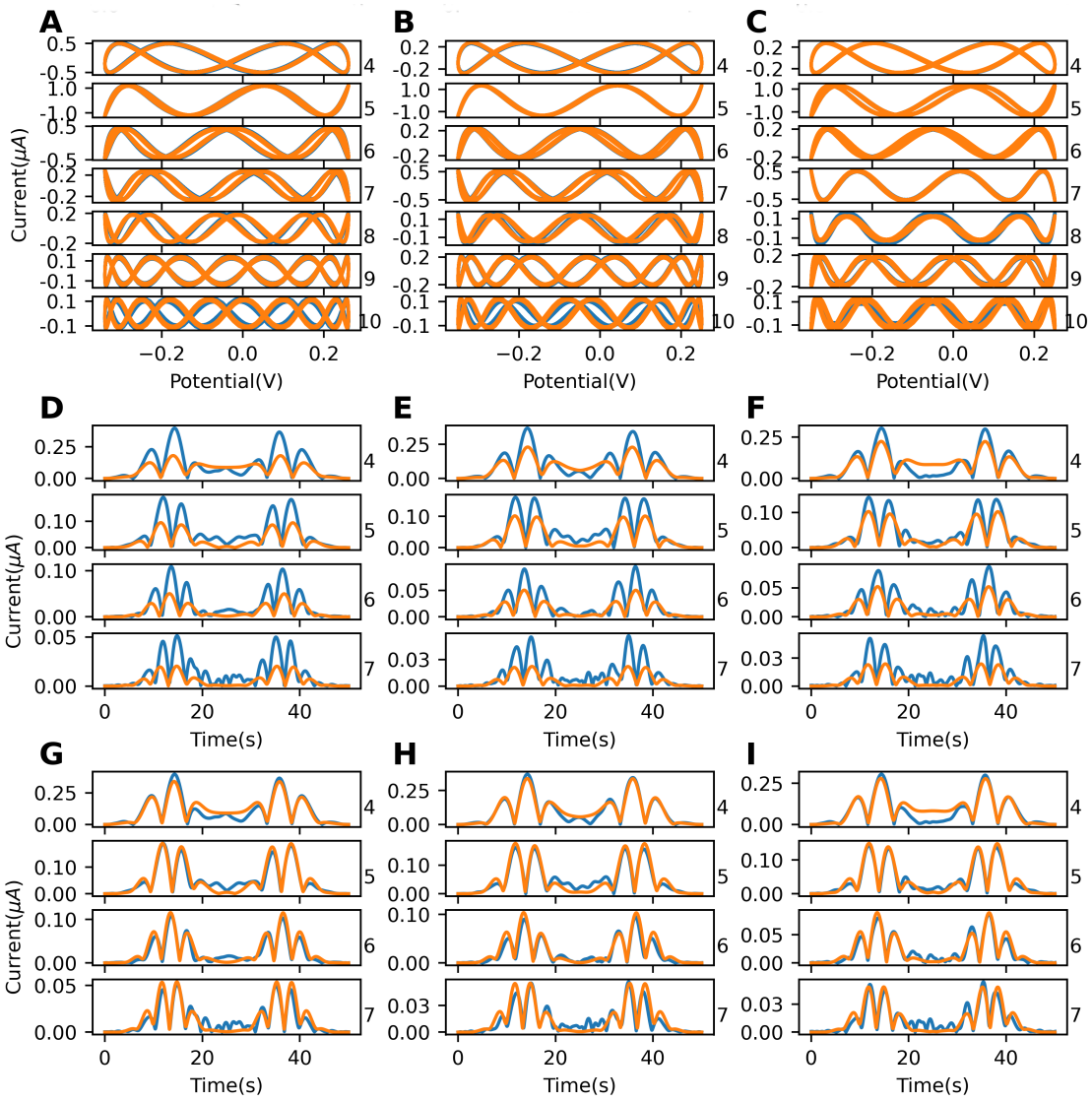


Figure 5.8: Best-fit simulations and data for harmonics 4-10 of PSV experiments 1-3 (figures A-C), using simulation parameters reported in table 5.3. Figures D-F show r-FTACV harmonics 4-7, from r-FTACV experiments 1-3 (performed using the same *Cj*X183 modified electrode as the appropriate PSV experiment), and simulations obtained using the same parameters as those used to generate the top row, except for the values of the phases (both of which were set to 0) and the input frequency (which was set to 8.96, 8.75 and 8.83 Hz respectively). Figures G-I show the same r-FTACV harmonic data as figures D-F, along with best-fit simulations, obtained using the parameters in brackets in table 5.3.

to a predicted 10+ hours for all 13 dimensions.

The multiple comparison approach for validating our "best-fit" reaction model parameters uses

both inter-technique and inter-experiment comparisons, such as the type performed in figure 5.8, which allows the removal of spurious parameter combinations, or those that provide a good fit to only one subset of the data. Consequently, the fact that the values we infer from PSV data can a) with a small level of alteration provide a good fit to r-FTACV harmonics obtained during the same set of measurements and are b) consistent across experiments conducted on different days, demonstrate that the inferred values are robust. We can also evaluate the PSV parameters in light of values obtained by other analysis methods. In this study, these methods are the estimates of the kinetic parameter from trumpet plot data, and surface coverage estimates from analysis of a single scan-rate DCV experiment peak charges, shown in figures 5.18 and 5.17 respectively. With regards to the former, the value of the kinetic parameter inferred from the trumpet plot was within 60 s^{-1} of the value inferred from the PSV currents, and as we show in figure 5.18, simulating trumpet plot data with the PSV-inferred value of $\sim$172s^{-1} does not significantly alter peak-position divergence point. With regards to the surface coverage, although the exact value inferred from the peak-integration technique is dependent on the subtraction approach, the values inferred from PSV are in agreement with the range of values inferred by this method.

In terms of the values that best report on the *Cj*X183 redox chemistry, because r-FTACV is more sensitive to parameter changes as demonstrated using Bayesian inference on synthetic noisy data in section 3.4, we believe that the values that generate the best-fit to the r-FTACV harmonics are those that best describe the properties of *Cj*X183

The midpoint potential of a cytochrome's heme Fe2+/Fe3+ redox couple can vary from -639 to +161 mV (vs. ref.) [79]. However, in the region of physiological pH, the values for E^0 tend to range from around -49 to +11 mV (vs. ref.) [80, 81]. The average value extracted for the midpoint potential mean and standard deviation as part of the best fit parameters was $\sim$-65 mV vs reference, and as such the inferred distribution of *Cj*X183 midpoint potentials is close to or within the suggested physiological range.

When rationalising the electron transfer rate constant measured in this chapter, it is important to consider that the *Cj*X183 protein under analysis is an artificially small fragment of a larger naturally occurring protein. However, the average best fit rate constant of $\sim$175 s^{-1} is fully consistent with the range of values determined in previous electrochemical studies of different cytochrome proteins. For example, the electron transfer rate constant for cytochrome-c

incorporated onto a calcium alginate film on a glassy carbon electrode was found to be 20.9 s^{-1} [82], the electron transfer rate constant for human, bovine and porcine cytochrome P450c17 on a PGE electrode were determined as 164, 157 and 153 s^{-1} respectively [83], and while more varied due to changing chain length, the electron transfer rate constant for rat heart cytochrome-c adsorbed onto gold electrodes modified with self-assembled monolayer has been determined to fall around 700 s^{-1} when the length of the carbon chain on the monolayer is between 4 and 7 carbon atoms in length [84].

Based on the geometric surface area of the working electrode (0.03 cm^2) and the width of the protein structure, by assuming spherical close packing of protein on the electrode, a theoretical maximum monolayer surface coverage of *Cj*X183 on the electrode can be calculated as 4.401 pmol cm^{-2}. The extracted best fit surface coverage parameter from the experimental data is 17.3 pmol cm^{-2}, this difference can be attributed to a non-spherical protein with better packing efficiency and a non-planar electrode surface (i.e. area greatly exceeding 0.03 cm^2 due to abrasive treatment of the graphite creating a rough surface). In general, protein film voltammetry experiments report pmol cm^{-2} coverage values [39]. Again, the constant separation between the η and C_{dl} η parameters noted in section 4.3.3 is also observed here, reinforcing the impression that this separation relates to an un-modelled physical process. What it is essential to note, however, is that the exact reported parameters are conditional on the modelling choices made. Two parameters in particular, α and C_{dl}, return the value of the upper bound. We shall discuss each of these in turn.

5.3.3.1 The effect of α

In all cases the reported α value is the value of the upper bound, and the solver converged to the value of the upper bound regardless of what this boundary was set to. This is not unexpected — as also noted in section 4.3.2, when the kinetic regime is approaching reversibility (reaction is approaching equilibrium on the timescale of the experiment), the effect of the symmetry factor is low. In this case, we did not model dispersion in the α parameter, as doing so did not stop $\alpha\mu$ from hitting whichever upper bound was set, which was not the case in section 4.3.2.1. We set the upper bound at 0.6 for reasons of chemical plausibility, but it is common practice to set the value of α to 0.5 under these conditions (i.e. not to fit α at all). We can show the impact of this modelling choice on the inferred parameters by progressively lowering the value of the α upper bound to 0.5. In table 5.4 we report best-fit parameters for the differing upper bounds, and the

Parameter	Symbol	Bounds	PSV 1	PSV 2	PSV 3
Midpoint potential mean	E^0_μ (V vs. ref.)	[-0.1, -0.04]	-0.072 (-0.061)	-0.067 (-0.063)	-0.065 (-0.061)
Midpoint potential s.d.	E^0_σ (V)	[1e-4, 0.06]	0.045 (0.033)	0.053 (0.036)	0.051 (0.035)
Rate constant	k_0 (s^{-1})	[50, 500]	173.8	176.5	172.9
Surface coverage	Γ (mol cm^{-2})	[9e-12, 9e-11]	1.35e-11 (1.68e-11)	2.05e-11 (1.83e-11)	1.79e-11 (1.45e-11)
Linear double-layer capacitance	C_{dl} (F)	[1e-7, 1e-5]	9.8-6	1.0e-5	1.0e-5
1^{st} order C_{dl}	C_{dlE1}	[-0.1, 0.1]	0.014	0.079	0.095
2^{nd} order C_{dl}	C_{dlE2}	[-0.05, 0.05]	0.04	0.021	0.045
3^{rd} order C_{dl}	C_{dlE3}	[-0.05, 0.05]	-5.6e-4	-4.4e-4	-3.8e-4
Uncompensated resistance	R_u (Ω)	[0, 900]	148.7	316.8	81.5
Potential frequency	ω (Hz)	[8.56, 9.46]	9.015 (8.96)	9.015 (8.75)	9.015 (8.83)
C_{dl} phase	C_{dl} phase (rads)	[3.77 , 5.65]	4.73	4.70	4.71
Phase	Phase (rads)	[3.77 , 5.65]	4.57	4.60	4.63
Symmetry factor	α	[0.4, 0.6]	0.6	0.6	0.6

Table 5.3: Best fit parameters for harmonics 4 and above of PSV experiments 1, 2 and 3. The resulting simulated PSV current is shown in figures 5.8(A-C). The same values were used to generate r-FTACV simulations shown in figures 5.8(D-F), and the values in brackets were used to generate the r-FTACV simulations in figures 5.8(G-I)

simulations and data (for the 6th harmonic) for the PSV and r-FTACV cases shown in figure 5.9. In effect, the same simulation is being returned each time, with different values reported, as driven by parameter compensation.

The main feature of the values in table 5.4 is the fact that, at an α value of 0.55, the values returned become totally reversible. To demonstrate this, the upper bound of k_0 was instead set to a value much greater than the one noted in table 5.3 — 3000 s^{-1}. This is incompatible with the trumpet plot data, which estimates a value for k_0 at around 100 s^{-1}. If the true value of k_0 was 3000 s^{-1}, then the expected peak potential divergence point would be at approximately 50000 mV s^{-1}. Although this kinetic value is not accurate, the parameter values in table 5.4 are useful in that they do also show the effect of parameter compensation; we see that, to achieve the same simulation output, a higher k_0 value is linked with a higher midpoint potential standard deviation and a lower resistance, which is corroborated by the 2D histograms in figure 5.13, 5.14 and 5.15. On the basis of this analysis, the true value of α is likely to reside in the region between

0.55-0.6; lower results in incompatible k_0 values, and higher is outside of the region of chemical plausibility. Although this approximate description of a key Faradaic parameter is not ideal, it is not a significant departure from the literature which frequently assumes a value of 0.5 — the workflow proposed in this chapter allows for a better understanding of the impact of such modelling decisions.

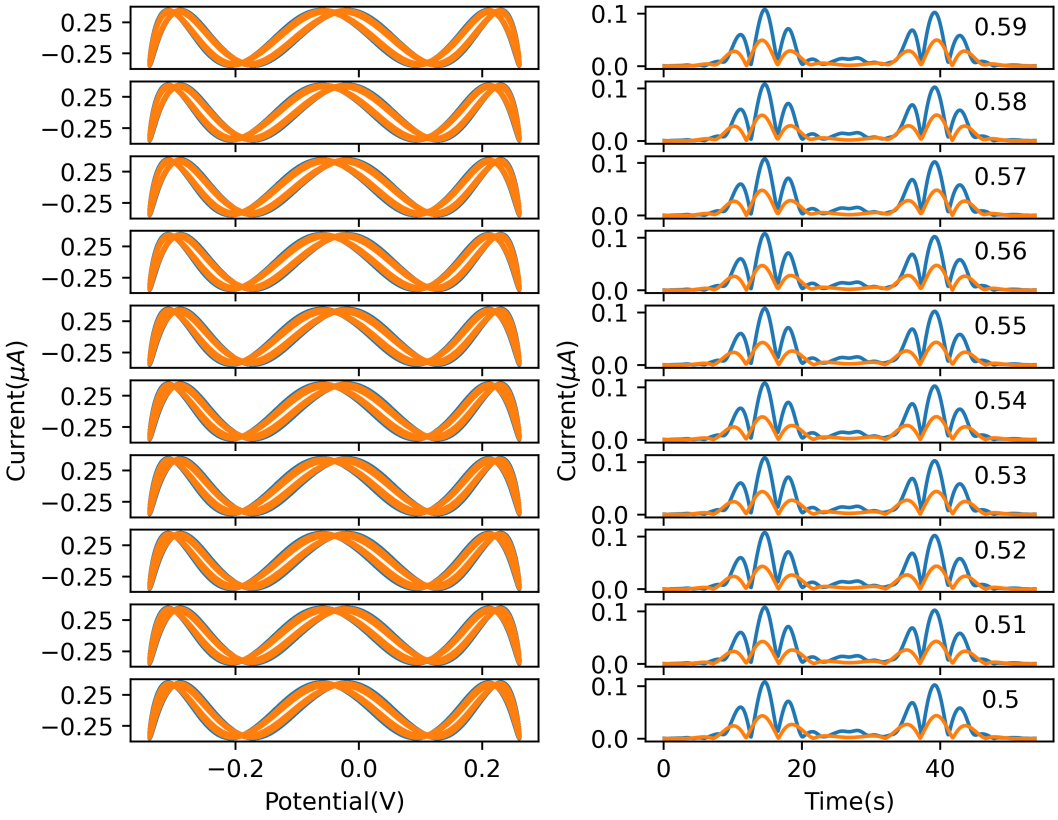


Figure 5.9: PSV and translated r-FTACV harmonic 6, where the upper bound of α was set at the value indicated in the r-FTACV plot. Simulation parameters can be found in table 5.4

5.3.3.2 The effect of C_{dl}

In terms of the capacitance, the value for all three experiments is at or close to the upper bound of 1e-5F. This may indicate that the capacitance values are not physically realistic. As we are interested in obtaining accurate estimates of the *Faradaic* parameters, our primary concern is that the unrealistic capacitance estimates are not affecting the accuracy of the other inferred parameters. For example, inaccurate estimates of the background current could lead to inaccurate estimates

Symbol	α Upper bound				
	0.59	0.58	0.57	0.56	0.55
E^0_μ (V vs. ref.)	-7.182E-2	-7.195E-2	-7.202E-2	-7.201E-2	-7.256E-2
$E^0\sigma$ (V)	0.045	0.046	0.046	0.048	0.053
k_0 (s^{-1})	167.129	162.553	168.189	204.224	2999.992
R_u (Ω)	137.714	125.403	108.95	86.855	39.516
C_{dl} (F)	9.996E-6	9.997E-6	9.606E-6	9.000E-6	8.581E-6
C_{dlE1}	0.016	0.018	0.021	0.023	-5.713E-3
C_{dlE2}	0.042	0.044	0.049	0.059	0.094
C_{dlE3}	-5.628E-4	-5.646E-4	-5.883E-4	-6.259E-4	-6.346E-4
Γ (mol cm^{-2})	1.355E-11	1.362E-11	1.369E-11	1.376E-11	1.388E-11
ω (Hz)	9.015	9.015	9.015	9.015	9.015
C_{dl} η	4.724	4.719	4.711	4.703	4.691
η	4.575	4.576	4.57	4.549	4.443
α	0.59	0.58	0.57	0.56	0.55

Symbol	α Upper bound				
	0.54	0.53	0.52	0.51	0.5
E^0_μ (V vs. ref.)	-7.293E-2	-7.295E-2	-7.297E-2	-7.265E-2	-7.301E-2
$E^0\sigma$ (V)	0.052	0.052	0.052	0.053	0.052
k_0 (s^{-1})	2998.856	2999.99	2999.99	2999.998	2999.997
R_u (Ω)	35.995	35.492	35.025	37.472	34.202
C_{dl} (F)	9.947E-6	9.040E-6	9.355E-6	9.977E-6	9.319E-6
C_{dlE1}	-1.292E-2	-1.425E-2	-1.406E-2	-6.136E-3	-1.442E-2
C_{dlE2}	0.085	0.094	0.092	0.083	0.093
C_{dlE3}	-5.466E-4	-6.015E-4	-5.814E-4	-5.466E-4	-5.839E-4
Γ (mol cm^{-2})	1.363E-11	1.364E-11	1.364E-11	1.393E-11	1.366E-11
ω (Hz)	9.015	9.015	9.015	9.015	9.015
C_{dl} η	4.692	4.692	4.692	4.69	4.691
η	4.444	4.443	4.443	4.443	4.443
α	0.54	0.53	0.52	0.51	0.5

Table 5.4: Simulation parameters used to generate the PSV simulations observed in figure 5.9, where the upper boundary for α was set to the value shown in the heading for each column.

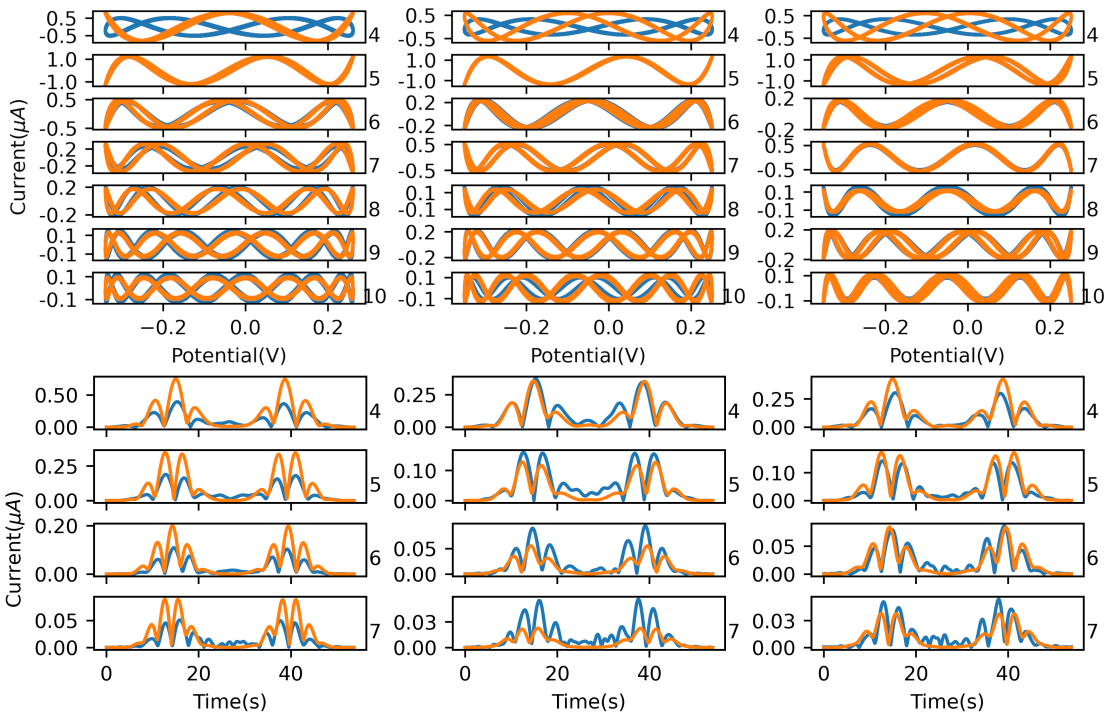


Figure 5.10: Two-step fits to PSV (top row) and r-FTACV (bottom) harmonics, using the parameters listed in table 5.5, expect in for the input frequency and phases for the r-FTACV case, which are as listed in table 5.3.

of the level of uncompensated resistance through the Ohmic drop effect. The uncompensated resistance is in turn known to be correlated with changes in the kinetic parameter (which we also demonstrate using our Bayesian inference analysis, in section 5.3.4), and consequently poor estimates of the background current could lead to poor estimates of the kinetic value. This is why we advocate a DCV trumpet plot measurement to set the bounds of k_0. Indeed, the key strength of the multi-experiment approach described here is to be able to address such concerns about spurious parameter combinations.

We can observe the impact of our capacitance modelling choices, by comparing to an alternative "two-step" fitting procedure. "Two-step" fitting refers to fitting capacitive current simulations (with zero Faradaic current) to total experimental PSV current, and then fitting Faradaic parameters in the Fourier domain, fixing the capacitive simulation parameters to the values obtained in the first step. When we attempted this for each of the three PSV experiments, good fits to harmonics 4-10 were not obtained in the second step, and so harmonics 5-10 were fitted. In

figure 5.10, we show the PSV harmonic fits (top row) and in the bottom row show what happens when the fits are translated to the corresponding r-FTACV harmonics. The PSV simulation parameters can be found in table 5.5. The inferred fits to PSV harmonics 5-10 are, by eye, not quite as good as the ones obtained by the single-step fitting process presented in figure 5.8. The inferred parameters are mostly of the same order of magnitude as those reported in table 5.3, although with some key differences. The first is that there is less consistency between the Faradaic values returned by fitting to the three experiments, particularly in the inferred mid-point potential and rate constant parameters. Secondly, the surface coverage values are larger, and greater than the values obtained using DCV charge-integration analysis as detailed below. The parameters from PSV 1 particularly appear to be in a different parameter regime, with a k_0 value approximately 3 times larger than the value inferred from the other two experiments, and an E_σ^0 value approximately 3 times smaller. However, what is interesting about the values inferred from PSV experiments 2 and 3 is that the E_σ^0 values are closer to the "corrected" values obtained for r-FTACV in table 5.3, and the inferred k_0 is slightly closer to the value inferred by analysis of the DCV trumpet plot. The "true" values certainly reside somewhere in the space defined by the two approaches, and resolving the discrepancy will be a topic of active research.

In figure 5.11, the parameter values for the single-step and two-step approaches are shown, where 0 is the lower and 1 the upper bound as defined in table 5.3. Additionally, for the simultaneous case, we use MCMC samples to calculate the necessary statistics for a box plot. There are a few interesting points to note. Firstly, the simultaneous parameters are more consistent (i.e. similar values across multiple experiments) for the Faradaic parameters and the two phases, whereas the two-step parameters are more consistent for the capacitive parameters. This can be interpreted in light of the fact that the two-step method has access to more information about the capacitance (by fitting to the total current) but by fixing these parameters, we limit the accuracy with which the Faradaic and phase parameters can be recovered. By the same token, only fitting in the filtered Fourier domain allows for more accurate recovery of the parameters that have a strong effect on this region, but reduces the amount of information available about the capacitance parameters.

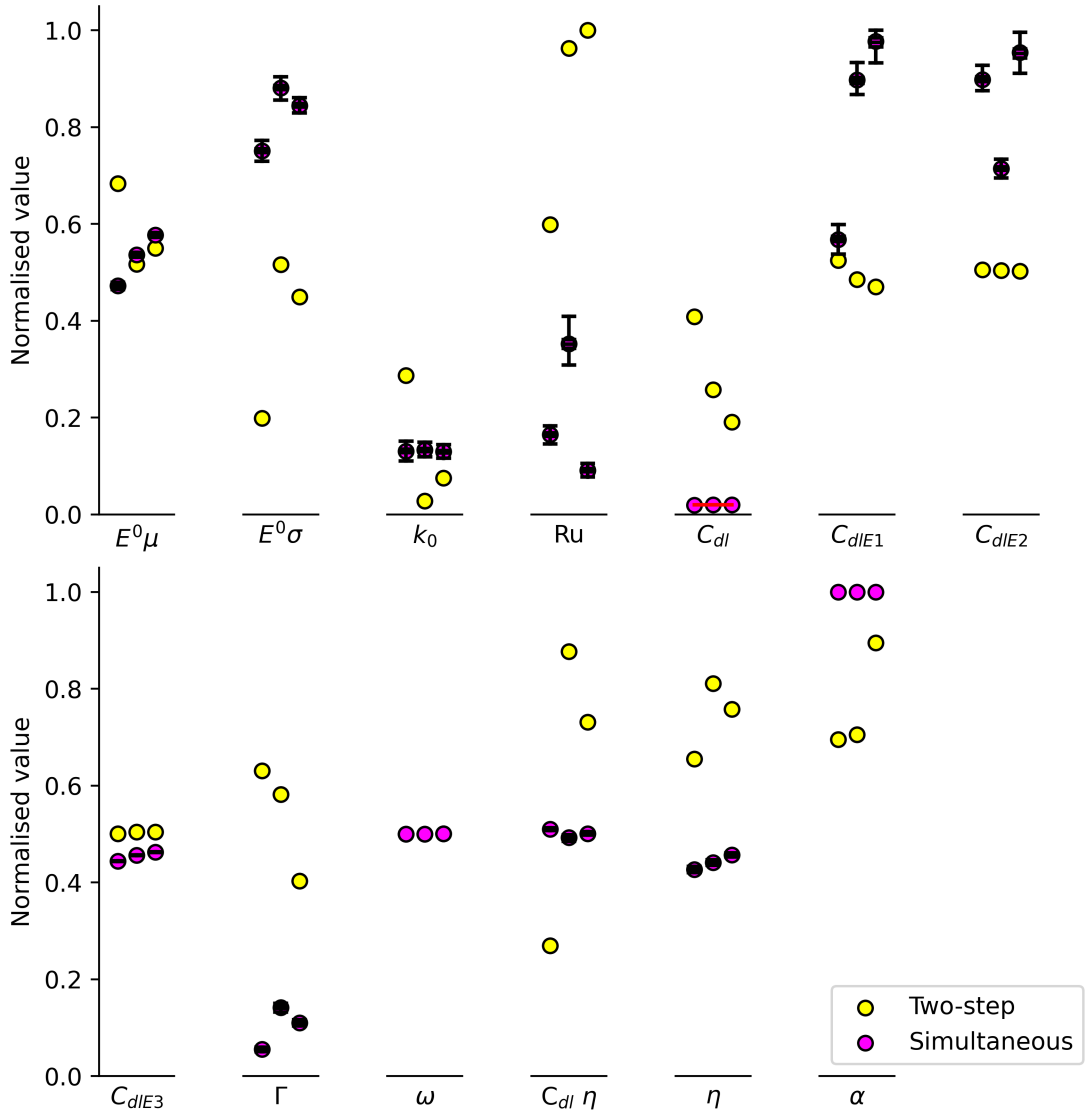


Figure 5.11: Values inferred using two methods for PSV experiments 1-3, as in table 5.5 and table 5.3, normalised so that the lower bound (also in table 5.3) is 0 and the upper bound is one. The box plot values are taken from MCMC-inferred distributions for the simultaneous case, where this was available. As the upper bound for C_{dl} was different for the two fitting approaches, we have recorded the upper bound for the simultaneous effort as a red line.

Parameter	Symbol	PSV 1	PSV 2	PSV 3
Midpoint potential mean	E_μ^0 (V vs. ref.)	-5.900E-2	-6.900E-2	-6.700E-2
Midpoint potential standard deviation	$E^0\sigma$ (V)	0.012	0.031	0.027
Rate constant	k_0 (s^{-1})	322.608	76.123	121.019
Uncompensated resistance	R_u (Ω)	538.88	866.475	900
Linear double-layer capacitance	C_{dl} (F)	2.043E-4	1.288E-4	9.557E-5
1^{st} order C_{dl}	C_{dlE1}	5.000E-3	-3.000E-3	-6.000E-3
2^{nd} order C_{dl}	C_{dlE2}	5.385E-4	3.566E-4	2.367E-4
3^{rd} order C_{dl}	C_{dlE3}	8.214E-6	4.391E-5	4.420E-5
Surface coverage	Γ (mol cm^{-2})	6.006E-11	5.612E-11	4.166E-11
Potential frequency	ω (Hz)	9.015	9.015	9.015
C_{dl} η	C_{dl} η	4.277	5.419	5.145
Phase	η	5.002	5.295	5.194
Symmetry factor	α	0.539	0.541	0.579

Table 5.5: Simulation parameters used to generate the PSV simulations in figure 5.10. Capacitance parameters, including R_u, were obtained by fitting to the total current. Faradaic parameters were obtained by fitting to harmonics 4-10, fixing the capacitance parameters obtained in the previous step, and bounding the uncompensated resistance to 20% of total-current inferred value.

5.3.4 Bayesian inference

Due to the speed at which PSV can be simulated, it is possible to quantitatively visualise parameter uncertainty using a Bayesian framework. In figure 5.12, we show inferred distributions for key parameters for PSV experiments 1-3, along with appropriate parameters as inferred from the trumpet plot (E^0 from the trumpet plot has been graphed alongside the E_μ^0 parameter as they both have the same effect on the appearance of the total current, as in section 3.3). The parameters C_{dl} and α are excluded because the MCMC algorithm does not converge if the chains get stuck at an upper or lower bound. It was not practically feasible to undertake a Bayesian inference attempt for r-FTACV, because of the high computational cost of the multiple $\sim$10+ hour fitting runs required for a many-parameter MCMC. Figure 5.12 demonstrates that despite the above discussed issues of parameter compensation and the possibility of spurious minima, the parameter values inferred for the three separate PSV experiments are very much in the same regime, as are parameters inferred using an entirely separate experimental technique. It might

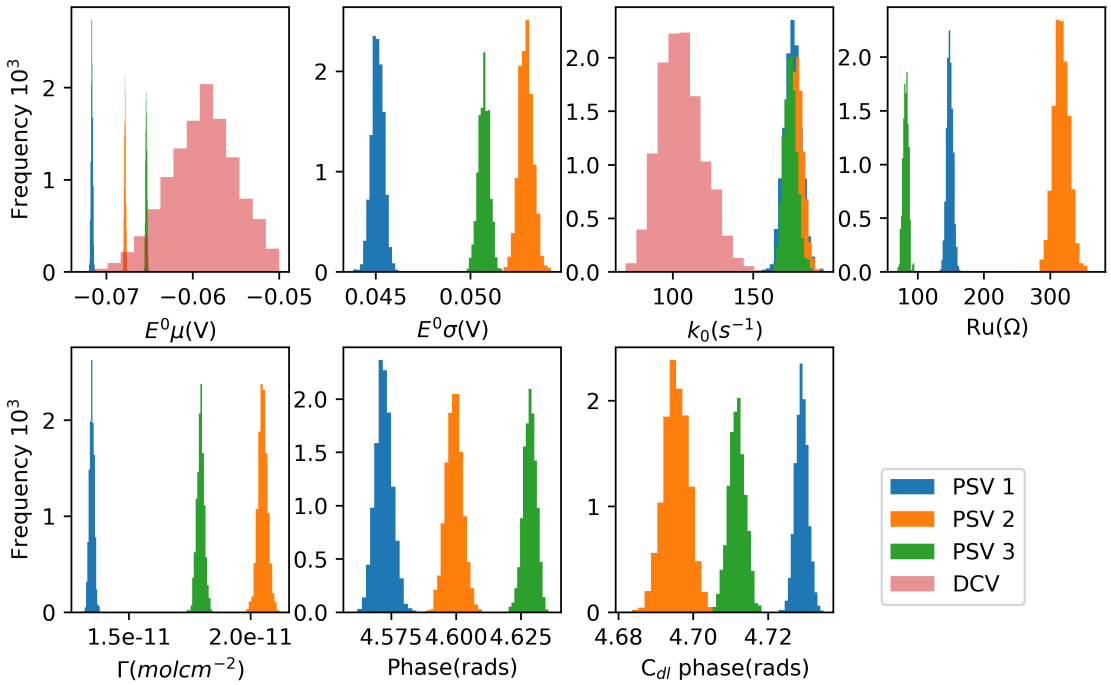


Figure 5.12: Inferred parameter histograms generated by pooling three independent MCMC chains, using harmonics 4 and above of the respective PSV experiments in the likelihood function, and using the parameters in table 5.3 as the starting point. C_{dl} and α were not included in this parameter inference approach for technical reasons mentioned in the text. The DCV histograms were as inferred from running an MCMC process on the trumpet data in figure 5.18, and as such we can only infer the parameters E^0 and k_0.

be expected that the posterior parameter histograms relating to the intrinsic chemical properties of the system (i.e. E^0 and k_0) would overlap for the different experiments. However, it is not particularly surprising that they do not, firstly as the composition of each protein mixture will be different, and secondly because of parameter compensation effects discussed below (driven by, for example, the differing amounts of surface coverage).

In figure 5.13, we assess the degree of correlation using MCMC-inferred posterior distributions for PSV experiment 1, with the results for experiments 2 and 3 shown in figures 5.14 and 5.15. The histogram of samples for each parameter is shown along the diagonal (with the parameter indicated on the x-axis), and a scatter plot for each pair-wise combination of parameters are shown on the off-diagonals, with the $y-$ and $x-$axes indicating which parameters are being plotted together. If two parameters are totally uncorrelated, then the histogram will be a circle. An

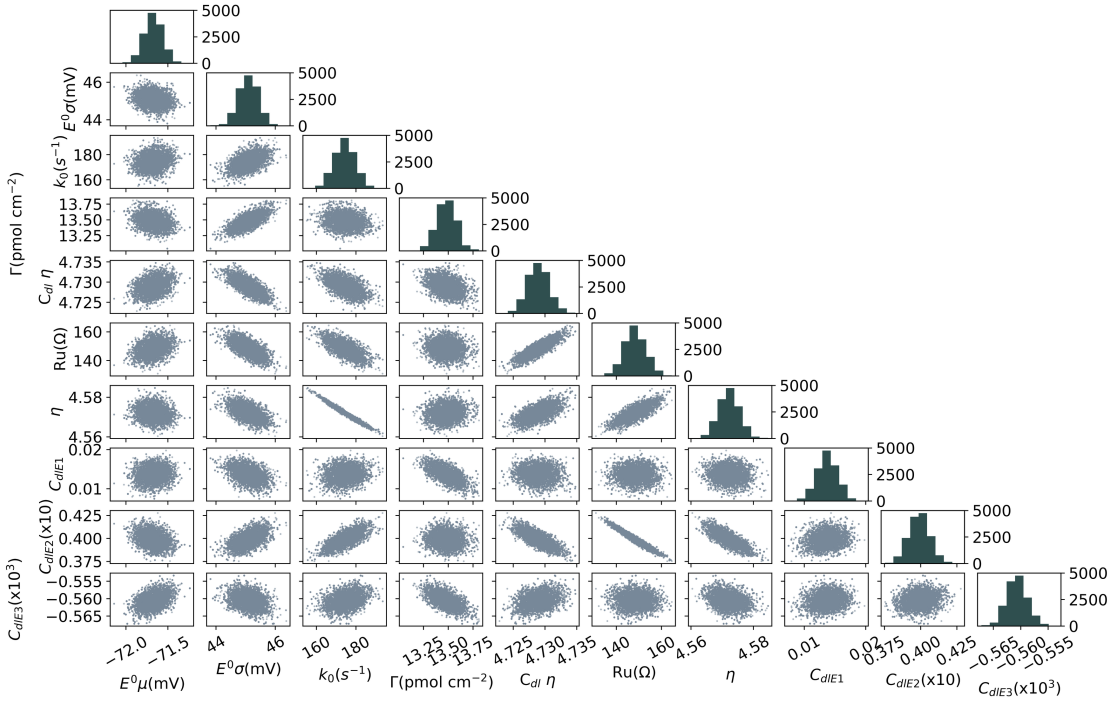


Figure 5.13: 2D histograms generated from the MCMC process for PSV experiment 1

ellipse angled upwards indicates positive correlation, i.e. an increase in the value of the x-axis parameter is associated with an increase in the $y-$axis parameter, and vice versa for negative correlation. The narrower the ellipse, the stronger the degree of correlation.

From the figure we can therefore see that many parameters are correlated with each other. In particular, of the reaction model parameters there are correlations between k_0, the uncompensated resistance, phase, C_{dl} phase and C_{dlE2}. Thus, it is clear how challenging it is to define the electron-transfer rate. The positive correlation between E^0_σ and Γ explains that the discrepancy between the best-fit PSV and r-FTACV values is driven by these parameter compensation effects. As the two-step fits presented in figure 5.10 above show, an increased C_{dl} value is associated with lowered E^0_σ and Γ values — indicating that our choice of C_{dl} bound (such that the returned value is relatively small) has led to slightly over-inflated predicted values for the positively correlated E^0_σ and Γ parameters. This again shows the utility of the multiple-experiment approach. For r-FTACV, this compensation between C_{dl}, E^0_σ and Γ is not of the same magnitude, which allows us to detect the slight over-estimation described above in figure 5.8. This is evidenced by the fact that two different sets of parameters can provide an identical PSV harmonic simulation, but

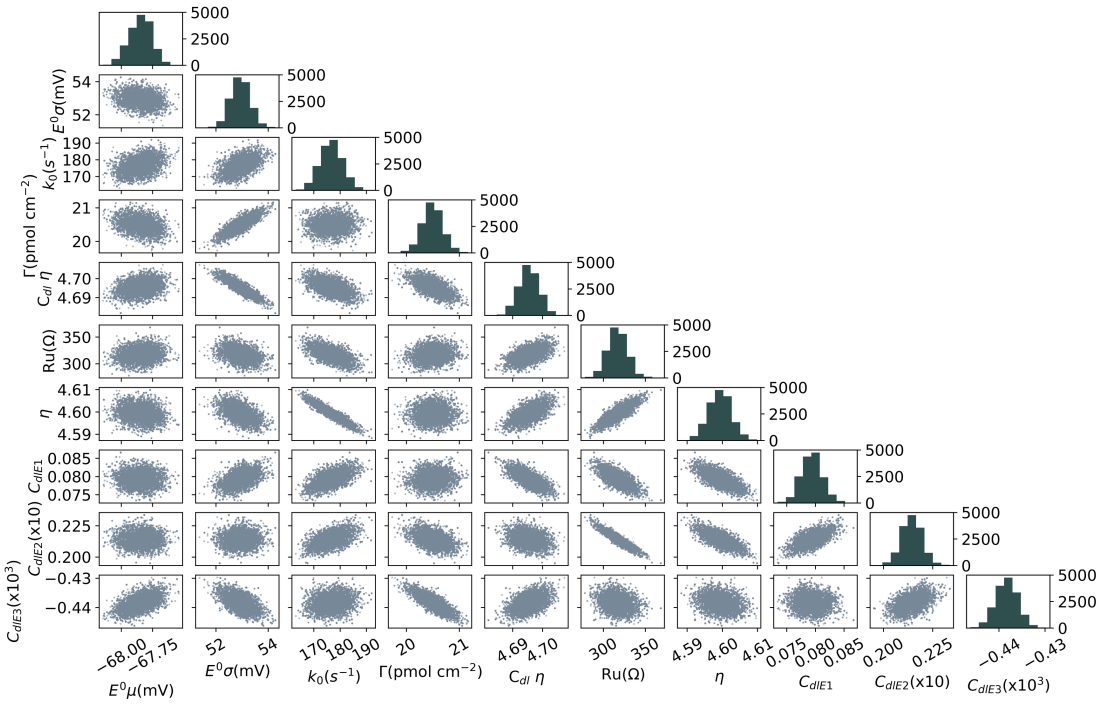


Figure 5.14: MCMC-inferred histograms for PSV experiment 2.

a different r-FTACV simulation, as shown by the low and high-C_{dl} simulations in figure 5.5 and table 5.2.

What figure 5.13 shows is that the choice of how to bound parameter space is not a neutral decision; because of parameter correlation, these choices affect every value returned. Consequently, although the MCMC analysis reports very high confidence in the inferred values, the exact values are contingent on the modelling choices we have made. However, the multi-experiment verification approach allows for verification of the inferred model parameters in a way that is not affected to the same extent by parameter compensations, the degree of which are peculiar to the specific data being analysed. Comparisons of experiments of the same type can be used to avoid the problem of fitting to noise in a single fitting run, and comparisons of different types of experiments provides an alternative perspective to as to spot parameter compensation effects, as with the example of C_{dl}, E^0_σ and Γ in the previous paragraph. The fact that we are able to have this highly granular discussion about the precise values of the returned parameters is an indication of the power of this framework. The use of PSV is essential to obtain this level of understanding, as the 600,000 forward problem simulations required for a single MCMC run can

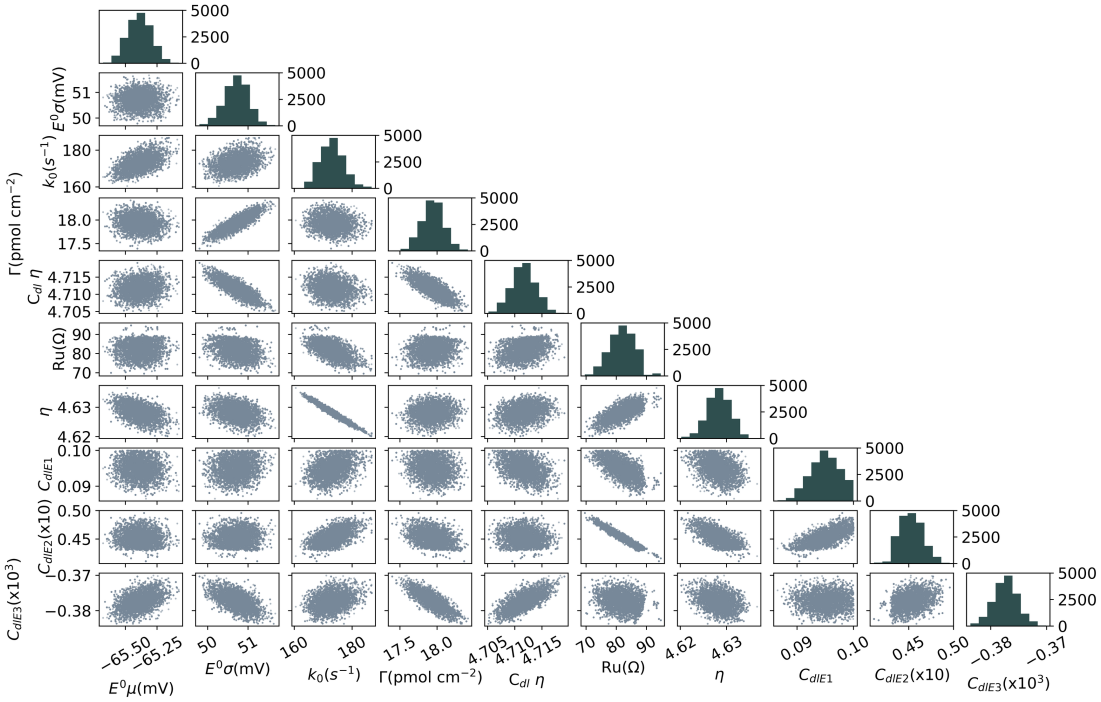


Figure 5.15: MCMC-inferred histograms for PSV experiment 3.

be completed in just over an hour.

5.3.5 DCV analysis

5.3.5.1 Single-scan rate analysis

In previous sections, our analysis has predominantly been on Fourier-transform voltammetry methods. For all three PSV experiments shown in table 5.3, we also collected DCV data at a scan rate of 30 mV s^{-1}, an example of which is shown in the leftmost plot of figure 5.16. In our proposed framework, insight from analysis of DCV experiments serves primarily to contextualise the results of the r-FTACV and PSV fitting results described above. In this section, we describe the various methods used to obtain these insights.

Although the level of signal shown in the total DCV current in the left-hand plot of figure 5.16 is extremely good for this type of experiment, with clearly visible Faradaic peaks, we found

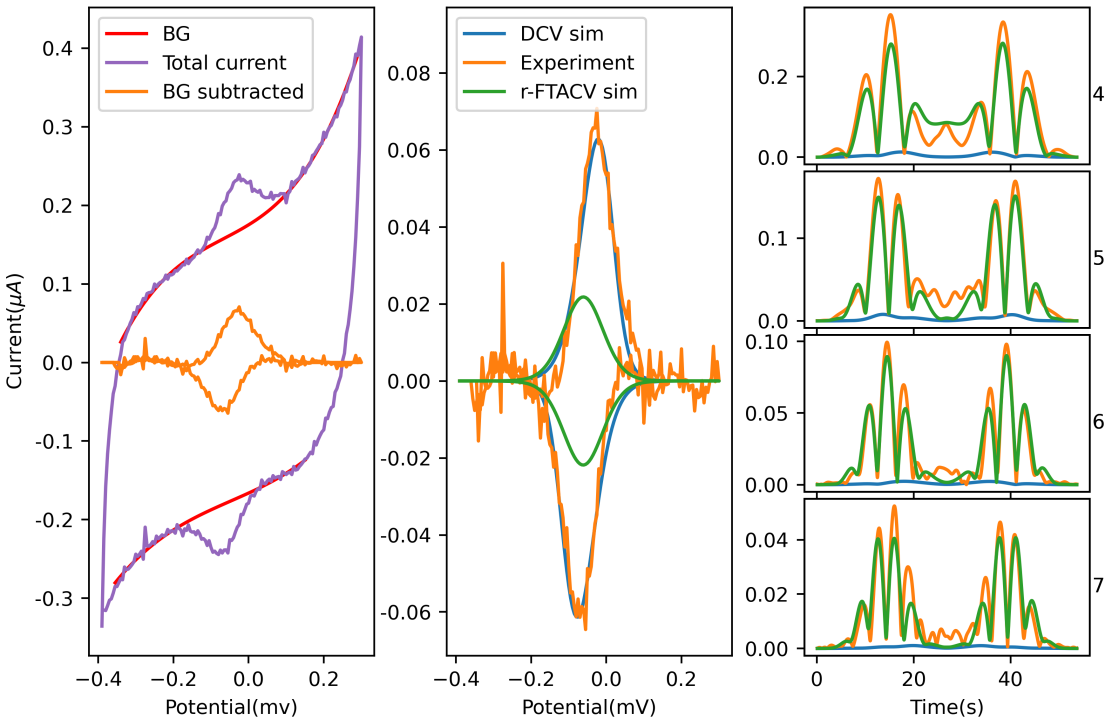


Figure 5.16: The leftmost plot is an example of the background subtraction procedure which attempts to isolate purely Faradaic current from a DCV experiment. We interpolate across the region of the total DCV current (purple) using a third-order polynomial (red), and subtract this polynomial from the total current to yield the background-subtracted current (orange). In the central plot, we show the same background-subtracted current with two purely Faradaic current simulations. The DCV simulation (blue) was generated using parameters obtained by fitting to the background-subtracted current, and are recorded in table 5.6. The r-FTACV simulation uses parameters listed in table 5.3, using the values in brackets where appropriate. In the leftmost plot, the parameters in table 5.6 fitted to the DCV experiment are used to generate r-FTACV harmonics 4-7, and plotted against r-FTACV data harmonics, alongside the harmonics from the r-FTACV simulation in green.

that parameters obtained through analysis of single-scan rate DCV data was contradictory to the results obtained by analysis of PSV and r-FTACV. The DCV analysis challenges were multi-faceted; the first issue was that our polynomial capacitance approximation could not accurately describe the capacitance current that occurred directly after the switching potential (from 0.25V to 0.15V in the DCV current in figure 5.16), partly because of the discontinuity in the potential input that occurs at this switching potential. One way to achieve the "curved" appearance was with extremely high resistances (>1e5Ω), as in work by Feldberg [85], which were incompatible with our PSV and r-FTACV fits, as well as with the observed DCV Faradaic signal. The standard tool to get around challenges in modelling the capacitive current is using "background

Parameter	Symbol	Single DCV	Trumpet
Midpoint potential mean	E^0_μ (V vs. ref.)	-4.92E-2	-5.928E-2
Midpoint potential s.d.	$E^0\sigma$ (V)	0.017	
Rate constant	k_0 (s^{-1})	0.589	109.4
Uncompensated resistance	R_u (Ω)	1.183E-7	*
Surface coverage	Γ (mol cm^{-2})	3.48E-11	
Symmetry factor	α	0.4806	
Separation	Sep (V)	*	0.032

Table 5.6: Table of inferred DCV parameters. The single DCV column was used to generate the purely Faradaic simulation plotted in the central graph of figure 5.16 in blue. The parameters for the 'trumpet" column were obtained by fitting to the data in figure 5.18, where we can only fit E^0 and k_0, along with a "separation" parameter.

subtraction", as specified in [43], and described above. Our alternative background subtraction methodology is shown by the plot on the left hand side of figure 5.16. The assumed background current (the third-order polynomials in red) is subtracted from the total current to obtain the background subtracted current in orange. In the middle plot, we show the same background-subtracted current, and purely Faradaic current simulations. The DCV current parameters, listed in table 5.6, were obtained by fitting in the time-domain to the background-subtracted current, and the r-FTACV parameters are as detailed in table 5.3 for PSV experiment 1. What is apparent is that there is a poor correspondence between being a good fit to DCV and a good fit to r-FTACV. The poor fit of the DCV parameters to the r-FTACV experimental data is explained by the fact that the inference process predicts a very small kinetic parameter when fitting to the DCV data. The reason for this is because, even though the scan rate is relatively slow, the peak position of the oxidative and reductive peaks is not the same in the background-subtracted current (a phenomenon that also observed for the data collected at multiple scan rates). At the scan rate of this experiment, the oxidative/reductive peak overlap (i.e. the point of complete kinetic reversibility at this scan rate) occurs at $5\ s^{-1}$, a value we know to be too low by inspection of r-FTACV harmonic waveshapes. The non-overlapping oxidative and reductive observed peaks in the background-subtracted DCV current is presumably because the subtraction process does not truly remove all the background current, and we are observing contamination from capacitive current. This issue has been identified in previous work, which suggested that this was driven by secondary kinetic processes. What this tells us is that the observed DCV current is not a simple

addition of capacitive and Faradaic current terms (they are convolved), and that consequently subtraction is not sufficient to separate the two contributions. Furthermore, any attempt to correct for this observed effect without a deeper understanding of the contributing kinetic processes will surely be entirely arbitrary, and we could not have confidence in the inferred values.

The poor translation of r-FTACV parameters to the DCV current is presumably driven by differences in surface coverage between the DCV and r-FTACV experiments (as the former was performed before the latter), but could imply that we are slightly underestimating the surface coverage in the inferred r-FTACV parameters, and overestimating the spread of midpoint potential values (i.e. E^0_σ). However, as r-FTACV is the more sensitive technique, the values inferred from fitting to r-FTACV carry greater weight.

Our conclusion from this comparison exercise is that our modelling methodology is not well-suited to obtaining information from a single DCV experiment, firstly because our capacitance model struggles to account for the non-idealities observed in actual DCV data, and secondly because the background-subtraction approach does not reduce the problem of capacitance to the same extent as filtering in the Fourier domain does. This conclusion appears to be supported by the literature, where all studies reviewed that attempted to infer kinetic parameters from DCV data used multiple DCV experiments in the form of a trumpet plot [8, 40, 41, 42, 44], which we discuss below.

5.3.5.2 Surface coverage estimation

As detailed above, it is possible to estimate the surface coverage by integrating over the area (or charge) of the Faradaic peak. In figure 5.17, we show that by slightly altering the window over which we exclude the Faradaic peak, in order to fit the polynomials in black, we can obtain a variation of 13 pmol cm^{-2} in the surface coverage estimate by applying this method to the background subtracted data in red. The lower value is consistent with the estimate obtained by fitting to the corresponding r-FTACV data, whereas the upper value is more consistent with the value obtained by fitting to the background subtracted data in figure 5.16. The conclusion was that this method can provide useful "order of magnitude" estimates, but further confirmation should come from other experiments.

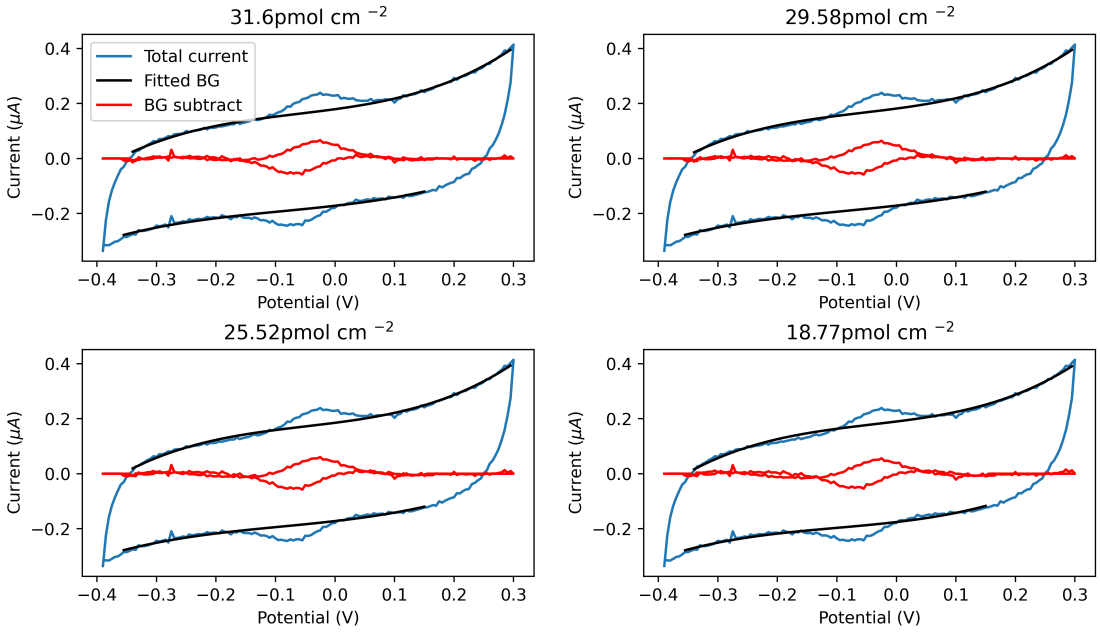


Figure 5.17: Four separate attempts to infer the surface coverage parameter from a DCV experiment, using different potential window sizes. The "fitted background" current was subtracted from the total current to obtain the background subtracted current. The area under the peak was then used to estimate the surface coverage, which is shown at the top of each plot.

5.3.5.3 Multiple-scan rate analysis

As described above, data from multiple DCV experiments conducted at different scan rates can be combined to form a trumpet plot, using the same protein solution as for PSV experiment 3, but different electrode films. In order to simulate trumpet plot data, multiple DCV experiments were simulated using the experimental scan rates (with all other parameters being held constant), and extracting the peak potentials. The only simulation parameters we used were E^0, which affects the value of the peak potential, k_0 which affects the scan rate at which the oxidative and reductive peak potentials cease to overlap, and a constant separation parameter. This was included, as a constant separation between peak potentials was observed, even at very slow scan rates. The value of the separation parameter was added to all oxidative peak potentials, and subtracted from all reductive peak potentials. This results in reasonably good fits to the experimental trumpet plot data, seen in figure 5.18, labelled "MCMC" using the simulation parameters found in table 5.6. It should be noted that the gradient of the experimental oxidative peak potentials in figure 5.18 is steeper than predicted by the simulation, but this artefact remains if we only fit to oxidative peak data. As there is a discrepancy between the kinetic value inferred from

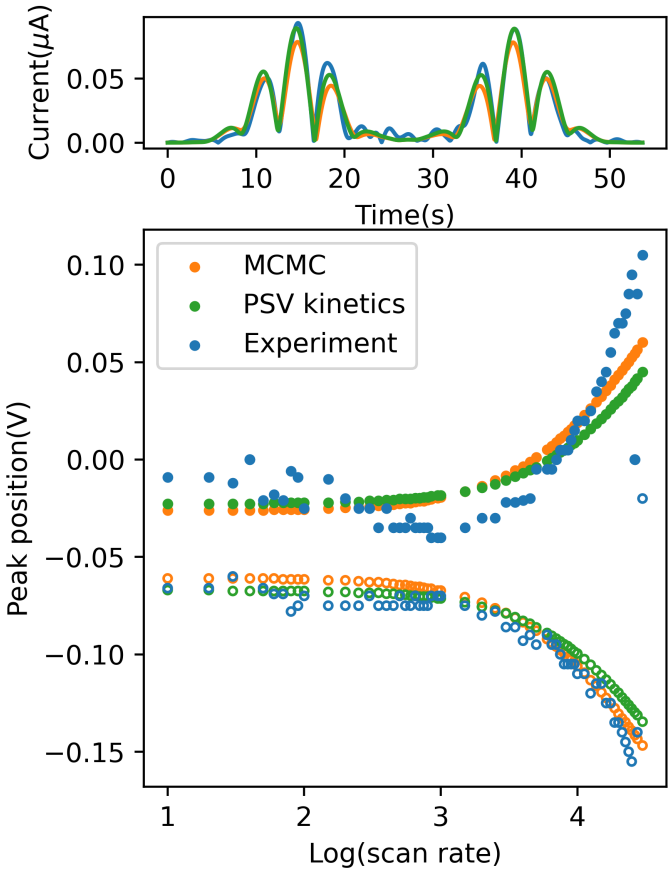


Figure 5.18: Bottom: Trumpet plot data and best fit simulation ("MCMC"). Filled circles indicate oxidative peak position, and hollow circles are for the reductive peak position. Additionally, we show the effect of altering the kinetic parameter to the value indicated for PSV 3 in table 5.3 ("PSV kinetics"). Top: r-FTACV harmonic 6, from experimental data, and using the best fit simulation parameters indicated in table 2. For the orange simulation corresponding to the "MCMC" simulation in the lower figure, the kinetic value was lowered to the one shown in table 5.6 (i.e. 109.4 s^{-1})

PSV (at 172.9 s^{-1} as inferred from PSV 3 vs at 109.4 s^{-1} as inferred by the trumpet plot), in the "PSV kinetics" trumpet simulation, we altered the k_0 value of the "MCMC" simulation to 172.9 s^{-1}. Although this results in a simulation that is a slightly worse fit to the data, the change does not result in a particularly large shift in the simulated peak potentials, indicating that the PSV-inferred kinetic regime is plausible. In the top box of figure 5.18, the 6th harmonic from PSV experiment 3 is plotted, along with r-FTACV simulations using the values from table 5.6 (green), and using the same values, except were the kinetic value was altered to 109.4 s^{-1} (orange). Again, the difference between the simulations is not large, indicating that we are in a

plausible kinetic regime.

5.4 Conclusions

In this chapter we have shown that it is possible to obtain a highly detailed picture of the electrochemistry of the single-electron redox reaction of *Cj*X183, to the extent that the parameters we have obtained are consistent across multiple experiments using different preparations of the same protein. The ability to robustly determine the electrical properties of metalloproteins allows for tracking how these properties change as a function of different physiological conditions, providing deep insight into the bioenergetics of living systems. We have shown how the multi-experiment approach, coupled with Bayesian inference, allows for increased confidence that the parameters we report are in a regime that is a good reflection of the underlying chemistry.

In chapter 4, we found that when inferring parameters from PSV total current-time data for a simple chemical system (ferrocene), there was one optimal parameter set which, when used to simulate a ramped experiment, also provided a good fit to r-FTACV data collected for the same electrode film. However, for the cytochrome data analysed in this chapter, the signal-to-background ratio is significantly worse, and we are forced to use filtering approaches to lessen the impact of background current, which was not necessary for the analysis of ferrocene. The challenge is that this filtering discards a large quantity of Faradaic information. As a result, we have found that we obtain multiple parameter vectors that generate an equally "good fit". This is exacerbated by the other weakness of PSV; because the filtered PSV current is largely uninterpretable by eye, we cannot *a priori* select one parameter vector over another. To address this problem, we have developed a protocol that uses preliminary DCV analysis and visual inspection of r-FTACV to appropriately bound parameter space for subsequent PSV fitting runs. The validity of this method is indicated by the fact that this approach enables us to infer parameters from PSV high harmonic current-time data, and the "best-fit" parameter set also provides a good fit to r-FTACV data collected for the same electrode film. We represent this process graphically in figure 5.8.

If the aim of understanding protein bio-electrochemistry by tracking changes in inferred parameters is to be realised, then parameter inference needs to be rapid, accurate and reproducible. The framework that we describe in this chapter fulfils these three criteria.

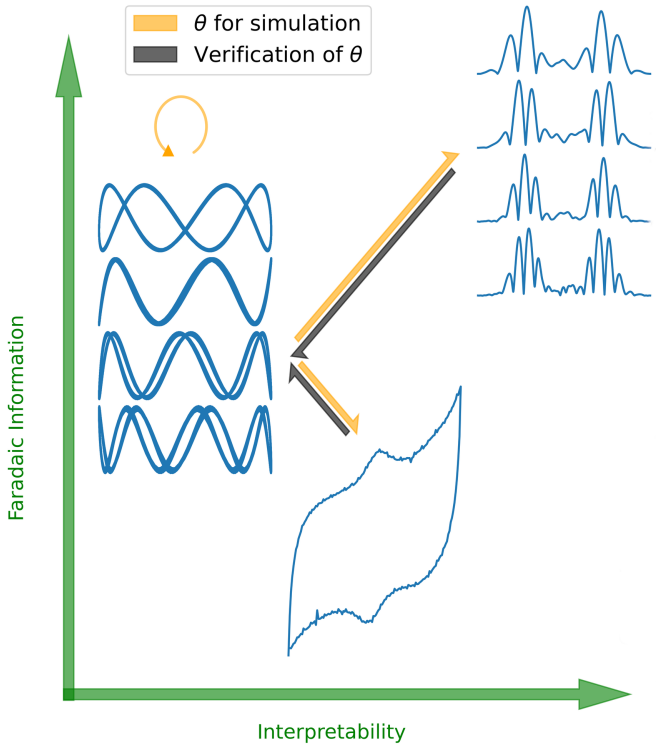


Figure 5.19: The three experiments analysed in this chapter, direct current voltammetry (DCV) as a current vs. potential plot, purely sinusoidal voltammetry (PSV) current harmonics 4-7 vs. potential, and ramped-Fourier Transform Alternating Current Voltammetry (r-FTACV) current harmonics 4-7 vs. time. The experiments are ranked according to how interpretable they are, and the amount of Faradaic information they provide. Orange arrows indicate simulation of current using a vector of chemical parameters ϕ, and black arrows indicate that the results of these simulations are used to assess the goodness-of-fit.

This framework is an attempt to leverage the advantages and mitigate the disadvantages of three voltammetry techniques to provide a highly accurate protein electron-transfer analysis toolkit.

In figure 5.19, we rank PSV, r-FTACV and DCV methods by the "interpretability" (ease of visual inspection) and Faradaic information content of the current output of a PFV experiment. The harmonic current for the PSV and r-FTACV cases is shown as this is the form of the data

that we analyse. In terms of Faradaic information, DCV is ranked last, because of the challenges associated with background subtraction; the highest ranked is r-FTACV, because it is highly sensitive to changes in redox reaction parameters, an effect we explored in section 3.2, and the capacitive and Faradaic current contributions can be clearly separated by filtering. In terms of interpretability, PSV is ranked last because the Faradaic contribution is difficult to discern in the oval total-current versus potential voltammogram and the appearance of the harmonics is non-intuitive; r-FTACV is ranked highest, as the effect of the distinct Faradaic parameters of interest can be determined through visual inspection.

The important contribution we aim to make in this work is to demonstrate how our new, integrated DCV, r-FTACV and PSV measurement and analysis protocol ultimately enables Bayesian statistical analysis to be utilised to visualise the potential pitfalls in the chosen modelling approach. This is an important step forward in enabling critical evaluation of the limitations in the chosen parameter inference process. We have formalised this methodology into the following "recipe" for robustly inferring parameters:

1. Collect experimental data. For each electrode functionalised with protein, we recommend collecting PSV, r-FTACV and DCV data (the latter at different scan rates), in that order. This is because there will naturally be some film-loss as a result of consecutive experiments, and consequently we have ranked the experiments in order of how important having good signals is for the purposes of analysis.

2. Obtain estimates for E^0 and k_0 bounds by fitting trumpet plot data, a relatively simple process. This information should then be used to judge parameters inferred from analysis of PSV and r-FTACV data. MCMC can also be used at this juncture to explore the confidence in the inferred values.

3. Define boundaries for fitting PSV data. Initial boundaries should encompass a reasonably large area of parameter space, but as a rule of thumb should not cover more than two orders of magnitude. If this scale of coverage is necessary, consider log-transformations.

4. Determine if it is feasible to fit PSV data in the time domain, without using dispersion. If inferred Faradaic parameters are highly divergent between different time-domain fitting runs, then we recommend fitting in the Fourier domain. If this is the case, inspect the harmonics of the blank PSV data to see what portions of the Fourier domain needs to be zeroed-out, as we demonstrate in figure 5.1.

5. Fit the form of the PSV data chosen above, using a simulation without dispersion. Using the parameters resulting from this inference process, generate a ramped simulation and compare to the r-FTACV data harmonics to assess the translatability of the parameters. It should be reasonably clear if you are neglecting thermodynamic dispersion, as the simulated harmonics will be narrower, and will not decrease in magnitude with harmonic number to the extent observed in the experimental data. If thermodynamic dispersion is present, you should go back to fitting the PSV data accordingly. If the kinetics of the system are irreversible/quasi-reversible then it may be worth considering kinetic dispersion as well, as discussed in section 3.3 and 4.3.3

6. Keep on comparing your PSV fits to the r-FTACV harmonics. Some other points to note:

    - If fitting to the Fourier domain, keep checking that the predicted total current simulation is not significantly greater in magnitude than the experimental current data, as we show in figure 5.5.
    - If a parameter is consistently hitting a defined boundary, then consider raising or lowering this bound as appropriate, unless this is outside of the realms of chemical plausibility. Beware of parameter compensation effects.
    - A good rule of thumb is that you will see a set of "good-fit parameters" multiple times in ten runs with random initialisations. Using the boundaries in table 5.3, we observed the values reported 2-4 times out of ten.

7. Choosing which parameters to report is something of a personal choice — our rationale was that the inferred parameters for the three experiments should be in the same regime while providing a good fit to each.

In the next and final results chapter of this thesis, we will show how we applied this methodology to capture changes in the electrochemical parameters of *Cj*X183 after mutations to the active site.

Chapter 6

Rational mutagenesis

6.1 Introduction

In chapter 5, we showed that it was possible to rapidly and reproducibly obtain parameters that described the chemical properties of wild-type *Cj*X183, and proposed a workflow to achieve this in other protein systems. In this chapter, we detail our efforts to apply this workflow to mutants of *Cj*X183 with altered electrochemical properties. In chapter 1, we stated that the primary motivation for this thesis was to understand biological electron-transfer; how metalloprotein structure controls the electrochemical properties of their cognate metal atoms. Creating mutants that alter these properties provides us with insight into how organisms can control thermodynamics and kinetics to fulfil biological function.

6.1.1 Heme properties - thermodynamics

The heme-containing cytochromes are incredibly versatile electron carriers, with recorded midpoint potentials for various cytochromes spanning from -400 to +400 mV vs. SHE, highly distinct from the midpoint potential of +771mV of Fe^{2+}/Fe^{3+} atoms in solution [86, 87]. This versatility is attributable to the sensitivity of the heme (also referred to as iron protoporphyrin IX [88]) to fine alterations to its chemical environment. Enumerating all of the physical processes known to affect recorded cytochrome midpoint potentials is the subject of some excellent, and lengthy, reviews [86, 87, 89], and in the following sections, we will briefly summarise how various physical processes can affect the thermodynamic properties of cytochrome electron-transfer.

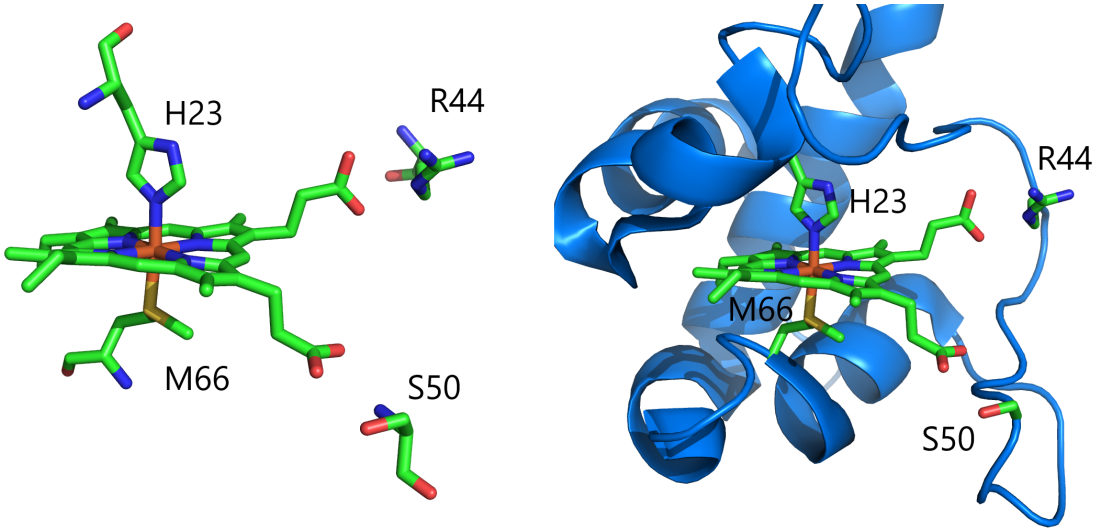


Figure 6.1: Left: structure of the heme from *Cj*X183, with several key residues included, including the axial/distal Fe ligands H23 and M66, and the propionate coordinating residues, R44 and S50. Nitrogen atoms are in blue, carbon in green, oxygen in red, iron in brown and sulphur in yellow. Right: the cognate heme in the context of the bulk structure of *Cj*X183, represented by a cartoon in blue [78]. Structure taken from protein data bank, accession code 7B21.

As a brief reminder, E^0 is a measure of the energy that electrons need to "jump" onto the species in question (i.e. to reduce it). Anything that preferentially stabilises the reduced state (which will have a lower charge, due to the presence of an additional negatively charged electron), or destabilises the oxidised state, will lower the required electron energy for reduction, and consequently increase the observed E^0. This set of relationships works in reverse, such that we can predict that an observed reduction in the E^0 value has been driven by destabilisation of the reduced state, or stabilisation of the oxidised state [6].

6.1.1.1 Co-ordination environment

Anything that donates or withdraws electron density in the vicinity of the Fe atom of the heme has the potential to affect E^0 in the manner described above. This particularly applies to the axial and distal ligands (in *Cj*X183 these are H23 and M66 respectively, as shown in figure 6.1) of the cytochrome [90, 91], which are in close proximity to the Fe atom of the heme, but modifications to E^0 can also be driven by hydrogen bonding or π-stacking interactions from more distant side-chains [86]. By a similar token, the functional side-chains of the porphyrin ring, such as the propionate side chains (as shown in co-ordination with R44 and S50 in figure 6.1), are capable of donating their electron density into the heme macrocycle. Consequently, the

degree of solvent exposure, hydrogen bonding or protonation state of these side chains will have an effect on the electronic environment experienced by the Fe atom, and therefore E^0 [88, 92, 93]. An extension to this is *ligand hardness*, which describes a set of properties related to the atomic radius, electronegativity and polarisability of a particular ligand [86] [1]. In the qualitative understanding, hard and soft ligands preferentially stabilise the harder Fe^{3+} and the softer Fe^{2+} respectively, which will affect E^0 according to the principles set out above.

6.1.1.2 Solvent environment

If solvent is excluded from the region of the protein containing the heme group, then, in addition to any resulting changes in the co-ordination environment, the relative permittivity (i.e. the ability to insulate charges from one another) experienced by the porphyrin changes significantly, from 80.4 (as in water [95]) to as low as an estimated 4 [96]. Consequently, this enhances the effects of the charges and electric dipoles that make up the environment around the heme. Anything that alters this relative permittivity, such as the degree of hydrophobicity of the pocket, will modulate this environment, again having an effect on E^0.

6.1.1.3 Molecular orbitals

In terms of the molecular orbitals of the Fe atom within the porphyrin, there are three d-orbitals (d_{xy}, d_{xz} and d_{yz}) which are lower energy, referred to collectively as t_{2g}, and two ($d_{x^2y^2}$ and d_{z^2}) which are higher energy, referred to collectively as e_g. The energy difference is a function of the orbital axis position; in an octahedral complex, ligands co-ordinate the metal on the x, y and z axis lines, and as such there is repulsion between the ligands and the orbitals that lie on these axis lines — $d_{x^2y^2}$ and d_{z^2} [97]. The relative distribution of electrons within these orbitals define whether the heme is in low- or high-spin state, and as such there are also "spin-exchange" transitions, with their own thermodynamics and kinetics, but the impact that these processes have on the observed E^0 defies simple explanation [98]. However, one established relationship is for the (low-spin) Fe^{3+} state, where five electrons are contained within the t_{2g} orbitals, versus six electrons for the low-spin Fe^{2+} state. Consequently, any effect that destabilises the t_{2g} orbitals will preferentially destabilise the Fe^{2+} state, decreasing E^0. One example of a physical process that destabilises t_{2g} is so-called heme "ruffling", which is a specific distortion of the planarity of the porphyrin ring (distinct from "saddling" or "doming") — the ring planarity is a function of

[1] It should be noted that there have been further quantitative characterisations, defining the hardness as the derivative of the chemical potential of a system [94].

the overall structure of the protein [88].

In sum, the thermodynamics of the redox reaction of a cytochrome is extremely sensitive to first, second and even third-order modifications to the electronic environment. This sensitivity undoubtedly contributes towards the wide usage of cytochromes in nature; allowing for minute tinkering of the thermodynamic properties of electron-exchange means that cytochromes can be leveraged for a wide variety of biological electron-transfer contexts.

6.1.2 Heme properties - kinetics

The literature on what affects the electron-transfer rates for the heme redox reaction is less comprehensive than that for reduction potentials, presumably as this parameter is more challenging to infer spectroscopically. Predictions for alterations to the rate constant value are predominantly informed using predictions from Marcus theory, using the degree of electronic coupling H_{ab} between the oxidant and reductant, as in equation 2.47 and 2.48 [86, 91]. Additionally, any change in the environment (such as solvent exposure, or ruffling) could also exert effects through modulation of the reorganisation energy λ[88]. Predictions informed by the relative stability of the Fe^{2+}/Fe^{3+} state as in section 6.1.1 are less useful, although for a homogeneous reaction, such as between two proteins, changes in the electron-transfer rate can be predicted using the differences in E^0, which contributes to ΔG^o [86].

6.2 Rational mutants

Two sets of single residue substitutions were made to wild-type *Cj*X183, both of which were predicted to alter the E^0 parameter. The first was a methionine-to-histidine translation at residue number 66 — M66H (M66 is indicated in figure 6.1). The second was a replacement of arginine with asparagine at position 44 — R44N. We shall discuss the predicted effect of each mutation on E^0 in turn.

6.2.1 M66H

The 66th residue of *Cj*X183 is the distal ligand that directly co-ordinates the iron atom of the porphyrin ring, and is therefore expected to exert a stronger effect over the thermodynamics of electron transfer compared to other residues. Histidine is expected to preferentially stabilise the

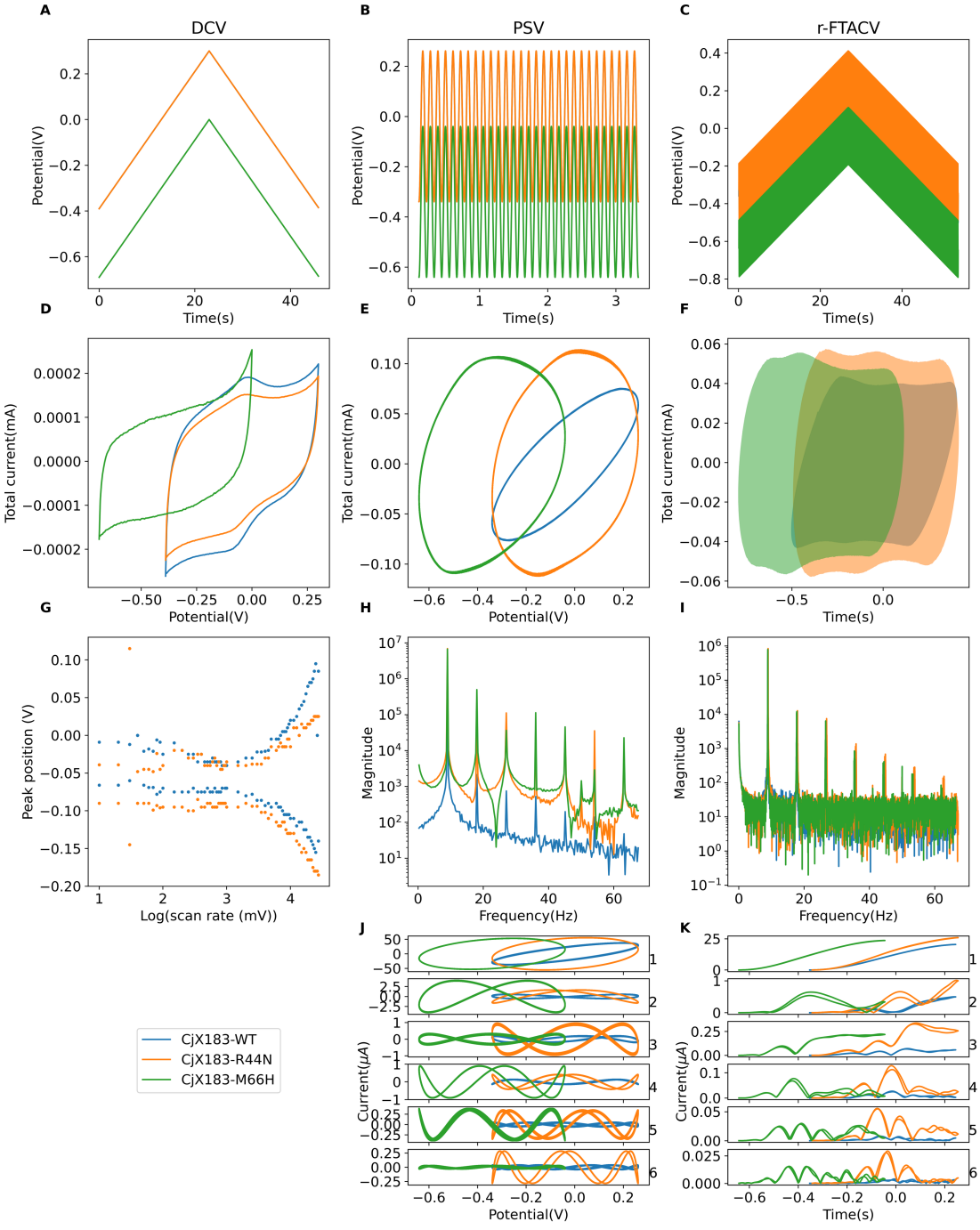


Figure 6.2: Voltammetry data collected for three different forms of *Cj*X183: WT (blue), R44N (orange) and M66H (green). Plots A-C show the potential input as a function of time and plots D-F show the resulting current as a function of potential. Plot G shows trumpet-plot data (DCV peak-position as a function of scan rate) for R44N, and WT data, although it should be noted that the wild-type data is as collected for PSV experiment 3 in section 5.3.5.3, as shown in figure 5.1, and is not part of the WT dataset shown in the rest of the figure. Plots H and I show the absolute Fourier spectrum corresponding to harmonics 0-7 of the PSV and r-FTACV total currents. Plot J shows PSV harmonics 1-6 as a function of the potential input, and plot K shows r-FTACV harmonics as a function of the linear ramp portion of the r-FTACV potential input.

Fe^{3+} state, thereby increasing the electron energy required to reduce Fe^{3+} to Fe^{2+}, consequently lowering E^0. There have also been some assessments on the effect of Met-to-His distal ligand mutations on the electron transfer rate in non-*Cj*X183 cytochromes, although the various efforts have found both increases [91] and decreases [99] in the apparent rate constant.

6.2.2 R44N

One of the heme propionate side chains discussed above interact with residue 44 in *Cj*X183. In the wild-type, the arginine side chain is positively charged at the experimental pH 7, and can form hydrogen bonds with the propionate, as indicated in figure 6.1. Mutation of the Arg44 residue to asparagine (which is negatively charged at pH 7) destabilises this interaction. As the negative charge on the propionate is no longer participating in a hydrogen bond, it is free to delocalise into the heme macrocycle [88], which has been shown to stabilise the Fe^{3+} state, and consequently lowering E^0, as discussed above for M66H.

6.3 Experimental methods

Wild-type, M66H and R44N variants were expressed and purified as in section 5.2. For DCV, r-FTACV and PSV the instrumentation was the same, and the reference electrode correction is +0.235 V. Voltammetry experiments were conducted in a slightly different order than as described in section 5.2. For all experiments, one single-scan rate DCV experiment was conducted at 20 mV s^{-1}, then one r-FTACV experiment at 22.5 mV s^{-1} and 8.94 Hz, then a PSV experiment at 9.015 Hz. Finally, multiple DCV scan rate experiments were conducted, again at scan rates varying from 10 to 30000 mV s^{-1}, although a signal was only recorded for R44N. For R44N and WT *Cj*X183, the DCV experiments were recorded from -0.39 V to 0.30 V vs. ref., the r-FTACV experiments were recorded from -345 mV and 255 mV and the PSV experiments from -344 mV to 254 mV. For M66H, the scan potentials were: -0.69 V to 0 V, -654 mV to - 45 mV and -644 mV to -44 mV for DCV, r-FTACV and PSV respectively.

6.3.1 Fitting time

As in chapter 5, solving a 13-dimension inverse problem for the PSV data took approximately 45 minutes. Solving a 4-dimensional inverse problem for r-FTACV data took approximately 2.5 hours. As before, each inverse problem was solved 10 times, and the best-scoring parameter

Parameter	Symbol	M66H	R44N
Midpoint potential mean	E^0_μ (V vs. ref.)	-0.509 (-0.495)	-0.102 (-0.086)
Midpoint potential s.d.	E^0_σ (V)	0.047 (0.037)	0.04 (0.036)
Rate constant	k_0 (s^{-1})	333.53	200.74
Uncompensated resistance	R_u (Ω)	315.57	807.5
Linear double-layer capacitance	C_{dl} (F)	3.000E-5	3.000E-5
1^{st} order C_{dl}	C_{dlE1}	0.0577	0.028
2^{nd} order C_{dl}	C_{dlE2}	-9.45E-4	0.003
3^{rd} order C_{dl}	C_{dlE3}	-1.003E-4	3.116E-5
Surface coverage	$\Gamma(molcm^{-2})$	8.933E-12 (6.76E-12)	8.09E-12 (6.89E-11)
Potential frequency	ω (Hz)	9.015 (8.866)	9.015 (8.959)
C_{dl} phase	C_{dl} η (rads)	3.922	6.24
Phase	η (rads)	4.66	4.78
Symmetry factor	α	0.6	0.4

Table 6.1: Best-fit PSV parameters as inferred by the CMAES algorithm for M66H/R44N data.

vectors observed multiple (2-4) times. The entire inference process relayed in this chapter was completed within two weeks.

6.4 Results and discussion

Using the methodology suggested in section 5.4, we collected a set of voltammetry data for the R44N and M66H *Cj*X183 mutants, along with additional wild-type (WT) *Cj*X183 data. The results of this data collection effort are shown in figure 6.2. However, due to time constraints and challenges associated with low protein concentration, the datasets collected were not as comprehensive as the ones presented in chapter 5. In particular, we were only able to collect trumpet-plot data for the R44N mutant (shown in figure 6.2 (G)), and present WT trumpet data from section 5.3.5.3 as a point of comparison.

However, despite these issues, we were still able to obtain a reasonably good fit to M66H PSV and r-FTACV harmonics and an excellent fit to the R44N PSV and r-FTACV harmonics; we could not obtain consistent fits to the WT data shown in figure 6.2, which we believe is driven

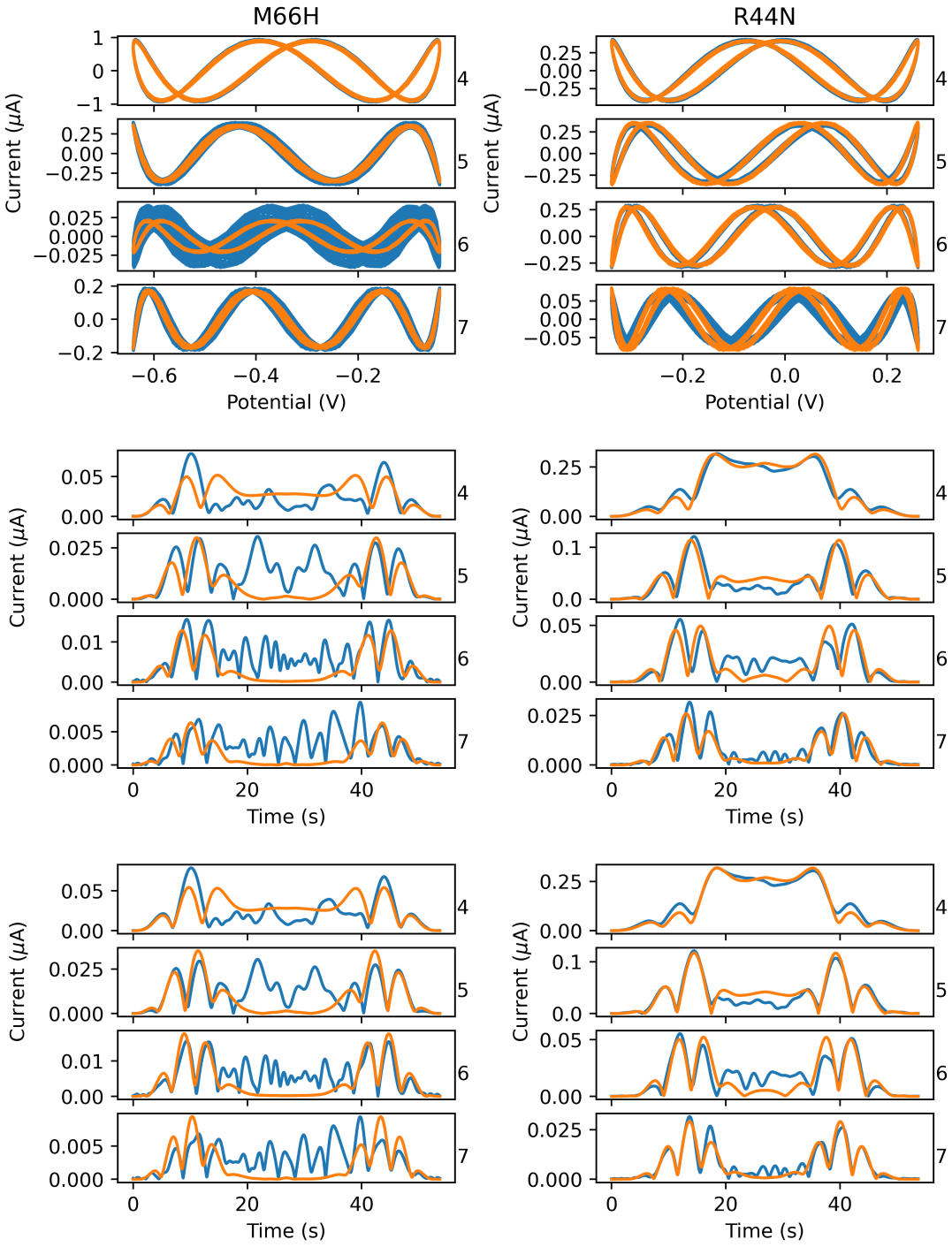


Figure 6.3: Fits to PSV and r-FTACV harmonics data for M66H (left column) and R44N (right column). Top row: simulated (orange) and experimental (blue) PSV harmonics 4-7, where the simulation parameters are as reported in table 6.1. Middle row: simulated (orange) and experimental (blue) r-FTACV harmonics 4-7, using the same parameters as for the PSV simulations, except the input frequency, which is written in brackets in table 6.1. Bottom row simulated (orange) and experimental (blue) r-FTACV harmonics 4-7 using parameters obtained by fitting to r-FTACV harmonics. The altered parameters are written in brackets in table 6.1

Parameter	Symbol	Lower bound	Upper bound
Midpoint potential mean	E^0_μ (V vs. ref.)	-0.55/-0.15	-0.4/-0.04
Midpoint potential standard deviation	E^0_σ (V)	1e-4	0.06
Rate constant	k_0 (s^{-1})	50	1000
Uncompensated resistance	R_u (Ω)	0	500
Linear double-layer capacitance	C_{dl} (F)	0	3E-5
1^{st} order C_{dl}	C_{dlE1}	-0.1	0.1
2^{nd} order C_{dl}	C_{dlE2}	-0.05	0.05
3^{rd} order C_{dl}	C_{dlE3}	-0.05	0.05
Surface coverage	Γ (mol cm^{-2})	1e-12	8e-11
Potential frequency	ω (Hz)	8.56	9.46
C_{dl} phase	C_{dl} η (rads)	3.14	6.28
Phase	η (rads)	3.14	6.28
Symmetry factor	α	0.4	0.6

Table 6.2: Parameter bounds passed to the CMAES when attempting to fit M66H/R44N PSV data. These boundaries were identical except for the E^0_μ parameter, with the values reported being in the order M66H/R44N.

by a low signal-to-noise ratio. The results of this fitting process are shown in figure 6.3, with fits to PSV shown on the top row, obtained by fitting to PSV harmonics 4-7 (using the simulation parameters in table 6.1). The same parameters were used to obtain an r-FTACV simulation, which is compared against r-FTACV harmonic data in the middle row. As in section 5.3.3, we then performed a limited optimisation on the parameters E^0_μ, E^0_σ, Γ and ω, fitting to r-FTACV harmonics 4-7, to obtain the simulations shown in the bottom row — the altered r-FTACV simulation parameters are written in brackets in table 6.1. Again as in chapter 5, accurately bounding parameter space was essential to obtaining a good PSV fit — these bounds can be found in table 6.2. So what do these results tell us?

6.4.1 M66H

Firstly for the case of M66H, the mutation has resulted in a large ($\sim$ 400 mV) decrease in the value of E^0_μ, which is in line with the predictions made in section 6.2.1, vindicating our mutation strategy. Additionally, there has been a change in the inferred k_0 value, from that of 175 s^{-1} in

section 5.3.3, to 333 s^{-1}. One would expect that the electron-transfer rate would be optimised to be as fast as possible in the wild-type, which does cast some doubt on the veracity of the inferred parameters, but an increased cytochrome electron-transfer rate after a single residue substitution has been reported before [100]. We should note that the M66H r-FTACV fits to harmonic 6 and particularly harmonic 7 are not very good, but this is at least partially attributable to the fact that the magnitude of the noise within these two harmonics is equivalent to the magnitude of the signal. This ambiguity does reinforce the utility of the trumpet plot for providing order-of-magnitude estimates for the value of k_0.

6.4.2 R44N

In terms of E_μ^0 for the R44N mutant, there does appear to be a very small ($\sim$ 30 mV) shift in the value of E_μ^0. If there is in fact a shift, this smaller magnitude relative to the M66H mutant would be explained in light of the fact that the distal ligand at position 66 has a greater impact over the co-ordination environment of the Fe atom than the second-order effects of the 44^{th} residue, as can be observed in figure 6.1. It is easier to see this small shift relative to the wild-type by normalising both harmonics to lie between 0 and 1 (written as $\max(I)$, as shown in figure 6.4). By looking at the lobes of each of the r-FTACV harmonics, we can see a very slight negative potential shift for R44N relative to WT. In terms of k_0, the value inferred is very similar to the value inferred from the wild-type (200 vs. 175 s^{-1} respectively). This small change is corroborated by the similarity between the reductive arms of the trumpet-plot data from the WT data presented in section 5.3.5.3, and the R44N trumpet-plot data (as shown in figure 6.2 (G)). It should be noted that the upper capacitance fitting bound was altered to 3e-5 F (as reported in table 6.2), and that this value was returned when fitting the R44N PSV data. If the C_{dl} bound was set to 1e-5 F when fitting the R44N PSV data (as in section 5.3.3), then the returned k_0 value was totally reversible. As this kinetic value is incompatible with the R44N trumpet plot data, the value of the C_{dl} bound was increased. For the sake of consistency, we also used this bound when fitting M66H data. Consequently, it is difficult to assess whether the apparent difference between the WT and R44N k_0 values is driven by this bounding choice, or by the mutation.

In section 5.3.3, we showed that using consistent capacitance bounds returned a consistent value for k_0 across multiple experiments when fitting WT *Cj*X183 data. If this relationship holds, then that means that the M66H mutant has a higher k_0 than the R44N mutant, which has similar

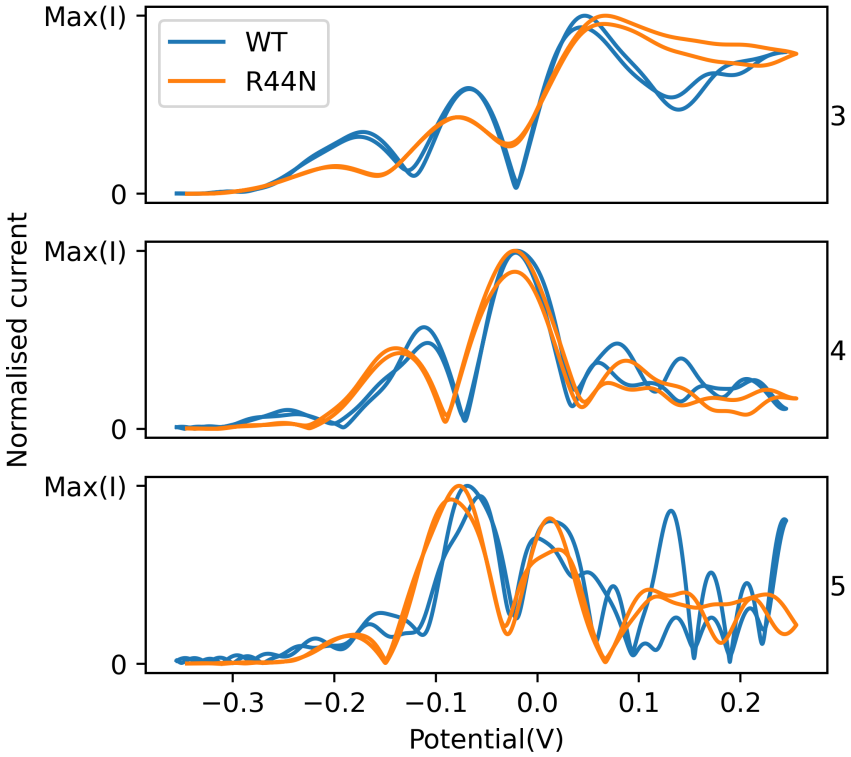


Figure 6.4: r-FTACV harmonics 3-5 from R44N and wild-type data, normalised so that the minimum value of each harmonic equals 0 and the maximum equals 1 (written as max(I) for clarity)

kinetics to the wild-type (as evidenced by the trumpet plot data), indicating that the Met-to-His mutation has increased the electron-transfer rate. However, what this mostly confirms is the issue of capacitance bounding when fitting PSV must resolved if we are to discuss more fine-scale changes in the apparent k_0. The inferred R44N surface coverage in table 6.1 is consistent with the value inferred using DCV peak-area analysis, and this could not be assessed for M66H as we could not determine the location of the Faradaic peak in the DCV data (as shown in figure 6.2 (D)). Interestingly, the α value hit the boundary at 0.4 rather than at 0.6, which has been the case for every PSV fitting attempt in chapters 4 and 5. It is also worthy of comment that the inferred E^0_σ parameter, when inferred by fitting to both PSV and r-FTACV *Cj*X183 data, has been highly consistent across multiple applications of protein to the electrode (within 10 mV for PSV and 4 mV for r-FTACV for all *Cj*X183 data analysed across chapters 4 and 5). Whether

this can be attributed to a consistent protein absorption process (resulting in a consistent distribution of E^0 values), or to some other peak-broadening process, such as Faradaic reactions that are better described by Marcus theory, is a topic of active enquiry.

6.5 Conclusions

Although there are some caveats about the quality and completeness of the experimental data, in this chapter we have shown that is possible to rapidly (on the order of two weeks) obtain good fits to rational mutants of *Cj*X183 using the methodology defined in section 5.4, once parameter space has been appropriately bounded. In future work, we will need to resolve how to model non-linear capacitance, because, as stated above, our choice of capacitance boundary does have an effect on the recovered k_0, which precludes discussion of fine-scale changes in the value of k_0 with mutation. Regardless, this is a set of encouraging results. In the first instance, the fact that we can consistently fit to *Cj*X183 data means that it will be possible to make deeper examinations of the underlying Faradaic reactions — such as resolving the interplay between dispersion, isotherms and Marcus theory that were discussed in section 2.2.7 and 3.3. Additionally, we set our sights on more complex biological electro-active molecules — ones exhibiting catalytic current, or with multiple interacting electro-active sites. The fact that we can consistently infer parameters about a biological Faradaic process (given reasonable parameter bounds) sets us in good stead for adapting to any additional modelling challenges these more complex processes will present.

Chapter 7

Conclusions

7.1 Summary

To conclude, we return to the research questions posed in chapter 1:

- Can we use PSV to obtain information about the electrochemical properties of a system, and does this information also predict how the system responds to r-FTACV and DCV experiments?
- How fast is this inference process, relative to r-FTACV analysis?
- Can the above analysis be applied to a biological system with poor Faradaic-to-background current ratios?
- To what extent can we capture fine-scale differences between biological systems using the above approach?

We will discuss how well we have answered these questions in turn. In chapter 4 we showed that we could fit PSV total data consistently when the Faradaic-to-background current ratio was favourable, with similar inferred parameters for ten sequential PSV runs performed on the same modified electrode, as shown in figures 4.2-4.6. These parameters also generated a simulation that was a good fit to r-FTACV current harmonics without modification, as shown in figure 4.10, and the surface-coverage values were consistent with those obtained from DCV analysis. The PSV inference process took significantly less time than would be required for r-FTACV — in the case of chapter 4, this was the difference between tens of minutes and several hours of computation time.

In chapter 5, we analysed PSV data from *Cj*X183, with less favourable Faradaic-to-background current ratios. As a result of this, it was no longer possible to obtain reasonable Faradaic parameters by fitting to PSV total current, and so we instead fitted to the higher harmonics of the Fourier transform of the PSV current. In comparison to the fitting process in chapter 4, we found multiple local minima in parameter space that provided a good fit to PSV harmonics. Consequently, comparison of the PSV-inferred parameters to r-FTACV harmonic data became an essential litmus test for the plausibility of a particular parameter vector. Initially, we did not find a single parameter vector that provided a good fit to PSV and r-FTACV harmonics simultaneously, and had to partially modify the most plausible PSV parameter vector to obtain good fits to the r-FTACV harmonics, as shown in figure 5.8. We believe that this discrepancy is driven by the way in which the linear capacitance dimension of parameter space was bounded, although this was necessary to prevent unrealistic solutions, such as presented in figure 5.5. Additionally, there was probably a contribution from the underlying inadequacy of our polynomial capacitance approximation, although the development of a more accurate model was beyond the scope of this thesis. Nevertheless, we were still able to obtain good fits to both r-FTACV and PSV harmonic data, over multiple different experiments; while the specific values reported are conditional on the modelling choices made, it is clear that the values reported are in tightly-constrained parameter space. Additionally, analysis of DCV observations, specifically the trumpet plot and charge-integration methods, was essential in providing order-of-magnitude estimates for the k_0 and Γ parameters. This strongly indicates that the more techniques incorporated into the workflow the better, especially if they allow for heuristic estimation of parameters in a way that is independent of the estimates of other parameters. The Bayesian inference approach was also invaluable for allowing an in-depth visualisation of the shape of the optimum in parameter space, as shown in figure 5.13. Bayesian inference will be a powerful tool for understanding the effects of parameter compensation and validating reaction models in future analyses. The chapter culminated in the development of a "recipe" for how to bound parameter space appropriately, and integrate observations from DCV, PSV and r-FTACV experiments.

With regards to the final research question, in chapter 5 we used two mutants of *Cj*X183 with altered thermodynamic properties to show that we could use the recipe defined in section 5.4 to capture fine-scale changes in the thermodynamic parameter E^0_μ, with a resolution of at least tens of millivolts. We also reported an increase in the inferred k_0 value of one mutant relative

to the other on the order of 60%, but noted that it was difficult to resolve to what degree this was driven by capacitance bounding choices. Nevertheless, this was a validation of the recipe in chapter 5; we were able to fit two datasets, and obtain some degree of insight into the effects of the mutations on the electrical activity of the wild-type, all of which was achieved within two weeks.

We also addressed several theoretical problems related to the forward and inverse problems, such as the problem of dispersion in 3.3, and an exploration of alternative approaches to inferring parameter values in appendix C. While these efforts were spurred by the need to continue working in the Covid-induced absence of experimental data, they were still useful exercises, and, particularly in the case of section 3.3, proved invaluable in contextualising the fitting results from experimental data.

7.2 Future work

7.2.1 Unresolved questions

7.2.1.1 Capacitance

The primary unresolved question in this thesis is the issue of bounding capacitance when fitting PSV harmonics. Although we did explore another approach to addressing capacitance in section 5.3.3.2 (the "two-step" method), we did not achieve the consistency between multiple experiments achieved by the simultaneous fitting approach. However, the simultaneous approach is clearly not ideal, as the extensive caveating about the k_0 parameters above shows. An essential next step is therefore to resolve this quandary, which could potentially be achieved by improving the two-step approach such that it becomes more consistent, for example by only fitting to portions of the experimental current known to contain purely capacitive current. However, this is only a stopgap for the true solution, which is to obtain a deeper understanding of the process of capacitance, using a physically-derived model (rather than the polynomial approximation used throughout this thesis) that can predict the non-linearity and phase dependency that we observe in real experimental data.

7.2.1.2 Uncompensated resistance

An additional area of interest is the accuracy of the inferred uncompensated resistance parameter. In chapters 4-6 we have used broad bounds when fitting the uncompensated resistance, and have assumed that the values returned have been accurate (as they have not hit either upper or lower bounds, and have all been of similar orders of magnitude). However, we have not validated these estimates. As a start, we can attempt to infer parameters from voltammetric current obtained with different electrolyte salt concentrations — as the concentration of salt decreases, the resistance should increase, and so at a minimum the estimates of the uncompensated resistance should be in the correct order. We would also like to incorporate electrochemical impedance spectroscopy into our toolkit for interrogating proteins, which is a technique used extensively to quantify the amount of resistance in an electrochemical system.

7.2.1.3 Dispersion

Another unresolved question is the interplay between thermodynamic dispersion, Marcus theory and isotherm theory. In response to broader-than-expected Faradaic peaks in experimental data, we have assumed this is explained by thermodynamic dispersion. However, peak broadening effects are also observed when using a Marcus-model of electrochemical kinetics, or modelling molecule self-interaction (i.e. not assuming a Langmuir isotherm). To assess the contribution from thermodynamic dispersion for example, we could attempt to infer parameters from an electrode modified with a mixture of WT and R44N *Cj*X183, and see if we can recover the true bimodal distribution of E^0 values. Additional experiments, such as increasing the viscosity of the electrolyte, could be used to assess the necessity of modelling Marcus, rather than Butler-Volmer kinetics.

7.2.2 More complex metalloproteins

The purpose of resolving the above modelling questions is to analyse progressively more complex metalloproteins. The first increase in complexity will be to infer parameters from proteins undergoing catalytic processes. This will be challenging, because FT voltammetry methods do not increase the amount of information about catalytic processes, as it does for Faradaic current arising from heterogeneous electron-transfer reactions. We will attempt to infer electrochemical parameters from a protein undergoing a single-electron redox process, and a catalytic process, such as the LPMOs discussed in the context of *Cj*X183 in chapter 5.

The second hurdle is that of multiple interacting electron-transfer centres, with E^0 values that are close within the potential range. Because the information about the two electron-transfer centres will be difficult to de-convolute, this will represent another challenge to our inferential procedures. A promising candidate for this is the WT/R44N protein mixture experiments proposed in section 7.2.1.3. Once we are satisfied we can infer accurate parameters from the two systems (catalytic and multiple electron-transfer centres), we will start finally to analyse a system which contains them both — the *Escherichia coli* hydrogenases. These enzymes are significantly more complex than *Cj*X183, containing four iron-sulphur clusters and undergoing catalytic reactions. Hyd-1 is oxygen tolerant, but produces hydrogen very slowly; Hyd-2 by contrast does not work in oxygen, but can catalyse the production of hydrogen at a rate equivalent to that of a platinum catalyst. The idea is to use site-directed mutagenesis approaches, coupled with our inferential workflow, to identify Hyd-1 mutants that can produce hydrogen at a higher rate while maintaining their oxygen tolerance.

This is a daunting task, and there will undoubtedly be many roadblocks along the way. To compensate for this, we hope firstly to expand the number of analytical techniques in our toolbox beyond DCV, PSV and r-FTACV, which both allows us to have more confidence in our inference results, and makes the workflow approach more attractive to scientists that use these techniques. This will include impedance (as above), and square-wave voltammetry, as these are widely used in the literature, and offer contrasting areas of insight to the other techniques we use. Secondly, we will build upon our preliminary theoretical results in optimising the features of the input potential to maximise the amount of information obtainable about various model parameters, developing a methodology for optimal experimental design. In terms of experimental design, there is also scope for optimising the waveform; for example, one might consider r-FTACV, PSV (and even EIS) as realisations of some input potential feature space. A natural question is therefore if we can sample directly for this feature space (varying parameters such as potential step, or number of overlapping sinusoids) to automatically design waveforms that offer improved resolution about parameters of interest.

In the long-term, once there is a toolkit to rapidly and reliably infer chemical information from protein voltammetry, a whole world of biological questions opens up. We can analyse how catalysis is driven by substrate movement, for example proton pumping, by inferring parameters at

different pHs. We can investigate how the catalytic properties of metalloenzymes are controlled by the co-ordination environment around the metal atoms, by scanning mutagenesis. The most exciting metalloenzymes are large and complex — the hydrogenases, nitrogenases and photosystems that can perform industrially relevant chemistries, such as producing hydrogen or fixing nitrogen, and do so without the need for expensive metal catalysts or large quantities of fossil fuels, all at ambient temperature and pressure. Rational mutagenesis could be used in conjunction with our workflow to make these proteins more stable, or functional in oxygen, whilst maintaining the desirable catalytic properties. There is also scope to uncover how sensitive the change in capacitance upon protein binding is to protein fold rearrangements.

7.3 Final thoughts

Electrochemistry is a vital force; for me, for you and for the whole of the living world besides. It will continue to be so for as long as there is life on this planet. The prospect of seeing a little part of that system — of holding a piece of that knowledge in one's hand — is tremendously exciting. The future is out there, and within each of us too.

Bibliography

[1] Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. "Electron transport and oxidative phosphorylation". In: *Molecular Cell Biology. 4th edition*. WH Freeman, 2000.

[2] Christine H Foyer, Jenny Neukermans, Guillaume Queval, Graham Noctor, and Jeremy Harbinson. "Photosynthetic control of electron transport and the regulation of gene expression". In: *Journal of Experimental Botany* 63.4 (2012), pp. 1637–1661.

[3] Jay R Winkler and Harry B Gray. "Electron flow through metalloproteins". In: *Chemical Reviews* 114.7 (2014), pp. 3369–3380.

[4] Antony R Crofts, J Todd Holland, Doreen Victoria, Derrick RJ Kolling, Sergei A Dikanov, Ryan Gilbreth, Sangmoon Lhee, Richard Kuras, and Mariana Guergova Kuras. "The Q-cycle reviewed: how well does a monomeric mechanism of the bc1 complex account for the function of a dimeric complex?" In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1777.7-8 (2008), pp. 1001–1019.

[5] Hope Adamson, Martin Robinson, Paul S Bond, Basem Soboh, Kathryn Gillow, Alexandr N Simonov, Darrell M Elton, Alan M Bond, R Gary Sawers, David J Gavaghan, and Alison Parkin. "Analysis of HypD disulfide redox chemistry via optimization of fourier transformed ac voltammetric data". In: *Analytical Chemistry* 89.3 (2017), pp. 1565–1573.

[6] Allen J Bard and Larry R Faulkner. *Electrochemical Methods: Fundamentals and Applications*. 2nd. Wiley, 2001.

[7] Rui Gao, Martin A Edwards, Joel M Harris, and Henry S White. "Shot noise sets the limit of quantification in electrochemical measurements". In: *Current Opinion in Electrochemistry* 22 (2020), pp. 170–177.

[8] Barry D Fleming, Jie Zhang, Darrell Elton, and Alan M Bond. "Detailed analysis of the electron-transfer properties of azurin adsorbed on graphite electrodes using dc and large-amplitude Fourier transformed ac voltammetry". In: *Analytical Chemistry* 79.17 (2007), pp. 6515–6526.

[9] Hope Adamson, Martin Robinson, John J Wright, Lindsey A Flanagan, Julia Walton, Darrell Elton, David J Gavaghan, Alan M Bond, Maxie M Roessler, and Alison Parkin. "Retuning the catalytic bias and overpotential of a [NiFe]-hydrogenase via a single amino acid exchange at the electron entry/exit site". In: *Journal of the American Chemical Society* 139.31 (2017), pp. 10677–10686.

[10] Alister R Dale-Evans, Martin J Robinson, Henry O Lloyd-Laney, David J Gavaghan, Alan M Bond, and Alison Parkin. "A voltammetric perspective of multi-electron and proton transfer in protein redox chemistry: Insights from computational analysis of Escherichia coli HypD Fourier transformed alternating current voltammetry". In: *Frontiers in Chemistry* 9 (2021), p. 424.

[11] Jaroslav Heyrovský and Jaroslav Kůta. *Principles of Polarography*. Academic Press: New York, 1966.

[12] Christopher G Bell, Costas A Anastassiou, Danny O'Hare, Kim H Parker, and Jennifer H Siggers. "Theory of high frequency, large-amplitude sinusoidal voltammetry for ideal surface-confined redox systems". In: *Electrochimica acta* 56.22 (2011), pp. 7569–7579.

[13] John T Long and Stephen G Weber. "Voltammetry in static and flowing solutions with a large-amplitude sine wave potential". In: *Electroanalysis* 4.4 (1992), pp. 429–437.

[14] Nicole E Hebert, Werner G Kuhr, and Sara A Brazill. "A microchip electrophoresis device with integrated electrochemical detection: A direct comparison of constant potential amperometry and sinusoidal voltammetry". In: *Analytical Chemistry* 75.14 (2003), pp. 3301–3307.

[15] Sara A Brazill, Sharin E Bender, Nicole E Hebert, John K Cullison, Eric W Kristensen, and Werner G Kuhr. "Sinusoidal voltammetry: a frequency based electrochemical detection technique". In: *Journal of Electroanalytical Chemistry* 531.2 (2002), pp. 119–132.

[16] Alan M Bond. "A perceived paucity of quantitative studies in the modern era of voltammetry: prospects for parameterisation of complex reactions in Bayesian and machine learning frameworks". In: *Journal of Solid State Electrochemistry* 24 (2020), pp. 2041–2050.

[17] Mark J Eddowes and H Allen O Hill. "Novel method for the investigation of the electrochemistry of metalloproteins: cytochrome c". In: *Journal of the Chemical Society, Chemical Communications* 21 (1977), 771b–772.

[18] Peter Yeh and Theodore Kuwana. "Reversible electrode reaction of cytochrome c". In: *Chemistry Letters* 6.10 (1977), pp. 1145–1148.

[19] Hope Adamson, Alan M Bond, and Alison Parkin. "Probing biological redox chemistry with large amplitude Fourier transformed ac voltammetry". In: *Chemical Communications* 53.69 (2017), pp. 9519–9533.

[20] Philip A Ash, Ricardo Hidalgo, and Kylie A Vincent. "Proton transfer in the catalytic cycle of [NiFe] hydrogenases: insight from vibrational spectroscopy". In: *ACS Catalysis* 7.4 (2017), pp. 2471–2485.

[21] Anup K Upadhyay, Doros T Petasis, Dave M Arciero, Alan B Hooper, and Michael P Hendrich. "Spectroscopic characterization and assignment of reduction potentials in the tetraheme cytochrome c554 from Nitrosomonas europaea". In: *Journal of the American Chemical Society* 125.7 (2003), pp. 1738–1747.

[22] Jessica H van Wonderen, Christopher R Hall, Xiuyun Jiang, Katrin Adamczyk, Antoine Carof, Ismael Heisler, Samuel EH Piper, Thomas A Clarke, Nicholas J Watmough, Igor V Sazanovich, Michael Towrie, Stephen R Meech, Jochen Blumberger, and Julea N Butt. "Ultrafast light-driven electron transfer in a Ru (II) tris (bipyridine)-labeled multiheme cytochrome". In: *Journal of the American Chemical Society* 141.38 (2019), pp. 15190–15200.

[23] Jessica H van Wonderen, Katrin Adamczyk, Xiaojing Wu, Xiuyun Jiang, Samuel EH Piper, Christopher R Hall, Marcus J Edwards, Thomas A Clarke, Huijie Zhang, Lars JC Jeuken, Igor V Sazanovich, Michael Towrie, Jochen Blumberger, Stephen R Meech, and Julea N Butt. "Nanosecond heme-to-heme electron transfer rates in a multiheme cytochrome nanowire reported by a spectrally unique His/Met-ligated heme". In: *Proceedings of the National Academy of Sciences* 118.39 (2021).

[24] Yingrui Dai, Yi Zheng, Greg M Swain, and Denis A Proshlyakov. "Equilibrium and kinetic behavior of Fe (CN) 63-/4- and cytochrome c in direct electrochemistry Using a film electrode thin-layer transmission cell". In: *Analytical chemistry* 83.2 (2011), pp. 542–548.

[25] Maurizio Bruschi, Raffaella Breglia, Federica Arrigoni, Piercarlo Fantucci, and Luca De Gioia. "Computational approaches to the prediction of the redox potentials of iron and copper bioinorganic systems". In: *International Journal of Quantum Chemistry* 116.22 (2016), pp. 1695–1705.

[26] Daniel Bim, Lubomir Rulisek, and Martin Srnec. "Computational electrochemistry as a reliable probe of experimentally elusive mononuclear nonheme iron species". In: *The Journal of Physical Chemistry C* 122.20 (2018), pp. 10773–10782.

[27] Dieter Britz and Jörg Strutwolf. *Digital simulation in electrochemistry*. Springer, 2005.

[28] David J Gavaghan and Alan M Bond. "A complete numerical simulation of the techniques of alternating current linear sweep and cyclic voltammetry: analysis of a reversible process by conventional and fast Fourier transform methods". In: *Journal of Electroanalytical Chemistry* 480.1-2 (2000), pp. 133–149.

[29] A C Fisher. *Electrode dynamics*. Oxford University Press, 1996.

[30] Keith Oldham and Jan Myland. *Fundamentals of electrochemical science*. Elsevier, 2012.

[31] Gerald B Folland. *Fourier analysis and its applications*. Vol. 4. American Mathematical Soc., 2009.

[32] James W Cooley and John W Tukey. "An algorithm for the machine calculation of complex Fourier series". In: *Mathematics of computation* 19.90 (1965), pp. 297–301.

[33] Pierre Duhamel and Martin Vetterli. "Fast Fourier transforms: a tutorial review and a state of the art". In: *Signal Processing* 19.4 (1990), pp. 259–299.

[34] Alan M Bond, Noel W Duffy, Si-Xuan Guo, Jie Zhang, and Darrell Elton. "Changing the look of voltammetry". In: *Analytical Chemistry* 77.9 (2005), 186–A.

[35] Richard I Masel. *Principles of adsorption and reaction on solid surfaces*. Vol. 3. John Wiley & Sons, 1996.

[36] Keng Yuen Foo and Bassim H Hameed. "Insights into the modeling of adsorption isotherm systems". In: *Chemical engineering journal* 156.1 (2010), pp. 2–10.

[37] Graham P Morris, Ruth E Baker, Kathryn Gillow, Jason J Davis, David J Gavaghan, and Alan M Bond. "Theoretical analysis of the relative significance of thermodynamic and kinetic dispersion in the dc and ac voltammetry of surface-confined molecules". In: *Langmuir* 31.17 (2015), pp. 4996–5004.

[38] Rubin Gulaboski, Valentin Mirčeski, Ivan Bogeski, and Markus Hoth. "Protein film voltammetry: electrochemical enzymatic spectroscopy. A review on recent progress". In: *Journal of Solid State Electrochemistry* 16.7 (2012), pp. 2315–2328.

[39] Christophe Léger and Patrick Bertrand. "Direct electrochemistry of redox enzymes as a tool for mechanistic studies". In: *Chemical Reviews* 108.7 (2008), pp. 2379–2438.

[40] Razvan C Stan, Alexander Kros, Namik Akkilic, Nusrat JM Sanghamitra, and Jeroen Appel. "Direct wiring of the azurin redox center to gold electrodes investigated by protein film voltammetry". In: *Journal of Electroanalytical Chemistry* 787 (2017), pp. 14–18.

[41] Fraser A Armstrong, Hendrik A Heering, and Judy Hirst. "Reaction of complex metalloproteins studied by protein-film voltammetry". In: *Chemical Society Reviews* 26.3 (1997), pp. 169–179.

[42] Fraser A Armstrong. "Insights from protein film voltammetry into mechanisms of complex biological electron-transfer reactions". In: *Journal of the Chemical Society, Dalton Transactions* 5 (2002), pp. 661–671.

[43] Nicholas D Yates, Mark R Dowsett, Phillip Bentley, Jack A Dickenson-Fogg, Andrew Pratt, Christopher F Blanford, Martin A Fascione, and Alison Parkin. "Aldehyde-mediated protein-to-surface tethering via controlled diazonium electrode functionalization using protected hydroxylamines". In: *Langmuir* 36.20 (2019), pp. 5654–5664.

[44] Judy Hirst and Fraser A Armstrong. "Fast-scan cyclic voltammetry of protein films on pyrolytic graphite edge electrodes: characteristics of electron exchange". In: *Analytical Chemistry* 70.23 (1998), pp. 5062–5071.

[45] Bernard Munge, Somes K Das, Robielyn Ilagan, Zeus Pendon, Jing Yang, Harry A Frank, and James F Rusling. "Electron transfer reactions of redox cofactors in spinach photosystem I reaction center protein in lipid films on electrodes". In: *Journal of the American Chemical Society* 125.41 (2003), pp. 12457–12463.

[46] Chong-Yong Lee, Barry D Fleming, Jie Zhang, Si-Xuan Guo, Darrell M Elton, and Alan M Bond. "Systematic evaluation of electrode kinetics and impact of surface heterogeneity for surface-confined proteins using analysis of harmonic components available in sinusoidal large-amplitude Fourier transformed ac voltammetry". In: *Analytica Chimica Acta* 652.1-2 (2009), pp. 205–214.

[47] Rubin Gulaboski, Valentin Mirčeski, and Milivoj Lovrić. "Square-wave protein-film voltammetry: new insights in the enzymatic electrode processes coupled with chemical reactions". In: *Journal of Solid State Electrochemistry* 23.8 (2019), pp. 2493–2506.

[48] Suzannah V Hexter, Thomas F Esterle, and Fraser A Armstrong. "A unified model for surface electrocatalysis based on observations with enzymes". In: *Physical Chemistry Chemical Physics* 16.24 (2014), pp. 11822–11833.

[49] Vincent Fourmond, Eric S Wiedner, Wendy J Shaw, and Christophe Léger. "Understanding and design of bidirectional and reversible catalysts of multielectron, multistep reactions". In: *Journal of the American Chemical Society* 141.28 (2019), pp. 11269–11285.

[50] Vincent Fourmond and Christophe Léger. "Modelling the voltammetry of adsorbed enzymes and molecular catalysts". In: *Current Opinion in Electrochemistry* 1.1 (2017), pp. 110–120.

[51] Judy Hirst. "Elucidating the mechanisms of coupled electron transfer and catalytic reactions by protein film voltammetry". In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1757.4 (2006), pp. 225–239.

[52] Christophe Léger, Anne K Jones, Simon PJ Albracht, and Fraser A Armstrong. "Effect of a dispersion of interfacial electron transfer rates on steady state catalytic electron transport in [NiFe]-hydrogenase and other enzymes". In: *The Journal of Physical Chemistry B* 106.50 (2002), pp. 13058–13063.

[53] Meriem Merrouch, Jessica Hadj-Saıd, Christophe Léger, Sebastien Dementin, and Vincent Fourmond. "Reliable estimation of the kinetic parameters of redox enzymes by taking into account mass transport towards rotating electrodes in protein film voltammetry experiments". In: *Electrochimica Acta* 245 (2017), pp. 1059–1064.

[54] Torsten Reda and Judy Hirst. "Interpreting the catalytic voltammetry of an adsorbed enzyme by considering substrate mass transfer, enzyme turnover, and interfacial electron transport". In: *The Journal of Physical Chemistry B* 110.3 (2006), pp. 1394–1404.

[55] W. Keith Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (1970), pp. 97–109.

[56] Christian P Robert and George Casella. "The Metropolis—Hastings Algorithm". In: Monte Carlo Statistical Methods. Springer, 1999, pp. 231–283.

[57] Heikki Haario, Eero Saksman, and Johanna Tamminen. "An adaptive Metropolis algorithm". In: *Bernoulli* 7.2 (2001), pp. 223–242.

[58] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Nov. 2013. DOI: 10.1201/b16018. URL: https://doi.org/10.1201/b16018.

[59] Martin Z Bazant, Kevin T Chu, and Bruce J Bayly. "Current-voltage relations for electrochemical thin films". In: *SIAM Journal on Applied Mathematics* 65.5 (2005), pp. 1463–1484.

[60] Martin Z Bazant, Katsuyo Thornton, and Armand Ajdari. "Diffuse-charge dynamics in electrochemical systems". In: *Physical Review E* 70.2 (2004), p. 021506.

[61] Ian Naismith Sneddon. *Special functions of mathematical physics and chemistry*. Oliver and Boyd, 1961.

[62] E Laviron. "General expression of the linear potential sweep voltammogram in the case of diffusionless electrochemical systems". In: *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* 101.1 (1979), pp. 19–28.

[63] Gary K Rowe, Michael T Carter, John N Richardson, and Royce W Murray. "Consequences of kinetic dispersion on the electrochemistry of an adsorbed redox-active monolayer". In: *Langmuir* 11.5 (1995), pp. 1797–1806.

[64] Rose A Clark and Edmond F Bowden. "Voltammetric peak broadening for cytochrome c/alkanethiolate monolayer structures: dispersion of formal potentials". In: *Langmuir* 13.3 (1997), pp. 559–565.

[65] Zhe Zhang and James F Rusling. "Electron transfer between myoglobin and electrodes in thin films of phosphatidylcholines and dihexadecylphosphate". In: *Biophysical chemistry* 63.2-3 (1997), pp. 133–146.

[66] Vincent Fourmond, Carole Baffert, Kateryna Sybirna, Thomas Lautier, Abbas Abou Hamdan, Sébastien Dementin, Philippe Soucaille, Isabelle Meynial-Salles, Hervé Bottin, and Christophe Léger. "Steady-state catalytic wave-shapes for 2-electron reversible electrocatalysts and enzymes". In: *Journal of the American Chemical Society* 135.10 (2013), pp. 3926–3938.

[67] Amol V Patil and Jason J Davis. "Visualizing and tuning thermodynamic dispersion in metalloprotein monolayers". In: *Journal of the American Chemical Society* 132.47 (2010), pp. 16938–16944.

[68] Jante M Salverda, Amol V Patil, Giulia Mizzon, Sofya Kuznetsova, Gerhild Zauner, Namik Akkilic, Gerard W Canters, Jason J Davis, Hendrik A Heering, and Thijs J Aartsma. "Fluorescent cyclic voltammetry of immobilized azurin: direct observation of thermodynamic and kinetic heterogeneity". In: *Angewandte Chemie International Edition* 49.33 (2010), pp. 5776–5779.

[69] Peter K Gessner and Mazen M Hasan. "Freundlich and Langmuir isotherms as models for the adsorption of toxicants on activated charcoal". In: *Journal of Pharmaceutical Sciences* 76.4 (1987), pp. 319–327.

[70] Gareth P Stevenson, Ruth E Baker, Gareth F Kennedy, Alan M Bond, David J Gavaghan, and Kathryn Gillow. "Access to enhanced differences in Marcus–Hush and Butler–Volmer electron transfer theories by systematic analysis of higher order AC harmonics". In: *Physical Chemistry Chemical Physics* 15.6 (2013), pp. 2210–2221.

[71] Eduardo Laborda, Martin C Henstridge, Christopher Batchelor-McAuley, and Richard G Compton. "Asymmetric Marcus–Hush theory for voltammetry". In: *Chemical Society Reviews* 42.12 (2013), pp. 4894–4905.

[72] Venkatasubramanian Viswanathan, Heine Anton Hansen, Jan Rossmeisl, Thomas Francisco Jaramillo, Heinz Pitsch, and Jens K Nørskov. "Simulating linear sweep voltammetry from first-principles: application to electrochemical oxidation of water on Pt (111) and Pt3Ni (111)". In: *The Journal of Physical Chemistry C* 116.7 (2012), pp. 4698–4704.

[73] Todd P Silverstein. "Marcus theory: thermodynamics CAN control the kinetics of electron transfer reactions". In: *Journal of Chemical Education* 89.9 (2012), pp. 1159–1167.

[74] John R Pladziewicz and James H Espenson. "Electron transfer reactions of ferrocenes". In: *The Journal of Physical Chemistry* 75.21 (1971), pp. 3381–3382.

[75] RG Bates and JB Macaskill. "Standard potential of the silver-silver chloride electrode". In: *Pure and Applied Chem* 50.11-12 (1978), pp. 1701–1706.

[76] Christophe Léger and Patrick Bertrand. "Direct electrochemistry of redox enzymes as a tool for mechanistic studies". In: *Chemical Reviews* 108.7 (2008), pp. 2379–2438.

[77] Leila Lo Leggio, Thomas J Simmons, Jens-Christian N Poulsen, Kristian EH Frandsen, Glyn R Hemsworth, Mary A Stringer, Pernille Von Freiesleben, Morten Tovborg, Katja S Johansen, Leonardo De Maria, Paul V Harris, Chee-Leong Soong, Paul Dupree, Theodora Tryfona, Nicolas Lenfant, Bernard Henrissat, Gideon J Davies, and Paul H Walton. "Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase". In: *Nature Communications* 6.1 (2015), pp. 1–9.

[78] Jessie Branch, Badri S Rajagopal, Alessandro Paradisi, Nick Yates, Peter J Lindley, Jake Smith, Kristian Hollingsworth, W Bruce Turnbull, Bernard Henrissat, Alison Parkin, Alan Berry, and Glyn R Hemsworth. "C-type cytochrome-initiated reduction of bacterial lytic polysaccharide monooxygenases". In: *Biochemical Journal* 478.14 (2021), pp. 2927–2944.

[79] Junjun Mao, Karin Hauser, and MR Gunner. "How cytochromes with different folds control heme redox potentials". In: *Biochemistry* 42.33 (2003), pp. 9829–9840.

[80] Richard Hendler and Gurmel Sidhu. "A new high potential redox transition for cytochrome aa3". In: *Biophysical journal* 54.1 (1988), pp. 121–133.

[81] Susanne J́unemann, John M Wrigglesworth, and Peter R Rich. "Effects of decyl-aurachin D and reversed electron transfer in cytochrome bd". In: *Biochemistry* 36.31 (1997), pp. 9323–9331.

[82] Shan Jian, Xingchao Liu, Hong Sun, and Shifeng Hou. "The electrochemical studies of cytochrome c incorporated in 3D porous calcium alginate films on glassy carbon electrodes". In: *RSC Advances* 4.12 (2014), pp. 6165–6172.

[83] Daniel L Johnson, Alan J Conley, and Lisa L Martin. "Direct electrochemistry of human, bovine and porcine cytochrome P450c17". In: *Journal of molecular endocrinology* 36.2 (2006), pp. 349–359.

[84] JJ Wei, Haiying Liu, K Niki, E Margoliash, and DH Waldeck. "Probing electron tunneling pathways: Electrochemical study of rat heart cytochrome c and its mutant on pyridine-terminated SAMs". In: *The Journal of Physical Chemistry B* 108.43 (2004), pp. 16912–16917.

[85] Stephen W Feldberg. "Effect of uncompensated resistance on the cyclic voltammetric response of an electrochemically reversible surface-attached redox couple: Uniform current and potential across the electrode surface". In: *Journal of Electroanalytical Chemistry* 624.1-2 (2008), pp. 45–51.

[86] Parisa Hosseinzadeh and Yi Lu. "Design and fine-tuning redox potentials of metalloproteins involved in electron transfer in bioenergetics". In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1857.5 (2016), pp. 557–581.

[87] Jing Liu, Saumen Chakraborty, Parisa Hosseinzadeh, Yang Yu, Shiliang Tian, Igor Petrik, Ambika Bhagi, and Yi Lu. "Metalloproteins containing cytochrome, iron–sulfur, or copper redox centers". In: *Chemical Reviews* 114.8 (2014), pp. 4366–4469.

[88] Kara L Bren. "Going with the electron flow: Heme electronic structure and electron transfer in cytochrome c". In: *Israel Journal of Chemistry* 56.9-10 (2016), pp. 693–704.

[89] Richard H Holm, Pierre Kennepohl, and Edward I Solomon. "Structural and functional aspects of metal sites in biology". In: *Chemical Reviews* 96.7 (1996), pp. 2239–2314.

[90] Adrienne L Raphael and Harry B Gray. "Axial ligand replacement in horse heart cytochrome c by semisynthesis". In: *Proteins: Structure, Function, and Bioinformatics* 6.3 (1989), pp. 338–340.

[91] A Dolla, L Florens, P Bianco, J Haladjian, G Voordouw, E Forest, J Wall, F Guerlesquin, and M Bruschi. "Characterization and oxidoreduction properties of cytochrome c3 after heme axial ligand replacements." In: *Journal of Biological Chemistry* 269.9 (1994), pp. 6340–6346.

[92] Diganta K Das and Okhil K Medhi. "The role of heme propionate in controlling the redox potential of heme: Square wave voltammetry of protoporphyrinato IX iron (III) in aqueous surfactant micelles". In: *Journal of Inorganic Biochemistry* 70.2 (1998), pp. 83–90.

[93] Victor Guallar and Brett Olsen. "The role of the heme propionates in heme biochemistry". In: *Journal of Inorganic Biochemistry* 100.4 (2006), pp. 755–760.

[94] Robert G Parr and Ralph G Pearson. "Absolute hardness: companion parameter to absolute electronegativity". In: *Journal of the American Chemical Society* 105.26 (1983), pp. 7512–7516.

[95] Francis Sears, Mark Zemansky, and Hugh Young. *University Physics*. 6th. Westley Publishing Company Reading, MA, 1982.

[96] Margareta RA Blomberg. "Active site midpoint potentials in different cytochrome c oxidase families: a computational comparison". In: *Biochemistry* 58.15 (2019), pp. 2028–2038.

[97] William W Porterfield. *Inorganic chemistry*. Academic press, 2013.

[98] Ehesan Ali, Biplab Sanyal, and Peter M Oppeneer. "Electronic structure, spin-states, and spin-crossover reaction of heme-related Fe-porphyrins: a theoretical perspective". In: *The Journal of Physical Chemistry B* 116.20 (2012), pp. 5849–5859.

[99] Gregory T Miller, Balli Zhang, John K Hardman, and Russell Timkovich. "Converting a c-type to a b-type Cytochrome: Met61 to His61 mutant of Pseudomonas Cytochrome c-551". In: *Biochemistry* 39.30 (2000), pp. 9010–9017.

[100] Matthew D Liptak, Robert D Fagerlund, Elizabeth C Ledgerwood, Sigurd M Wilbanks, and Kara L Bren. "The proapoptotic G41S mutation to human cytochrome c alters the heme electronic structure and increases the electron self-exchange rate". In: *Journal of the American Chemical Society* 133.5 (2011), pp. 1153–1155.

[101] Mark A Richards. "The discrete-time Fourier transform and discrete Fourier transform of windowed stationary white noise". In: *Georgia Institute of Technology, Technical Memorandum* (2013).

[102] Nikolaus Hansen. "The CMA evolution strategy: A tutorial". In: *arXiv preprint arXiv:1604.00772* (2016).

[103] Rubin Gulaboski. "Theoretical contribution towards understanding specific behaviour of "simple" protein-film reactions in square-wave voltammetry". In: *Electroanalysis* 31.3 (2019), pp. 545–553.

[104] Rubin Gulaboski, Valentin Mirčeski, and Milivoj Lovrić. "Critical aspects in exploring time analysis for the voltammetric estimation of kinetic parameters of surface electrode mechanisms coupled with chemical reactions". In: *Macedonian Journal of Chemistry and Chemical Engineering* 40.1 (2021), pp. 1–9.

[105] Valentin Mirčeski and Milivoj Lovrić. "Split square-wave voltammograms of surface redox reactions". In: *Electroanalysis* 9.16 (1997), pp. 1283–1287.

[106] Valentin Mirčeski, Eduardo Laborda, Dariusz Guziejewski, and Richard G Compton. "New approach to electrode kinetic measurements in square-wave voltammetry: amplitude-based quasireversible maximum". In: *Analytical Chemistry* 85.11 (2013), pp. 5586–5594.

[107] Valentin Mirčeski, Šebojka Komorsky-Lovrić, and Milivoj Lovrić. *Square-wave voltammetry: theory and application*. Ed. by F Scholz. Springer Science & Business Media, 2007.

Appendices

Appendix A

The discrete Fourier transform of Gaussian noise

In this appendix, we show that the discrete Fourier transform (DFT) of a Gaussian random independent and identically distributed complex variable v_n itself has a Gaussian distribution [101]. We define v_n such that

$$v_n \sim \mathcal{N}_{\mathcal{C}}(0, \sigma^2) \tag{A.1}$$

where $\mathcal{N}_{\mathcal{C}}$ is the complex normal distribution and σ is the standard deviation. As in equation 2.40, although with the normalisation factor N^{-1}, the DFT of v_n is

$$\mathscr{F}(k) = \frac{1}{N} \sum_{n=0}^{N-1} v_n e^{-i\frac{2\pi}{N}nk}. \tag{A.2}$$

By using Euler's identity ($e^{ix} = \cos(x) + i\sin(x)$), we can rewrite equation A.2 in terms of the real R, and imaginary I components of $\mathscr{F}[k]$

$$\mathscr{F}[k] = \frac{1}{N} \sum_{n=0}^{N-1} \big(R(v_n) + iI(v_n)\big)\big(\cos\left(\frac{2\pi}{N}nk\right) + i\sin(\frac{2\pi}{N}nk)\big). \tag{A.3}$$

We initially focus on the expression for the real part of the DFT, $R(\mathscr{F}[k])$, which can be written as

$$R(\mathscr{F}[k]) = \frac{1}{N} \sum_{n=0}^{N-1} R(v_n)\cos\left(\frac{2\pi}{N}nk\right) + \frac{1}{N} \sum_{n=0}^{N-1} I(v_n)\sin\left(\frac{2\pi}{N}nk\right). \tag{A.4}$$

To obtain information about how the computed Fourier components are distributed, we require some properties of the regular and complex Gaussian distributions,

1. $x \sim \mathcal{N}(0, \sigma_1^2)$, $y \sim \mathcal{N}(0, \sigma_2^2)$, $x + y \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$,
2. $R(v_n) \sim \mathcal{N}(0, \frac{\sigma^2}{2})$, $I(v_n) \sim \mathcal{N}(0, \frac{\sigma^2}{2})$,
3. $cv_n \sim \mathcal{N}(0, c^2\sigma^2)$,
4. $\sum_{n=0}^{N-1} v_n \sim \mathcal{N}(0, \sum_{n=0}^{N-1} \sigma^2)$.

To illustrate, a single element of the first sum in equation A.4 is distributed according to

$$R(v_n)\frac{1}{N}\cos\left(\frac{2\pi}{N}nk\right) \sim \mathcal{N}(0, \frac{1}{2N^2}\cos\left(\frac{2\pi}{N}nk\right)^2\sigma^2), \tag{A.5}$$

and the full expression for the distribution of the first sum in equation A.4 is

$$\frac{1}{N}\sum_{n=0}^{N-1} R(v_n)\cos\left(\frac{2\pi}{N}nk\right) \sim \mathcal{N}(0, \frac{1}{2N^2}\sum_{n=0}^{N-1}\cos\left(\frac{2\pi}{N}nk\right)^2\sigma^2). \tag{A.6}$$

Using Euler's formula, we can show that

$$\sum_{n=0}^{N-1}\left(\cos\left(\frac{2\pi}{N}nk\right) + i\sin\left(\frac{2\pi}{N}nk\right)\right) = \sum_{n=0}^{N-1} e^{\left(i\frac{2\pi}{N}k\right)^n}, \tag{A.7}$$

and since the last term in equation A.7 is a geometric series, and consequently can be written as

$$\sum_{n=0}^{N-1} e^{\left(i\frac{2\pi}{N}k\right)^n} = \frac{e^{\left(i\frac{2\pi}{N}k\right)^N} - 1}{e^{\left(i\frac{2\pi}{N}k\right)} - 1} = 0. \tag{A.8}$$

This can be used to obtain

$$\sum_{n=0}^{N-1}\cos\left(\frac{2\pi}{N}nk\right)^2 = \sum_{n=0}^{N-1}\frac{\left(1 + \cos\left(\frac{2\pi}{N}nk\right)\right)}{2} = \frac{N}{2}, \tag{A.9}$$

to reduce the distribution in equation A.6 down to

$$\frac{1}{N}\sum_{n=0}^{N-1} R(v_n)\cos\left(\frac{2\pi}{N}nk\right) \sim \mathcal{N}\left(0, \frac{\sigma^2}{4N}\right). \tag{A.10}$$

We can repeat the same process for the second sum in equation A.4, noting that

$$\sum_{n=0}^{N-1} \sin\left(\frac{2\pi}{N}nk\right)^2 = \frac{N}{2}. \tag{A.11}$$

and consequently

$$R(\mathscr{F}[k]) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2N}\right). \tag{A.12}$$

Because the Gaussian distribution is symmetrical, its real and imaginary components are identically distributed, and therefore

$$\mathscr{F}[k] \sim \mathcal{N}_{\mathcal{C}}\left(0, \frac{\sigma^2}{N}\right). \tag{A.13}$$

Appendix B

The CMAES algorithm

B.1 Overview

Given a multi-parameter optimisation problem, the covariance matrix optimisation evolutionary strategy (CMAES) algorithm optimises the mean and covariance of a multivariate normal distribution, such that the shape defined by the covariance matrix, a d-dimensional ellipsoid, approximates the contours of the function to be minimised. An example of this approach for a 2D problem can be seen in figure B.1. CMAES has many advantages for challenging optimisation problems, as it is designed to find the minima of multi-modal and rugged functions with high dimensionality, which often present challenges for other optimisation methods. For each step of the algorithm, a number of search points, each comprising a parameter vector, is drawn from a multivariate Gaussian distribution. The best candidates (i.e. those with the lowest scores) are used to update the mean and the covariance matrix of the distribution, such that this update increases the chance of taking steps in parameter space that result in higher scores for the likelihood function. In addition, the covariance matrix is updated using information from prior iterations (the evolution path), and adjusted to alter the scale, changing the size of the "steps" the algorithm takes in parameter space. The CMAES algorithm was selected for a number of reasons; it does not require the calculation of derivatives and it was explicitly designed to find a global optimum in high-dimensional parameter space. A simplified version of the algorithm is shown in algorithm 2.

$g \leftarrow 0$;

$\lambda \leftarrow$ Population size;

$\mathbf{m}^{(0)} \leftarrow$ Initial distribution mean;

$\mathbf{C}^{(0)} \leftarrow I$;

$\mathbf{p}_\sigma{}^{(0)} \leftarrow 0$;

$\mathbf{p_c}^{(0)} \leftarrow 0$;

$\sigma^{(0)} \leftarrow$ Initial step size;

while *termination condition not reached* **do**

`/* Sample new solutions */`

for $i = \{1..\lambda\}$ **do**

$\mathbf{x}_i^{(g+1)} \sim \mathcal{N}(\mathbf{m}^{(g)}, \sigma^{(g)2}\mathbf{C}^{(g)})$;

$f_i \leftarrow \mathcal{L}_t(x_i^{(g+1)})$;

end

`/* Sort sample solutions according to score */`

$s \leftarrow$**argsort**$(f_{1...\lambda})$;

$\mathbf{x}_{1...\lambda}^{(g+1)} \leftarrow \mathbf{x}_{s(1)...s(\lambda)}^{(g+1)}$;

`/* Select and recombine best scoring solutions */`

$\mathbf{m}^{(g+1)} \leftarrow$**update_m**$(\mathbf{x}_1^{(g+1)}, ..., \mathbf{x}_\lambda^{(g+1)})$;

`/* Update cumulation path */`

$\mathbf{p_c} \leftarrow$**update_pc**$(\mathbf{p_c}^{(g)}, \sigma^{(g)^{-1}}(\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}))$;

`/* Update covariance matrix */`

$\mathbf{C}^{(g+1)} \leftarrow$**update_covariance**$(\mathbf{C}^{(g)}, \mathbf{p_c}^{(g+1)}, \sigma^{(g)^{-1}}(\mathbf{x}_{(1,...,\lambda)}^{(g+1)} - \mathbf{m}^{(g)}))$;

`/* Update step-size path */`

$\mathbf{p}_\sigma{}^{(g+1)} \leftarrow$**update_ps**$(\mathbf{p}_\sigma{}^{(g)}, \sigma^{(g)^{-1}}\mathbf{C}^{(g)^{-\frac{1}{2}}}\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)})$;

`/* Update covariance matrix scale */`

$\sigma^{(g+1)} \leftarrow$**update_sigma**$(\sigma, ||\mathbf{p}_\sigma{}^{(g+1)}||)$;

$g \leftarrow g + 1$

end

Algorithm 2: The Covariance Matrix Adaptation Evolutionary Strategy (CMAES) algorithm, using the time-domain likelihood function defined in equation 2.54

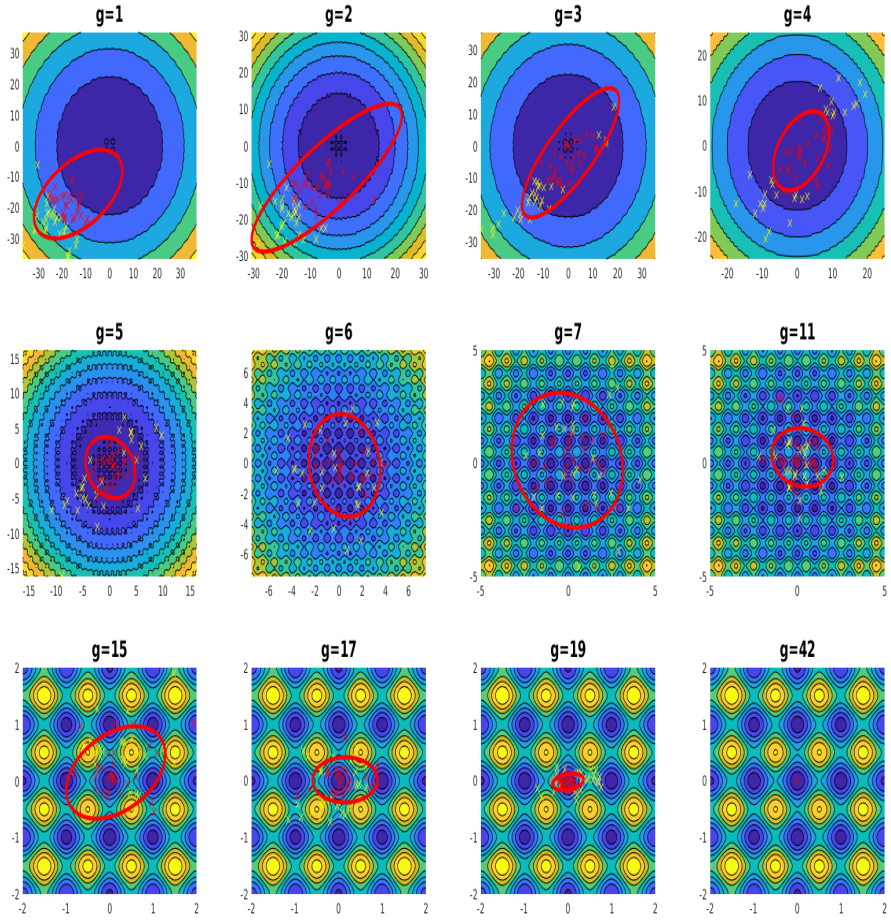


Figure B.1: Iterations of the CMAES algorithm finding the minimum of the 2D Rastrigin function, which has a minimum at [0,0], shown here as a filled contour plot, with blue areas and yellow areas representing local maxima and minima respectively. Crosses in red indicate the highest scoring points that were used as part of the calculation of the covariance matrix $\mathbf{C}^{(g+1)}$, shown as a red ellipse with a scale factor of 0.95. $\mathbf{C}^{(g+1)}$ was then used to generate points for the next iteration. Crosses in yellow were discarded and not used to update the covariance. The initial mean was set at [-10, -10].

The state variables of the algorithm are:

- $\mathbf{m}$, the mean of the current distribution
- $\mathbf{C}$, the covariance matrix

- $\mathbf{p_c}$, the anisotropic evolution path
- $\mathbf{p}_\sigma$, the isotropic evolution path
- σ, the step-size, which is used to scale the covariance matrix

We will discuss the update of each of these state variables in turn.

B.2 Updating the mean

At each generation g of the algorithm, λ n-dimensional vectors (where n is the problem dimension), are drawn randomly from the current multivariate Gaussian distribution, $\mathbf{x}_i^{(g+1)} \sim \mathcal{N}(\mathbf{m}^{(g)}, \sigma^2\mathbf{C}^{(g)})$ for $i = 1, ..., \lambda$. These are obtained by drawing samples z_i from the isotropic Gaussian distribution,

$$z_i \sim \mathcal{N}(0, \mathbf{I}), \tag{B.1}$$

where I is the identity matrix. Scaling the result appropriately,

$$\mathbf{x}_i^{(g+1)} = \mathbf{m}^{(g)} + \sigma^{(g)}\mathbf{B}\mathbf{D}z_i \sim \mathbf{m}^{(g)} + \sigma^{(g)} + \mathcal{N}(0, \mathbf{C}^{(g)}), \tag{B.2}$$

where $\mathbf{B}$ is an orthogonal matrix, each column of $\mathbf{B}$ is a normalised eigenvector of $\mathbf{C}^{(g)}$, and $\mathbf{D}$ is a diagonal matrix where the diagonal entries are the square roots of eigenvectors of $\mathbf{C}^{(g)}$. This arises from the eigendecomposition of $\mathbf{C}^{(g)}$, such that $\mathbf{C}^{(g)^{\frac{1}{2}}} = \mathbf{B}\mathbf{D}\mathbf{B}^T$. Each solution vector is evaluated according to the objective function (i.e. either equation (2.54) or (2.55)), and the μ lowest scoring candidates are selected, where $\mu \leq \lambda$. The μ vectors are then sorted according to their score, and used to calculate the mean for the next generation $\mathbf{m}^{(g+1)}$. Each solution is weighted, so that the lowest scoring candidates have a higher impact on the updated mean,

$$\mathbf{m}^{(g+1)} = \mathbf{m}^g + c_m \sum_{i=1}^{\mu} w_i(\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}). \tag{B.3}$$

$\mathbf{x}_{i:\lambda}$ has been sorted according to the value of the likelihood function f, such that $f(\mathbf{x}_{1:\lambda}^{(g+1)}) \geq f(\mathbf{x}_{2:\lambda}^{(g+1)}) \geq ... \geq (\mathbf{x}_{\lambda:\lambda}^{(g+1)})$, and the weights are defined such that

$$\sum_{i=1}^{\mu} w_i = 1, w_1 \geq w_2 \geq ...w_\mu > 0. \tag{B.4}$$

Finally, c_m is a learning parameter — if $c_m = 1$ then $\mathbf{m}^{(g+1)}$ is dependent only on the current μ best solutions.

B.3 Updating the covariance matrix

An update to the covariance matrix $\mathbf{C}^{(g+1)}$ is calculated using the best sample *transitions* of the generation $g+1$ is as in equation (B.5), with weights defined as in equation (B.4). This biases the update of the covariance matrix towards lower-scoring solutions. Then we have,

$$\mathbf{C}^{(g+1)} = \sum_{i=1}^{\mu} w_i (\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)})(\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)})^T. \tag{B.5}$$

In equation (B.5), we are estimating the covariance of the selected steps, or the deviation from the mean of the previous generation. By using other reference mean values (and appropriate normalisations), we could estimate the covariance of every sample point in the new generation ($\mathbf{x}_{1,\ldots,\lambda}^{(g+1)}$), with the reference mean being $\frac{1}{\lambda}\sum_{j=1}^{\lambda} = \mathbf{x}_j^{(g+1)}$, or the covariance of the selected sample only ($\mathbf{x}_{1,\ldots,\mu}^{(g+1)}$) by using $\mathbf{m}^{(g+1)}$ as the reference. The intention behind using $\mathbf{m}^{(g)}$ as the reference is that the updated covariance matrix $\mathbf{C}^{(g+1)}$ will tend to produce *steps* that result in better scores more often. This can be observed in the first iteration of figure B.1, where the covariance matrix points towards the origin more often than would be expected if we were merely estimating the covariance of the red points.

The rest of the state variables — the step size and two evolutionary path variables $\sigma, \mathbf{p_c}$ and $\mathbf{p}_\sigma$ — are included to modify the covariance matrix to improve convergence to the minimum, robustness of search or speed of the algorithm.

B.4 Rank-μ update

To ensure fast searching, the number of evaluated solutions λ must be relatively small. However, this will decrease the reliability of equation (B.5) as an estimator of the covariance matrix. In order to remedy this, previous covariance matrices are used as part of the updating process. This is expressed as

$$\mathbf{C}^{(g+1)} = (1 - c_\mu)\mathbf{C}^{(g)} + c_\mu \sum_{i=1}^{\mu} w_i \left(\frac{\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}} \right) \left(\frac{\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}} \right)^T. \tag{B.6}$$

$\sigma^{(g)}$ is included to normalise the covariance matrix, so that two covariance matrices can be compared between generations with different σ values. c_μ is a learning parameter and defines the relative contribution of each generation of the covariance matrix to the updated value. This weighting of different generations is referred to as exponential smoothing [102].

B.5 Rank-one update

In the calculation of $\mathbf{C}^{(g+1)}$ in equation (B.6), the sign is neglected when updating the covariance matrix. To reinsert this sign information, $\mathbf{p_c}$, the anisotropic evolution path, is introduced into the update of the covariance matrix. For generation g, the evolution path records the transitions between each mean vector $\mathbf{m}^{(g)}$,

$$\sum_{i=1}^{g} = \frac{\mathbf{m}^{i+1} - \mathbf{m}^{i}}{\sigma^{i}}. \tag{B.7}$$

The state variable $\mathbf{p_c}$ incorporates exponential smoothing into the calculation of the path, such that

$$\mathbf{p_c}^{(g+1)} = (1 - c_c)\mathbf{p_c}^{(g)} + \sqrt{c_c(2 - c_c)\mu_{eff}}\frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}. \tag{B.8}$$

Where $\sqrt{c_c(2 - c_c)\mu_{eff}}$ is a normalisation constant. The evolution path is used to update the covariance matrix

$$\mathbf{C}^{(g+1)} = (1 - c_1)\mathbf{C}^{(g)} + c_1\mathbf{p_c}^{(g+1)}\mathbf{p_c}^{(g+1)^T}, \tag{B.9}$$

using c_1 as a learning parameter, as for c_μ previously. The update to the covariance matrix incorporates both the rank-μ and the rank-one update. The updated covariance, $\mathbf{C}^{g+1}$, includes information about variance of the previous generation in the form of $\mathbf{C}^{(g)}$, and sign information from every generation from the evolution path $\mathbf{p_c}$, such that more recent generations are weighted more highly. The full expression for the generation of $\mathbf{C}^{(g+1)}$ is

$$\mathbf{C}^{(g+1)} = (1 - c_1 - c_\mu)\mathbf{C}^{(g)} + c_1\mathbf{p_c}^{(g+1)}\mathbf{p_c}^{(g+1)^T} + c_\mu \sum_{i=1}^{\mu} w_i \left(\frac{\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}\right)\left(\frac{\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}\right)^T. \tag{B.10}$$

The terms of this equation are, from left to right, the current covariance matrix with weighting, the rank-one update containing the anisotropic evolution path, and the rank-μ update which constructs a covariance matrix from the best solutions in the current generation. The purpose of these updates is to create a covariance matrix which increases the chance of the algorithm taking steps in parameter space that result in better scoring points.

B.6 Step-size control

The final state variable left to discuss is p_σ, the second evolutionary path. The previous expression for the update of the covariance matrix in equation (B.10) changes the shape, but not the scale of $\mathbf{C}^{(g+1)}$. By scaling the covariance matrix, σ changes the size of steps taken in parameter space. The scaling process encourages the CMAES algorithm to take steps of "optimal" size. For

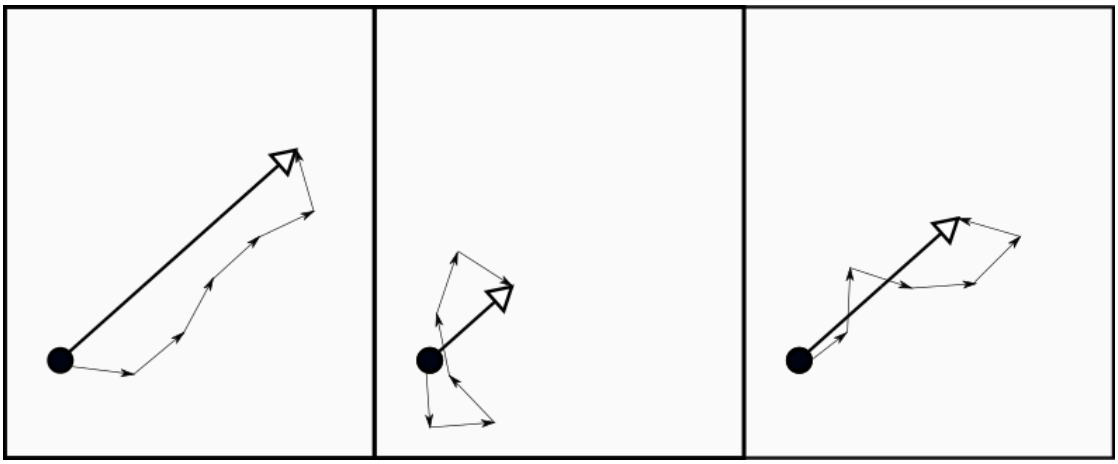


Figure B.2: Illustration of different path lengths with CMAES, where the individual steps for each figure are the same length. Left, steps are correlated, and the step size should be increased. Middle, steps are anticorrelated and cancelling each other out, and so step size should be decreased. Right, steps broadly uncorrelated, no correction of step size required. Inspiration was taken from [102]

a certain number of generations, if there are many steps taken but the overall distance travelled is small, this implies that the step size is too large (figure B.2, middle). On the flipside, if many steps are taken but these are all approximately in the same direction, then one larger step would have sufficed, and consequently the step size is too small (figure B.2, left). The ideal step-size is in between these two extremes (figure B.2, right). This can also be thought of in terms of correlation — when the step-size is too small, the steps taken are approximately correlated, and correspondingly anti-correlated when the step size is too large. Consequently, the ideal steps will be *uncorrelated*. In order to assess the length of steps relative to optimality, we compare the length of the evolutionary path $\mathbf{p}_\sigma$ to what we would expect under random movement. $\mathbf{p}_\sigma$ is recorded using

$$\mathbf{p}_\sigma{}^{(g+1)} = (1 - c_\sigma)\mathbf{p}_\sigma{}^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_{eff}}\mathbf{C}^{(g)^{-\frac{1}{2}}} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}, \tag{B.11}$$

where c_σ is a learning parameter which incorporates the previous state of the path, and $\sqrt{c_\sigma(2-c_\sigma)\mu_{eff}}$ is for normalisation. Equation (B.11) is very similar to the method for recording the evolutionary path for sign information, shown in equation (B.8). Both expressions incorporate the current path using exponential smoothing and normalise the current average step. The only difference is the multiplication of the vector of mean steps by the inverse square root of the current covariance matrix $\mathbf{C}^{(g)^{-\frac{1}{2}}}$. Because the vector of average steps $\frac{\mathbf{m}^{(g+1)}-\mathbf{m}^{(g)}}{\sigma^{(g)}}$ is distributed according to $\mathcal{N}(0, \mathbf{C}^{(g)})$, multiplying the vector by $\mathbf{C}^{(g)^{-\frac{1}{2}}}$ means the result is distributed according to $\mathcal{N}(0, \mathbf{I})$, and so is $\mathbf{p}_\sigma{}^{(g)}$, as long as $\mathbf{p}_\sigma{}^{(0)} \sim \mathcal{N}(0, \mathbf{I})$. After this normalisation, we can meaningfully compare the steps taken over the path of the recent generations of the algorithm to the distance expected during random search, also distributed according to $\mathcal{N}(0, \mathbf{I})$. The full expression for this is

$$\ln \sigma^{(g+1)} = \ln \sigma^{(g)} + \frac{c_\sigma}{d_\sigma}\left(\frac{||\mathbf{p}_\sigma{}^{(g+1)}||}{E[||\mathcal{N}(0,\mathbf{I})||]} - 1\right). \tag{B.12}$$

Here, d_σ is a damping parameter and $E||\mathcal{N}(0, \mathbf{I})||$ is the expected Euclidean norm of a vector distributed randomly according to $\mathcal{N}(0, \mathbf{I})$. Hence, if $||\mathbf{p}_\sigma{}^{(g+1)}|| = E[||\mathcal{N}(0, \mathbf{I})||]$ then $\sigma^{(g+1)}$ will remain unchanged. Larger or smaller values will scale the value of $\sigma^{(g+1)}$ accordingly. The comparison is performed on the log scale, as the $E[\ln \sigma^{(g+1)}|\sigma^{(g)} = \ln \sigma^{(g+1)}$ when $p_\sigma^{(g+1)} \sim \mathcal{N}(0, \mathbf{I})$. $\sigma^{(g+1)}$ is calculated by taking the exponent of equation (B.12), to obtain

$$\sigma^{(g+1)} = \sigma^{(g)} \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{||\mathbf{p}_\sigma{}^{(g+1)}||}{E||\mathcal{N}(0,\mathbf{I})||} - 1\right)\right). \tag{B.13}$$

We can list the expression for each function of algorithm 2:

- **update_m**$(\mathbf{x}_1^{(g+1)}, ..., \mathbf{x}_\lambda^{(g+1)}) \rightarrow$ equation (B.3)
- **update_pc**$(\mathbf{p_c}^{(g)}, \sigma^{(g)^{-1}}(\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)})) \rightarrow$ equation (B.8)
- **update_covariance**$(\mathbf{C}^{(g)}, \mathbf{p_c}^{(g+1)}, \sigma^{(g)^{-1}}(\mathbf{x}_{(1,...,\lambda)}^{(g+1)} - \mathbf{m}^{(g)})) \rightarrow$ equation (B.10)
- **update_ps**$(\mathbf{p}_\sigma{}^{(g)}, \sigma^{(g)^{-1}}\mathbf{C}^{(g)^{-\frac{1}{2}}}\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}) \rightarrow$ equation (B.11)
- **update_sigma**$(\sigma, ||\mathbf{p}_\sigma{}^{(g+1)}||) \rightarrow$ equation (B.13)

Appendix C

Assessing the impact of noise on heuristic parameter inference approaches

There were numerous challenges associated with collecting experimental data because of the Covid-19 pandemic, and consequently some adaptation to planned DPhil work was required, that could be achieved remotely without the need to collect real physical data. This appendix corresponds to one of these efforts. In it, we analyse how so-called "heuristic" methods, which rely on fitting to some extracted feature of an experimental voltammogram, perform in the presence of noise compared to the MCMC approach laid out in chapter 2. Content in this section has been re-purposed from the following reference: Henry Lloyd-Laney, Martin Robinson, Alison Parkin, and David Gavaghan, 2021. Theoretical assessment of the impact of noise on heuristic parameter inference methods for surface-confined non-catalytic voltammetry experiments (in press) ***Journal of Electroanalytical chemistry***

For the experimentalist, there is a large array of inference approaches to choose from. An alternative, such as the method detailed in section 3.1 is to extract some feature from the current data, often as a function of varying experimental conditions, such that the extracted feature has a predictable relationship to one or two parameters of interest. Taking inspiration

from our previous benchmarking approach, we consider how these methods are affected by the presence of independent and identically-distributed (i.i.d.) noise, the simplest form of noise model, in the most ideal of cases — purely Faradaic current with no capacitance or dispersion in the thermodynamic or kinetic parameters, for a single-electron surface-confined redox process with no catalytic activity. In keeping with the rest of the thesis, we have focused on the purely Faradaic case, but it should be noted that there is extensive interest in the literature in systems with catalytic activity [47, 103, 104]. Feature relationships can be experimentally [105, 106] or analytically [6] derived, and are used as an alternative to total current analysis, the approach we use, detailed in chapter 2, for a number of reasons. One such reason is ease of analysis, as the derived relationships are often mathematically simpler than the necessary simulation procedures for predicting the total current. They often take the form of a linear relationship between the extracted feature and a dependent variable, and consequently calculating the gradient or intercept of the line of best fit is all that is required to obtain a value for a parameter of interest; a process that can be trivially implemented in spreadsheet software such as excel. Another reason is that often the total experimental current contains non-idealities such as dispersion and non-linear capacitance that would require additional considerations to incorporate into a model of the total current, but that have a small effect on the extracted feature. Additionally, in the past when performing large numbers of total current simulations was computationally infeasible, these approaches were far more achievable on available hardware. In this section we quantify the impact of adding varying amounts of noise when attempting to perform one of 5 different inference techniques from the literature, two using direct current or cyclic voltammetry (DCV) and three using square wave voltammetry (SWV). We compare these results to an automatic approach which analyses either the same set of extracted features or the total current, and allows us to quantify the level of uncertainty inherent to either approach.

C.1 Approach

C.1.1 Heuristic inference

In order to determine the robustness of the various inference processes described below to the presence of i.i.d. noise, we generated multiple noisy voltammograms, using the simulation procedures for DCV and SWV described in chapter 2, and an excellent textbook on square wave voltammetry by Lovrić and coauthors respectively [107]. Noise was subsequently added using

the process described in the "Noise" section. We then applied the various inference techniques to attempt to recover the appropriate parameter used in the simulation of the synthetic voltammogram. This process involves extracting some feature(s) from the synthetic current, or set of synthetic currents, and using those features to determine the value of the appropriate parameter. This value was recorded for each attempt, and the set of inferred parameters was used to assess the robustness of the technique.

C.1.2 Markov-chain Monte Carlo

To benchmark the robustness of the various inference approaches, we compare the range of values obtained using our heuristic inference approach to the appropriate inferred parameter distribution obtained using Bayesian inference, as detailed in chapter 2. The distributions generated by this process are not exactly comparable to the process of benchmarking the inference methods described in the previous section, since with the MCMC algorithms we only analyse one noisy set of data. However, if we were to repeat the MCMC process multiple times, it would not significantly alter the inferred parameter distributions (as the level of noise would be the same).

As long as we are consistent with how the parameters are used to generate a given simulation, it is possible to perform inference on either the total current from a single voltammogram, or the extracted features used for the heuristic approaches. This allows us to obtain both an explicit comparison between the MCMC and heuristic approaches (when fitting to sampled features) and allows us to observe the impact of feature extraction (by performing MCMC on the total current from a single voltammogram).

For MCMC, we have to construct a log-likelihood function. For the total current case, assuming Gaussian i.i.d. error, this is

$$\log \mathcal{L}(\theta, \sigma | I_{data}) = -\frac{N_t}{2} \log 2\pi - N_t \log \sigma - \frac{1}{2\sigma^2} \sum_{j=1}^{N_t} (I_{data}(t_j) - I_{sim}(\theta, t_j))^2 \tag{C.1}$$

where N_t is the number of timepoints, and σ is the standard deviation. When considering the likelihoods for the heuristic MCMC case, we define the set of observed features f, such that

$$\log \mathcal{L}(\theta, \sigma | f_{data}) = -\frac{N_t}{2} \log 2\pi - N_t \log \sigma - \frac{1}{2\sigma^2} \sum_{j=1}^{N_{obs}} (f_{data}(j) - f_{sim}(\theta, j))^2, \tag{C.2}$$

where N_{obs} is the number of observed features. We should note that for the extracted feature MCMC case, there will be some bias in the estimated probabilities. This is because we are assuming normally distributed errors on the observations, when in fact the normally-distributed noise was added to the total current, and the features then extracted from this current. Consequently, we have applied some non-linear function to the Gaussian noise, and so assuming Gaussian i.i.d. is not an accurate assumption. This potentially explains why bias is observed for the heuristic MCMC parameter inference attempts in figures C.4 and C.6 in the results below, as well as the non-monotonic relationship between noise level observed in table C.3.

C.1.3 Noise

When creating artificially noisy voltammograms, our procedure for the addition of i.i.d. noise is simply

$$I_{noisy}(t) = I_{sim}(t) + \epsilon(t), \quad \epsilon(t) \sim \mathcal{N}(0, (\max(I_{sim})p)^2), \tag{C.3}$$

such that $\mathcal{N}(0, (\max(I_{sim})p)^2)$ indicates a normally-distributed random variable with 0 mean, and a standard deviation proportional to $\max(I_{sim})$, and p is a scaling factor in the range of 0.005-0.05, or 0.5-5% of the maximum. We have found that a noise level of 0.05% of the maximum current is usually a good approximation to the noise observed in actual experimental voltammetry data.

C.1.3.1 A note on noise in square-wave voltammetry

Initially, when generating noisy SW voltammograms, we used the magnitude of the total current (as opposed to the sampled) to calculate the level of noise. In other voltammetric experiments, the magnitude of the current arising from the signal to be analysed (in DCV, for example, the reductive and oxidative peaks) is usually on a similar order of magnitude as the value of the maximum current. In the case of SWV, the current is sampled at the end of a pulse, which is effectively the *minimum* magnitude of the current on the timescale of that pulse. Consequently, any set of parameters generates a current which contains a large discrepancy between the maximum total current and the average magnitude of the sampled current will result in a magnitude of ϵ in equation C.3 that is sufficient to obscure any signal observed in the sampled current, sometimes two or three orders of magnitude larger. As the impact of noise does not seem to be a particularly high concern among SWV practitioners, and after observation of real SWV data, we have decided, for the purposes of this exercise, to use the maximum of the sampled SWV net

current rather than the maximum of the total current to calculate the noise value.

C.2 DCV

DCV is one of the simplest and most widely used voltammetry techniques. A single DCV experiment uses a potential input that increases or decreases linearly (i.e. with a constant potential scan rate) to a switching potential, and then returns to the starting potential.

C.2.1 Surface coverage

The surface coverage parameter Γ (as defined in chapter 2) can be determined from the area under either the reductive or oxidative peak, such that, for a purely Faradaic current beginning at time $t = 0$ and ending at time $t = t_N$, where F is Faraday's constant, i_f is the total Faradaic current, n the number of electrons transferred, v is the scan rate and a the electrode area [6]

$$FAnv\Gamma = \int_0^{t_N} i_f \tag{C.4}$$

In practice, the integral in equation C.4 is calculated using numerical quadrature, for example by using Simpson's rule [6]. An example of this process is shown graphically in figure C.1

C.2.2 Rate constant

Trumpet plots, where multiple DCV experiments are performed at different scan rates and the potential position $E_p(ox)/E_p(red)$ of the peak current is recorded, can be used to determine the rate constant k_0. These plots are popular as the relationship between the peak position and the scan rate is not affected by background currents arising from double-layer capacitance to the same extent as the rest of the voltammogram. Consequently, analysis of the trumpet plot is less dependent on the ability to correctly subtract away background current. By rearranging the Butler-Volmer equations (as defined in chapter 2), we obtain an expression for the potential of the oxidative/reductive peak as a function of the logarithm of the scan rate $\ln(v)$ and k_0, such that

$$E_{p(ox)} - E^0 = -\frac{RT}{\alpha nF}\ln(v) - \frac{RT}{\alpha nF}\ln\left(\frac{\alpha nF}{RTk_0}\right), \tag{C.5}$$

and

$$E_{p(red)} - E^0 = \frac{RT}{(1-\alpha)nF}\ln(v) + \frac{RT}{(1-\alpha)nF}\ln\left(\frac{(1-\alpha)nF}{RTk_0}\right), \tag{C.6}$$

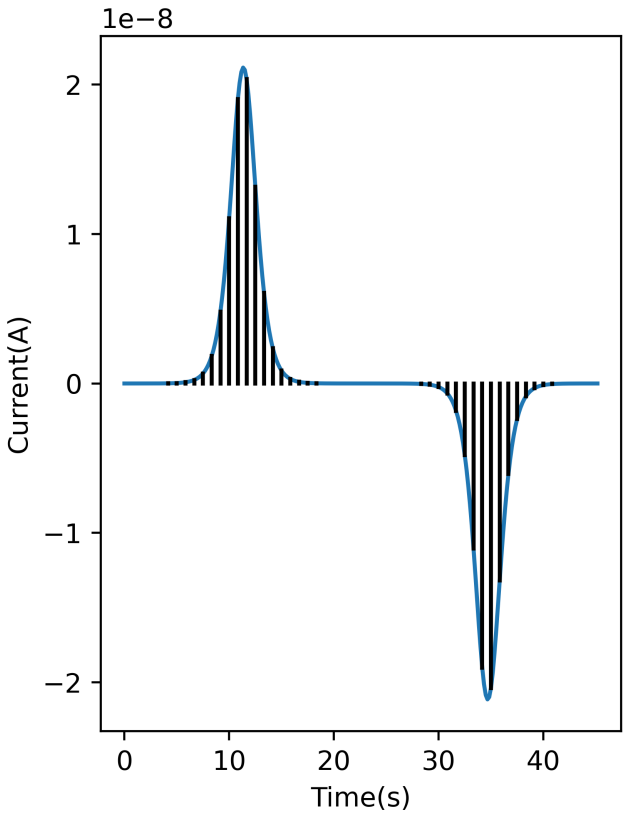


Figure C.1: Purely Faradaic DCV experiment vs. time, with black lines indicating how the integral is approximated using numerical quadrature

where F is Faraday's constant, T is the temperature, R the ideal gas constant and n the number of electrons. A linear function of the form $f(x) = mx + c$ can consequently be fit to the the linear regions of an appropriately normalised trumpet plot, which represents $E_{p(ox/red)} - E^0$ as a function of $\ln(v)$, as shown on the right hand plot of figure C.2. Consequently, the recovered gradient m can be used to calculate the value of α, and subsequently the recovered intercept c can be used to determine the value of k_0 [6].

C.3 SWV

Square-wave voltammetry uses a square-wave potential superimposed over a staircase potential — jumps between potential values are hypothetically instantaneous, and the potential is then maintained at a constant value until the subsequent jump. The current is sampled at the end of

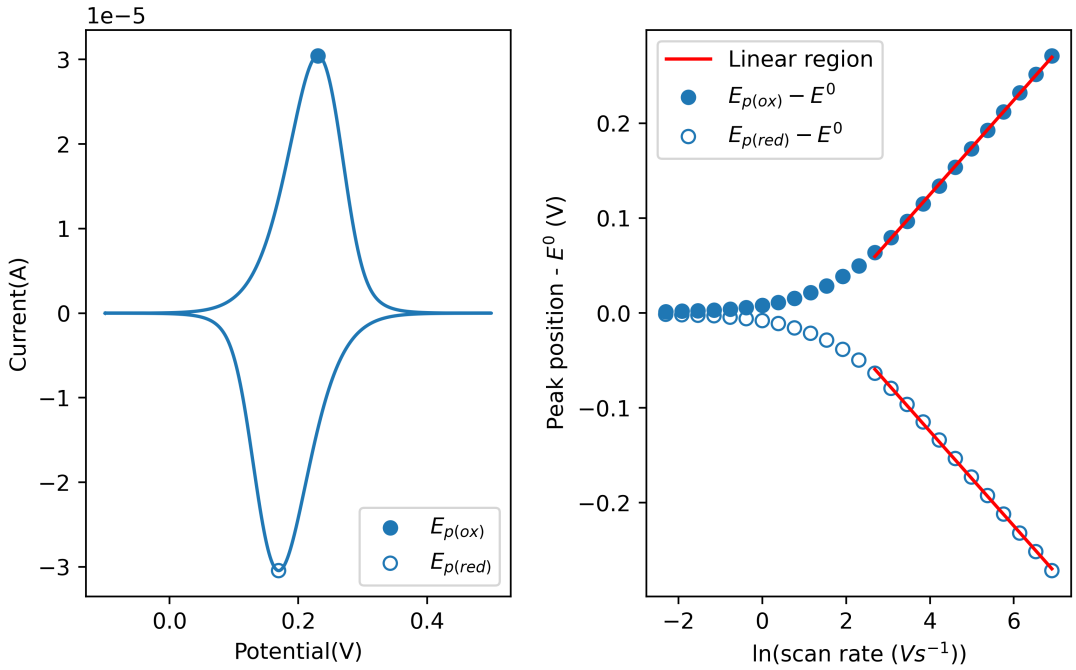


Figure C.2: Left: Purely Faradaic DCV current vs. potential, where $v = 5\ Vs^{-1}$ and $k_0 = 100\ s^{-1}$. Right: Trumpet plot showing normalised peak positions of oxidative and reductive peaks, and the linear region through which we draw a line of best-fit (red)

each pulse, and this sampling is separated into "forwards current", where the sampling occurs at the maximum of each square-wave oscillation, "backwards current" which is sampled at the minimum, and the net current, which is simply the backwards current subtracted from the forwards [107]. The sampling procedure is shown graphically in figure C.3.

C.3.1 Reversible rate constants

At sufficiently high values of K (i.e. the dimensionless kinetic parameter $\frac{k_0}{f}$, where f is the square-wave frequency) for a surface-confined redox process interrogated using a square-wave potential, a phenomenon known as "peak-splitting" is observed, which refers to the sampled net current exhibiting two distinct peaks. Lovrić et al. determined that the separation between these two peaks (ΔE_p) was a function of the square wave amplitude in mV (E_{sw}) and K, such that $\Delta E_p = 5\frac{k_0}{f} + 2nE_{sw} - 50mV$ [105]. This expression can be rearranged into the form $y = mx + c$, such that $mx = \frac{k_0}{f}$, and consequently the value of k_0 can be determined from the fitted value of m.

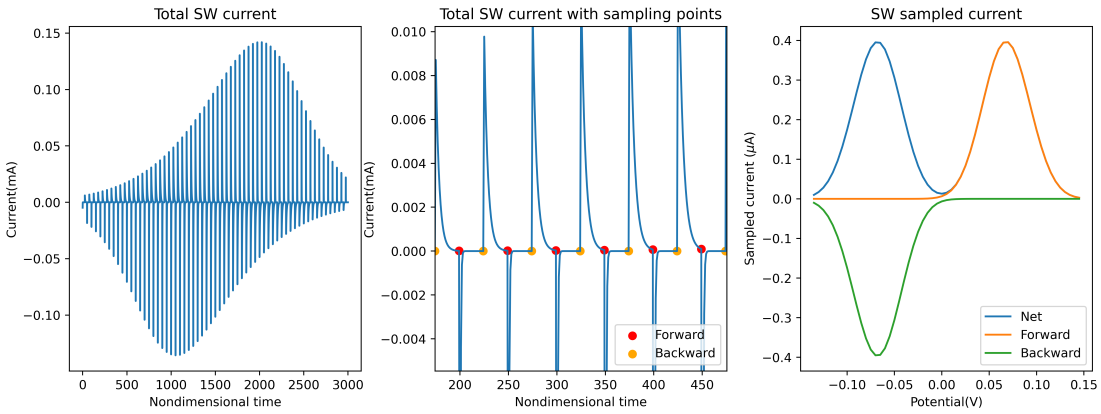


Figure C.3: Demonstration of the procedure for obtaining sampled current from a SWV simulation. The total current vs. the non-dimensionalised time (left) is sampled using "forwards" and "backwards" points, shown on top of a section of the total current in the central figure. These sampled points are then plotted as orange and green traces in the rightmost figure, against the potential, as well as the net current which is the backwards subtracted from the forwards current.

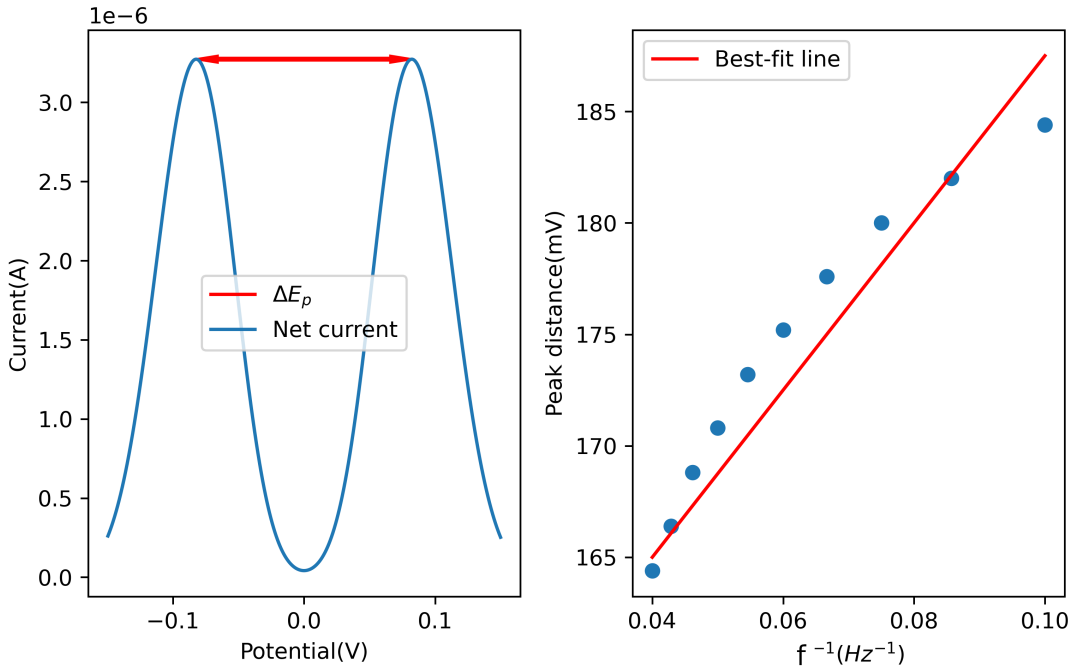


Figure C.4: Left: net SWV current, generated using $f = 25$ Hz and $k_0 = 75\ s^{-1}$, with the peak distance metric shown in red. Right: Normalised peak distances plotted as a function of the inverse frequency (blue dots), with the best-fit line used to calculate k_0 shown in red.

C.3.2 Symmetry factor

In the same paper, Lovrić et al. also describe a method to obtain an estimate of the symmetry factor α, with similar assumptions about the value of $\frac{k_0}{f}$, by calculating the ratio of the two peak heights, such that

$$\alpha = -\frac{1}{3.4606} \ln\left(\frac{|I_f|}{5.614|I_b|}\right), \tag{C.7}$$

where I_f and I_b are the maximum currents for the sampled forwards and backwards SWV currents respectively, and the coefficients are taken directly from the work by Lovrić et al. [105]. The process of obtaining I_f and I_b is shown on the left hand side of figure C.5. The right hand shows the value of alpha calculated from SW voltammograms simulated with differing values of α (and a $\frac{k_0}{f}$ value of 6), along with the predictions from equation C.7, indicating very good agreement

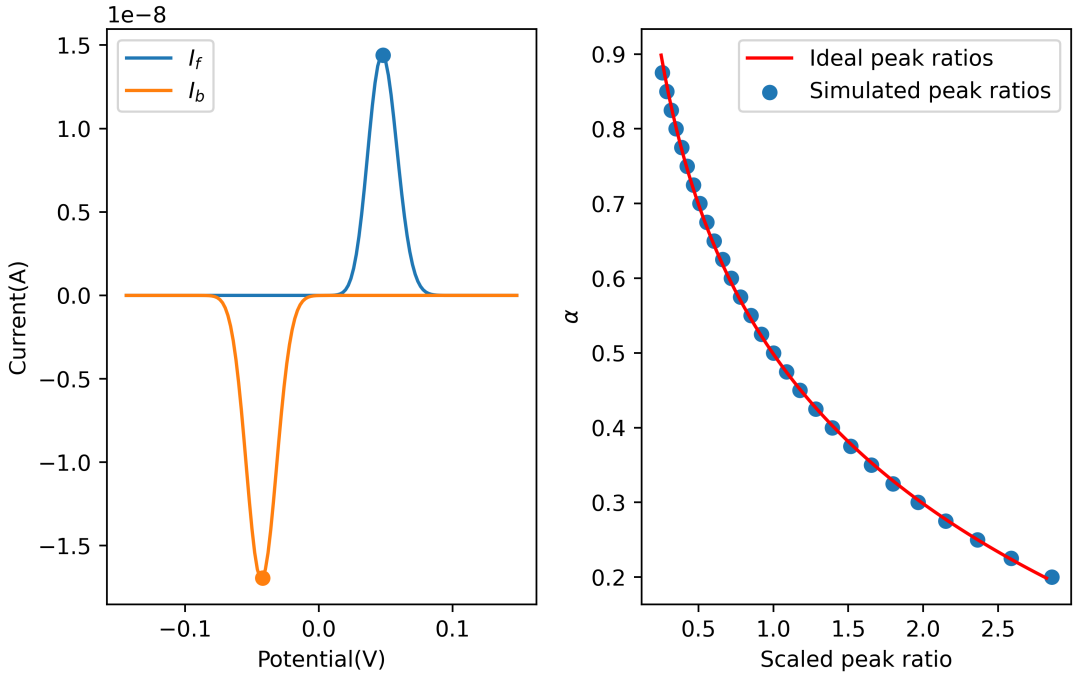


Figure C.5: Left: Forward and backward currents from a SWV experiment $f = 1$ Hz, $k_0 = 6$ s^{-1} and $\alpha = 0.5$, with the key features I_f and I_b indicated. Right: Peak ratios obtained from sampled current simulated with different values of α vs the values of α calculated using equation C.7 (blue) and "ideal" peak ratios calculated by rearranging equation C.7 (red)

C.3.3 Quasi-reversible rate constants

The peak-splitting inference approach for k_0 is valid only under conditions where peak-splitting is observed, i.e. $\frac{k_0}{f} > 2$. Compton and coauthors described an approach for determining the rate constant when $0.01 < \frac{k_0}{f} < 0.5$. They termed this approach the "quasi-reversible maximum". Multiple SW voltammograms are measured at varying frequencies f and amplitudes E_{sw}. For each experiment, the peak net current, normalised by E_{sw} is noted. The maximum of these maxima I_{QRM} is recorded as a function of frequency, which is then used to determine k_0 using the following relationship, where the coefficients come from [106]

$$\frac{k_0}{f} = 10^{\frac{I_{QRM} - 0.8}{-148.08}}. \tag{C.8}$$

As before, we determine the gradient m of the normalised I_{QRM} values vs. the inverse frequency, giving us a value for k_0. This process is shown in figure C.6.

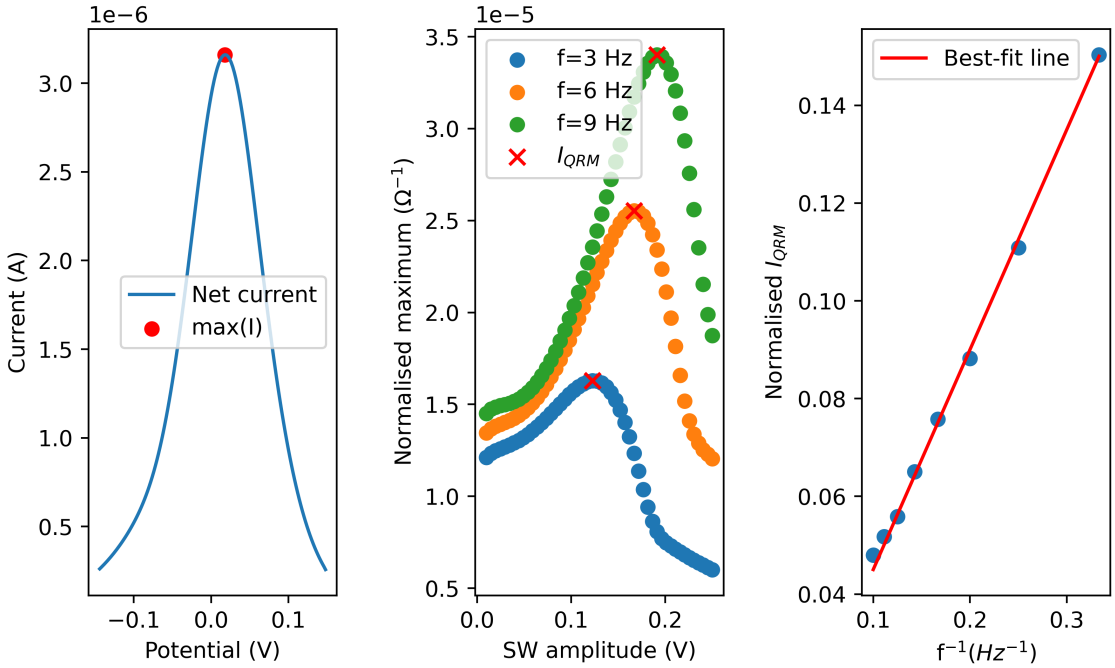


Figure C.6: Left: Net SWV current, simulated using $f = 7$ Hz, $k_0 = 0.5\ s^{-1}$ and $E_{sw} = 0.13\ V$, with the maximum current indicated by the red dot. Centre: maximum current (normalised by the value of E_{sw}), plotted against the varying E_{sw} value for three different frequencies. The I_{QRM} value is indicated by the red cross. Right: I_{QRM} plotted as a function of the inverse frequency, with the best-fit line drawn in red.

Parameter	DCV Γ	DCV k_0
Start potential (V)	-0.1	-0.1
Reverse potential (V)	0.5	0.5
Scan rate (V s^{-1})	0.001-31	0.03
Rate constant (s^{-1})	100	100
Reversible potential (V)	0.2	0.2
Surface coverage mol cm^{-2}	1e-11	1e-11
Symmetry factor	0.5	0.5
Number of features	1	2505 (25)
Simulation time (s)	0.013 (N/A)	0.47 (0.012)
MCMC time (min)	1.0 (N/A)	35 (1.1)

Table C.1: Simulation parameters used to generate synthetic current data for both heuristic and MCMC inference approaches for the DCV case. Values in brackets refer to single-experiment simulations, and non-bracketed values refer to extracted feature simulations.

C.4 Simulation information

The parameters used for the simulation of the synthetic voltammograms can be found in table C.1 and C.2 for the DCV and SWV cases respectively, along with information about the number of features, and simulation times. Simulation parameters were chosen to be in a regime where the corresponding heuristic inference approach was maximally effective.

C.5 Results and Discussion

C.5.1 DCV: Surface coverage

The first heuristic method, inferring the surface coverage in a surface-linked DCV experiment by determining the area under the oxidative and reductive peaks, was extremely robust to the addition of I.I.D. noise. This can be observed in figure C.7 and in table C.3, where even with 5% noise, the standard deviation of inferred Γ values is less than 1% of the true value. This can be explained by the fact that, because the mean of the noise in equation C.3 is zero, the overall impact on the area under the curve of the current is minimal, leading to the very accurate estimates observed. Because of this accuracy, we did not believe it was necessary to compare

Parameter	SWV k_0 (R)	SWV α	SWV k_0 (QR)
Start potential (V)	0.15	0.15	0.15
Potential window (V)	0.3	0.3	0.3
Scan increment (V s^{-1})	0.005	0.005	0.005
Square-wave amplitude (V)	0.05	0.05	0.12-0.22 [25]
Square wave frequency (Hz)	10-25	1	5-10 [10]
Rate constant (s^{-1})	75	5	0.5
Reversible potential (V)	0.0	0.0	0.0
Surface coverage mol cm^{-2}	1e-11	1e-11	1e-11
Symmetry factor	0.5	0.5	0.5
Number of features	20 (72)	1 (72)	10 (72)
Simulation time (s)	0.0034 (0.039)	N/A (0.0038)	0.0038 (0.87)
MCMC time (min)	2.9 (0.5)	N/A (0.4)	71 (0.5)

Table C.2: Simulation parameters used to generate synthetic current data for both heuristic and MCMC inference approaches for the SWV case. Values in square brackets indicate the number of values within the range. Simulation times are from a Dell laptop with 8 GB RAM and an Intel i5–8350U CPU @ 1.70 GHz. MCMC times were calculated by observing the time taken for 20 iterations with three chains and scaling for 5000, and is faster than would be expected from simulation alone as a result of parallelisation. Values in brackets refer to single-experiment simulations, and non-bracketed values refer to extracted feature simulations

Method	Noise level	DCV Γ	DCV k_0	SWV k_0 (R)	SWV α	SWV k_0 (QR)
Heuristic	0.5%	0.06%	4.04%	1.99%	0.28%	1.73%
	1%	0.13%	5.8%	2.88%	0.51%	2.23%
	2%	0.26%	7.98%	3.82%	0.98%	2.9%
	3%	0.38%	9.83%	4.67%	1.33%	3.65%
	4%	0.5%	13.25%	5.39%	1.68%	4.1%
	5%	0.64%	15.95%	5.8%	2.14%	4.51%
MCMC	0.5%	N/A	0.97%(0.05%)	3.26%(0.02%)	N/A(0.12)%	2.74%(0.05%)
	1%		1.2%(0.11%)	2.97%(0.04%)	N/A(0.24)%	3.64%%(0.11%)
	2%		1.83%(0.22%)	3.43%(0.07%)	N/A(0.51)%	2.55%(0.18%)
	3%		2.41%(0.32%)	3.73%(0.11%)	N/A(0.73)%	3.86%(0.27%)
	4%		2.3%(0.43%)	4.06%(0.16%)	N/A(1.1)%	4.32%(0.4%)
	5%		2.68%(0.55%)	4.13%(0.18%)	N/A(1.17)%	6.33%(0.52%)

Table C.3: Ratio of the standard deviation of the inferred parameter values for each inference approach to the true value of that parameter, with varying amounts of added noise. Where appropriate, the values in brackets were inferred from fitting to the total current, and the other values are from fitting to the extracted features.

this technique to MCMC. Although it is outside of the scope of this section (which focuses on the impact of i.i.d. noise), we should note that this heuristic is highly impacted by the efficacy with which the background current is subtracted, an issue we noted in chapter 5. In figure C.7, and all subsequent histograms, "frequency" refers to the number of samples of that parameter value returned by the MCMC algorithm

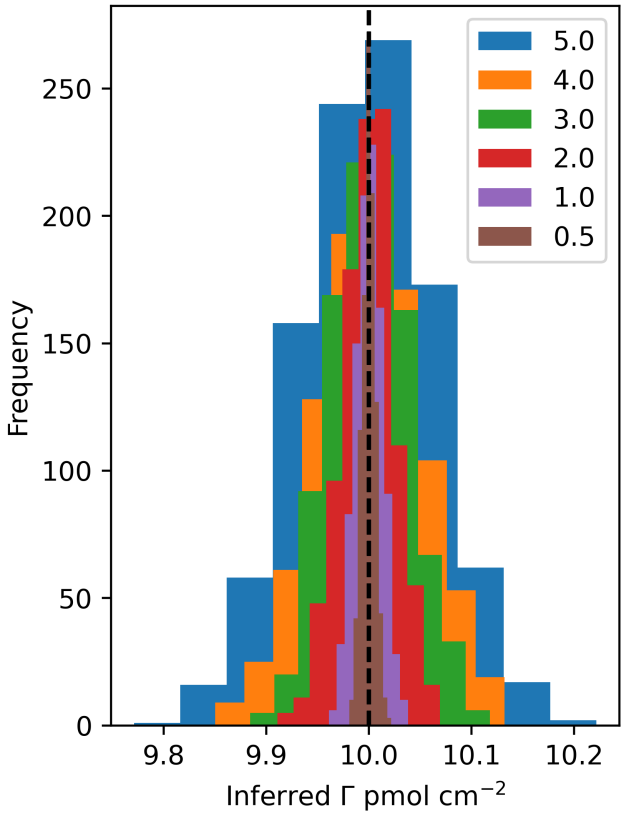


Figure C.7: Values of the surface coverage parameter Γ, where the true value was 10 pmol cm^{-2}, inferred using the approach described in the heuristic section, with varying amounts of noise. The noise percentage is indicated in the legend. We did not compare the results of this to MCMC in this case, because the values inferred by this method were so close to the true value.

C.5.2 DCV: Kinetic parameter

The method of inferring the kinetic parameter from trumpet plots was by far the least robust to the presence of i.i.d. noise of all the techniques assessed. The spread of values, as communicated in table C.3, and shown in figure C.8 is 4-8 times worse than when using MCMC using the

same (peak-position) observations, and 1-2 orders of magnitude worse when fitting to the total current. We believe that the reason for the poor performance of this method is that it requires accurate estimates of both the intercept and gradient of the best-fit line, as opposed to the peak-splitting and quasi-reversible maximum methods (discussed below) that only require an estimate of the gradient. As the method requires normalisation of the peak potentials using the reversible potential, we also assessed the impact of incorrect estimates of E^0, but found that it did not significantly alter the results presented in table C.3 and figure C.8.

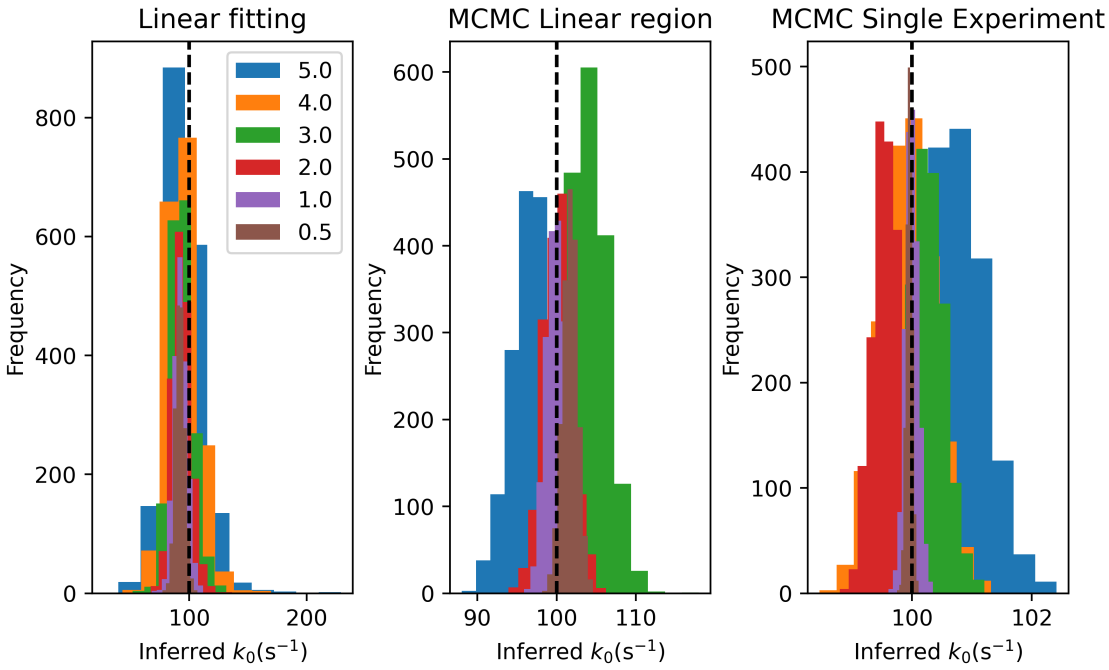


Figure C.8: Left and centre: Kinetic parameter values inferred by fitting to the linear region of a trumpet plot for varying amounts of noise, where the noise percentage is indicated by the legend, and the true value of k_0 was 100s^{-1}. For the linear fitting, the inferred anodic and cathodic values were pooled. Right: Inferred parameter values for k_0 using MCMC, fitted to a trumpet plot.

C.5.3 SWV: Kinetic parameter, reversible case

The performance of the MCMC algorithm when fitting to ΔE_p recovers values with an accuracy that has the same order of magnitude as the heuristic approach, although it is more affected relative to the heuristic approach at the lower noise percentages, but less affected at the higher percentages. This may reflect the fact that the majority of the uncertainty in this approach is

driven by the amount of information that is discarded at the feature extraction stage (i.e. the representation of a SWV experiment as a single ΔE_p value). This is reinforced by the fact that the values inferred when analysing the total current from a single SWV experiment using MCMC have a spread around the true value that is $\sim 1-2$ orders of magnitude smaller, as shown in figure C.9.

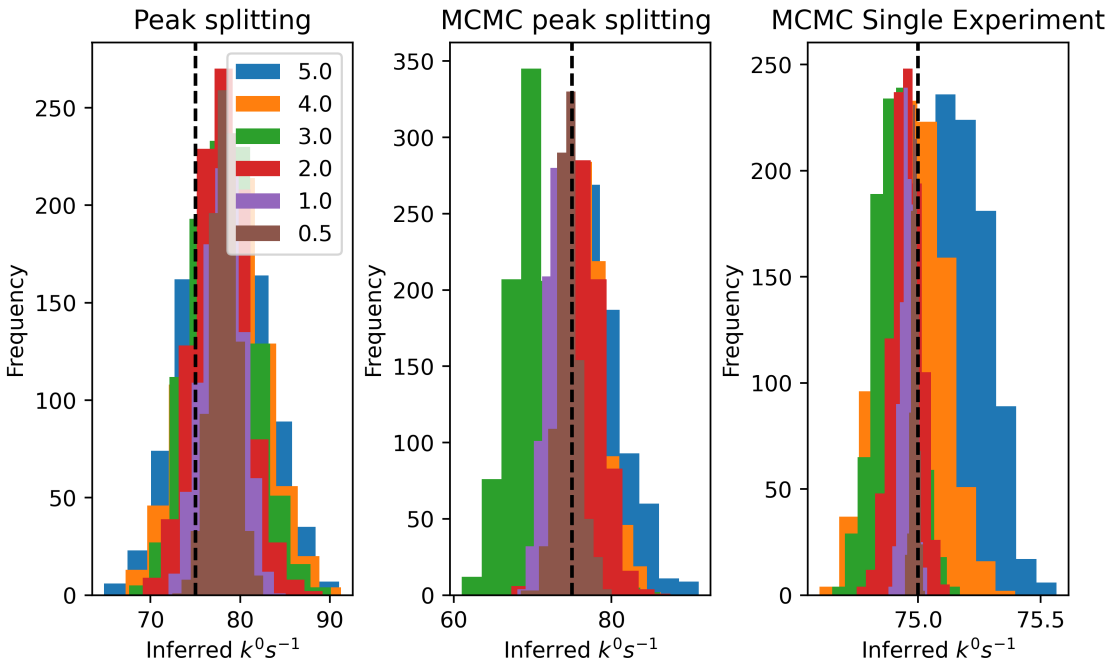


Figure C.9: Left: Values of the kinetic parameter inferred from the net sampled SWV current using the peak splitting method described in the heuristic inference section, for varying amounts of noise where the noise percentage is shown in the legend. The true value of k_0 was 75s^{-1}. Centre and right: Values of the kinetic parameter inferred using MCMC, using sampled net current (centre) and total current (right).

C.5.4 SWV: Symmetry factor

The method of estimating α from the ratio of peak heights present in the net SWV current was the second most effective heuristic approach when judged by having the smallest range of inferred values around the true parameter, with the spread being only 2-3 times larger than for the MCMC case, as observed in figure C.10. We did not use MCMC on the peak ratio data, as this would have required fitting to a single datapoint, which is not appropriate for MCMC.

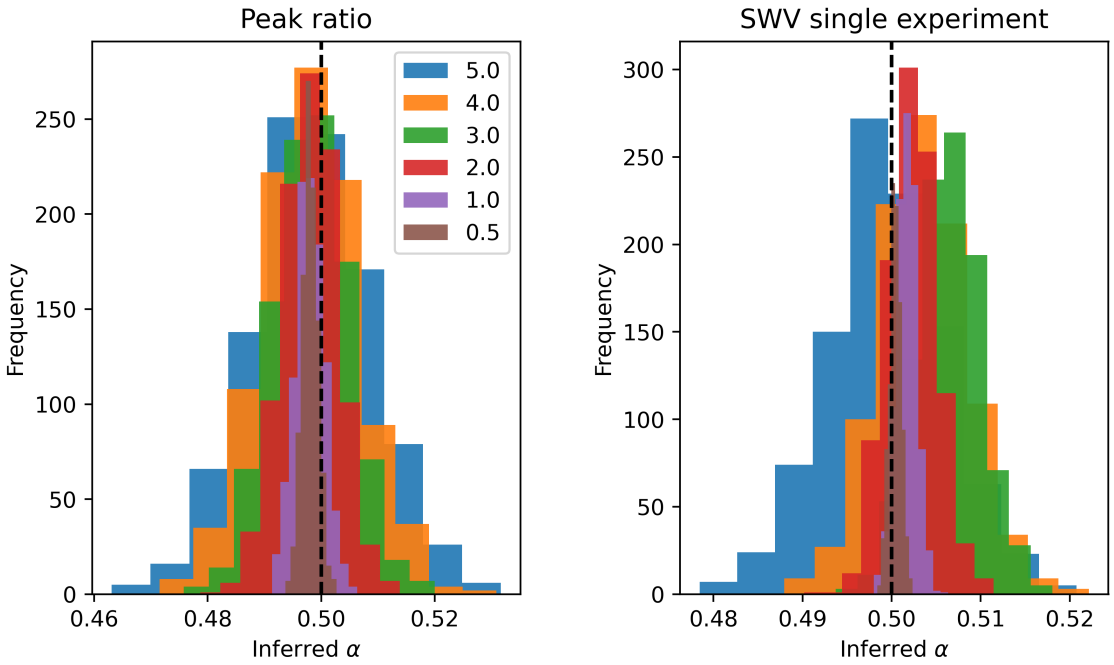


Figure C.10: Left: α values inferred from noisy sampled forwards and backwards SWV currents using the peak ratio approach described in the heuristic inference section, for different values of noise, where the noise percentage is shown in the legend. The true value of α was 0.5. Right and centre: α values inferred from noisy sampled net current (centre) and total SWV current (right)

C.5.5 SWV: Kinetic parameter, irreversible case

As with the peak-splitting case, the performance of the heuristic approach and MCMC when applied to the heuristic features is of a similar order of magnitude, with the values inferred using the total current analysed by MCMC being 1-2 orders of magnitude more accurate, as shown in figure C.11.

C.6 Conclusions

These results indicate that, under the conditions analysed here, methods which involve the extraction of features from the total current are outperformed by analysing the total current itself. The worst-performing heuristic method, using the linear region of a trumpet plot, performs worse than both fitting to the total current and fitting to the extracted features, presumably because it requires accurate estimates of both the gradient and intercept of the fitted line. The approach

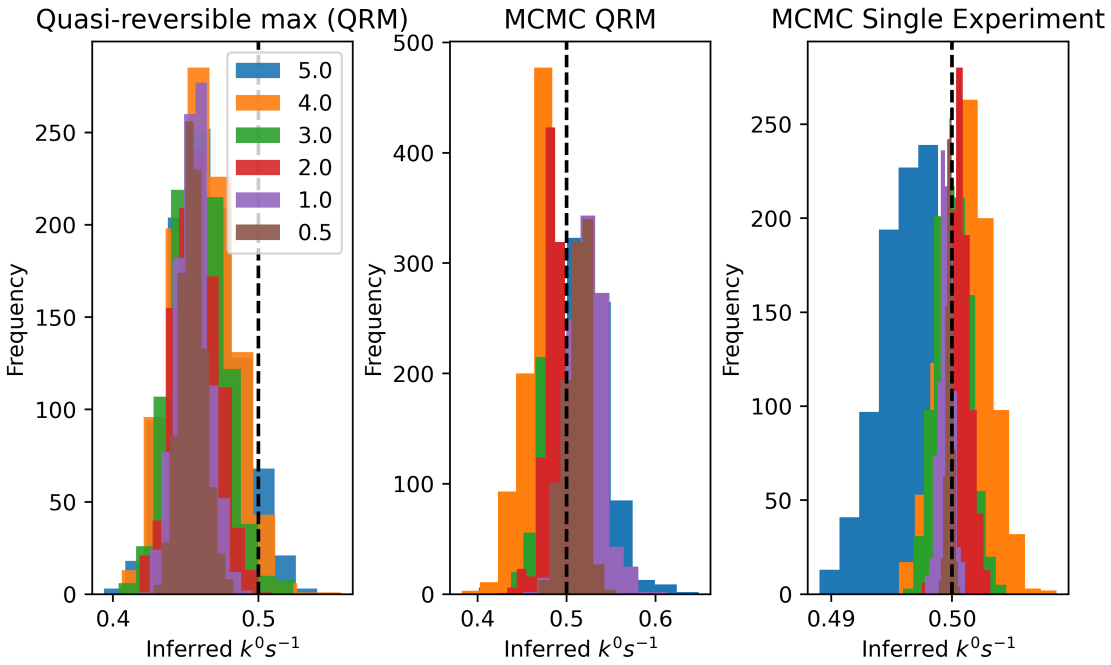


Figure C.11: Left: Kinetic parameter values inferred from sampled net SWV current using the quasi-reversible maximum method, for different values of noise where the noise percentage is shown in the legend. The true value of k_0 was $0.5 s^{-1}$. Right and centre: Kinetic parameter values inferred using MCMC using noisy sampled (centre) and total (right) SWV current.

with the best accuracy (DCV surface coverage) did not use a linear fit at all. The analysis of SWV data is more interesting - the performance of the heuristic approaches are comparable to the results obtained from using MCMC to fit to the same set of features, implying that the linear fitting (requiring only an estimate of the gradient) is as robust as the MCMC approach. In addition, when using MCMC to fit to features for the peak-splitting and QRM approaches, the standard deviation of the recovered values does not increase monotonically with the added noise (as it does with MCMC fitting to the total current) — these two observations suggest that the majority of the uncertainty with these techniques is driven by the feature extraction process, where information is discarded in order to compress a voltammetric experiment into a single data-point, rather than the actual fitting process. Interestingly, increasing the number of extracted features for the peak splitting case from 20 (as used to generate the values in figure 8) to 70, which is consistent with the number of points in the sampled total current, reduced the width of the kinetic distribution inferred by MCMC analysis of ΔE_p, which is expected as this is increasing the amount of data "visible" to the MCMC. For the 0.5% noise case, the standard

deviation of the inferred kinetic distribution decreased from 2.74% to 0.5% of the true value (of 75s^{-1}) as a result of this increase in the amount of data. However, this standard deviation obtained from fitting to the ΔE_p, was still 10 times larger than the MCMC-inferred standard deviation of the kinetic parameter for the total current at 0.05%. This implies that the difference between the standard deviations of distributions inferred by fitting to extracted features vs. total current using MCMC is partially, but not totally due to a simple discrepancy in the number of observations. The remainder of the difference is presumably the result of intrinsically greater information present in the total current relative to the extracted features. Additionally, as detailed above, the assumption of iid noise for the heuristic case may have led to some biasing of the results.

It should also be noted that the non-monotonic behaviour may be the result of inaccurate assumptions about the noise model, which we discuss in the supplementary information. Additionally, all simulation of synthetic current was performed in the absence of background current, and that the presence of the background current is one of the reasons discussed for using the feature extraction methods in the first place. However, even when fitting to the extracted features instead of the total current, MCMC methods at minimum offer similar levels of accuracy. This is in addition to their other advantages, in particular the ability to provide a measure of confidence for multiple parameters simultaneously, while the heuristic approach only returns a single estimate for one or two parameter values. Of the three reasons discussed for using the heuristic approach, namely ease of implementation, smaller impact from non-idealities and computational feasibility, only the second motivation is really valid, as it is becoming consistently easier to implement statistical approaches such as MCMC, and computing power is no longer a limiting factor. However, even in this case, the extracted features can still be analysed using an MCMC approach, rather than using a linear fit.

Appendix D

Phase in electrochemical impedance spectroscopy

D.1 Introduction

Although electrochemical impedance spectroscopy (EIS) was not used in this thesis, the theory of EIS can provide some important context, especially with regards to the concept of phase. As such, in this appendix, we cover some basic theories of impedance, and how they correspond to observed time-domain phase shifts. Content in this section was primarily informed by the following reference [6].

D.2 Impedance components

The two simplest electrical components are a resistor, with the relationship between current i and potential E is linked to the resistance R by Ohm's law

$$E(t) = i(t)R. \tag{D.1}$$

The second is a capacitor, where the charge Q is proportional to the voltage across the element, characterised by a constant capacitance C

$$Q(t) = CE(t) \tag{D.2}$$

$$i(t) = \frac{dQ}{dt} = C\frac{dE}{dt} \tag{D.3}$$

If the potential oscillates sinsuoidally, the current through a resistor will be a sinusoid in phase with the potential. For a capacitor, the current will take the form of the derivative of the input, and consequently the form of a cosine, or a sine-wave phase shifted (η) by $\frac{\pi}{2}$. This can be generally represented as

$$E(t) = E_0 \sin(\omega(t)) \tag{D.4}$$

$$i(t) = i_0 \sin(\omega(t) + \eta) \tag{D.5}$$

The relationship between current and potential is known as the *impedance Z*

$$Z(t) = \frac{E(t)}{i(t)} = \frac{E_0}{i_0}\frac{\sin(\omega(t))}{\sin(\omega(t) + \eta)} \tag{D.6}$$

In the complex exponential representation, the magnitude of the impedance has a magnitude and a phase, such that

$$Z = Z_0 e^{j\eta}. \tag{D.7}$$

Accordingly, Z is a complex number, and can be also be represented as having real and imaginary parts

$$Z = Z_{re} + iZ_{im}, \tag{D.8}$$

where i is the imaginary unit. The conversions back to equation D.7 are

$$|Z_0|^2 = |Z_{re}|^2 + |Z_{im}|^2, \tag{D.9}$$

and

$$\tan\eta = \frac{Z_{im}}{Z_{re}}. \tag{D.10}$$

Electrochemical impedance spectroscopy interrogates the system under investigation at a range of frequencies ω to obtain a characteristic impedance response $Z(\omega)$. The impedance in the Laplace domain $Z(s)$ can be represented as the ratio of the Laplace transforms of the time-domain representations of the current and the potential, such that

$$Z(s) = \frac{\mathcal{L}\{V(t)\}}{\mathcal{L}\{I(t)\}}. \tag{D.11}$$

were $s = \sigma + j\omega$. For a non-damped sinusoidal input, $s = j\omega$. We can then obtain the impedance of a resistor and a capacitor from their time-domain formulas.

For a resistor

$$Z_R(s) = \frac{R\mathcal{L}\{I(t)\}}{\mathcal{L}\{I(t)\}}, \tag{D.12}$$

and therefore

$$Z_R(\omega) = R. \tag{D.13}$$

For a capacitor

$$Z_C(s) = \frac{\mathcal{L}\{V(t)\}}{\mathcal{L}\{CdV(t)/dt\}}, \tag{D.14}$$

$$Z_C(s) = \frac{\mathcal{L}\{V(t)\}}{sC\mathcal{L}\{V(t)\}}, \tag{D.15}$$

$$Z_C(\omega) = \frac{1}{j\omega C}. \tag{D.16}$$

The constant phase element (CPE) is a modification of equation D.16, defined as

$$Z_{CPE}(\omega) = \frac{1}{Qj\omega^{\alpha}}, \tag{D.17}$$

where Q is the pseudo-capacitance, and α takes a value between 0 and 1. When $\alpha = 1$, then the CPE representation is identical to Z_C, and when $\alpha = 0$, then the CPE is identical to Z_R. These dynamics can be visualised with a Nyquist plot, as shown in the top row of figure D.1, which plots the real against the imaginary component of $Z(\omega)$ (with Z_{Im} being scaled by -1 for convenience). In the leftmost plot, we see the impedance of a resistor for different values of the constant R. As the resistor has no imaginary impedance response, and no frequency dependence, the data in the Nyquist plot shows a single point on the real axis, at the value of the resistor. In the time-domain, the current is in-phase with the potential. For the capacitor, the impedance has an imaginary impedance component that is scaled by the frequency, with no real component (and as such, the projection on the real axis is 0). In the time-domain, the current is shifted by 90 degrees — this phase angle can also be observed in the Nyquist plot, as the angle between the real axis and the impedance response. These two observations characterise the rightmost plot, that of the constant phase element. The value of α varies between 0 and 1, varying the phase angle between that of a resistor (0) and a capacitor (90). Consequently, at $\alpha = 0.5$, the phase angle is 45 degrees. This variable phase angle also manifests itself in the time-domain response.

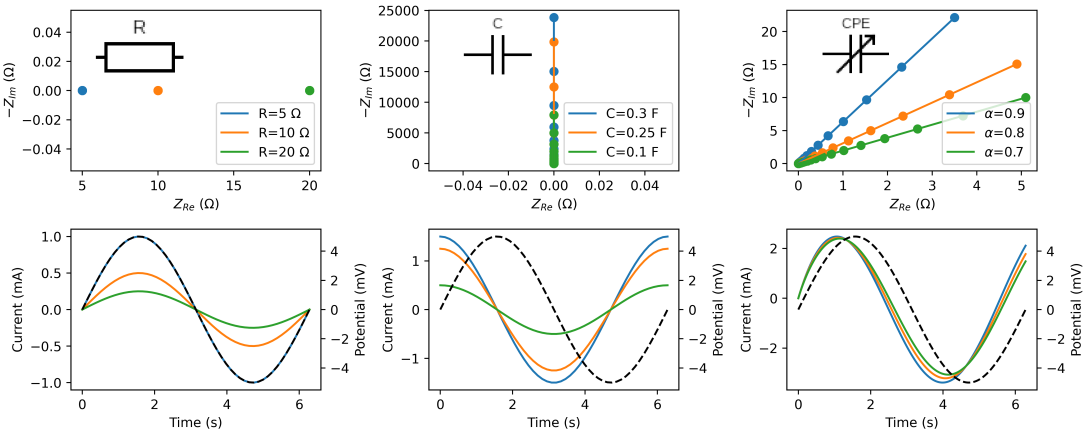


Figure D.1: Top: Nyquist plots of three components, from left to right: a resistor, a capacitor and a constant phase element. The range of frequencies used was $10^{-1.5}$ - 10^{5} Hz. Bottom: The simulated time-domain current response to a sinusoidal potential with an amplitude of 5 mV and a frequency of 0.16 Hz, as shown in black.

D.3 Equivalent circuit analysis

It is possible to model a very large quantity of impedance responses $Z(\omega)$ by combining the above elements in series or in parallel, using the appropriate circuit rules. For two elements $Z_1(\omega)$ and $Z_2(\omega)$ in series, the total impedance $Z_3(\omega)$ is

$$Z_3(\omega) = Z_1(\omega) + Z_2(\omega), \tag{D.18}$$

and in parallel, the relationship is

$$\frac{1}{Z_3(\omega)} = \frac{1}{Z_1(\omega)} + \frac{1}{Z_2(\omega)}. \tag{D.19}$$

D.4 Equivalent circuit models for a redox process

Although impedance modelling was not undertaken during the course of the DPhil, it is possible to make some educated guesses based on the electrochemical literature, and derive some conclusions thereof. One of the most famous equivalent circuits is the Randles circuit, which models a generic electrochemical cell in a solution-phase voltammetry experiment, in the absence of catalysis. This has 4 components. R_s represents the solution resistance, C_{dl} the double layer capacitance, R_{ct} the resistance to charge transfer (i.e. a Faradaic process) and a Warburg element

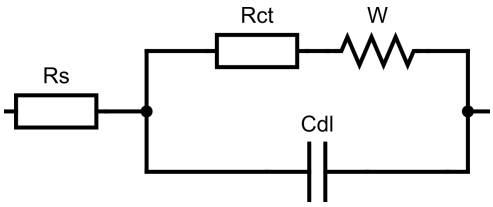


Figure D.2: Equivalent circuit drawing of the Randles circuit

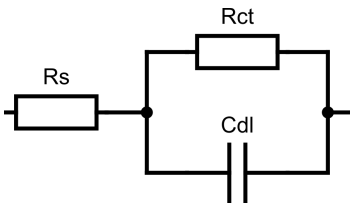


Figure D.3: Suggested equivalent circuit for a surface-confined voltammetry experiment

W which is introduced to model diffusive processes. The intuition behind the construction of this circuit is that the charge-transfer process is controlled by the diffusive process (and R_{ct} is hence in series with W). These two processes are occurring in parallel with C_{dl}, and are modelled as such.

Using this intuition, we can suggest a circuit that occurs for a surface confined redox process, by removing the element that models diffusion. Consequently, in the ideal case (as is also predicted by a Butler-Volmer modelling approach), the phase of the capacitive current will be shifted by 90 degrees relative to the Faradaic current.