Design Space Exploration in Ion Channels Using Fine Grained Brownian Dynamics

by

May Siksik

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2020

© May Siksik, 2020
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Design Space Exploration in Ion Channels Using Finite Grained Brownian Dynamics

submitted by  May Siksik  in partial fulfillment of the requirements for
the degree of  Doctor of Philosophy
in  Electrical and Computer Engineering

Examinining Committee:

Andre Ivanov
Supervisor
Matthew Hoyles
Supervisory Committee Member
Karen Cheung
Supervisory Committee Member
David Fedida
University Examiner
Purang Abolmaesumi
University Examiner

Additional Supervisory Committee Members:

Eric Accili
Supervisory Committee Member
Abstract

Design Space Exploration (DSE) in the context of ion channels refers to the systematic exploration of a design space defined with the dimensions of the space corresponding to channel characteristics. The goal of DSE here is to find points within the space that maximize a figure of merit related to conduction. Finding an efficient means to performing DSE for ion channels holds promise in several application areas such as nano-medicine and drug development where it is commonly desirable to efficiently “reverse-engineer” drugs or channels in order to determine which channel structures or drug lead to a particular conduction behavior. One example of DSE would have the dimension of the design space specified by dielectric constants throughout the channel, and with a figure-of-merit defined by the conduction. The primary roadblock to using DSE for ion channels is computational complexity as evaluating each channel characteristic (design point) requires $10^{10}$ simulation iterations. If, for example, the design space is defined by 5 parameters each having 10 possible values, the process of evaluating all possible combinations of these parameters exhaustively would require $10^5 \times 10^{10} = 10^{15}$ simulation iterations. Depending on the time it takes for each iteration, a DSE study could take years or even decades. As a result, it is critical that the approach used for evaluating each design point is both fast and efficient in order to save on simulation time and computational resources. This thesis proposes two approaches for improving the efficiency of DSE for ion channels. First, it proposes an approach for improving the speed of DSE by systematically reducing the design space size using statistical-based inference. It shows how this methodology can be utilized to reduce the design space size by orders of magnitude for two different scenarios: with and without the presence of a drug in the channel. Second, it proposes a novel Fine Grained Brownian
Dynamics framework for evaluating design points. Using both approaches together, the framework achieves accuracy that is consistent with Molecular Dynamics (with $R^2 = 82\%$), a significantly higher resolution modeling technique, at a fraction of the cost.
Lay Summary

Ion channels are water filled nano-pores in the cell membrane. Their malfunction causes diseases such as cancer, epilepsy, diabetes, and several types of autoimmune disease [1]. This research work proposes a methodology that allows for deducing channel characteristics that would result in a particular channel behavior. Such a tool would allow an unrestricted environment for experimentation to elucidate how these channels function at a molecular level, providing a much needed insight for drug development, implantable medical devices development, and other nano-medicine applications. This characterization of ion channel properties/function has been infeasible due to the computational resources required. The proposed framework allows such characterization by addressing two issues:

- Reducing the number of required combinations of channel characteristics (i.e. the size of the design space).
- Reducing the time required for the evaluation of combinations of channel characteristics while maintaining evaluation accuracy.
Preface

All of the work presented henceforth was conducted in the Statistical Signal Processing Group at the University of British Columbia, Point Grey campus.

A version of Chapter 3 and Section 7.1 has been published in the Eighth IEEE International Workshop on High Performance Computational Biology [Siksik, M. and Krishnamurthy, V., Stochastic Multi-particle Brownian Dynamics Simulation of Biological Ion Channels: A Finite Element Approach, DOI: IPDPS.2009.5160932], and the Annual International Conference of the IEEE Engineering in Medicine and Biology Society [Siksik, M. and Krishnamurthy, V., An Asymmetric Approach to Modeling Ion Channels Using Finite Element Analysis, DOI: IEMBS.2009.5332636]. The modeling technique proposed in Chapter 3 improves upon an existing Brownian Dynamics approach developed by Dr. Shin-Ho Chung’s research group at the Australian National University College of Science [2]. I was the lead investigator, responsible for all major areas of concept formation, data collection and analysis, as well as manuscript composition. Dr. Vikram Krishnamurthy was the supervisory author on this project and was involved throughout the project in concept formation and manuscript edits.

A version of Chapter 4, Chapter 5, and Section 7.3 has been published in the IEEE Transactions on Nanobioscience [Siksik, M. and Krishnamurthy, V., Multi-dielectric Brownian Dynamics and Design Space Exploration Studies of Permeation in Ion Channels, DOI:TNB.2017.2723002]. I was the lead investigator, responsible for all major areas of concept formation, data collection and analysis, as well as manuscript composition. Dr. Vikram Krishnamurthy was the supervisory author on this project and was involved throughout the project in concept formation and manuscript edits.
A version of Chapter 6, Section 7.2, Section 7.4, and Section 7.5 is being prepared for publication submission. I was the lead investigator, responsible for all major areas of concept formation, data collection and analysis, as well as manuscript composition. Dr. Andre Ivanov and Dr. Vikram Krishnamurthy were the supervisory authors on this project. Dr. Vikram Krishnamurthy was involved in the early stage of the project in concept formation and manuscript edits. Dr. Andre Ivanov was involved in the later stage of the project in manuscript edits.
# Table of Contents

Abstract ................................................................. iii

Lay Summary ......................................................... v

Preface ................................................................. vi

Table of Contents ............................................... viii

List of Tables ....................................................... xii

List of Figures ...................................................... xiii

Abbreviations ....................................................... xvi

Acknowledgements ................................................... xvii

Dedication ............................................................. xviii

1 Modeling Permeation in Ion Channels ......................... 1
   1.1 Ion channels .................................................. 1
   1.2 Understanding the Permeation Mechanisms .................. 2
   1.3 Computational Modeling of Permeation .................... 3
   1.4 Design Space Exploration ................................. 4
   1.5 Design Space Exploration of Permeation in Ion Channels 5
   1.6 What Makes A Good Computational Model .................. 5
   1.7 Computational Modeling Approaches ...................... 7
      1.7.1 Quantum Mechanics .................................. 7
      1.7.2 Full Atomic Molecular Dynamics ................... 7
      1.7.3 Coarse-Grained Molecular Dynamics ................. 8
      1.7.4 Brownian Dynamics ................................. 9
      1.7.5 Continuum Modeling Techniques ................. 9

2 Research Statement ............................................... 11
   2.1 Modeling of Ion Channels: Challenges .................... 11
   2.2 Significance of Modeling Dielectric Heterogeneity in BD 13
   2.3 The Need for DSE in the Context of Ion Channels .......... 15
   2.4 Issues with Using Current Approaches for DSE Studies of Permeation in Ion Channels 17
5.6 Design Space Reduction .................................................. 59  
5.6.1 Sensitivity Analysis ...................................................... 59  
5.6.1.1 Overview .............................................................. 59  
5.6.1.2 PEP Sensitivity to Each Design Parameter ....................... 60  
5.6.1.3 Optimization Based on Sensitivity Analysis Data ................. 62  
5.6.1.4 Hierarchical Clustering ............................................. 68  
5.6.1.5 Association Analysis .............................................. 69  
5.6.1.6 Correlation Analysis .............................................. 70  
5.7 Results and Discussion .................................................. 71  
5.7.1 Estimated Dielectric Map Based on Golden Reference ............... 71  
5.7.2 Model Validation: DSE Estimation is Consistent With MD Computations . 73  

6 DSE Study 2 ................................................................. 74  
6.1 DSE Study Overview .................................................... 74  
6.2 Background ............................................................... 75  
6.3 Numerical Set-up .......................................................... 76  
6.4 DSE Cost Function and Design Parameters ............................ 76  
6.5 Design Space Reduction .................................................. 77  
6.5.1 Sensitivity Analysis .................................................... 78  
6.5.1.1 PEP Sensitivity to Each Design Parameter ....................... 78  
6.5.1.2 Optimization Based on Sensitivity Analysis Data ................. 83  
6.5.1.3 Hierarchical Clustering ............................................. 88  
6.5.1.4 Association Analysis .............................................. 89  
6.5.1.5 Correlation Analysis .............................................. 91  
6.6 Results and Discussion .................................................. 92  
6.6.1 Impact of Dielectric Heterogeneity on Channel Model ............... 92  
6.6.2 Finalizing Estimation .................................................. 93  
6.6.3 Accuracy of FGBD in Comparison to Homogeneous BD and MD ....... 93  
6.7 Relevance of Dielectric Heterogeneity in Drug Design ................ 98  

7 Computational Feasibility ............................................... 100  
7.1 Scalability Limits of FGBD .............................................. 100  
7.1.1 Limits of Spatial and Temporal Resolutions ......................... 100  
7.1.2 Comparing scalability of FEM and BEM Based PEP Calculations .... 104  
7.2 Performance Improvement of BD Over MD for A Single Point in the Design Space 106  
7.3 Runtime Complexity Improvement for DSE ............................ 109  
7.4 Overview of FGBD in Comparison to Other Relevant Approaches ....... 110  
7.5 Scalability of Design Space Exploration ................................ 111  
7.6 DSE Throughput using High Performance Computing and Cloud Computing 113  

8 Conclusion ................................................................. 117  
8.1 Problem Addressed ...................................................... 117  
8.2 Research Summary ...................................................... 118  
8.3 Limitations ............................................................... 122  
8.4 Future Work .............................................................. 123
8.4.1 Introduction of Machine Learning ........................................ 123
8.4.2 Correlating Structural Changes and Dielectric Value ................ 125
8.4.3 Improving the Resolution of Dielectric Heterogeneity ................ 125
8.4.4 Incorporating Asymmetric Channels Models ............................. 125
8.4.5 Investigating Drugs ......................................................... 126

Bibliography ................................................................. 127
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Typical resource usage for the KcsA biological ion channel case study.</td>
<td>39</td>
</tr>
<tr>
<td>3.2</td>
<td>Computational cost of BEM vs. FEM for both homogeneous and heterogeneous domains.</td>
<td>40</td>
</tr>
<tr>
<td>5.1</td>
<td>Disjoint design space regions defined by z-axis ranges.</td>
<td>62</td>
</tr>
<tr>
<td>5.2</td>
<td>Comparing the dielectric map from solution 1 which is estimated using DSE given the golden reference PEP by Chung et al. [3] to the dielectric map computed using MD, and the associated error in solution 1.</td>
<td>72</td>
</tr>
<tr>
<td>5.3</td>
<td>Comparing the dielectric map from solution 2 which is estimated using DSE given the golden reference PEP by Chung et al. [3] to the dielectric map computed using MD, and the associated error in solution 2.</td>
<td>72</td>
</tr>
<tr>
<td>6.1</td>
<td>Disjoint design space regions defined by z-axis ranges.</td>
<td>87</td>
</tr>
<tr>
<td>6.2</td>
<td>Comparing DSE estimated dielectric maps for the TEA-occupied KcsA channel using MD computed PEP by Crouzy et al. [4] as a golden reference and using the PEP that reproduces the experimentally determined inhibitory constant computed by Heginbotham et al. [5] as a golden reference to the dielectric map computed using MD for the empty channel. The presence of the TEA molecule dampens the dielectric constant in and around the region where the molecule resides. Dielectric map estimated using MD computed PEP deviates from the dielectric map estimated using the PEP that reproduces experimental data in the lower filter region.</td>
<td>96</td>
</tr>
<tr>
<td>7.1</td>
<td>Problem size and matrix size results for BEM and FEM formulations of the KcsA channel.</td>
<td>106</td>
</tr>
<tr>
<td>7.2</td>
<td>Comparison of algorithm attributes and performance for FGBD, BD, and MD.</td>
<td>112</td>
</tr>
<tr>
<td>7.3</td>
<td>Estimates of runtime for 100k iterations of a single design point evaluation for the reference workstation (used in the computational cost table in Section 3.3.2), a 2019 workstation, and four example cloud computing instances currently available on Amazon Elastic Computing (EC2) [6].</td>
<td>115</td>
</tr>
</tbody>
</table>
# List of Figures

2.1 Permeation in the KcsA channel [7] .......................................................... 15

3.1 Fine-Grained Brownian Dynamics framework ............................................. 26

3.2 A top level description of the molecular geometry generator for experimentally- derived molecular structures .......................................................... 27

3.3 Symmetric Channel Geometry Generation Algorithm ................................. 29

3.4 KcsA channel structure generated using the scripting-based geometry generator, . 30

3.5 A conceptual diagram illustrating the fragmentation issue when generating the channel surface: Channel boundary gets fragmented as the surface grows beyond those atoms closest to the z-axis at a high MCA threshold. Fragmentation can be avoided using radial boundary conditions that stops the surface from growing beyond the atoms closest to the z-axis. ......................................................... 32

3.6 KcsA channel structure generated using the Asymmetric Channel Geometry Generation Algorithm: The KcsA biological ion channel. .......................... 33

4.1 Types of DSE using proposed framework. .................................................. 46

4.2 DSE methodology (DSE Engine) ............................................................... 49

5.1 The 6 dielectric regions of the KcsA channel which represent the different design parameters. ................................................................. 58

5.2 PEP for upper reservoir dielectric region values 50-80. ............................... 60

5.3 Rate of change in potential energy as a function of the rate of change of the upper reservoir dielectric. .............................................................. 61

5.4 PEP for upper channel dielectric region values 1-20. .................................. 62

5.5 Rate of change in potential energy as a function of the rate of change of the upper channel region dielectric. ......................................................... 63

5.6 PEP for filter dielectric region values 16-40. .............................................. 63

5.7 Rate of change in potential energy as a function of the rate of change of the filter region dielectric. ................................................................. 64

5.8 PEP for pore dielectric region values 48-72. .............................................. 64

5.9 Rate of change in potential energy as a function of the rate of change of the pore region dielectric. ................................................................. 64

5.10 PEP for protein dielectric region values 2-16. ......................................... 65

5.11 Rate of change in potential energy as a function of the rate of change of the protein region dielectric. ............................................................... 66

5.12 Hierarchical clustering for the KcsA channel indicating the design parameters that can be optimized together. ......................................................... 68
5.13 The design parameters, $X$, and their impact on the cost function, $f(x)$ as they vary for a range of values.

5.14 A scatterplot showing the relationship between design parameters and the cost function for a range of values to reveal any linear relationships with the cost function: The cost function varies significantly with changes in the dielectric for the protein, upper channel and filter; and has a monotonic relationship with the filter dielectric.

5.15 Linear correlation of the design parameters with the cost function for the KcsA DSE case-study. 1 indicates maximum positive correlation and 0 indicates no correlation.

5.16 A comparison of the PEP computed using FGBD with MD-computed dielectric map, computed using FGBD using a homogeneous dielectric, computed using FGBD with DSE deduced dielectric map, and the golden reference PEP from Chung et al. [3] which reproduces I/V curve reported in Heginbotham et al. [5].

6.1 Two of the four subunits of the KcsA protein are represented as ribbons. Water molecules are shown in gold. The circled region of the protein, containing the selectivity filter is enlarged on the right-hand side and shown in a stick-and-ball model, together with three water molecules and two potassium ions (gold) [2].

6.2 Dielectric regions in TEA-occupied KcsA channel.

6.3 PEP for upper upper channel dielectric region values 1, 6, 10, 12, and 20. A higher dielectric value causes the wells to become shallower.

6.4 Rate of change in potential energy as a function of the rate of change of the upper upper channel dielectric region.

6.5 PEP for lower upper channel dielectric region values 2, 5, 12, and 20. A higher dielectric value causes the wells to become shallower.

6.6 Rate of change in potential energy as a function of the rate of change of the lower upper channel dielectric region.

6.7 PEP for upper filter dielectric region values 4, 8, 16, 20, and 25. An increase in the dielectric value causes a significant increase in the barrier height.

6.8 Rate of change in potential energy as a function of the rate of change of the upper filter dielectric region.

6.9 PEP for lower filter dielectric region values 8, 10, 12, 16, and 20. An increase in the dielectric value causes the wells to become shallower.

6.10 Rate of change in potential energy as a function of the rate of change of the lower filter dielectric region.

6.11 PEP for upper pore dielectric region values 48, 50, ... 72. The change in dielectric value does not have a significant impact on the PEP.

6.12 Rate of change in potential energy as a function of the rate of change of the upper pore dielectric region.

6.13 PEP for lower pore dielectric region values 48, 50, ... 72. The change in dielectric value does not have a significant impact on the PEP.

6.14 Rate of change in potential energy as a function of the rate of change of the lower pore dielectric region.

6.15 PEP for protein dielectric region values 1, 2, 3, 4, 5, and 6. A higher dielectric value in the protein regions causes the wells to become shallower.

6.16 Rate of change in potential energy as a function of the rate of change of the protein dielectric region.
6.17 Hierarchical clustering for the TEA-occupied KcsA channel indicating the design parameters that can be optimized together. .......... 89
6.18 The design parameters, X, and their impact on the cost function, f(x) as they vary for a range of values. ................................. 90
6.19 A scatterplot showing the relationship between design parameters and the cost function for a range of values to reveal any linear relationships with the cost function; the protein, lower upper channel, and upper filter dielectric regions values have a significant impact on the cost function, whereas all other dielectric regions have a minimal impact. Unlike the protein, the lower upper channel dielectric and the upper filter’s both have a monotonic and linear relationship with the cost function. .......... 90
6.20 Linear correlation of the design parameters with the cost function. 1 indicates maximum positive correlation and 0 indicates no correlation ................................. 91
6.21 Manipulating the PEP through dielectric heterogeneity through the channel. The legend shows the dielectric values for each region in the form [protein lower reservoir upper reservoir lower pore upper upper channel upper pore lower upper channel lower filter upper filter]. ......................................................... 94
6.22 Manipulating the PEP through dielectric heterogeneity in the upper part of the channel ................................. 94
6.23 Current as a function of TEA concentration for both BD simulated (solid circles) and experimentally determined (open diamonds) [2] with experimental results reproduced from [5]. The applied potential and the ionic concentration used for the simulations are 196 mV and 150 mM, respectively. The depth of the energy-well encountered by the TEA molecule is 13kT. ......................................................... 97
6.24 Geometric model of the KcsA channel showing the TEA binding pocket and the region for which the PEP is compared ................................. 97
6.25 A comparison in the TEA binding pocket region of the KcsA channel between the PEP computed using FGBD with DSE-estimated dielectric map, the PEP computed using FGBD with homogeneous dielectric, and the golden reference PEP computed directly using MD by Crouzy et al. [4]. The PEP computed using FGBD with homogeneous dielectric (similar to classic BD) deviates significantly from the golden reference computed using MD. ......................................................... 98
6.26 A comparison between the PEP estimated using FGBD with the DSE-estimated dielectric map and the golden reference PEP from [2] that reproduces the inhibitory constant determined experimentally by Heginbotham et al. [5]. The depth of the well in the PEP computed using the framework is very close (12.5KT) to that of the golden reference (13KT). ......................................................... 99

7.1 Memory footprint corresponding to the different grid sizes in a KcsA channel. .......... 101
7.2 The error in electric field estimation as a function of the table size. ............... 102
7.3 Aggregate error in the interpolated electric field as a function of the number of ions in the channel and the memory footprint. ............... 103
7.4 The z-position of an ion trajectory for temporal resolutions, Δt = 2fs, 5fs, ..., 100fs. 104
7.5 Memory-usage cost as a function of granularity of heterogeneity. ............... 106
7.6 Impact of DSE methods used on DSE cost in terms of number of points visited. 110
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD</td>
<td>Brownian Dynamics</td>
</tr>
<tr>
<td>BEM</td>
<td>Boundary Element Method</td>
</tr>
<tr>
<td>CGMD</td>
<td>Coarse Grained Molecular Dynamics</td>
</tr>
<tr>
<td>DSE</td>
<td>Design Space Exploration</td>
</tr>
<tr>
<td>FEM</td>
<td>Finite Element Method</td>
</tr>
<tr>
<td>FGBD</td>
<td>Finite Grained Brownian Dynamics</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MGGT</td>
<td>Molecular Geometry Generator Tool</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>PEP</td>
<td>Potential Energy Profile</td>
</tr>
<tr>
<td>PEPC</td>
<td>Potential Energy Profile Calculator</td>
</tr>
<tr>
<td>PMF</td>
<td>Potential of Mean Force</td>
</tr>
<tr>
<td>PNP</td>
<td>Poisson Nernst Planck</td>
</tr>
</tbody>
</table>
Acknowledgements

Thanks to Dr. Andre Ivanov for his support and guidance. Thanks to Dr. Matthew Hoyles for all the valuable discussions and insights throughout the project. Thanks to Dr. Shin-Ho Chung for allowing me access to his group’s BD program (classic BD), and for providing me with the MD prepared structure for KcsA. Thanks to Dr. Steven Plotkin for providing me with the molecular dynamics-computed dielectric map for the KcsA ion channel. Thanks to my committee members for their thorough review of my work.
Dedication

To my beautiful family.
Chapter 1

Modeling Permeation in Ion Channels

1.1 Ion channels

The transfer of ions in and out of biological cells is needed for various activities including growth and reproduction. The cell membrane however is hydrophobic and, as a result, is a very effective barrier to such transfer. Biological cells have evolved to have pores in their membrane that facilitate this transfer. These water-filled pores, called ion channels, traffic the flow of ions in and out of the cell by opening and closing in response to a stimulus such as the change in the membrane potential in the case of voltage gated ion channels.

Ion channels play a wide range of critical biological functions [8–10]. Different types of channels exist in living organisms, each having its own properties that are designed for a specific physiological role. Consequently, the set of ion channels found in a certain cell will define its characteristic electric signature. For example, ion channels found in heart muscle determine the repetitiveness and duration of the cardiac action potentials. On the other hand, ion channels found in certain nerve cells determine whether these cells fire at a high frequency rate or only transiently [11]. As a result, biological ion channels are responsible for the diversity in electrical signaling found in all excitable cells such as nerve and muscle cells, and are the driving force behind all movement, sensation and thought processes in all living organisms.
In humans, failure of ion channels to function properly is known to cause a broad range of diseases including cancer, diabetes, hypertension, epilepsy, and chronic pain [1, 12-14]. At least 55 different medical conditions are known to be caused by ion channel malfunction [10, 15]. As a result, understanding their structural and functional properties can provide much needed insight into not only these diseases, but all disorders of cellular electrical activity, the influence of drugs and hormones on the cell, properties of the muscular system, and the unique properties of the nervous system. In addition to the better understanding of medical disorders, reverse-engineering ion channels has substantial benefits to applications such as nano-biotechnology and nano-electronics [9], and bio-sensor design.

1.2 Understanding the Permeation Mechanisms

Permeation in ion channels is a widely researched area. However, there remains a number of important issues that are still under debate and require a better understanding. These include the significantly different behavior that structurally similar channels display (e.g. significantly varying sensitivity to a particular blocker), the precise mechanisms behind potassium channels’ ability to be highly selective for potassium ions over sodium despite of having identical charge and the role of ion hydration in the permeation process, etc. [16].

To understand the behavior of ion channels and relate their structure to function, it is important to examine the microscopic mechanisms that lead to the channel’s macroscopic behavior observable in physiological experiments (i.e. conduction). A modeling approach must capture ions’ dynamics at a femtosecond time scale to avoid loss of accuracy when tracing these ions’ trajectory for many microseconds in order to reproduce experimentally observed conduction results. This requires billions of simulation iterations making the development of an accurate ion channel simulation model a challenging task (see section 2.1 for more details).
1.3 Computational Modeling of Permeation

Structural studies performed using techniques such as X-ray crystallography and Cryo-electron microscopy can only provide a static picture of the system. This is limiting because the permeation process is inherently dynamic and thus it is challenging to gain understanding of channel behavior with only static pictures. Wet-lab experiments such as the patch clamp method allow researchers to watch some aspects of ion channel dynamics. In the patch clamp method, a small patch of membrane is isolated electrically using a very small glass pipette that is filled with a suitable electrolyte solution and is pressed against the membrane such that the seal between pipette and membrane become so tight that no ions can flow between the pipette and the membrane. The patch clamp can be used to investigate various ionic currents including single and multi-channel current and the current resulting when certain drugs are present. Depending on the goal of the study, the patch clamp can be applied in different configurations. For example, a single channel current can be recorded by attaching the pipette to a single channel such that all the ions that flow when a single ion channel opens must flow into the pipette. The current flowing into the patch can then be measured with a sensitive amplifier connected to the pipette. Although the patch clamp technique allows for watching certain ion channel dynamics, it is not possible to observe the dynamics at the time scale required to understand the mechanisms behind conduction.

Conversely, numerical simulations provided through computational modeling can reveal microscopic behavior and dynamics of the channel over a period of time. To most effectively simulate the dynamics of an ion channel, a computational model should consider the following aspects: the membrane bi-layer, protein atoms that form the channel, water molecules inside the channel, and ions that go through the channel. Modeling each of these components poses its own challenges such as the correct representation of electrostatic and electro-dynamic properties of the water molecules and ions inside the channel, the construction of the system boundaries, and polarization effects.

Computational modeling of the permeation process is an active area of research with a variety of approaches being explored by a number of research groups (e.g., [16–28]). The quality of the information obtained from such simulations is reliant on the computational model used. To date,
there has not been a computationally feasible software model that is capable of modeling varying polarizability in the channel.

### 1.4 Design Space Exploration

Design space exploration [29–34] (DSE) is the process of exploring and evaluating design alternatives in order to determine the set of design parameters for a system such that a particular metric is optimized. This is typically done during system development and can make implementation significantly more effective and efficient. DSE has been widely used in many engineering applications where there are a large number of parameters to optimize such as embedded systems design, computational fluid dynamics, neural networks design, cyber security, rapid prototyping, and multi-component system configuration and integration. For example, in cyber security, design parameters could include the architecture of the controller, the choice of algorithm used, and the sampling rate/clock speed.

DSE is challenging because of the sheer number of design alternatives to be explored. A large complex system may have millions, if not billions of design alternatives. In some cases, the number of design alternatives may be large enough to be considered infinite from a computational perspective. Manual ad-hoc approaches to DSE are tedious, inefficient, and might lead to inferior solutions if optimal design parameters combinations are not explored. Exhaustive approaches, on the other hand, are computationally infeasible.

The set of all design alternatives for an optimization problem is referred to as a design space with each point in the space representing a unique set of parameter values. To explore the design space is the iterative process of evaluating each point in the space to determine how optimal it is. For optimization parameters with a large number of parameters, and thus a large design space, DSE includes a method of simplification by which the design space is reduced through a set of approximations.
1.5 Design Space Exploration of Permeation in Ion Channels

Computational modeling is an ideal tool for observing molecular interactions. However, characterizing the channel characteristics/function relationship additionally requires the determination of parameter values that lead to a target behavior. The computational effort of DSE is not just dictated by the size of the space, but also the cost of evaluating each point visited in the design space. For ion channels, this is particularly costly because the evaluation of each point requires $10^{10}$ simulation iterations (see Section 2.1 for more details).

The goal of DSE is to find points in the design space that minimize a cost function. In this thesis, the cost function is the mean square error between a given simulated “permeation characteristic” computed using the proposed framework and the expected result of the same "permeation characteristic". The expected permeation characteristic can be an ion channel function measured through physiological experiments (e.g. channel conductance), or computed using other models (e.g. electric potential energy across the channel), or set by the user (e.g. channel blockage time when evaluating drug molecules). This expected characteristic is called the golden reference.

Chapter 4 provides more details on the DSE problem of permeation in ion channels, and describes a methodology for reducing the size of the design space and determining a set of “near optimal modeling parameters”.

1.6 What Makes A Good Computational Model

Generally speaking, modeling permeation involves the following steps:

1. A working structure has to be established from the topological information. First the system’s topological information is obtained either from the Protein Data Bank [35–38], or through computational models such as homology modeling [39, 40]. A “working structure” is then prepared using complex Molecular Dynamics simulations which involve building, aligning, and
combining the membrane and protein; adding water molecules (solvating) to the channel; performing energy minimization and equilibration with the protein constrained then equilibration with the constraints released. Energy minimization is done to relax the system such that any unnatural stress (e.g. atoms too close to each other), which might result in large forces that can possibly blow up the system, is adjusted so that the system can reach the closest local energy minimum. Equilibration is then done to further relax the system (i.e. water molecules that were created as a crystal lattice can rearrange around the solute).

2. A geometric model of the structure is then generated along with grid points through out the channel. This is described further in Chapter 3

3. Electrostatic forces acting on each ion have to then be computed. This is done by solving Poisson’s differential equation numerically to compute the Potential Energy Profile (PEP) experienced by each ion for all possible ions’ positions in the channel. This is described further in Chapter 3

4. Finally, ions trajectories are computed using Langevin’s stochastic differential equation at each time iteration of the simulation. The channel current or conductance can then be computed by tracking the number of ions that enter and leave the channel. This is described further in Chapter 3

Often, models are developed to address specific steps of the permeation modeling process described in 1.6. For example, this research work addresses steps 2 and 3. Any modeling technique that is suitable for DSE should satisfy two criteria: an acceptable level of accuracy, and computational feasibility. A modeling technique that is appropriate for modeling permeation dynamics in ion channels, on the other hand, has to satisfy three criteria: accuracy, computational feasibility, and flexibility. Computational feasibility is critical because billions of simulation iterations are required to model permeation. Flexibility is required so that the model can be used to study a variety of ion channels. Section 1.7 describes the most common computational modeling approaches and discusses the issue of model flexibility further.
To address the criteria required for modeling permeation dynamics, many computational models have been proposed. These models range from highly accurate, highly detailed with a high computational cost (e.g. molecular modeling) to abstract, less accurate, and as a result computationally inexpensive models (e.g. continuum electrostatics). In practice, detailed models are limited by the size of systems they can model and the possible simulation period. It is critical that a computational model can be optimized in terms of complexity, accuracy, and cost such that it can give the required level of accuracy for the simulation period and system size of interest. The next section highlights the most common modeling approaches.

1.7 Computational Modeling Approaches

1.7.1 Quantum Mechanics

Quantum mechanics models such as ab initio [41, 42] provide the most detail and the highest level of accuracy but at the greatest computational cost. For these models, system properties are computed from first principles by solving the Schrodinger equation numerically, and as a result, are highly computationally-intensive. Semi-empirical models still model at the electronic scale but use simplified versions of equations from ab initio models.

1.7.2 Full Atomic Molecular Dynamics

Classical Molecular Dynamics (MD) [43–52] provide a level of abstraction for which modeling is accomplished at the atomic scale. The functional form and parameters that define how the potential can be computed are called the force field model. The force fields used in the MD simulation play an important role in defining the structural model for the molecular system. These force field models are typically developed by combining experimental data and ab initio computations for small models that form larger systems (e.g. atom models) [53–55]. These models are then parametrized to include several atom types which describe various configurations of the same atoms/functional groups. As
a consequence of this approach, MD simulations are as accurate as the degree to which the force field models being applied have captured the "reality" of the forces working on the atoms of the system being studied. An additional limitation caused by this approach is that force field models are not easily transferable (i.e. for studying different systems) thus resulting in overall modeling inflexibility.

MD has a speed improvement over quantum mechanics and consequently it can be used to simulate larger systems than quantum mechanics models; however, MD is still too computationally demanding to be used for simulating those time periods required to compute macroscopic (experimentally observable) properties (e.g. current in biological ion channels). Without the use of specialized hardware acceleration, it not possible to use MD for DSE experiments which require a very large number of simulation iterations.

Another significant limitation of classical MD is its inability to model certain important quantum effects such as electronic polarization, which is relevant in many important processes (e.g. Oxygen binding to hemoglobin [53, 54]). An example where MD failed to compute the electrostatics that reproduces experimentally determined results is the Gramicidin channel [25–27]. In [25], the authors altered the Potential of Mean Force (PMF), from which a Potential Energy Profile (PEP) can be computed, in order to arrive at a PEP that would reproduce experimental results. In recognition of this problem, various polarizable force field models for MD simulations have been developed (e.g. AMOEBA, CHARMM, and AMBER [56–58]). Unfortunately, polarizable force fields are in general slower and still suffer model inflexibility.

1.7.3 Coarse-Grained Molecular Dynamics

A variant of MD, called Coarse-Grained MD (CGMD), is an approach intended to expedite full-atomic MD by grouping atoms into groups called beads [59–61]. A disadvantage of coarse-grained MD that is worth noting is that its quality-of-result is heavily dictated by its beading granularity (i.e. the number of atoms grouped together to form a bead) and other application-specific beading
decisions, making this model even more inflexible and less transferable for studying other molecular systems (e.g. ion channels) other than what they have been optimized for.

### 1.7.4 Brownian Dynamics

Meso-scale modeling approaches such as Brownian Dynamics (BD) [2, 62–75] provide another level of abstraction that reduces computational complexity allowing for longer simulation periods. BD is considered to be a middle-of-the-road approach that balances simulation accuracy and computational complexity allowing for macroscopic properties to be reproduced in simulations. The downside to BD, however, is that such abstraction can lead to some degree of freedom being averaged out as large groups of atoms are treated as single entities. For example, when BD is applied to ion channels, the solvent is modeled as a dielectric and only permeating ions are modeled explicitly (as point charges). More explicitly, three approximations are made: first, protein atoms are assumed to be rigid while in reality they vibrate; second, the dynamics of water molecules are not simulated and the net effect of collisions between water molecules and other particles are represented as frictional and random forces; and third, protein atoms and water molecules are represented as continuum dielectrics.

When BD is used to model ion channels, Poisson’s equation has to be solved every time step in order to compute the potential energy and electric field acting on each ion. This information is then used to compute the position of the ion using the Langevin stochastic differential equation.

### 1.7.5 Continuum Modeling Techniques

Continuum modeling techniques assume that matter is continuous and system properties are treated as fields. This significantly reduces the computational cost providing the ability to handle systems of larger sizes and longer time scales. The downside to these models is that they lack molecular detail and can thus be insufficient for explaining the molecular origins of macroscopic behavior. PNP and Reaction Rate theories are common examples of these techniques:
• **Poisson Nernst Planck (PNP) Theory:** PNP \([76–81]\) theory provides a higher level of abstraction based on the mean field approximation such that permeating ions are no longer treated explicitly and instead are represented as charge densities. In this approach, the flux of ions is determined using the Nernst Planck equation.

• **Reaction Rate Theory:** This technique \([8, 18, 82]\) makes further simplifications where an ion channel is represented by a series of binding sites separated by energy barriers. The probability of ions hopping from one site to another is dependent on the height of the particular energy barrier.
Chapter 2

Research Statement

2.1 Modeling of Ion Channels: Challenges

To verify the validity of any ion channel model, the current, or the number of ions that pass through the channel per unit of time, computed using the model must be consistent with the current observed experimentally (i.e. simulated current is within the defined acceptable experimental error margin as measured in lab experiments using patch clamp neurophysiology). Channel current is typically measured in several to tens of picoamps range. A picoamp is 6 electrons per microsecond. As a result, to reproduce the channel current using a model, simulations must run for at least several microseconds. And in order to get statistically sensible measurements, these simulations must be run for tens of microseconds. On the other hand, in order to accurately represent relevant molecular interactions with the model, a femtosecond resolution is required for the simulation time step. This is because in the narrow part of the channel, where the ionic movement is Newtonian as opposed to diffusive, as a result, larger ionic movements due to larger time steps can result in infinite repulsive forces that cannot be integrated by the motion algorithm (the Berendsen Langevin equation discretization algorithm). This causes ions to go through the channel protein wall or “through” other ions. Further, larger time steps can result in loss of accuracy in the ion’s position.
The systematic deviation of an ion from its trajectory for time steps of 100 femtosecond or larger has been shown in [83] and is discussed in Chapter 6 of this thesis as well (See Figure 7.4).

Because computing channel current that reproduces experimental results requires simulations for tens of microseconds (e.g. $10^{-5}$ seconds), and a femtosecond time-resolution (e.g. $10^{-15}$ seconds) is required for proper modeling of channel dynamics, the number of simulations iterations required $= 10^{-5}/10^{-15} = 10^{10}$ iterations. Simulating the electrodynamics of a minimal representation of a channel for billions of iterations is challenging as multiple forces acting on each possible pair of the thousands of fixed and induced changes in the channel must be considered for each iteration.

The cost of modeling such a system can be computationally infeasible depending on the size and complexity of the ion channel being studied, the number of simulation iterations required to simulate the channel characteristic of interest, and the level of resolution/accuracy required. As discussed in Chapter 1, available modeling approaches make various abstraction choices to reduce the modeling cost. To date, there has not been a computational modeling approach that is considered to be the “standard” or agreed upon approach for modeling ion channels. BD is a middle of road approach in terms of its level of abstraction and computational cost, and it has been possible to reproduce experimental data for various permeation scenarios using BD. Dielectric heterogeneity is not modeled in classic BD, and as a result, BD might not be appropriate for modeling more complex permeation scenarios. Chapter 6 for example, shows that classic homogeneous BD fails to model the ion channel system when a blocker drug molecule (Tetraethylammonium or TEA) is present in the KcsA channel.

This represents a great computational challenge considering the fact that the channel has to be simulated at micromolar concentrations. On a spatial scale, the whole system is a nanometer in size and has to be resolved with resolution in the order of Angstroms, with thousands of atoms and many interacting forces that must be considered in order to avoid loss of model accuracy.

Introducing dielectric heterogeneity significantly adds to the computational cost of a BD model using classic Boundary Element Method-based BD approaches. Chapter 3 proposes a novel BD-based framework, called Fine Grained Brownian Dynamics (FGBD), that can model dielectric heterogeneity while substantially cutting computational cost. Chapter 6 shows how FGBD computations, for
the case of KcsA channel block by the Tetraethylamonium (TEA) drug where classic BD fails to model the channel, are consistent with those computed directly using MD.

2.2 Significance of Modeling Dielectric Heterogeneity in BD

The dielectric constant is a measure of the material’s polarizability and its ability to align its dipoles in the field direction. A recent MD study showed that the dielectric in an ion channel system is heterogeneous [84]. Other BD studies such as [3], however, found that in particular scenarios, channel conductance that is in agreement with experimental data [8] corresponds to a homogeneous dielectric value for the water in the channel. This discrepancy raises questions about the sensitivity of the channel model and its computations to dielectric heterogeneity.

Intuitively, it would be sensible to assume that the dielectric strength would be affected by the type of atoms surrounding the permeating ion in the channel. For example, the dielectric should be different when the ion is close to hydrophobic atom groups as opposed to polar groups. This dielectric heterogeneity can become significant in the case of ion channels since they are water filled pores and water is a strong dielectric, and as a result, polarization effects are of a greater extent. This is especially important in BD models where water molecules are not modeled explicitly and consequently the selection of an appropriate dielectric map for the channel can be crucial for model accuracy.

Permeation in the KcsA [85, 86] potassium ion channel provides a good example that illustrates the importance of modeling dielectric heterogeneity. Figure 2.1 shows the KcsA channel structure which contains four identical protein subunits (only two are shown in Figure 2.1) that fold into four transmembrane helices. The length of the channel is about 45 Å and it has three main components: a large water-filled cavity, a narrow part called the selectivity filter, and an upper channel opening where typically drugs bind. The crystal structure for this channel was resolved by Dr. Rod MacKinnon’s group in 1998 [86]. The channel opens in response to depolarization of the membrane potential and is highly selective favoring potassium ions (non-hydrated radius 1.33 Å) over sodium
ions (non-hydrated radius, 0.95 Å) with a ratio of 10000:1. When the channel opens, potassium ions flow causing hyperpolarization of the membrane potential. This high selectivity for potassium ions is interesting given that both the potassium and sodium ions have the same charge and the selectivity is for the larger ion that has a diameter that is wider than parts of the selectivity filter.

The precise mechanisms behind the selectivity process are still not fully understood and remain under constant debate. It has been suggested that the electrostatic interaction of the carbonyl groups in the selectivity filter with the potassium ions is much more energetically favorable than the sodium ions. This is because Carbonyl oxygens in the selectivity filter are too far away from the smaller sodium ion to compensate for the energy expense associated with the loss of water molecules required for entry [16, 87].

Aside from selectivity, the function of the KcsA channel is paradoxical: although the protein atoms that make the channel are hydrophobic (have a low dielectric constant) which implies a high energy barrier, it achieves a high throughput of potassium ions close to diffusion limit (10^8 per second) implying a low energy barrier. There has been several hypothesis regarding this behavior, for example, [86, 88, 89] have shown that the geometry of the KcsA channel and the water filled cavity help in the permeation process.

Unlike the cavity which is lined primarily with hydrophobic amino acid residues (low dielectric environment), the selectivity filter is formed by a five residue sequence, TVGYG (Threonine, Valine, Glycine, Tyrosine, Glycine), called the signature sequence, which are primarily polar carbonyl groups. Although the cavity is lined with low dielectric hydrophobic amino acid residues, it is polar because it is filled with water (about 60 water molecules) which is a strong dielectric. The upper channel opening also has a different dielectric environment due to its shape and size.

As shown in Figure 2.1, when an ion enters the water filled cavity, it becomes hydrated and remains hydrated as it goes through the cavity. However, for the ion to go through the selectivity filter, it must shed its hydration shell. The backbone carbonyl oxygen in the selectivity filter form a cage that fits the potassium ion perfectly replacing the water molecules that formed the hydration
shell. The helix dipoles (see Figure 2.1) also stabilize the potassium ion making the passage of the potassium ion through the channel much easier.

This variation in polarity in the environment around the ion as it goes through the channel has strong implications on channel dynamics, and modeling dielectric heterogeneity can be an effective and efficient way of capturing the effects of this variation.

2.3 The Need for DSE in the Context of Ion Channels

DSE of permeation in ion channels can be a valuable tool in many applications such as drug design, artificial pore design, nano-medicine, ion channel modeling optimization, predicting unknown ion channel structures, and others. When used for model optimization, DSE can be utilized to investigate the impact of modeling abstractions (e.g. the impact of particular approximations of geometry, the value of dielectrics given to water or channel wall, etc.) on model accuracy. Through DSE, the level of abstraction is adjusted such that the impact of the modeling accuracy is maintained. When used to predict unknown channel structures, various channel structures and their impact on channel function can be evaluated in order to deduce unknown channel structures from known structures. The “permeation characteristic” here could be the expected channel conductance. When DSE is
performed in the context of designing drugs or artificial nano-pores, it can be used to investigate
the impact of certain channel characteristics (called design parameters) on its function. Examples
of these design parameters include:

1. The number of ions in the channel
2. The potential across the channel
3. Drug candidate compound types
4. Types of mutations in the channel
5. Strength of particular dipoles/amino acids
...

For example, when used in the context of drug design, a possible DSE experiment would be to
investigate the best combination of size of drug molecule and charge given a set of other channel
parameters such as the membrane potential, and the number of ions. The result of this experiment
would be the particular drug size and charge that would give optimal (in this case the required)
channel blocking time. An example DSE study that involves a drug is given in Chapter 6. This
study has a different objective where it investigates the optimal set of modeling parameters that
would reproduce a particular channel blocking behavior. The value of this DSE study is the optimal
model configuration that is estimated using DSE which could be used in other studies.

An exhaustive search of the ion channel system design space requires simulating channel function
for all possible combination of values (i.e. design points) for these parameters/characteristics in
order to find optimal design points. The computational infeasibility of an exhaustive search of the
ion channel system design space where each design point is evaluated by determining the simulated
behavior and comparing to a target behavior motivates the need to address two issues:

- Reducing the size of the design space (or the number of points in the design space).
- Reducing the time required for the evaluation of each point in the design space while main-
taining evaluation accuracy.
2.4 Issues with Using Current Approaches for DSE Studies of Permeation in Ion Channels

The discussion in this section will focus on MD- and BD-based approaches as they have been widely used due to the benefits that they offer in terms of modeling resolution, accuracy, and cost optimization for a particular application. Other modeling techniques discussed in Chapter 1 are either computationally prohibitive, inflexible, or are not likely to provide a sufficient level of modeling resolution/accuracy.

2.4.1 MD:

2.4.1.1 Computational Feasibility:

MD has a significantly higher computational cost when compared to BD. This is because in MD, all atoms (and their interactions) are modeled explicitly, while in BD, water is replaced by a continuum. The complexity of the ion channel model, the computational cost of each DSE time-iteration of all-atom Molecular Dynamics (MD) [90–97], and the large number of time-iterations of MD needed to evaluate each point in the design space, make DSE using MD computationally prohibitive. While fast implementations of MD have been developed in hardware such as Anton [98, 99], these implementations are inflexible making adding new features time-consuming, costly, and not always possible. MD approaches can be used to achieve higher levels of accuracy once the space has been narrowed down to a few candidates. BD [71, 100–107] is the more suitable approach for the majority of the exploration effort.

The cost of an MD simulation depends on the algorithm used. This cost is typically $O(n^2 \times m)$, where $n$ is the total number of atoms in the system, and $m$ is the number of time steps.

The cost of a BD simulation is also $O(n^2 \times m)$ when using the look up tables method [108] where the PEP information is stored and then fetched during simulations. $n$ in this case is the number
of ions as opposed to the number of atoms in the case of MD (around a hundred or less ions as opposed to thousands of atoms in MD).

Additionally, the time step in MD is typically 50x shorter than the time step in BD. For example, if the ion channel system including membrane and water has 100 ions and 50000 atoms, and assuming the simulation is divided into 1000 10 Åx 10 Åx 10 Å boxes used to compute short range interactions, and that the algorithm used for computing the long range interactions takes roughly the same amount of time as the short range interactions algorithm does, the cost of the MD simulation would be $2 \times 1000 \times 50^2 \times 50$, and the cost of the BD simulation would be $100^2 \times 1$. In this case, the MD algorithm cost in comparison to the BD algorithm is 25000 to 1.

2.4.1.2 Model Flexibility:

MD is much more sensitive to details in channel structure than BD. For example, designing a force field model and finding the correct atomic structure that would reproduce the properties of a particular channel after a simple variation in the atomic coordinates would be a very time-consuming task. As a result, DSE would be very challenging using an MD-based modeling approach. MD-based DSE using today’s computational resources is not feasible for many permeation DSE problems. Aside from the computational cost associated with evaluating a single design point (e.g. simulating channel conductance or blocking behavior), it also would be a complex and time-consuming task to finalize a channel structure that would “work” given a particular parameter variation (see Section 1.6. MD is better suited when a near-final solution is known, and a detailed simulation is required. For the purpose of DSE, when the values of parameters are not known, a faster solution (such as BD) is needed.
2.4.2 BD:

2.4.2.1 Accuracy:

A concern with classic BD is the level of modeling resolution and consequently accuracy. Treating the water and protein as continuous dielectric media is not an accurate representation of the molecular interactions in the ion channel system. This abstraction becomes a significant issue in the case of narrow ion channels (e.g. potassium channels) where in the narrow part of the channel, the selectivity filter, a permeating ion is completely dehydrated. In such scenarios, polarization issues become significant. Further, in classic BD, typically, the dielectric values used for water and protein are assigned using trial and error which is another criticism of classic BD.

2.4.2.2 Model Flexibility:

BD offers great flexibility. Abstraction of water molecules and assuming that the protein atoms are rigid makes a BD model insensitive to details in the structure and can as a result be used to model a wide range of ion channel systems.

2.5 DSE As A Tool For Understanding Channel Function and the Impact of Drugs On It

The DSE framework proposed in this thesis aims at providing a tool that can be utilized to understand channel function and how it is influenced by the presence of drugs (e.g. [2]). The proposed DSE framework uses FGBD which improves upon an existing BD framework developed by Dr. Shin-Ho Chung’s research group at the Australian National University College of Science [2]. Unlike previous models, FGBD does not assume dielectric homogeneity of the water or the protein. Further, the model is capable of importing a channel MD-computed “dielectric map” using the method in [84]. The resolution of the dielectric map can be adjusted (by dividing it into “dielectric regions”) as required considering the level of accuracy needed and the computational resources.
available. Representing the channel system as a heterogeneous dielectric according the MD computations improves the model’s resolution and accuracy as shown in Chapters 5 and 6. Chapters 3 through 7 demonstrate how DSE can be performed using the proposed framework. Chapter 3 discusses the design techniques used to significantly cut the computational cost of FGBD. Chapter 4 discusses the proposed DSE methodology which uses various statistical inference techniques that utilize knowledge of a channel structure to reduce the size of the design space making DSE feasible using FGBD. Chapters 5 and 6 provide two DSE studies that verify and demonstrate the use of the framework. And Chapter 7 discusses the computational feasibility of the proposed framework.

2.6 Relevant Work

The focus of research work in this area has been on improving modeling accuracy and performance. Various groups have worked on improving different aspects of modeling. For example, improving the accuracy and resolution of molecular structures of channels, improving performance of computational models through hardware acceleration, and improving accuracy of modeling approaches by improving modeling resolution of these approaches.

The importance of modeling accuracy and performance is perhaps best demonstrated by the development of the Anton processor by ED Shaw [67–69, 109]. This group is focused on commercializing efficient ion channel simulation hardware in order to allow drug companies to efficiently explore the function/structure design space. The group was heavily funded to build a custom processor with the goal of providing pharmaceutical companies with the ability to explore the impact of new drugs on the structure/function relationship of different channels. Despite the substantial speed improvement, this approach would still not be suitable for DSE. This is because the working structure of the channel would have to be reproduced for each design point. This process is not automated and as a result is not suitable for DSE.

Previous work has focused on the speed/accuracy trade off of a single point in the design space. There is no known work that focuses on reducing the total number of simulations needed to optimize
a set of parameters (i.e the design space). The need for an efficient DSE is evident by the fact that the cost of simulations is high, and the number of simulations needed to find the correct set of parameters for a particular problem is high.

Addressing dielectric effects is still an open question. For example, [64] uses a very simplified Brownian dynamics model to investigate the impact of variations in channel shape as well as molecule size and charge on permeability. This BD model, however, does not consider the atomic detail of the channel or the dielectric variation.

The authors in [70] are also interested in exploring the impact of a channel characteristic/design parameter on the channel behavior. In particular, they conduct an MD study that investigates the impact of the motion of a polar side-chain on permeation.

The authors in [19] highlight the fact that dielectric and steric effects are crucial issues when it comes to ion permeation in narrow and crowded ion channels. The authors propose a three-dimensional BD ion channel simulator which uses the Boundary Element Method to compute the PEP that improves on the work in [110] by using a faster Poisson’s equation solver that can compute the electric field at each step of the simulations as the Langevin simulator runs, and thus avoiding the need to solve Poisson’s equation offline and store the electric field information in tables to be used later in the Langevin simulator. The authors criticize how most BD approaches treated water in the channel as water in bulk by assigning the same dielectric value to both. The authors acknowledged Dr. Shin-Ho Chung’s group’s work where water is inside the channel is assigned a different dielectric value than in bulk (while the dielectric is still homogeneous within the channel). They criticized Dr. Chung’s group approach by showing concerns that interpolation is needed with the tables method and using tables for storing asymmetric channel information might not be feasible. Chapter 7 investigates the computational feasibility and accuracy of the proposed framework and shows that reasonable levels of accuracy are achieved given the resolution of tables and the resulting interpolation. Further, it shows that asymmetric channel information can be stored using six-dimensional tables and that this method is computationally feasible. The authors in [111] address the dielectric issue by simply
assigning a different dielectric to water inside the channel than bulk water. However, the water in
the channel is still treated as a homogeneous dielectric.

The DSE experiment discussed in Chapter 5 is similar in motivation to [112] in which the dielectric
constants of the ion channel are estimated. In [112], the dielectric is assumed to be homogeneous
across the water and protein in the channel. Further, the problem addressed in [112] is to determine
the optimal dielectric constant value of the protein and water such that the simulated current best
matches the experimentally derived current. In [112], the simulated current is treated as a random
variable and thus the authors solve a stochastic optimization problem. The DSE experiments
provided in Chapters 5 and 6 differ in that the channel is represented using a spatially varying
dielectric. The dielectric constant values are optimized such that the BD-derived PEP best matches
the golden reference. In this case, the PEP is not represented as a random variable and thus the
problem is a combinatorial optimization problem. Because the dielectric constant across the channel
is not assumed to be homogeneous and can vary with position, the size and complexity of the design
space are greater thus motivating the need for the proposed statistical design space size reduction
techniques.

2.7 Research Rationale and Contributions

Allowing DSE studies of permeation in ion channels requires accurate models that are computationally feasible. This research addresses this using a two fold strategy: improving on classic BD models to improve modeling resolution and accuracy and proposing a DSE methodology that significantly reduces the size of the design space.

The main research contributions are as follows:

1. **A Molecular Geometry Generator Tool (MGGT):** MGGT constructs molecular surfaces for enclosed structures (e.g. molecules) as well as channels (e.g. ion channels). Results using this tool were published in [113].
2. **Finite-Grained Brownian Dynamics (FGBD) Framework:** FGBD improves on the framework developed by Dr. Shin-Ho Chung’s group at the Australian National University College of Science. The following improvements were made:

(a) Unlike the previous BD model, FGBD uses the Finite Element Method (using COMSOL) to solve Poisson’s Equation and compute the Potential Energy Profiles (PEP), and subsequently the forces acting on an ion.

(b) Hydration/polarization effects are taken into account by representing the channel as heterogeneous dielectric with the ability to adjust the resolution of the dielectric variation (i.e. the size of dielectric regions).

(c) FGBD is capable of importing an MD-computed dielectric map for the channel.

It was shown in the studies conducted that, in a complex scenario where the KcsA channel is occluded by a blocker, the model computes a PEP consistent with MD-computed PEP. This is significant because the model can achieve resolution similar to MD’s at a fraction of the cost of MD. Results were published in Study [113].

3. **Design Space Exploration Methodology:** The proposed DSE methodology utilizes knowledge of channel structure as well as statistical inference techniques (Sensitivity Analysis, Hierarchical Clustering, Association Analysis, and Correlation Analysis) to reduce the size of the design space. Results from DSE experiments conducted in Chapters 5 were published in [113], and results from Chapter 6 will be submitted for publication.

### 2.8 Example Application: The Pharmaceutical Industry

Recent studies reported that 15% of all drugs are targeted to ion channels with a worldwide sale of ion channel drugs being over 12 billion [114], [15]. To date, ion channels are still largely unexplored for drug development. This is due to several reasons such as the high cost and complexity of drug discovery and development processes, and the high failure rates of new investigation drugs
due to safety issues during development or lack of efficacy in clinical trials [15]. The high cost of drug development along with the high failure rate of new investigation drugs motivate the need for a low-cost solution for investigational drug leads prior to lab testing. This allows for the initial identification of drug leads and for the optimization of such leads to achieve early proof-of-concept, which would increase probability of success after further clinical investigation and cut the cost of drug development [15].

A crucial factor to consider when developing ion channels-related drugs is molecular interactions with candidate drugs. Accounting for these interactions can result in more effective drugs, the reduction of drug side-effects, and most importantly providing insight into the function of ion channels and their related diseases. Detailed knowledge of these molecular interactions can be obtained by characterizing the relationship between particular channel characteristics/parameters (e.g. the presence of a drug, channel structure, mutation, ionic concentration, etc.) and channel function. This can be done through DSE studies where drug development can be done through an iterative process of introducing various blocking schemes (i.e. drug candidates) while evaluating the simulated behavior of the channel using each scheme. In this case, the "permeation characteristic" used for the cost function could either be the minimum or maximum blocking time. A hypothetical experiment example is using DSE to deduce the dielectric map that would reproduce a particular channel blocking behavior. The dielectric dampening effect can then be determined by comparing this dielectric map to the dielectric map of the channel when the drug is not present. This information is valuable in deciding on the drug charge strength and size as that determines the strength of the field around the blocker drug and as a result the dielectric strength.

Chapter 6 provides a DSE study that explores the optimal modeling parameters (here, dielectric map) required to create a model for TEA block of the KcsA channel that would reproduce experimental data. TEA is an experimental drug for which the mechanisms of action are still being investigated. It was shown that it is possible to estimate close to optimal values of these modeling parameters (dielectric map) using the proposed DSE methodology.
Chapter 3

Fine Grained Brownian Dynamics

3.1 Overview

FGBD is based on Brownian Dynamics [71, 100–107]. Unlike previous BD-based modeling approaches, FGBD models the channel as a spatially-varying dielectric medium. The framework achieves results consistent with those computed using full atomic MD while maintaining the cost of classic BD models. As a result, this framework is suitable for performing DSE studies of ion channels.

FGBD consists of three components: a channel geometry generator tool that constructs molecular surfaces, a potential energy profile calculator tool (PEP Calculator tool) that is based on a Finite-Element Method (FEM) solver to compute the electric-potential across the channel, and a Langevin Dynamics Simulator that models the permeation process using the electric potential information generated by the PEP Calculator tool. The Langevin simulation results used in this project were produced using the Langevin Simulator developed by Dr. Shin-Ho Chung’s research group at the Australian National University College of Science.

In BD, water molecules inside the channel are modeled implicitly and only ions are modeled explicitly. Poisson’s equation is solved numerically at different time steps to compute the electric-potential
across the channel. The gradient of this potential, the electric field, is subsequently used to compute the ions’ trajectories using Langevin’s stochastic differential equation.

Figure 3.1 summarizes the key flow of information in the simulation framework:

1. The topological information for the ion channel is first read. The dielectric map for the channel can be then computed in MD using the methodology described in [84] or deduced using DSE. The atomic information is also used to generate the channel molecular surface.

2. Poisson’s partial differential equation is then solved numerically using a Finite-Element Method (FEM) at different time steps to compute the electric potential across the channel.

3. The gradient of this potential, the electric field, can then be used to compute ion trajectories using Langevin’s stochastic differential equation.

The next few sections describe the details of each of these steps.
3.2 Molecular Geometry Generation Tool

3.2.1 Generating the Channel Boundary

The Geometry Generation Algorithm generates a mesh for the molecular surface which is defined by the boundary between the water and protein atoms. This boundary is a consequence of a number of forces acting between the protein atoms, the water molecules in the channel, and the ions that flow through the channel. Practically speaking, it acts as a hard boundary for the ions and is thus modeled as such. Thus, the boundary must be constructed as a continuous and smooth hard-boundary dictated by the locations of the protein atoms. The Geometry Generator constructs a channel wall that defines this boundary in three dimensions. Once the geometry is generated, two large cylindrical reservoirs are then attached to the ion channel. These reservoirs mimic the extracellular and intracellular space.
Figure 3.2 shows a top level diagram of the Geometry Generator. The first step is reading topological information for the channel of interest from the Protein Databank [35]. The channel structure has to be in a “working state” to allow for simulating ion permeation, so the topology file has to be processed using a molecular dynamics simulation package. For example, this can be done using the well known MD program CHARMM which can generate a residue topology file (RTF) and a coordinate file (CRD) [115]. CHARMM is a popular molecular dynamics simulation and analysis package. CHARMM uses classical (empirical and semi-empirical) and quantum mechanical (ab initio) energy functions to perform MD simulations, and analyzes the simulation results to determine structural, equilibrium, and dynamic properties. Once the topological information of a working structure of the channel has been read, a molecular boundary is then defined based on the atoms’ locations. This boundary can be either a representation that does not make any approximations about the shape of the channel as in the asymmetric model or an approximation where it is assumed that the channel’s geometry is made of rings stacked on top of each other as in the symmetric model. The asymmetric model uses a modified version of the Marching Cubes Algorithm to construct a molecule or channel geometry. The symmetric Model, on the other hand, uses a moving average algorithm and can only be used for channels. Both schemes will be discussed in detail in the next two sections.

**Important Note**  The "Working Structure" of the KcsA Channel used in all studies performed in this research work was prepared by Dr. Shin-Ho Chung’s group at the Australian National University College of Science.

### 3.2.2 Symmetric Model

The basic steps of the moving average algorithms are summarized below:
Figure 3.3: Symmetric Channel Geometry Generation Algorithm

1. Read channel's topological information
2. Define a window size \( n \)
3. For every value along the z axis of the channel \( Z_i \):
   - Compute the average radius size in this window \( \bar{r}_{av} = \frac{r_{Zi} - r_{Zi-n}}{n} \)
   - Assign the radius at this z value to the window average radius: \( r_{Zi} = \bar{r}_{av} \)
   - if \( r_{Zi} < \text{IonicRadius} \) then \( r_{Zi} = \text{IonicRadius} + \Delta \) where \( \text{IonicRadius} \) is the radius of the largest ion in the system
4. Construct the two dimensional (2D) boundary according to the computed radii
5. Construct the reservoirs that represent the extracellular and intracellular spaces according to a set of parametric equations
6. Append the reservoirs to the protein boundary
7. Revolve the 2D boundary with a resolution \( \theta \) to generate the 3D structure of the system

This algorithm is illustrated in Figures 3.3. Figure 3.4 shows an illustration of the resulting 3-dimensional symmetric channel geometry.

The number of operations required per sample for generating the protein boundary using this algorithm is reduced to one addition, one subtraction and one division. Because the formulation is independent of the window size \( n \), the runtime complexity is constant (i.e. \( O(1) \)).

These steps are similar to those found in [116]; however, the channel geometry is targeted for use in FEM analysis rather than an iterative solver. There are two major distinctions between the
approach taken in [116] and this approach. Rather than tiling the boundary in order to iteratively assign surface charges, the boundary is simply defined by a set of rings. Both approaches retain the same level of channel resolution and both assume cylindrical symmetry. A second distinction is that the approach taken here allows for the channel pores to be divided into several geometric regions each with their own dielectric strength.

3.2.3 Asymmetric Model

The tool defines this boundary using the electric field strength of the protein atoms as a guide. This is explained later in this section. The tool is designed such that it is not only specific to ion channels. We divide molecular geometries into two categories:

- Surface enclosures: Given atom locations, generate the surface surrounding these atoms.
- Channel structures: Given atom locations, generate the channel formed by these atoms.

In both cases, the first step to generate the molecular geometry is reading atomic coordinates. This information is then used as input to a modified Marching Cube (MC) algorithm [117] which
constructs the molecular structure. The goal is to generate a smoothed surface that "molds" the shape enclosed by the atoms. To achieve this, the MC algorithm is modified to include a threshold test function based on electric field equations for these molecules with the constraint that the generated surface must be fully connected; in this way, the surface will emulate the shape of a contour of constant electric field. As the threshold test value is varied, the distance between the surface and the atoms will vary accordingly. The Modified MC algorithm can be used without any other constraints when generating surface enclosures.

Unlike surface enclosures, when generating molecular channels, the MC algorithm has to adhere to other physical constraints. First, the channel should take on a cylindrical shape with only two openings at either end of the z-axis. Second, the surface of the boundary must be continuous and smooth. Third, the boundary must be closer to the z-axis than all atoms for a given value of the z-axis. We need to guarantee that the channel is cylindrical with openings at the top and bottom along the z-axis. At low threshold test values, the channel would then appear as a perfect cylinder around the z-axis. As the threshold is raised, the radius of the cylinder surface generated by the MC algorithm would continue to grow. Because the atoms are also part of the threshold test equation, their contribution towards the shape of the channel would increase. As the cylinder grows, it will "mold" against the molecules/atoms that form the channel walls thus taking on a smoothed shape of the channel boundary.

If the threshold becomes too large, the surface has the potential to grow beyond those atoms closest to the z-axis thus fragmenting the channel boundary (see Figure 3.5). This is a violation of the physical constraint that the channel must remain continuous and smooth. To fix this problem, we apply an additional "legalization" condition for each atom such that there is a line of constant voltage extending radially from the location of the atom to infinity. This has the effect of preventing fragmentation of the surface (see Figure 3.5). It is important to note that this additional condition helps in the construction of a more realistic geometry for modeling and should not be considered a physically-accurate boundary condition.
Figure 3.5: A conceptual diagram illustrating the fragmentation issue when generating the channel surface. Channel boundary gets fragmented as the surface grows beyond those atoms closest to the z-axis at a high MCA threshold. Fragmentation can be avoided using radial boundary conditions that stops the surface from growing beyond the atoms closest to the z-axis.

The basic steps of the algorithm are the following:

1. Create a grid of cubes large enough to encapsulate the molecular system.

2. For each cube in the grid:
   - Compute the value of the threshold test function at the eight corner points of the cube (i.e. compute the electric field at all eight corner points due to all atoms).
   - Determine if the value of the test function is consistently greater or less than the threshold test value for all corner points. If not, then the boundary must pass through the cube.
   - If the boundary passes through the cube, replace the cube with an appropriate set of polygons based on an 8-bit pattern where each bit corresponds to a corner point and indicates whether or not that point is above or below the threshold value.

3. Combine all polygons generated in order to form a surface defined by the threshold function.

Figure 3.6 is an example channel geometry of the KcsA channel generated using the proposed
molecular geometry generation tool. In this figure, the dots represent the atoms that make the channel walls and the green structure represents the generated channel boundary.

3.2.4 Important Note: DSE Using the Asymmetric Channel Structure

It was not possible to use the asymmetric structure for the DSE studies performed in this project. This is because the Comsol-based Finite Element Method mesh generator created meshes of excessively fine granularity in order to model the asymmetric channel structure (which was generated by the asymmetric geometry generator tool discussed in section 3.2.3). The required increase in mesh granularity was significantly greater than that of the symmetric channel structure. This problem was further exacerbated by the addition of multiple dielectric domains. This resulted in significant increase in run-time and memory-usage with the tool often hanging. Attempts were made to relax the granularity of the FEM meshing tool but this resulted in an overall degradation of modeling accuracy thus diminishing any benefits of the asymmetric model over the symmetric model.
3.3 Heterogeneous Dielectric PEP Calculator

3.3.1 FEM-Based Solver

The PEP Calculator (PEPC) is a FEM-Based Heterogeneous Poisson’s PDE solver that computes the electric potential and resulting electric field across the channel to be used by the simulator kernel described in Section 3.4. The Comsol-based [118] solver is capable of importing a dielectric "map" of the channel and approximating the map into a set of dielectric regions. It then computes the electric potential acting on ions in a heterogeneous environment where the dielectric constant varies with position using:

\[ \nabla.(\epsilon(x)\nabla \varphi(x)) = -\frac{\rho(x)}{\epsilon_o} \tag{3.1} \]

where:

- \( \epsilon(x) \) is the relative dielectric constant and varies with location \( x \in R^3 \)
- \( \varphi(x) \) is electric potential
- \( \rho \) is the charge density
- \( \epsilon_o = 8.85 \times 10^{12}[F/m] \) is the dielectric constant of vacuum

Solving Poisson’s equation gives the electric potential throughout space, and the gradient of this potential is the electric field: \( E(X) = -\nabla \varphi(x) \).

In practice, ion positions will rarely reside exactly on a grid point defined by the FEM solver and thus interpolation must be used for every step of the simulation algorithm and for every ion in the system using electric potential information computed by the FEM solver. A multi-linear interpolation that has an implementation that scales well for higher-dimension problems is used. The generalized expression for a multi-linear interpolation in \( N \) dimensional space reads,
\[ E_{N,est}(x) = \sum_{\mathbf{q} = (q_1, q_2, \ldots, q_N): q_j \in \{0, 1\}} w^{(k)}_s(x) E_N(k + \mathbf{q}) \]  
\text{(3.2)}

where \( x = (x_1, x_2, \ldots, x_N) \) specifies the \( N \)-dimensional coordinates of the point at which the electric field is being computed. \( k = (k_1, k_2, \ldots, k_N) \) is an \( N \)-dimensional list denoting the indices of the neighboring grid points to \( x \). Each term in the above summation is the weighted electric field, \( E_N(k + \mathbf{q}) \), at one of the \( 2N \) neighboring grid points indexed by \( k + \mathbf{q} \) where \( q_j \in 0, 1 \) for \( j \in 0, 1, \ldots, N \). The weight function is defined as,

\[ w^{(k)}_s(x) = \prod_{j = 1}^{n} ((1 - q_i) + (2q_i - 1)(x_j - k_j)) \]  
\text{(3.3)}

This expression can be derived by induction, see [119].

### 3.3.2 Computational Rationale Behind the Use of FEM

There are several numerical methods for approximating solutions to Poisson’s equation (equation (3.1)) with the most commonly-used being the Finite-Element Method (FEM), the Boundary-Element Method (BEM), and the Finite-Difference Method (FDM).

BEM finds solutions by using boundary conditions to substitute boundary values into the corresponding integral equation; this integral equation can then be used to numerically compute the solution for any desired location within the interior. Thus, a BEM problem is represented by a 2-D mesh over the modeled surface. FEM differs from BEM in that it approximates the governing differential equations by locally approximating the problem over many small sub-domains and then systematically combining these to form a global solution. Local approximations typically include an iterative process of fitting trial functions in a differential equation while minimizing error. The FEM problem is represented by a 3-D mesh of elements each representing a sub-domain.

BEM is often considered to be computationally more efficient than FEM because the FEM problem is defined by elements spanning the volume rather than the surface thus implying that the size of a
problem for FEM is the square of that for BEM. The computational complexity of BEM becomes an issue for the ion channel geometry though as the size of the problem tends to grow quadratically because all grid element-to-element relationships must be considered to generate a solution. In the context of the ion channel model, the formulation includes asymmetric fully populated matrices and thus the computational time and the space requirements grow quadratically with the problem size. FEM, on the other hand, is banded (elements are only locally connected) and thus the solution is generated based on local connects and the computation time and space requirements grow linearly with the problem size.

In addition to being computationally more efficient for domains with large surface area to volume ratios, FEM is better at capturing local effects. This includes heterogeneous domains with dissimilar properties throughout the domain or problems in which there are varying desired granularity throughout the domain. In addition, FEM is also more effective when solving problems in which the model structure lacks smoothness.

Another alternative to FEM is the Finite-Difference-Method (FDM). FDM works by generating a grid of points and then approximating the derivative at each point based on the function values at neighboring points. This works well if the grid is regular. If, on the other hand, the grid can only be irregular to properly capture the shape of the object being modeled, then it becomes difficult to implement FDM while maintaining high-order accuracy. The grid specification for FEM, on the other hand, can be easily constructed from irregular shapes and can be used to model any geometric shape as accurately as required. In fact, the density of the mesh can be varied such that small grid elements are used where the electric potential changes rapidly and bigger elements elsewhere. This is what makes FEM well-suited for complex structures such as ion channels.

3.3.3 Decoupling of the FEM Solver from Simulation

The FEM solver is computationally intensive (see table 3.1) and consequently best decoupled from the “inner-loop” of the DSE algorithm. The solver is spatially-based and does not model time explicitly; however, it must be executed for each time-iteration of the Langevin dynamics simulator.
used for computing ions’ trajectories. As a result, computational cost can be significantly reduced by decoupling the FEM solver from the simulator. This requires that all possible solutions for the FEM solver corresponding to all possible ion positions in the channel to be stored in memory to be subsequently accessed by the simulator. This approach was previously used in [108].

To compute the memory required to store the electric-potential across the channel, we consider the size of permeation in an ion channel problem: if there are \( n \) ions and the channel is divided into \( m \) grid points at which ions can reside (with all other possible ion locations being interpolated) then the total number of entries required to store the electric potential across the channel is \( \binom{m}{n} = \frac{m(m-1)\cdots(m-n+1)}{n(n-1)\cdots1} \). Thus, for a system with 64 ions and 200 grid points, the number of entries \( \binom{200}{64} \approx 1.7 \times 10^{53} \). This number is prohibitive both in the computational effort required to determine all entries and the memory needed to store them. To solve this problem, the superposition principle is used such as in [108] thus making electric-potential at any point in the channel the sum of potentials due to each source - in this case, each contribution due to a source is an individual entry and thus the number of entries is far smaller.

Using the superposition principle, the electric potential can be computed by adding together the contributions from each ion. The electric potential at the position of an ion is the sum of the electric potential due to the other ions, the applied field, the surface charges induced by these, and the surface charges induced by that particular ion. This means that we only have to compute the electric potential at every position once and store that information in a table.

The potential \( V_i \) experienced by an ion \( i \) in the system can be expressed as:

\[
V_i = V_{S,i} + V_{X,i} + \sum_{j \neq i} V_{I,ij} + \sum_{j \neq i} V_{C,Ij}
\]  

(3.4)

where \( V_{S,i} \) represents the self-potential due to the surface charge induced by ion \( i \) on the channel boundary, \( V_{X,i} \) represents the external potential due to the applied field, the fixed charges in the protein wall, and the charges induced by these fixed charges. The third and fourth terms in equation (3.4) represent the potential caused by ion \( j \) acting on ion \( i \). The third term is the image potential.
due to the surface charges induced by ion \( j \) and the fourth term is the Coulomb potential due to ion \( j \).

The Electric field \( E_i \) experienced by an ion \( i \) in the system can be expressed similarly as:

\[
E_i = E_{S,i} + E_{X,i} + \sum_{j \neq i} E_{I,ij} + \sum_{j \neq i} E_{C,Ij}
\]  

(3.5)

To compute the potentials in Equation 3.4, we solve Poisson’s equation in two different scenarios:

1. Computing the self potential and external potential: The first two terms in Equation 3.4 are due to the effect of external charges and fixed charges upon channel. To compute these terms, the solver is run with no ions in the system and with the fixed charges present. From a single run (i.e. solving Poisson’s equation once), the potential is extracted for all grid points. The results from this run represent the values of the electrical potential at each possible ion location in the channel due to the external applied field, and fixed charges in the protein wall. Electric potential contributions are direct, and indirect through induced charges in the channel wall.

2. Computing the image and Coulomb potentials: The last two terms in Equation 3.4 are due to the effect of ions upon themselves and others. To compute the last two terms in Equation 3.4, the solver is run with no fixed charges or externally applied potential and with one ion. Poisson’s equation is solved for all possible grid locations of the ion. For each location (position \( i \)), results for electric potential are determined for all grid points in the channel. For given grid point position \( j \), the electrical potential is that which is produced by the ion at ion \( j \) both directly through coulomb potential and indirectly by the induced charge on the channel wall (image potential). This represents the impact on one ion upon another. In the case when the position \( i \) is the same location as the grid point \( j \) (i.e. \( i = j \)), the electric potential represents the self potential of the ion onto itself through an induced charge on the boundary wall.
### Computing resources used for case study

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.06 GHz Xeon quad-core E5450</td>
<td></td>
</tr>
<tr>
<td>Memory footprint (FEM solver)</td>
<td>4 GB</td>
</tr>
<tr>
<td>Run time (FEM solver)</td>
<td>14 hours</td>
</tr>
<tr>
<td>Memory footprint (BD solver)</td>
<td>293 MB with 64 ions</td>
</tr>
<tr>
<td>Run time (Simulator)</td>
<td>18 seconds with 64 ions (100,000 iterations)</td>
</tr>
</tbody>
</table>

**Table 3.1:** Typical resource usage for the KcsA biological ion channel case study.

The results from Scenario 1 solution are saved in a 3-D table since the information needed only depends on the position of the ion. The results from Scenario 2 solutions depend on the position of both ion $i$ and ion $j$ and as a result are saved in a 6-D table (or a 5-D table in the case of symmetric model). The electric potential results determined in Scenario 1 and Scenario 2 combine to form the electric potential for on an ion $i$ at position $j$ using the superposition principle.

#### 3.3.4 Computational Cost of FEM and BEM Ion Channel Formulations

The theoretical comparison between the computational cost of a FEM derivation and a BEM derivation of the ion channel geometry can be made for a homogenous domain with a few assumptions. Assuming that the separation between surface grid elements for a BEM formulation is $\Delta s$, the number of elements in x-dimension is $K = (X_{max} - X_{min})/\Delta s$ where $X_{max}$ is the maximum coordinate of the domain being modeled and $X_{min}$ is the minimum coordinate. The size of grid element then is $\Delta s^2$ and the BEM problem size is $N_{BEM} = K^2$ for a homogenous domain. All element-to-element relationships must be considered when using a BEM formulation and thus requires storage capacity in the order of, $O(n) = O(N_{BEM}^2) = K^4$. In the case of a 3-D mesh FEM formulation, the problem size is $N_{FEM} = K^3$. In contrast to BEM, the matrices for FEM formulations tend to be sparse symmetric matrices because elements are locally connected only. As a consequence, matrices tend to grow linearly with the problem size and the storage capacity is in the order of, $O(n) = O(N_{FEM}) = K^3$. For heterogeneous systems, BEM formulations scale poorly as a complex set of surfaces must be constructed to divide the domain into the many sub-domains required to vary
<table>
<thead>
<tr>
<th>Description</th>
<th>BEM</th>
<th>FEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogenous Problem Size (N)</td>
<td>$K^2$</td>
<td>$K^3$</td>
</tr>
<tr>
<td>Homogenous Matrix Size (S)</td>
<td>$K^4$</td>
<td>$K^3$</td>
</tr>
<tr>
<td>Heterogeneous Problem Size (N)</td>
<td>$K^4$</td>
<td>$K^3$</td>
</tr>
<tr>
<td>Heterogeneous Matrix Size (S)</td>
<td>$K^6$</td>
<td>$K^3$</td>
</tr>
</tbody>
</table>

Table 3.2: Computational cost of BEM vs. FEM for both homogeneous and heterogeneous domains.

one or more properties of the problem. Assuming a case in which the granularity of heterogeneity is the same as the grid itself, the problem size for the BEM formulation becomes $N_{BEM} = K^3$. As a result, the storage requirements are in the order of $O(n) = O(N_{BEM}) = K^6$ which is significantly greater than that of its homogenous counterpart. Because FEM formulations are already based on a 3-D grid representation and interactions are local, heterogeneous systems scale at the same rate as homogenous systems. Thus, the problem size and storage requirements for a FEM formulation remains unchanged (See Table 3.2).

3.3.5 MD-Computed Heterogeneous Dielectric Map

The framework has been designed such that there are no restrictions on the homogeneity of the dielectric in the channel model. Further, to improve modeling resolution and consequently accuracy, the dielectric across the channel can be computed directly using MD (i.e. channel dielectric map) using the method in [84]. Introducing dielectric heterogeneity to the ion channel model significantly impacts the computational complexity of the problem. To overcome such computational burden while maximizing accuracy, the channel can be broken into different "regions" that are an approximation of the dielectric map where the dielectric is homogeneous within a region. The regions reflect the significant variation in the dielectric value in the map.

3.3.6 Significance of the Potential Energy Profile

The PEP represents the potential energy in one dimension (along the z-axis). It is a useful way for visualizing the barriers and wells that an ion experiences as it moves along the z-axis. The width and
depth of the well and height of barrier in the PEP have a strong implication on the channel behavior such as whether the channel will conduct or will be blocked, how long the blocking period is, or what the conduction rate is. As a result, the PEP provides valuable insights into channel function. For example, once it has been established at what depth/width of well (and/or width/height of barrier) the channel will conduct or will be blocked, one can then try different conditions (i.e. different design parameters such as ionic concentration or type of blocker etc.) for which we generate new PEPs and compare against target PEPs which produce the desired channel behavior. The PEP can then be examined. If the well is too deep, conductance will be reduced. In the case of a blocker, if the well is not deep enough, the channel won’t be blocked. Such information is very relevant in DSE experiments. In a drug evaluation or drug design DSE experiment, this would give insight into information such as how strong the blocker charge should be.

Langevin simulations used to compute conductance (see Section 3.4) are done in three dimensions and as a result, the potential (which is the potential energy divided by the charge) at each grid point in the channel is used. It is important to note that the PEP resulting from solving Poisson’s equation can directly be used to compute the electric field to be used in Langevin simulations. However when the PEP is computed using MD, it in fact is one dimensional free energy and as a result must be "patched" into the Langevin Simulation algorithm. This involves subtracting off the part of the PEP that is due to short range forces from the macroscopic PEP (i.e. the PEP computed by solving Poisson’s equation) then adding the MD PEP to the macroscopic PEP.

### 3.4 Langevin Simulations

The framework is designed to work with a Langevin Dynamics stochastic simulation kernel. An example is the simulation kernel described in [3]. The conduction computations referred to in this project were done by colleagues from Dr. Shin-Ho Chung’s research group at the Australian National University who used the simulation kernel described in [3] to produce these computations. In this kernel, the channel is modeled as a large scale multi-particle stochastic dynamical system in which the ions interact with the atoms that make the ion channel wall. The medium in which ions
move (i.e. water) is modeled as a dielectric. Further, the protein atoms that make up the boundary wall are assumed to be rigid and thus are modeled as a dielectric boundary as well. Only ions are treated explicitly and the collisions between ions and water molecules are treated as frictional and random forces. This stochastic dynamic system evolves every time step according to Langevin’s equation of motion. For an ion at position $i$ with charge $q_i$ and mass $m_i$:

$$m_i \frac{dv_i}{dt} = -m_i \gamma_i v_i + F_R(t) + q_i E_i$$  \hspace{1cm} (3.6)

The first two terms on the right-hand side of equation 3.6 describe the effects of collisions of ions with the surrounding water molecules: the first term represents an average frictional force and the second term $F_R(t)$ represents the random part of the collisions and rapidly fluctuates around a zero mean. $m_i \gamma_i$ is the frictional coefficient where $\frac{1}{\gamma_i}$ is the relaxation constant of the system. The frictional and random forces in the equation are connected through the fluctuation-dissipation theorem [120] which relates the frictional coefficient to the autocorrelation function of the random force,

$$m_i \gamma_i = \frac{1}{KT} \int_{-\infty}^{+\infty} \langle F_R(0)F_R(t) \rangle \, dt \hspace{1cm} (3.7)$$

where $K$ is the Boltzmann constant and $T$ is the temperature in degrees Kelvin. The last term in equation 3.6 represents the systematic force acting on an ion at location $X_i = (x_i, y_i, z_i)$ and is computed by numerically solving Poisson’s equation in three dimensions for each Cartesian component ($x, y, z$) of the velocity. The electric field $E_i$ in this term denotes the three-dimensional total electric field at the position of the ion due to the following interactions:

- **Inter-ion Long-Range Interaction:** This is the electric field acting on ion $i$ at location $X_i$ due to ion $j$ at location $X_j$. This includes the direct Coulomb interaction of each ion $j$ on ion $i$, ...
the indirect interaction of ion \( j \) on ion \( i \) by induced surface charges on the dielectric boundary interface, and the interaction of ion \( i \) on itself by induced surface charges on the same interface.

- **Inter-ion Short-Range Interaction**: Short-range interactions are a repulsive force caused by the overlap the electron clouds when two ions are within close proximity.

- **Ion-boundary Short Range Interaction**: Similar to the inter-ion short range interaction, this is the short range interaction between the ion and the boundary when an ion \( i \) is within close proximity of the boundary wall at position \((x_i, y_i, z_i)\).

- **Fixed Charges in the Boundary Wall**: The charges in the boundary wall induce attractive and repulsive forces to different ion species.

- **External Field**: This is the uniform external field that is applied across the channel.

Equation 3.6 is implemented in computer simulations by first showing that, by integrating (3.6), we obtain the following expressions for position and velocity at time \( t_n \) [121]:

\[
x(t_{n+1}) = x(t_n)(1 + e^{-\tau}) - x(t_{n-1})e^{-\tau} + \frac{F(t_n)}{m\gamma^2}\tau(1 - e^{-\tau}) + \frac{\dot{F}(t_n)}{m\gamma^3} \left( \frac{\tau^2}{2} (1 + e^{-\tau}) - \tau(1 - e^{-\tau}) \right) + X_n(\Delta t) - X_n(-\Delta t)e^{-\tau} \tag{3.8}
\]

\[
v(t_n) = \frac{2\gamma}{\sinh \tau} \left[ x(t_{n+1}) - x(t_{n-1}) + \frac{2}{m\gamma^2} \left( \frac{F(t_n)}{m\gamma^2} - \frac{\dot{F}(t_n)}{m\gamma^3} \times (\sinh \tau - \tau) - X_n(\Delta t) + X_n(-\Delta t) \right) \right] \tag{3.9}
\]
where \( \Delta t \) is the time step, \( \tau = \gamma \Delta t \), \( X_n(\Delta t) \) is a random variable and is explained in detail in [121]. Here, Equations (3.8) and (3.9) are used to compute the trajectories of the interacting ions using the following steps:

1. Compute the force \( F(t(n)) \) in (3.8) and (3.9) acting on each ion at time \( t(n) \) using the electric field information: \( F(t) = q_i E \).

2. Compute the derivative of the systematic force using: \( \dot{F}(t_n) = [F(t_n) - F(t_{n-1})]/\Delta t \).

3. Compute the stochastic force \( F_R(t_n) \) in (3.6) using a sampled value from a Gaussian distribution with zero mean and width described in [121].

4. Compute the position of each ion at time \( t_n + \Delta t \) from (3.8).

5. Compute the velocity of each ion at time \( t_n + \Delta t \) from (3.9).

6. Repeat the above steps for the required simulation period.

The main advantage of this time discretization technique is that the integration time step \( \Delta t \) is not limited by the condition \( \Delta t \ll (\gamma^{-1}) \) where \( \gamma^{-1} \) is the relaxation time constant. Putting such a constraint can restrict the applicability of the time discretization algorithm. For example, the relaxation time constant for sodium ions is 10 femtoseconds and as a result the simulation time step \( \Delta t \) would have to be much less than 10 femtoseconds. The constraints that the algorithm puts on \( \Delta t \) are the following:

- The average distance an ion traverses in each time step must be small compared to the dimensions of the system
- The time derivative of the systematic force should be small relative to the absolute magnitude of the force
Chapter 4

Design Space Exploration Methodology

4.1 Types of DSE in the Context of Ion Channel Modeling

The goal of DSE is to find a set of parameters that would minimize the cost function. Figure 4.1 shows the types of DSE that can be performed using the proposed framework:

1. DSE using the PEP as the golden reference: The evaluation of each design point here only requires computing the PEP. Using the DSE engine shown in Figure 4.2 which uses various design space reduction techniques, we iteratively try different design parameters values and compute the resulting PEP using the PEP calculator (which uses the Comsol-based Poisson’s solver) until the difference between the simulated and target value is minimum within the design space.

2. DSE using channel conductance as the golden reference: The evaluation of each design point here requires computing the PEP as well as the channel conductance for the set of parameters (particular design point). Using the DSE engine shown in Figure 4.2, we iteratively try different design parameters values and compute the resulting PEP and channel conductance until the difference between the simulated and target value is minimum within the design space.
DSE of the first type reduces the cost of DSE substantially by saving on the time spent when visiting each point in the design space. This is because DSE is only based on solutions of Poisson’s Equation, and ion dynamics are no longer simulated thus cutting $10^{10}$ simulation iterations for each design point evaluation. This type of DSE can be conducted when a target PEP that is computed using MD is available or a PEP that leads to experimentally-determined conductance is available.

In this thesis, DSE using the first type (i.e. using PEP as a golden reference) is used for both DSE studies provided. This is sensible as in the first DSE study, the MD-computed PEP was available, and in the second DSE study, both the MD-computed PEP as well as the PEP that reproduces experimental data were available.
4.2 The DSE Problem Formalization

In an ion channel DSE problem, a design parameter is a discrete variable representing a physical attribute of the ion channel that takes its value from an ordered, finite set. A design space is a multi-dimensional space such that each dimension represents a design parameter and each point in the space corresponds to a design configuration. A cost function over the design space is a function that maps each design configuration to a unique real value. The cost function is the mean squared error between the target channel behavior computed by the proposed framework and a target value.

A DSE problem over a design space \( S(x) \) spanned by parameter set \( X = \{x_1, x_2, ..., x_N\} \), is the process of defining the cost function \( f(X) \) at the minimum number of design points (design configurations) so that it is possible to estimate the cost function at any arbitrary point in the design space \( S(X) \). In other words, can we find \( X_{\text{opt}} \) such that \( f(X) \) is optimized while having to know \( f(X) \) for as few points as possible in \( S(x) \). This is important for DSE of ion channels, where the channel is represented by a number of grid points, and the cost of knowing each point in the design space requires one computation of the PEP at each of the channel grid points using the Poisson Solver.

For the ion channel DSE problem, the number of configurations (i.e. the size) of a design space with \( N = |X| \) dimensions is,

\[
R = \prod_{i=1}^{N} R_i
\]  

(4.1)

where \( R_i \) represents the number of possible values for design parameter \( i \) and \( R \) represents the total number of design configurations that need to be computed in order to find the optimal configuration using a brute-force approach.

In the DSE studies provided in this project, the cost function used is, \( f(X) \), where \( f(X) \) is the Mean Squared Error (MSE) measure between the PEP computed using the proposed framework and a “golden reference” PEP. The cost function is defined as,
\[ f(X) = \frac{1}{m} \sum_{t=0}^{m-1} (\alpha_t - \beta_t)^2 \] (4.2)

where \( \alpha = \{\beta_1, \beta_2, ..., \beta_m\} \) represents the PEP over \( m \) z-coordinates computed using the framework and \( \alpha = \{\alpha_1, \alpha_2, ..., \alpha_m\} \) is the golden reference PEP. The use of MSE as a cost function here is sensible. This is because MSE gives more weight to larger variations. This is important in this case where we are estimating the ion channel behavior corresponding to particular design parameters and large variations in channel behavior highlight possible “unintuitive” channel design parameters that cause unexpected channel behavior.

The sheer size of the design space makes an exhaustive search to find the point that minimizes \( f(X) \) computationally prohibitive. To address this issue, we use statistical inference techniques to reduce the size of design space and as a result the time required for DSE. Figure 4.2 shows the various stages of the two main components of the DSE methodology used where the size of the design space can be reduced using the design space reduction techniques introduced in the next section.

Figure 4.2 shows the different techniques involved in the DSE methodology (which are discussed in details in the following section), once the design space has been reduced, we can perform DSE where the design parameters values can be set and the channel characteristic of interest can be computed and compared to the target value. If the difference is minimum within the design space, then the design parameters values used would be optimal. Otherwise, a new DSE is required were new values of the design parameters are set and evaluated and so forth.

4.3 Design Space Reduction Methodology

4.3.1 Overview

The following subsections describe how the design space can be simplified using various statistical inference techniques to improve the run-time complexity of DSE. These techniques along with domain knowledge can be used to significantly reduce the size of the design space that needs to be
evaluated to find the optimal point. The proposed design space reduction techniques are generic and can be used to conduct DSE for any ion channel. Only the number and type of design parameters/variables will differ depending on the type of channel being modeled and type of DSE being conducted.

It is important to note that regardless of whether PEP or conductance is used as a golden reference, the size of the resulting design space after design reduction using the techniques discussed in this section remains unchanged. This is because variation in any of the design parameters/variables (in the studies provided, the dielectric values corresponding to the dielectric regions in the channel represent the design parameters) that impacts channel conductance would also impact the PEP. Further, the effectiveness of these techniques is dependent on the particular ion channel system and scenario being studied, and the amount of design space size reduction will vary from case to case.
4.3.2 Sensitivity Analysis

A sensitivity analysis determines the extent to which the cost function, $f(X)$, at each z-coordinate responds to each of the design parameters, $x_i$. The design parameters in the DSE studies discussed in this project are the dielectric constants corresponding to the different dielectric regions. In doing so, we can gain knowledge about the “shape” of the design space. By understanding the design space “shape”, we can apply various minimization techniques to reduce the complexity of a DSE such as curve-fitting and the removal of redundant or irrelevant dimensions.

Each of the design parameters can be explored along with the range of values explored and the average cost function achieved for each range of values. The parameters that the cost function is most sensitive to can then be identified. Also, regions for which the largest variation in the cost function as a function of a design parameter occur can also then be identified. A sensitivity analysis plot of the corresponding cost function for each design parameter should then be examined. The goal is to divide the optimization problem such that sub-problems can be optimized independently and with different possibilities for design-space reduction. This requires subdividing the cost function into different sub-functions that can be optimized independently.

4.3.3 Hierarchical Clustering

Clustering groups similar dimensions together into “clusters”. This clustering of data elements based on a measure of similarity examines correlations between design parameters and exposes redundancy among them. Redundant design parameters can then be pruned to control the size of the design space. The measure of Similarity is $1 - r$, where $r$ is the Spearman correlation coefficient which determines the strength and direction of the monotonic relationship between two design parameters. The Spearman correlation coefficient can range from $+1$ to $-1$, where $1$ is maximum positive correlation, $0$ is no correlation, and $-1$ is maximum negative correlation. For example, a value that is greater than $0.5$ indicates a strong positive relationship between the clusters. Similarly, a value that is less than $-0.5$ indicates a strong negative relationship between the clusters.
The Spearman correlation coefficient $r$ between the rank design parameters is computed using:

$$r = \frac{\sum_{r=1}^{R} (A_{i,r} - \bar{A}_i)(A_{j,r} - \bar{A}_j)}{\sqrt{\sum_{r=1}^{R} (A_{i,r} - \bar{A}_i)^2} \sqrt{\sum_{r=1}^{R} (A_{j,r} - \bar{A}_j)^2}}$$

(4.3)

where $A_{i,r}$ is the rank of the $r$th value of the $i$th design parameter and $A_{j,r}$ is the rank of the $r$th value of the $j$th design parameter (in the DSE studies provided, these are the one of the possible dielectric values corresponding to each of the dielectric regions).

The clustering algorithm is first initialized by assigning each design parameter to its own cluster followed by a series of iterations in which the most similar pairs of clusters are grouped into a single cluster. The similarity between any two clusters is the maximum similarity between any two design parameters in each cluster. Hierarchical clustering of this type identifies potential redundancy in the data set. Design parameters of the design space that are highly correlated and are classified in the same cluster can be represented by a single design parameter.

Agglomerative Hierarchical Clustering is used to form hierarchical clusters of the design parameters. The algorithm steps are as follows:

1. Consider all data data points as individual clusters

2. Construct the distance matrix $D$, where the number in the $i$-th row $j$-th column is the distance between the $i$-th and $j$-th clusters.

3. Find the most similar clusters

4. Merge similar clusters together to form a single cluster: Merge clusters $a$ and $b$, for which $D(a,b)$ is the minimum

5. Go to the next level

6. Compute the distance between the new cluster and old cluster, and update the distance matrix $D$
7. If the data points are in one cluster then stop, otherwise go to step 3.

The hierarchical clustering relationships between similar sets of data can be shown using a dendrogram. Figures 5.12 and 6.17 are example dendrogram. The horizontal axis of the dendrogram represents the objects and clusters, and the vertical axis represents the distance between clusters using the Spearman correlation coefficient as a measure of dissimilarity.

### 4.3.4 Association Analysis

Design parameters and cost function are associated if knowing the value of a parameter implies something about the value of the cost function. Association analysis is done to find out how design parameters are associated with the cost function by uncovering hidden relationships and the significance of these relationships. To do that, a scatterplot of the cost function can be established. The scatterplot qualitatively captures relationships such as monotonicity or linearity.

### 4.3.5 Correlation Analysis

Correlation analysis can be done to quantify the linear correlation using the Pearson correlation coefficient $\rho$ [122]. The coefficient has values between +1 and -1 inclusive, where 1 is maximum positive correlation, 0 is no correlation, and -1 is maximum negative correlation. The Spearman correlation coefficient is related to the Pearson correlation coefficient, it is defined as the Pearson correlation coefficient between the rank parameters.

The Pearson correlation coefficient $\rho$ [122] between the cost function and a design parameter is computed using:

$$\rho_i = \frac{\sum_{r=1}^{R}(f_r - \bar{f})(x_{i,r} - \bar{x}_i)}{\sqrt{\sum_{r=1}^{R}(f_r - \bar{f})^2}\sqrt{\sum_{r=1}^{R}(x_{i,r} - \bar{x}_i)^2}} \quad (4.4)$$

where $f$ is the cost function, $x$ is the design parameter.
4.4 DSE Studies

In the DSE studies provided in Chapter 5 and Chapter 6, the dielectric values corresponding to the various dielectric regions in the channel represent design parameters, and the closeness of the PEP computed using the framework to the golden reference PEP represents the cost function. These studies were performed with the following goals in mind:

1. Demonstrating DSE framework usage scenarios. This included the following:

   - Evaluating the impact of modeling decisions on model accuracy: this refers to the impact of certain modeling decisions or abstractions applied to optimize cost on the accuracy of the modeling approach used. More specifically, DSE is used to explore the impact of modeling the channel with homogeneous vs. heterogeneous dielectric on modeling accuracy. Studies provided in Chapters 5 and 6 investigate the impact of representing water as a homogeneous dielectric on simulation accuracy. It was found that in a simple case (drug-free channel used in the DSE study in 5), the model was not sensitive to such abstraction, and a particular combination of homogeneous dielectric values for the protein (2) and water (60) lead to the same correct channel behavior as did the MD-computed heterogeneous map. However, in a more complex scenario (when a blocker was present in the KcsA channel) in the DSE study provided in Chapter 6, the homogeneous dielectric representation no longer resulted in correct results.

   - Estimating modeling parameters (i.e. a dielectric map) that would result in a particular channel behavior: The DSE study provided in Chapter 6 uses DSE to investigate the correct channel dielectric map which leads to correct channel function in the presence of TEA blocker, and the impact of the presence of the blocker on the dielectric map.

2. Validating the DSE methodology: In both DSE studies, it is shown that the size of the design space was reduced by over 70%.
3. Validating the FGBD framework: It was not possible to compute the PEP which reproduced experimental results in [2] using classic BD which models the water as a homogeneous dielectric. This motivated the need for FBGD which improves upon the modeling accuracy of classic BD. It was shown in the DSE study provided in Chapter 6 that it is possible to compute the PEP using FGBD that reproduces experimental results.
Chapter 5

DSE Study 1

5.1 DSE Study Overview

In this study, the framework is used to estimate the dielectric map for the blocker-free KcsA ion channel that would result in a PEP that reproduces conductance consistent with experimental results reported in [5]. The golden reference here is the PEP computed in [3] using the Langevin Simulator which leads to the I/V curve reported in Figure 3 in [5]. The I/V curve demonstrates that the relationship between the channel current and applied potential approaches a linear model when the applied potential value is between -80 to 100 mV, and the current begins saturating at electric potential values larger than 100mV and smaller than -80 mV. These characteristics of the experimental I/V curve are reproduced by the simulation results reported in [3] using the golden reference PEP.

The goal for this DSE study is to answer the following questions:

1. Can FGBD compute the correct PEP (i.e. PEP that reproduces experimentally determined results) using the MD-computed dielectric map?

2. Does the MD-computed dielectric map using the method in [84] result in a PEP that reproduces experimentally determined results?
3. Is there more than one dielectric map that would result in the same PEP?

4. To what degree can the design space size be reduced?

5.2 Background

The KcsA ion channel belongs to the family of potassium ion channels found in almost all organisms. These channels have diverse functions including osmotic regulation and neuronal signaling. KcsA is found in the soil bacteria Streptomyces lividans and was the first channel for which the crystal structure was resolved [123]. The channel is pH activated and is highly selective favoring potassium ions. The amino acid sequence that makes the selectivity filter of KcsA is highly conserved among many potassium channels, and as a result, KcsA has become a model for studying other potassium channels found in the human body and has been studied extensively. Research on KcsA has provided important insights on the mechanisms behind selectivity and conduction. The number of water molecules in each of the different regions (e.g. the selectivity filter) of the channel (see Figure 2.1) significantly impacts the permeation process. In classic BD, water throughout the channel is modeled as a homogeneous dielectric medium. In other words, the dielectric constant representing the water molecules in the very narrow region of the channel that is only wide enough for an ion is the same as the dielectric constant representing the water molecules in the wide part of the channel, the pore. One of the primary goals of this DSE study is to investigate, using DSE, the optimal dielectric combination (called dielectric map) for the various KcsA channel regions, and examine how that compares to the homogeneous dielectric combination.

5.3 Numerical Set-up

All experiments were conducted on the Western Canadian Research grid\(^1\). The KcsA potassium channel experimental protein structure used for DSE is the structure file 1BL8 from the Protein

\(^1\)Using a 4-core 2.66GHz 4GB Xeon E5430 workstation
Data Bank which is determined using X-ray crystallography. The 1BL8 structure of the channel is in the closed state. The channel structure was opened using MD simulations as described in [124]. This was done by Dr. Shin-Ho Chung’s group at the Australian National University (ANU).

Two cylindrical reservoirs are connected to the top and bottom of the channel; these reservoirs mimic the intracellular and extracellular spaces in biological ion channels. The influence of membrane potential is included by applying a uniform electric field across the channel where the potential drop occurs only across the channel and is uniform in the reservoirs.

5.4 MD Computed Dielectric Map

The dielectric map for the KcsA channel was computed by Dr. Steven Plotkin’s group in Biophysics at the University of British Columbia (UBC) according to the method described in [84]. Simulations were done using an all-atom classical MD simulation using the CHARM force field where proteins were solvated in a box of 10 Å on all sides. The simulation took 12 hours on an 8-core 2.5GHz Xeon workstation. The resolution of their mapping had the width of the channel to be roughly 40 samples. To limit the computational cost, the dielectric map computed using MD was approximated such that the channel is represented using six dielectric regions representing regions where the significant variation in dielectric occur: protein, upper channel, filter, pore, and upper and lower reservoirs (see Figure 5.1). These values are reported in tables 5.2 and 5.3. For a given region, an average value of a collection of samples was taken in the center of region. The width/height of the sample space was roughly 0.5 of the width/height of the region.

5.5 DSE Cost Function and Design Parameters

The cost function used in this study, $f(X)$, is defined in Equation 4.2. The dielectric regions in the channel structure represent the design parameters where $X = \{x_1, x_2, ..., x_N\}$ is a vector
Figure 5.1: The 6 dielectric regions of the KcsA channel which represent the different design parameters.

representing the dielectric values for the different dielectric regions. There is a total of 6 design parameters (see Figure 5.1):

1. Upper Reservoir
2. Upper Channel
3. Filter
4. Pore
5. Lower Reservoir
6. Protein
5.6 Design Space Reduction

The following subsections describe how the design space can be simplified using statistical inference techniques described in Chapter 4. It is important to note that regardless of whether PEP or conductance is used as a golden reference, the size of the resulting design space after design reduction using the techniques discussed in this section remains unchanged. This is because variation in any of the design parameters/variables (here, the dielectric values corresponding to the dielectric regions in the channel) that impacts channel conductance would also impact the PEP.

5.6.1 Sensitivity Analysis

5.6.1.1 Overview

In the sensitivity analysis, each parameter in the design space is explored independently while all others are fixed. This gives us a rough indication of the sensitivity of the cost function to each parameter. For each parameter explored, all other parameters are fixed to a default value which corresponds to an initial estimate that seeds the design space search.

Here, each of the design parameters has to be explored along with the average cost function achieved for each range of values for each parameter. For simplicity, the change or difference in the potential energy was used (instead of the difference between the computed potential energy and the golden reference potential energy) as that is a direct indication of the cost function. The rate of change in the PEP as a function of the rate of change of each of the design parameters $i$ (dielectric constant corresponding to each of the six channel regions per z-coordinate), $\Delta E/\Delta \epsilon_i$ was computed. It was found that the only significant rates of change occurred for the protein and upper channel regions. The following subsections show the data obtained for each design parameter.
5.6.1.2 PEP Sensitivity to Each Design Parameter

The plots in this section show the resulting PEP when varying a particular dielectric region while setting the other regions to a seed value.

Figures 5.2 and 5.3 show the various PEPs corresponding to variations in the upper reservoir dielectric region. Because these reservoirs represent the extracellular and intracellular spaces, the dielectric should be the same as bulk water. Water is a strong dielectric so only values between 50 and 80 (50, 52, 54, 56, ..., 72, 74, 76, 78, 80) were explored. The PEP was not sensitive to variations in this dielectric region and the rate of change in potential energy as a function of the rate of change of the upper reservoir dielectric was close to zero.

The PEP and rate of change of PEP plots are not included for lower reservoir dielectric region as they are very similar to the upper reservoir region.
Figures 5.4 and 5.5 show the results for upper channel region. Values explored for the upper channel region are 1, 2, 3, ..., 20. It is evident that the $\Delta E/\Delta \epsilon_{UCH}$ region is significant in the z-coordinate range of 0 to 40Å.

Figures 5.6 and 5.7 show results for the filter region. Values explored for the filter region are 16, 18, 20, ..., 40. Even though the filter dielectric region has a slightly larger impact than the upper and lower reservoirs do on the PEP, the PEP is still relatively insensitive to variations in the filter region.

Figures 5.8 and 5.9 show the results for the pore region. Values explored for the pore region are 48, 50, 52, ..., 72. Similar to the filter region, the impact of variations in the pore region on the PEP is insignificant. The rate of change of PEP potential energy as a function of the rate of change of the pore region dielectric is close to 0.

Figures 5.10 and 5.11 show the results for the protein region. Values explored for the protein region are 2, 4, 6, ..., 16. Similar to the upper channel region, the rate of change of the potential energy as a function of the rate of change of the dielectric value in the protein region is significant in the z-coordinate range of −40 to 40Å.
5.6.1.3 Optimization Based on Sensitivity Analysis Data

Based on the sensitivity results in figures 5.11 and 5.5, the design space is sub-divided into 4 regions, \( f(X) = \{f_1(X), f_2(X), f_3(X), f_4(X)\} \) defined by non-overlapping ranges of the z-axis and as a result can be optimized independently. Because it was observed that different z-ranges of the PEP are dependent on different dielectric constants, we can optimize these design space regions corresponding to these z-ranges independently. Table 5.1 specifies the ranges of z-axis values used to partition the design space.

<table>
<thead>
<tr>
<th>Cost Function</th>
<th>z-axis range</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1(X) )</td>
<td>-60 to -40</td>
</tr>
<tr>
<td>( f_2(X) )</td>
<td>-40 to 0</td>
</tr>
<tr>
<td>( f_3(X) )</td>
<td>0 to 40</td>
</tr>
<tr>
<td>( f_4(X) )</td>
<td>40 to 60</td>
</tr>
</tbody>
</table>

Table 5.1: Disjoint design space regions defined by z-axis ranges.
Figure 5.5: Rate of change in potential energy as a function of the rate of change of the upper channel region dielectric.

Figure 5.6: PEP for filter dielectric region values 16-40.
**Figure 5.7**: Rate of change in potential energy as a function of the rate of change of the filter region dielectric.

**Figure 5.8**: PEP for pore dielectric region values 48-72.
Figure 5.9: Rate of change in potential energy as a function of the rate of change of the pore region dielectric.

Figure 5.10: PEP for protein dielectric region values 2-16.
Figure 5.11: Rate of change in potential energy as a function of the rate of change of the protein region dielectric.

Because $f_1(X), f_2(X), f_3(X)$ and $f_4(X)$ are disjoint design space regions, they can be optimized independently and thus the design space can be simplified to,

$$f(X) = f_1(X) + f_2(X) + f_3(X) + f_4(X). \tag{5.1}$$

such that,

$$f_1(X) = \chi_{(-60,-40)} f(X),$$
$$f_2(X) = \chi_{(-40,0)} f(X),$$
$$f_3(X) = \chi_{(0,40)} f(X),$$
$$f_4(X) = \chi_{(40,60)} f(X).$$

where,

$$\chi_A(x) = \begin{cases} 
1, & \text{if } x \in A \\
0, & \text{if } x \notin A 
\end{cases}$$
Given this division of the design space, figures 5.5 through 5.11 indicate that $\Delta(f_2(X))/\Delta(\epsilon_{prot}) > 0$ and $\Delta(f_2(X))/\Delta(\epsilon_f) > 0$ for the dielectric constants of the protein and filter regions, respectively. It was similarly found that $\Delta(f_2(X))/\Delta(\epsilon_i) \approx 0$ for $i \in \{lr, ur, pore, uch, f\}$. As a result,

$$f_2(\epsilon_{lr}, \epsilon_{pore}, \epsilon_f, \epsilon_{prot}, \epsilon_{uch}, \epsilon_{ur}) = f_2(\epsilon_{prot}), \quad (5.2)$$

Similarly, it was found that $\Delta(f_3(X))/\Delta(\epsilon_i) \approx 0$ for $i \in \{lr, pore, ur\}$. As a result,

$$f_3(\epsilon_{lr}, \epsilon_{pore}, \epsilon_f, \epsilon_{prot}, \epsilon_{uch}, \epsilon_{ur}) = f_3(\epsilon_{prot}, \epsilon_f, \epsilon_{uch}), \quad (5.3)$$

Also from figures 5.5 to 5.11, it is evident that $\Delta(f_1(X))/\Delta(\epsilon_i) \approx 0$ and $\Delta(f_4(X))/\Delta(\epsilon_i) \approx 0$ for $i \in \{prot, f, lr, pore, uch, ur\}$ and thus,

$$f_1(\epsilon_{lr}, \epsilon_{pore}, \epsilon_f, \epsilon_{prot}, \epsilon_{uch}, \epsilon_{ur}) = 0, \quad (5.4)$$

and

$$f_4(\epsilon_{lr}, \epsilon_{pore}, \epsilon_f, \epsilon_{prot}, \epsilon_{uch}, \epsilon_{ur}) = 0. \quad (5.5)$$

Thus, equation 5.1 becomes,

$$f(X) = f_1(X) + f_2(X) + f_3(X) + f_4(X) = f_2(\epsilon_{prot}) + f_3(\epsilon_{prot}\epsilon_f, \epsilon_{uch}). \quad (5.6)$$

And the size of the DSE problem in equation 4.1 can be reduced to:

$$R = R_{prot} + R_{prot}R_{f}R_{uch}. \quad (5.7)$$
5.6.1.4 Hierarchical Clustering

For the KcsA channel, figure 5.12 illustrates hierarchical clustering of the DSE design parameters. In this figure, the y-axis represents the Spearman correlation coefficient, $r^2$ which is a measure for similarity [125]. Larger values of the Spearman Correlation coefficient represent a high degree of correlation up to a maximum value of 1.0 and smaller values represent a low degree of correlation to a minimum value of 0.0. As shown in the figure, the dielectric constant of the protein region is highly correlated with the dielectric constant of the upper channel region. As a consequence, these two design parameters will be optimized together. The filter, lower reservoir, upper reservoir, and pore domains were found to be weakly correlated with all other domains. As a consequence, the size of the space needed to be explored in equation 5.7 can be reduced to,

$$R = R_{prot} + R_{prot,uch}R_f$$

such that $R_{prot,uch}$ is the number of values explored for $x_{(prot,uch)}$ which is a single design parameter that simultaneously varies the values of $x_{prot}$ and $x_{uch}$.

**Figure 5.12:** Hierarchical clustering for the KcsA channel indicating the design parameters that can be optimized together.
5.6.1.5 Association Analysis

Figure 5.13 shows the impact of varying the design parameters for a range of values on the cost function. A scatterplot of the cost function is shown in figure 5.14. From figure 5.14, it can be observed that the protein, upper channel, and filter dielectric constants have a significant impact on the cost function, whereas all other dielectric regions have a minimal impact. Unlike the protein and upper channel dielectrics, the filter’s dielectric constant has a monotonic and linear relationship with the cost function.

<table>
<thead>
<tr>
<th></th>
<th>$\varepsilon$</th>
<th>Cost Func.</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein</td>
<td>(2,4)</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>2.06</td>
</tr>
<tr>
<td></td>
<td>(8,16)</td>
<td>3.67</td>
</tr>
<tr>
<td>upper channel</td>
<td>(1,10)</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>(11,19)</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td>1.58</td>
</tr>
<tr>
<td>filter</td>
<td>(16)</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>(16,28)</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>(30,40)</td>
<td>1.93</td>
</tr>
<tr>
<td>upper reservoir</td>
<td>(50,68)</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>(70,78)</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>(80)</td>
<td>1.91</td>
</tr>
<tr>
<td>lower reservoir</td>
<td>(50)</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>(52,68)</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>(70,78)</td>
<td>1.90</td>
</tr>
<tr>
<td>pore</td>
<td>(48)</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>(50,68)</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>(68,72)</td>
<td>1.93</td>
</tr>
</tbody>
</table>

Figure 5.13: The design parameters, $X$, and their impact on the cost function, $f(x)$ as they vary for a range of values.
Figure 5.14: A scatterplot showing the relationship between design parameters and the cost function for a range of values to reveal any linear relationships with the cost function: The cost function varies significantly with changes in the dielectric for the protein, upper channel and filter; and has a monotonic relationship with the filter dielectric.

5.6.1.6 Correlation Analysis

Correlation analysis results are shown in Figure 5.15. As evident from the figure, the dielectric constant of the filter has a strong linear correlation with the cost function. Therefore, this design parameter can be modeled with a linear regression model and only two values of the filter dielectric need to be explored to approximate all other values. As a consequence, the size of the DSE problem becomes,

\[ R = 2R_{prot,uch} + R_{prot} \]  \hspace{1cm} (5.9)
Thus, if we assume that $R_{prot} = R_1$ and $R_{uch} = R_2$, and that the size $R_{12}$ is the larger of $R_1$ and $R_2$, then the use of reduction techniques reduced the size of the design space from:

$$R = \prod_{i=1}^{6} R_i$$

...to

$$R = 2R_{12} + R_1$$

where $R_{12} = \max\{R_1, R_2\}$.

**Figure 5.15:** Linear correlation of the design parameters with the cost function for the KcsA DSE case-study. 1 indicates maximum positive correlation and 0 indicates no correlation.

5.7 Results and Discussion

5.7.1 Estimated Dielectric Map Based on Golden Reference

Using the simplified design space, DSE experiments were used to estimate the optimal set of dielectric constants given the PEP reported in [3]. Because it was shown in previous studies such as [3] that it is possible to compute a PEP that reproduces results consistent with experimental data using a homogeneous dielectric model, one of the DSE experiments was restricted such that only homogeneous values for the water and protein were tried. Final results show that DSE estimates two dielectric maps that result in very similar PEPs that are consistent with the golden reference PEP from [3]. The two resulting maps are compared to dielectric values computed directly using MD in
tables 5.2 and 5.3. The first DSE-estimated dielectric map closely matches that computed using MD for all regions but one (the error in approximating $\epsilon_{uch}$ was 40%). On average, the estimated values match those from MD with a 7% error. The other DSE-estimated map is identical to the map used in [3] and is inconsistent with MD computations with a total error of 74%. The PEPs resulting from the various dielectric maps are shown in Figure 5.16.

<table>
<thead>
<tr>
<th></th>
<th>Computed Using MD</th>
<th>Estimated Using DSE: Solution 1</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_{prot}$</td>
<td>3</td>
<td>3</td>
<td>0%</td>
</tr>
<tr>
<td>$\epsilon_{uch}$</td>
<td>25</td>
<td>35</td>
<td>40%</td>
</tr>
<tr>
<td>$\epsilon_{f}$</td>
<td>30</td>
<td>30</td>
<td>0%</td>
</tr>
<tr>
<td>$\epsilon_{pore}$</td>
<td>50</td>
<td>50</td>
<td>0%</td>
</tr>
<tr>
<td>$\epsilon_{lr}$</td>
<td>80</td>
<td>80</td>
<td>0%</td>
</tr>
<tr>
<td>$\epsilon_{ur}$</td>
<td>80</td>
<td>80</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 5.2: Comparing the dielectric map from solution 1 which is estimated using DSE given the golden reference PEP by Chung et al. [3] to the dielectric map computed using MD, and the associated error in solution 1.

<table>
<thead>
<tr>
<th></th>
<th>Computed Using MD</th>
<th>Estimated Using DSE: Solution 2</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_{prot}$</td>
<td>3</td>
<td>2</td>
<td>33%</td>
</tr>
<tr>
<td>$\epsilon_{uch}$</td>
<td>25</td>
<td>60</td>
<td>140%</td>
</tr>
<tr>
<td>$\epsilon_{f}$</td>
<td>30</td>
<td>60</td>
<td>200%</td>
</tr>
<tr>
<td>$\epsilon_{pore}$</td>
<td>50</td>
<td>60</td>
<td>20%</td>
</tr>
<tr>
<td>$\epsilon_{lr}$</td>
<td>80</td>
<td>60</td>
<td>25%</td>
</tr>
<tr>
<td>$\epsilon_{ur}$</td>
<td>80</td>
<td>60</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 5.3: Comparing the dielectric map from solution 2 which is estimated using DSE given the golden reference PEP by Chung et al. [3] to the dielectric map computed using MD, and the associated error in solution 2.

It is interesting that one of the dielectric maps that resulted in a PEP that is close to the golden reference which reproduces the correct channel conductance is a homogeneous map. This raised the question of the relevance of modeling the channel as a heterogeneous medium. It is possible that the model in this scenario is not sensitive to dielectric heterogeneity and that it would be more sensitive in other more complex scenarios. This motivated the need for conducting another DSE study to investigate a more complex DSE scenario. It is important to note that one of the two DSE estimated dielectric maps is in agreement with the dielectric map computed directly using MD and that this dielectric map does, in fact, result in the correct channel conductance. This validates the accuracy
of the framework by showing that using the MD-computed dielectric map, FGBD computes a PEP that consistent with the PEP that reproduces experimentally determined results.

5.7.2 Model Validation: DSE Estimation is Consistent With MD Computations

It is significant that the heterogeneous dielectric map estimated by performing DSE using FGBD is consistent with MD computations. The implications is that the framework provides a reliable tool for estimating the correct dielectric map given the desired golden reference. Further, this confirms that the dielectric map computed using MD does in fact result in computations that are consistent with experimental data, and as a result, this is a validation for the method proposed in [84].
Chapter 6

DSE Study 2

6.1 DSE Study Overview

This DSE study explores a more complex scenario in which the KcsA channel is occluded by the Tetraethylammonium (TEA) molecule outside the selectivity filter on the extracellular side of the channel (see Figure 6.1 [2]).

In this DSE study, two DSE experiments are performed:

1. **DSE Experiment 1**: the MD-computed PEP reported in [4] is used as the golden reference to estimate the dielectric map for the TEA-occupied KcsA ion channel.

2. **DSE Experiment 2**: the PEP that reproduces the KcsA TEA external block inhibitory constant \( K_i \) (computed using the Langevin Simulator [2]) that is consistent with the experimentally determined results reported in [5] (see Figure 6.23) is used as the golden reference.

The goal of these DSE studies is to answer the following questions:

1. Can FGBD be used to compute the correct PEP (i.e. PEP that reproduces experimental determined results) in a more complex scenario such as when a drug is present in the channel?
2. Can the homogeneous dielectric model lead to the correct PEP as the case in the previous DSE study?

3. How does the presence of the TEA drug molecule impact the dielectric strength? In other words, how does the TEA-occupied channel dielectric map look in comparison to the TEA-free channel dielectric map?

4. To what degree can the size of the design space be reduced in this case?

6.2 Background

TEA is an experimental drug that is a very effective blocker for potassium channels and has been widely used to probe the function and dynamics of potassium channels [8]. The sensitivity of different potassium channels to TEA varies considerably (within 700-fold [126]). This sensitivity also varies for the same channel if the channel structure changes slightly. For example, the inhibitory constant $k_i$, which is a measure of the blocker’s effectiveness (defined as the concentration required to produce half maximum inhibition), for the Shaker K1 channel is 27 mM. However, the channel’s TEA binding affinity increases significantly when the threonine residue at position 449 is substituted with the tyrosine or phenylalanine residue (3,4). In KcsA and Kv2.1, where four tyrosine residues (at position 82 or 380), located just external to the selectivity filter, form a binding pocket for TEA,
the current is halved when its concentration reaches between 1.7 and 4.5 mM [21, 127–129]. The external TEA affinity in KcsA is substantially reduced when tyrosine at position 82 is substituted with cysteine or threonine [21]. The mechanisms behind this TEA binding affinity in different channels and relevance to structural changes are not fully understood.

In DSE Experiment 1, we compare the modeling accuracy of FGBD against classic BD which uses a homogeneous dielectric. In DSE Experiment 2, using DSE, we estimate the dielectric map that results in the golden reference PEP which reproduces the experimentally determined inhibitory constant for TEA reported in [5] (see Figure 6.23). We examine the impact the presence of TEA has on the dielectric map by comparing the blocker-occupied channel dielectric map to the blocker-free channel dielectric map.

6.3 Numerical Set-up

The numerical set-up for DSE study 2 is similar to that in DSE study 1, with the difference that a TEA molecule is placed at the entrance of the selectivity filter on the extracellular side of the channel.

6.4 DSE Cost Function and Design Parameters

The cost function, $f(X)$, is defined in Equation 4.2. The dielectric regions in the channel structure represent the design parameters where $X = \{x_1, x_2, ..., x_N\}$ is a vector representing the dielectric values for the different dielectric regions.

Various studies have shown that the dielectric constant is reduced when an ion is placed in a solution in the region surrounding the ion [130, 131]. As a result, to determine the number of design parameters in this DSE study (i.e. the number of dielectric regions in the TEA-occupied channel), the position of the TEA molecule in the channel has to be considered: a higher resolution has to be provided in the regions surrounding the ion, and as a result, the TEA-occupied channel dielectric
map was approximated such that the channel is represented using nine dielectric regions (see figure 6.2):

1. Upper Reservoir
2. Upper Upper Channel
3. Lower Upper Channel
4. Upper Filter
5. Lower Filter
6. Upper Pore
7. Lower Pore
8. Lower Reservoir
9. Protein

The MD-computed dielectric map was discretized to represent the dielectric regions specified above. The dielectric values corresponding to these regions are reported in Table 6.2.

6.5 Design Space Reduction

A similar analysis as in Chapter 5 is performed to reduce the size of the design space using the proposed DSE methodology. Similar to Chapter 5, the following subsections represent the sequential set of steps of the analysis. Each subsection describes how and to what extent the size of the design space can be reduced.
6.5.1 Sensitivity Analysis

6.5.1.1 PEP Sensitivity to Each Design Parameter

The sensitivity analysis in this chapter was performed in a similar manner to study 1. More explicitly, each variable in the design space is explored independently while all others are fixed to a seed value.

As in DSE study 1, values between 50 and 80 (50, 52, 54, 56, ..., 72, 74, 76, 78, 80) were explored for the upper and lower reservoirs dielectric regions. The dielectric variations within the upper and lower reservoirs regions had a negligible impact on the PEP, and as a result, the rate of change of the potential energy as a function of the rate of change of the dielectric value in both regions is close to zero.

Values (1, 2, 3, 4, ..., 20) were explored for both the upper upper channel and the lower upper channel regions. Figures 6.5 and 6.3 show sample PEPs that illustrate the impact of the change in
dielectric on the PEP. It is shown that the maximum value of $\Delta E/\Delta \epsilon_{\text{uce}}$ is greater than 1.5 (see Figure 6.4), and the maximum value of $\Delta E/\Delta \epsilon_{\text{uc}}$ is under 1 (see Figure 6.6).

Values (4, 5, 6, ...., 45) were explored for the upper and lower filter dielectric regions. Figures 6.9 and 6.7 show sample PEPs that illustrate the impact of the change in dielectric on the PEP. The rate of change of the potential energy as a function of the rate of change of the dielectric value in the upper filter region is significant in the mouth of the channel where the TEA molecule resides (see Figure 6.8). The rate of change of the potential energy as a function of the rate of change of the dielectric value in the lower filter region is not as significant, with the maximum of this ratio being almost 2 in the filter and upper channel regions (see Figure 6.10).

Values (48, 50, 52, ...., 72) were explored for the upper and lower pore dielectric regions (see Figures 6.11 and 6.13). The impact of the rate of change of the dielectric values in both the upper and
**Figure 6.4:** Rate of change in potential energy as a function of the rate of change of the upper upper channel dielectric region.

**Figure 6.5:** PEP for lower upper channel dielectric region values 2, 5, 12, and 20. A higher dielectric value cause the wells to become shallower.
**Figure 6.6:** Rate of change in potential energy as a function of the rate of change of the lower upper channel dielectric region.

**Figure 6.7:** PEP for upper filter dielectric region values 4, 8, 16, 20, and 25. An increase in the dielectric value causes a significant increase in the barrier height.
Figure 6.8: Rate of change in potential energy as a function of the rate of change of the upper filter dielectric region.

Figure 6.9: PEP for lower filter dielectric region values 8, 10, 12, 16, and 20. An increase in the dielectric value causes the well to become shallower.
lower pore regions on the rate of change in the PEP is not significant with this rate being less than 1 (see Figures 6.12 and 6.14).

Values (1, 2, 4, 6, ..., 16) were explored for the protein dielectric region. As shown in Figure 6.15, the PEP is very sensitive to variations in the protein region dielectric $\epsilon_{prot}$ for the z-coordinate range of $-4 \times 10^{-9}$ to $2 \times 10^{-9}$ Å (see Figure 6.16).

6.5.1.2 Optimization Based on Sensitivity Analysis Data

Based on the sensitivity results in figures 6.4 - 6.16, the design space is sub-divided into 4 regions, $f(X) = \{f_1(X), f_2(X), f_3(X), f_4(X), f_5(X)\}$ defined by non-overlapping ranges of the z-axis and as a result can be optimized independently. Similar to DSE study 1, it was observed that different z-ranges of the electric potential curve are dependent on different dielectric constants, as a result, we can optimize the cost function corresponding to these z-ranges independently. Table 6.1 specifies the ranges of z-axis values used to partition the design space.
Figure 6.11: PEP for upper pore dielectric region values 48, 50, ... 72. The change in dielectric value does not have a significant impact on the PEP.

Figure 6.12: Rate of change in potential energy as a function of the rate of change of the upper pore dielectric region.
Figure 6.13: PEP for lower pore dielectric region values 48, 50, ... 72. The change in dielectric value does not have a significant impact on the PEP.

Figure 6.14: Rate of change in potential energy as a function of the rate of change of the lower pore dielectric region.
Figure 6.15: PEP for protein dielectric region values 1, 2, 3, 4, 5, and 6. A higher dielectric value in the protein regions causes the wells to become shallower.

Figure 6.16: Rate of change in potential energy as a function of the rate of change of the protein dielectric region.
<table>
<thead>
<tr>
<th>Cost Function</th>
<th>z-axis range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1(X)$</td>
<td>$-60$ to $-40$</td>
</tr>
<tr>
<td>$f_2(X)$</td>
<td>$-40$ to $10$</td>
</tr>
<tr>
<td>$f_3(X)$</td>
<td>$10$ to $18$</td>
</tr>
<tr>
<td>$f_4(X)$</td>
<td>$18$ to $30$</td>
</tr>
<tr>
<td>$f_5(X)$</td>
<td>$30$ to $60$</td>
</tr>
</tbody>
</table>

**Table 6.1**: Disjoint design space regions defined by z-axis ranges.

Because $f_1(X), f_2(X), f_3(X), f_4(X),$ and $f_5(X)$ are disjoint functions, they can be optimized independently and thus the design space can be simplified to,

$$f(X) = f_1(X) + f_2(X) + f_3(X) + f_4(X) + f_5(X).$$

(6.1)

such that,

$$f_1(X) = \chi_{(-60,-40)}f(X),$$

$$f_2(X) = \chi_{(-40,10)}f(X),$$

$$f_3(X) = \chi_{(10,18)}f(X),$$

$$f_4(X) = \chi_{(18,30)}f(X),$$

$$f_5(X) = \chi_{(30,60)}f(X).$$

where,

$$\chi_A(x) = \begin{cases} 
1, & \text{if } x \in A \\
0, & \text{if } x \notin A 
\end{cases}$$

Given this division of the design space, and examining figures 6.4 - 6.16, it is possible to see that,

$$f_1(\epsilon_{ur}, \epsilon_{lr}, \epsilon_{up}, \epsilon_{lp}, \epsilon_u f, \epsilon_l f, \epsilon_{prot}, \epsilon_{uuch}, \epsilon_{luc}) \approx 0$$

(6.2)

$$f_2(\epsilon_{ur}, \epsilon_{lr}, \epsilon_{up}, \epsilon_{lp}, \epsilon_u f, \epsilon_l f, \epsilon_{prot}, \epsilon_{uuch}, \epsilon_{luc}) = f_2(\epsilon_{prot}),$$

(6.3)
\[ f_3(\epsilon_{ur}, \epsilon_{lr}, \epsilon_{up}, \epsilon_{lp}, \epsilon_{uf}, \epsilon_{lf}, \epsilon_{prot}, \epsilon_{uuch}, \epsilon_{luc}) = f_3(\epsilon_{prot}, \epsilon_{lf}), \]  

(6.4)

\[ f_4(\epsilon_{ur}, \epsilon_{lr}, \epsilon_{up}, \epsilon_{lp}, \epsilon_{uf}, \epsilon_{lf}, \epsilon_{prot}, \epsilon_{uuch}, \epsilon_{luc}) = f_4(\epsilon_{prot}, \epsilon_{uf}, \epsilon_{uuch}), \]  

(6.5)

\[ f_5(\epsilon_{ur}, \epsilon_{lr}, \epsilon_{up}, \epsilon_{lp}, \epsilon_{uf}, \epsilon_{lf}, \epsilon_{prot}, \epsilon_{uuch}, \epsilon_{luc}) \approx 0 \]  

(6.6)

Thus, equation 6.1 becomes,

\[ f(X) = f_1(X) + f_2(X) + f_3(X) + f_4(X) + f_5(X) = f_2(\epsilon_{prot}) + f_3(\epsilon_{prot}, \epsilon_{lf}) + f_4(\epsilon_{prot}, \epsilon_{uuc}, \epsilon_{uf}) \]  

(6.7)

And the size of the DSE problem in equation 4.1 can be reduced to,

\[ R = R_{prot} + R_{prot}R_{lf} + R_{prot}R_{uuc}R_{uf}. \]  

(6.8)

### 6.5.1.3 Hierarchical Clustering

Similar to DSE study 1, the clustering algorithm is used to determine which design variables are highly correlated and can be considered redundant. Any pair of variables considered redundant can be treated as one variable with respect to optimization. The clustering results shown in Figure 6.17 indicates the strongest correlation to be between the protein and upper upper channel regions. In this case, the Spearman Correlation coefficient is 0.96. The next strongest correlation is between the lower pore, upper pore, and lower filter regions with a Spearman Correlation coefficient of 0.70. In equation 6.8, we are interested in any redundancies between variables in the same term. More explicitly, we are interested in determining if there are any redundancies between the following
pairings: protein/lower upper channel, protein/lower filter, upper filter/lower upper channel. As shown in the cluster graph, none of these pairing is sufficient to warrant clustering for the purpose of collapsing the design space for exploration. Thus, the size of design space as expressed by equation 6.8 remains unchanged.

6.5.1.4 Association Analysis

A scatterplot of the cost function is shown in figure 6.19. From figure 6.19, it can be noticed that the protein, lower upper channel, and upper filter dielectric regions values have a significant impact on the cost function, whereas all other dielectric regions have a minimal impact. Unlike the protein, the lower upper channel dielectric and the upper filter’s dielectric constant both have a monotonic and linear relationship with the cost function.

Figure 6.17: Hierarchical clustering for the TEA-occupied KcsA channel indicating the design parameters that can be optimized together.
Figure 6.18: The design parameters, $X$, and their impact on the cost function, $f(x)$ as they vary for a range of values.

<table>
<thead>
<tr>
<th>Region</th>
<th>$\varepsilon$</th>
<th>Cost Func.</th>
</tr>
</thead>
<tbody>
<tr>
<td>upper upper-channel</td>
<td>(1,10)</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>(11,19)</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td>5.6</td>
</tr>
<tr>
<td>lower upper-channel</td>
<td>(1,10)</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>(11,19)</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td>6.6</td>
</tr>
<tr>
<td>protein</td>
<td>(20)</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>(4,10)</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>(12,16)</td>
<td>29.8</td>
</tr>
<tr>
<td>pore</td>
<td>(48)</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>(50,68)</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>(70,72)</td>
<td>7.8</td>
</tr>
<tr>
<td>upper filter</td>
<td>(16)</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>(18,28)</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>(30,40)</td>
<td>8.3</td>
</tr>
<tr>
<td>lower filter</td>
<td>(16)</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>(18,28)</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td>(30,40)</td>
<td>14.1</td>
</tr>
</tbody>
</table>

Figure 6.19: A scatterplot showing the relationship between design parameters and the cost function for a range of values to reveal any linear relationships with the cost function: the protein, lower upper channel, and upper filter dielectric regions values have a significant impact on the cost function, whereas all other dielectric regions have a minimal impact. Unlike the protein, the lower upper channel dielectric and the upper filter's both have a monotonic and linear relationship with the cost function.
6.5.1.5 Correlation Analysis

Correlation analysis results are shown in Figure 6.20. As evident from the figure, the dielectric constant values of both the upper filter and upper upper channel regions have a strong linear correlation with the cost function. Therefore, these design parameters can be modeled with a linear regression model and only two values of each of these parameters need to be explored to approximate all other values. As a consequence, the size of the DSE problem becomes,

\[ R = R_{prot} + R_{prot}R_{lf} + 2 \times R_{prot} \times 2 = 5R_{prot} + R_{prot}R_{lf} = R_{prot}(5 + R_{lf}) \] (6.9)

Thus, if we assume that \( R_{prot} = R_1 \) and \( R_{lf} = R_2 \), then the use of reduction techniques reduced the size of the design space from:

\[ R = \prod_{i=1}^{9} R_i \] to \( R = R_1(5 + R_2) \).

**Figure 6.20:** Linear correlation of the design parameters with the cost function. 1 indicates maximum positive correlation and 0 indicates no correlation.
6.6 Results and Discussion

6.6.1 Impact of Dielectric Heterogeneity on Channel Model

The authors in [132] have shown that the shape of the energy wells and barriers in the PEP provide important insights about channel conduction behavior, and that conductance is dependent on the depth and width of the energy wells. For example, a deeper well results in the ion staying in the well for a longer amount of time than a shallower well, given that the two wells have the same width. This results in a decrease in the current. Similarly, a wider well will cause the ion to stay in it for longer than a narrower well given that the two wells have the same well depth. This also results in a decrease in the current. As a result, the ability to manipulate the PEP in a controlled manner based on design parameters would be valuable. For example, the following general trends were observed in this DSE study:

- Lowering the dielectric constant for the upper filter region causes the well to be shallower and shifts it towards the center of the channel.
- Lowering the dielectric constant for the lower filter region causes the well to be deeper.
- Increasing the dielectric constant for the lower filter region causes the well to be shallower and shifts it to towards the top of the channel.
- Lowering the dielectric constant for the upper upper channel region causes the well to be slightly deeper.

This implies that by varying these dielectric combinations, we are effectively influencing how fast an ion goes through the channel, and as a result, the strength of the current flow. This is significant because it allows us to perform studies such as investigating selectivity where we adjust the dielectric regions values such that conduction matches a golden reference. For example, an experiment that would help investigate the selectivity of potassium over sodium, we could find the dielectric combination that would slow down the ionic flow by a factor of 10000. This dielectric combination
would be the dielectric map that should be used to compute the PEP for sodium ions as they permeate through the channel.

Figure 6.21 demonstrates how it is possible to manipulate the PEP by adjusting the design parameters. The figure shows the PEPs computed using FGBD that correspond to sample combinations of dielectric values for the different dielectric regions. The legend in the figure shows the dielectric values for each region in the form [protein lower reservoir upper reservoir lower pore upper upper channel upper pore lower upper channel lower filter upper filter]. Figure 6.22 shows a zoomed-in version of Figure 6.21 showing two of the higher resolution dielectric PEPs at the mouth of the channel region where the channel gets occluded by the drug molecule. It can be seen that different regions of the PEP can be “manipulated” by adjusting the dielectric values. It is also possible to see how similar dielectric values within the same region for certain regions (here the upper filter region) produce similar PEPs. The curve labeled as “Lower resolution dielectric: [1] [80] [80] [50] [25] [50] [25] [30] [30]” shows the PEP when the dielectric region breakdown is the same as in DSE study 1 in which case the filter and upper channel regions are not further broken into two regions. In this case, the PEP in the mouth of the channel appears as a straight line and there is no “resolution” within this part of the channel.

6.6.2 Finalizing Estimation

It is possible that more than one PEP produce a similar channel behavior (e.g. conductance, inhibitory constant). When this occurs, selection of design parameters should be constrained by chosen features of the PEP computed using BD or MD PEP. In the DSE studies provided in Chapters 5 and 6, the final selection of the design parameters was constrained by the depth of the well.

6.6.3 Accuracy of FGBD in Comparison to Homogeneous BD and MD

In this section, the FGBD and the DSE methodology effectiveness is validated by examining a small part of the PEP which is the upper part of the channel that has the binding pocket where the TEA
**Figure 6.21:** Manipulating the PEP through dielectric heterogeneity through the channel. The legend shows the dielectric values for each region in the form [protein lower reservoir upper reservoir lower pore upper upper channel upper pore lower upper channel lower filter upper filter].

**Figure 6.22:** Manipulating the PEP through dielectric heterogeneity in the upper part of the channel.
molecule binds just exterior to the selectivity filter (see Figure 6.24). Zooming in on this significant but small region of the channel is a test of FGBD’s modeling resolution by testing the framework’s sensitivity to changes in this small part of the channel.

Figure 6.25 shows a comparison in the TEA binding pocket region of the KcsA channel between the PEP computed using FGBD with DSE-estimated dielectric map, the PEP computed using FGBD with homogeneous dielectric, and the golden reference PEP computed directly using MD by Crouzy et al. [4].

It is evident from the figure that the homogeneous model deviates significantly from the correct PEP. FGBD, on the other hand, provides a much closer match to the PEP computed directly using MD with $R^2 = 82\%$ while $R^2$ drops to 65% when using homogenous dielectric BD. It is important to note that it was not possible to compute a PEP that provides a close match to the PEP computed using MD without increasing the channel dielectric heterogeneity by further breaking the filter and upper pore regions into sub-regions.

The DSE estimated dielectric map closely matched the dielectric map computed directly from MD for the ion-free channel with the exception that the dielectric was significantly reduced in the region where the TEA ion resides and the regions surrounding it including the protein (table 6.2). This confirms the finding in other studies that the dielectric constant is reduced when an ion is placed in a solution in the region surrounding the ion [130, 131]. This also implies that it is possible that dielectric map determined for a blocker-occupied channel can be used to model a blocker-free channel. This map can be adjusted slightly such that the dielectric constant in and around the region where the drug molecule resides is dampened. This dampening effect can be explored using DSE.

It was shown in [2] that the best agreement between simulation results produced using the Langevin simulator [2] and experimental results [5] (see Figure 6.23) is when the depth of the PEP seen by TEA is assumed to be 13 kT. Figure 6.23 shows the resulting channel current as well as the attenuation
Dielectric map computed using MD for ion-free channel

DSE estimated using MD computed PEP by Crouzy et al.

DSE estimated using the PEP that reproduces the experimentally determined inhibitory constant computed by Heginbotham et al. as a golden reference

<table>
<thead>
<tr>
<th></th>
<th>Dielectric map computed using MD for ion-free channel</th>
<th>DSE estimated using MD computed PEP by Crouzy et al.</th>
<th>DSE estimated using the PEP that reproduces the experimentally determined inhibitory constant computed by Heginbotham et al. as a golden reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_{prt}$</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\epsilon_{uuch}$</td>
<td>25</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$\epsilon_{luch}$</td>
<td>25</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\epsilon_{uf}$</td>
<td>30</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$\epsilon_{lf}$</td>
<td>30</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>$\epsilon_{up}$</td>
<td>50</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>$\epsilon_{lp}$</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>$\epsilon_{lr}$</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>$\epsilon_{ur}$</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 6.2: Comparing DSE estimated dielectric maps for the TEA-occupied KcsA channel using MD computed PEP by Crouzy et al. [4] as a golden reference and using the PEP that reproduces the experimentally determined inhibitory constant computed by Heginbotham et al. [5] as a golden reference to the dielectric map computed using MD for the empty channel. The presence of the TEA molecule dampens the dielectric constant in and around the region where the molecule resides. Dielectric map estimated using MD computed PEP deviates from the dielectric map estimated using the PEP that reproduces experimental data in the lower filter region.

in this current as a result of the concentration of the TEA molecule in the channel. The figure compares the simulated current using the Langevin Simulator as a function of TEA concentration (solid circles) against experimentally determined results (open diamonds) [2]. It is shown that there is an exponential relationship between the channel current and the concentration of the TEA molecule. This relationship is also present in the percentage attenuation of the channel current. It is also shown that the simulated value of the inhibitory constant which is the concentration required to produce half maximum inhibition is consistent with the experimentally determined value. Figure 6.26 shows a comparison between the PEP estimated using FGBD with the DSE-estimated dielectric map and the golden reference PEP that reproduces the inhibitory constant determined experimentally in [5]. The depth of the well in the PEP computed using the framework is very close (12.5KT) to that of the golden reference (13KT).

The estimated dielectric map that would result in this PEP (which as a 13kT well depth) is also reported in Table 6.2. The table compares that map to the MD-computed dielectric map for the blocker-free channel as well as the dielectric map estimated based on MD-computed PEP for the
Figure 6.23: Current as a function of TEA concentration for both BD simulated (solid circles) and experimentally determined (open diamonds) [2] with experimental results reproduced from [5]. The applied potential and the ionic concentration used for the simulations are 196 mV and 150 mM, respectively. The depth of the energy-well encountered by the TEA molecule is 13kT.

Figure 6.24: Geometric model of the KcsA channel showing the TEA binding pocket and the region for which the PEP is compared.

TEA-occupied channel. Results show that the presence of the TEA molecule reduces the dielectric constant in the region where it resides and the areas in close proximity to the molecule, and that to deepen the well further to 13kT, the dielectric would have to drop in the lower filter region where the ion resides by 84%.
These results show the significant improvement in modeling resolution achieved using FGBD as it was shown that the PEP computed using the framework for the complex scenario of the TEA-occupied is consistent with the PEP computed directly using MD with a $R^2 = 82\%$. The impact of dielectric heterogeneity on modeling resolution is evident from the fact that the PEP computed using a homogenous model deviates from that computed using MD with $R^2$ dropping to 65%.

### 6.7 Relevance of Dielectric Heterogeneity in Drug Design

Figures 6.25 and 6.26 show the PEP in the upper channel region. It is significant that a BD-based model (FGBD) is sensitive to changes in this small part of the channel. Classic BD is not sensitive to such changes as evident from Figure 6.25.

The PEP in these figures show that there is a well towards the end of the selectivity filter on the extracellular side of the channel where the TEA molecule binds. The presence of this well is
Figure 6.26: A comparison between the PEP estimated using FGBD with the DSE-estimated dielectric map and the golden reference PEP from [2] that reproduces the inhibitory constant determined experimentally by Heginbotham et al. [5]. The depth of the well in the PEP computed using the framework is very close (12.5KT) to that of the golden reference (13KT).

Also evident in the significant drop in the upper filter region and the lower upper channel regions dielectric values when TEA is present. The channel blocking in terms of frequency and period, and as a result, conductance, are very much dependent on the depth of this well. The probability that TEA will be blocking the channel is exponentially related to the depth of the well [2]. This impacts how long TEA binds for and whether the unbinding process will be possible, which would significantly impact the operation of the channel. Using DSE to design the shape and size of the well would be useful in applications such as the design of other blockers or the design of artificial pores that rely on similar mechanisms.
Chapter 7

Computational Feasibility

7.1 Scalability Limits of FGBD

7.1.1 Limits of Spatial and Temporal Resolutions

This section investigates the limits of spatial and temporal resolutions required to achieve acceptable levels of accuracy for FGBD while achieving memory and runtime scalability.

As discussed previously, the large number of electric field entries used to perform BD constitutes the “bottleneck” of the simulation framework with respect to memory. To accommodate the large number of simulation iterations needed to acquire relevant electric field information, this set of matrices must be entirely stored in memory and must provide constant runtime access to the BD simulator.

This section investigates the feasibility of FGBD. It shows that reasonable levels of accuracy can be achieved with modest table sizes (2.2 GB).

The Heterogeneous Dielectric PEP Calculator was used for six different levels of grid spacing (Figure 7.1). Each level of spacing corresponds to five different properties which were adjusted in unison. These properties affect how the mesh is constructed when various obstacles are encountered such
as small curved parts of the geometry. For simplicity, the levels of grid spacing are denoted by $\beta_1$, $\beta_2$, ..., $\beta_6$.

Each level of grid spacing corresponds to a different number of total grid points per channel thus affecting the size of the table needed to store pre-computed electric field information. $\beta_1$ denotes the highest resolution number and $\beta_6$ denotes the lowest resolution number. For example, for $\beta_5$, each dimension of the 5-D table is of size 27 so the total size of the table in bytes is $27^5$ table entries x 6 numbers/entry x 8 bytes/number = 657 MB assuming the use of double-precision numbers. The set of table sizes corresponding to $\beta_1$ through $\beta_6$ range from 350 MB to 22.5 GB.

<table>
<thead>
<tr>
<th>Grid Level</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid Points/Channel</td>
<td>22004</td>
<td>11130</td>
<td>7030</td>
<td>5128</td>
<td>3816</td>
<td>2765</td>
</tr>
<tr>
<td>Table Size (MB)</td>
<td>22164</td>
<td>5671</td>
<td>2262</td>
<td>1204</td>
<td>657</td>
<td>350</td>
</tr>
</tbody>
</table>

**Figure 7.1:** Memory footprint corresponding to the different grid sizes in a KsA channel.

Using the PEP Calculator, a table was constructed for each level of grid spacing. The accuracy of levels $\beta_2$ through $\beta_6$ were compared against $\beta_1$ on a point by point bases. To facilitate such a
comparison, all matrices were constructed using the same number of grid points. In the case of level $\beta_1$, 100% of the table entries were computed using FEM. For levels $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$, and $\beta_6$, only some of the matrices were computed by the Heterogeneous Dielectric PEP Calculator and the remaining entries were interpolated. Effectively, the accuracy of interpolation is being determined as a function of the grid spacing. This is important because BD simulations must interpolate electric field information based on the matrices each time an ion moves into a new location within the channel. This occurs every time step of the simulation and for all ions in the system.

![Figure 7.2](image)

**Figure 7.2:** The error in electric field estimation as a function of the table size.

Figure 7.2 shows the error in interpolated electric field values as a function of the table size. The shape of this curve appears to follow that of an exponential decay such that the proportional reduction in error diminishes as we construct larger matrices. Thus, if we choose grid spacing on the "knee" of the curve (i.e., $\beta_3$) then we can be relatively certain that larger table sizes will not result in significant improvements in accuracy. $\beta_3$ falls within the 0.5 mV error range. Further, $\beta_3$ corresponds to a table size of 2.2 GB which is certainly within the reach of currently available
workstations.

Figure 7.3: Aggregate error in the interpolated electric field as a function of the number of ions in the channel and the memory footprint.

Figure 7.3 provides results for the aggregate error for electric field as a consequence of interpolation during simulation. It represents the most immediate indicator of simulator accuracy. Results show the aggregate error as a function of both the memory footprint and the number of ions in the channel. As expected, more closely-packed grid points require a larger memory footprint and results in increased interpolation accuracy. This trend is more strongly pronounced when there are more ions in the system. This is because the error is compounded due to the addition of electric field when applying the superposition computation.

The temporal resolution, $\Delta t$, of the simulator has a linear effect on runtime and has no significant effect on memory. As discussed previously, one advantage of the Gunsteren and Berendsen algorithm is that temporal resolution is not limited to the condition $\Delta t \ll 1/\gamma$. The algorithm is instead limited by the average distance traveled per time step which should be small compared to the dimensions of the system.
A series of experiments with different temporal resolutions, $\Delta t = 2\text{fs}, 5\text{fs}, \ldots, 100\text{fs}$, were conducted. The average displacement for this range of resolutions ranged from $0.00393\text{Å}$ for $\Delta t = 2\text{fs}$ to $0.296\text{Å}$ for $\Delta t = 100\text{fs}$ which is an order of magnitude less than the narrowest part of the channel (1.443Å).

For all resolutions, the trajectory of a single ion was logged for 500fs, the corresponding $z$-positions are plotted in Figure 7.4. In this figure, the trajectories of the $\Delta t = 5\text{fs}$ and $\Delta t = 10\text{fs}$ simulations follow that of the $\Delta t = 2\text{fs}$ simulation closely. In comparison, the simulations corresponding to $\Delta t = 50\text{fs}$ and $\Delta t = 100\text{fs}$ exhibit trajectories that appear to diverge from that of $\Delta t = 2\text{fs}$.

For the KcsA channel, the above mentioned results suggest that temporal resolutions of less or equal to $\Delta t = 20\text{fs}$ are appropriate. Similar results can be generated for any channel in order to calibrate and minimize simulation time while maintaining accuracy.

### 7.1.2 Comparing scalability of FEM and BEM Based PEP Calculations

This section compares FEM and BEM based models to illustrate the rationale behind using FEM when modeling the channel as a heterogeneous dielectric medium. To achieve this, both models of the KcsA ion channel were constructed to determine their runtime and storage scalability. For the
case of BEM, the channel was constructed in slices in a similar fashion to [116]. In both cases, the separation of grid elements was $\Delta s = 1 \text{ Å}$ for the vestibules and 0.5 Å for the neck of the channel. The channel was 68.4 Å in length and the top and bottom vestibules had a height of 16 Å and a radius of 33.96 Å. For a homogenous channel, the number of grid points required for the BEM and FEM formulations were $N_{BEM} = 11,887$ and $N_{FEM} = 143,630$, respectively, as provided in Table 7.1, matrix sizes required to solve the BEM and FEM formulations were $S_{BEM} = 141,300,769$ and $S_{FEM} = 143,630$, respectively. Although the problem size is smaller for BEM, the matrix size is significantly larger as they are asymmetric and fully-populated whereas the FEM matrices are sparse and thus scale linearly. In the case of the heterogeneous system, the difference between BEM and FEM is significantly greater. As shown in Table 7.1, the matrix sizes for BEM and FEM are $7.081E+10$ and 143,630, respectively. The size of the FEM problem in heterogeneous systems is the same as for homogenous systems because the 3-D grid required to distinguish between heterogeneous sub-domains is the same 3-D grid already in place for the homogenous formulation.

The results in Table 7.1 were generated under the assumption that the granularity of heterogeneity for the heterogeneous formulation was the same as the 3-D grid used to solve FEM (i.e. $\Delta s = 1$ Å for the vestibules and 0.5 Å for the neck). In Figure 7.5, a plot is provided showing the matrix size of both BEM and FEM formulations as a function of the granularity of heterogeneity. As the granularity is varied from $\Delta s = 16$ Å to 1 Å the matrix size of the BEM formulation grows exponentially from $0.5x10^8$ to $7.0x10^{10}$. For FEM, on the other hand, matrix size remained constant as long as the granularity is equal or larger to that of the 3-D grid used to define the elements. Also shown for reference in this plot are the matrix sizes for BEM and FEM for a homogenous system. It is clear from this plot that FEM formulations are significantly more scalable than BEM formulations.
Figure 7.5: Memory-usage cost as a function of granularity of heterogeneity.

<table>
<thead>
<tr>
<th></th>
<th>BEM</th>
<th>FEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogenous Problem Size (N)</td>
<td>11,887</td>
<td>143,630</td>
</tr>
<tr>
<td>Homogenous Matrix Size (S)</td>
<td>1.41x10^8</td>
<td>143,630</td>
</tr>
<tr>
<td>Heterogeneous Problem Size (N)</td>
<td>266,099</td>
<td>143,630</td>
</tr>
<tr>
<td>Heterogeneous Matrix Size (S)</td>
<td>7.1x10^10</td>
<td>143,630</td>
</tr>
</tbody>
</table>

Table 7.1: Problem size and matrix size results for BEM and FEM formulations of the KcsA channel.

7.2 Performance Improvement of BD Over MD for A Single Point in the Design Space

This section gives a rough comparison of the time required to compute the PEP using MD, using DSE based on conductance as a golden reference, using DSE based on MD-computed PEP as a golden reference, and directly using FGBD that uses an MD-computed dielectric map.

To compare the total time required to compute the PEPs for a set of \( n = (n_1, n_2, \ldots, n_k) \) ions in the system with \( m \) grid points (where there are \( n_1 \) number of ions of type 1, \( n_2 \) number of ions of type 2,
etc.) using MD and using BD, we first compute the total number of possible position configurations for all ions, \(N_{\text{config}}\), using,

\[
N_{\text{config}} = \frac{m!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k! \cdot (m - n_1! - n_2! - \ldots - n_k!)}
\]  
(7.1)

As a result, the time required to compute the PEP in MD for all for a given set of \(n = (n_1, n_2, \ldots, n_k)\) ions, \(t_{MDtotal}\), can be computed using,

\[
t_{MDtotal} = t_{MD} \cdot N_{\text{config}}
\]  
(7.2)

where \(T_{MD}\) is the time required for one PEP computation in MD = 100 hours [133].

This is a significant disadvantage over BD in which ions are modeled as point charges when integrating Langevin equation of motion, and as a result, the superposition principle is utilized in order to reduce the number of times that Poisson’s Equation (which is used to compute the PEP) has to be solved. The electric field can be computed for a system of many ions by adding together the contributions from each ion. The electric field at the position of an ion is the sum of the electric field due to the other ions, the applied field, the surface charges induced by these, and the surface charges induced by that particular ion. As a result, the electric field at every position has to be computed only once:

Time required to compute the PEP in BD for a given set of \(n = (n_1, n_2, \ldots, n_k)\) ions, \(t_{BDtotal}\), can be computed using,

\[
t_{BDtotal} = t_{BD} \cdot m
\]  
(7.3)

where \(t_{BD}\) is the time required for one PEP computation using the FEM solver (see Chapter 3) in BD \(\approx 3.5\) minutes.
Thus, the speed improvement in BD over MD for the PEP computations, $IM$, can be computed using,

$$IM = \frac{t_{MD}/t_{BD}}{(m-1)!/n_1!\cdots/n_k! \cdot (m-n_1-n_2-\cdots-n_k)!}$$  \hspace{1cm} (7.4)$$

For a system with 20 grid points and with one type 1 ions and two type 2 ions residing in the channel, the speed up of BD over MD would be 1454X faster.

The time that it would take the deduce the PEP by performing DSE that uses conductance would be substantially more time consuming than DSE that uses PEP as each DSE iteration includes computation of PEP (more particularly the electric field) and computation of channel conductance.

Time required to compute the PEP for all ions in the system by performing DSE that uses conductance as a golden reference, $DSE_{PEP}$, can be computed using,

$$DSE_{cond} = N_{PEP trials} \cdot (t_{BD total} + t_{cond})$$  \hspace{1cm} (7.5)$$

where $N_{PEP trials}$ is the number of trials of PEP needed to produce the correct conductance, and $t_{cond}$ is the time it takes to compute the conductance.

Time required to compute the PEP for all ions in the system by performing DSE that uses MD-computed a PEP as a golden reference, $DSE_{PEP}$, can be computed using,

$$DSE_{PEP} = N_{DM trials} \cdot t_{BD total}$$  \hspace{1cm} (7.6)$$

where $N_{DM trials}$ is the number of trials of dielectric maps needed to produce the correct PEP.

The time required to compute the PEP for all ions in the system using FGBD that uses MD-computed dielectric map, $t_{PEPDM}$, can be computed using,
\[ t_{PEPDM} = t_{MDDM} \ast t_{BD} \]  \hspace{1cm} (7.7)

where \( t_{MDDM} \) is the time it takes to compute the dielectric map using MD (several hours).

It is important to note that computing conductance requires \( 10^{10} \) simulation iterations regardless of whether this is done using BD or MD. A simulation iteration in MD requires a PEP and one iteration of applying Newton’s law to move the atoms for one time step, about 100 hours [133]. One iteration of simulations of BD is only applying Langevin Equation of motion for one time step and fetching and possibly interpolating the value of PEP/electric field (fraction of a second). Further, applying DSE that uses conductance implies that one iteration of DSE involves \( 10^{10} \) iterations. As a result, using a golden reference PEP provides a much faster scheme for doing DSE.

### 7.3 Runtime Complexity Improvement for DSE

The reduction in the size of the design space, that is - the total amount of points that have to be evaluated - using the DSE reduction methodology in the two DSE studies discussed in Chapters 5 and 6 was substantial. The design space size reduction achieved is dependent on the shape of the design space and as a result will be case dependent. Figure 7.6 shows an example of the reduction of design space size achieved using each of the DSE reduction methods in the case of study 1. The figure illustrates the reduction achieved using each of the DSE reduction methods in terms of the resulting total number of points visited. It compares the number of points that a DSE would have to be evaluated for an exhaustive search of the design space to three cases in which different degrees of design space reduction are used. For this graph, "DSE" represents an exhaustive search of the space, "DSE+SA" includes reductions in the design space possible after sensitivity analysis (SA) was used, "DSE+SA+HC" includes reductions possible after hierarchical clustering (HC) was used, and "DSE+SA+HC+LR" includes reductions possible after a linear-regression (LR) model was used. Two sets of data are presented representing the cases when \( Ri = 5 \) values per dimension are evaluated and \( Ri = 10 \) values per dimension are evaluated.
The results presented in Figure 7.6 demonstrate the significant reduction in the number of points visited during DSE. For this sample DSE study where the dielectric constants of KcsA were found, high levels of reduction were possible due to certain structural properties of the channel. For example, the dielectric values of one region had a very little impact on the potential energy in other regions. This degree of reduction is case dependent, and such significant reduction may not be possible in certain cases.

![Figure 7.6: Impact of DSE methods used on DSE cost in terms of number of points visited.](image)

### 7.4 Overview of FGBD in Comparison to Other Relevant Approaches

Table 7.2 provides a high level comparison of algorithm attributes and performance for FGBD, BD, and MD. The following should be noted:

1. Estimating the dielectric constants for water and protein semi-empirically involves adjusting experimentally-determined values in bulk to match experimental conductance for a molecular system.
2. Adjusting the molecular structure refers to the ability to perform structural changes that are required for example to provide an open channel structure that can be used for simulating conduction. This would be required for a channel that was crystallized in the closed channel state such as the case for the KcsA biological ion channel that was used in the DSE studies discussed in Chapters 5 and 6. Changing the molecular structure is possible in BD and FGBD, however, unlike MD, it would be based on estimations of atomic locations and would lack the atomic level simulations required to consider the interatomic interactions needed to support any structural changes.

3. The last row reports only a rough comparison of the simulated time for 200,000 simulation iterations using Langevin dynamics which corresponds to approximately 10 nanoseconds of simulation time of channel function using molecular systems that contain a similar number of atoms, and using a single processor. This does not give a true comparison as that would require using the same molecular system (here, ion channel) and using identical computational resources.

7.5 Scalability of Design Space Exploration

Poison solvers such as those based on FEM and BEM are both parallelizable algorithms to varying degrees depending on their implementation. Parallelism for such algorithms is principally based on the subdivision of the ion channel into regions each of which is assigned to a different thread of execution. The Gunsteren and Berendsen implementation of Langevin Dynamics can also be parallelized with parallelism applied to per-ion motion calculations. In this thesis, parallelism within both the Poison solver and Langevin Dynamics algorithms were not explored. For these, we assumed a single-thread implementation. For comparisons made in this section, focus was placed on studying the benefits of DSE-level parallelism. DSE-level parallelism was achieved by simulating separate runs of the Poison solver or Langevin Dynamics solver (depending on the type of DSE).
Modeling Resolution

- Protein atoms and ions are modeled explicitly
- Water molecules are modeled implicitly
- Supports a spatially varying dielectric

Model parameters required

- Protein atoms locations and charges
- Membrane potential
- Types and numbers of ions
- Dielectric map

Model parameters computations

Dielectric map can be computed using MD (≈ 10 hours) or determined using DSE (with a golden reference PEP)

Accuracy in the case of TEA blockage of the KcsA channel

- Computed PEP is consistent with the correct PEP

Model sensitivity to change in channel characteristics

- No calibration required: Model can be reused with new values for channel parameters

Ability to adjust molecular structure while considering interatomic forces

- Not possible

CPU hours for one PEP

- 0.05

CPU hours for 10 ns simulated time

- 0.01

Table 7.2: Comparison of algorithm attributes and performance for FGBD, BD, and MD.
simultaneously for different points in the design space in parallel as separate threads on separate cores.

High levels of parallelism are achievable for the DSE optimization methodology proposed in this thesis. As described in Section 4, the first step for optimization is to reduce the size of the design using a number of statistical approaches. After design space reduction, an exhaustive search of the design space is conducted for all remaining design points. Since the methodology calls for all remaining points to be simulated exhaustively, there are no dependencies between the simulation of these points as there would for a heuristic-based algorithms or gradient-based algorithm. Thus, all points in the reduced design space can be run in parallel.

7.6 DSE Throughput using High Performance Computing and Cloud Computing

Recent trends in computing paradigms have focused heavily on high-performance computing for which computationally intensive tasks are distributed amongst many CPUs and nodes. An example of high performance computing that has grown significantly in recent years is cloud computing with Amazon AWS and Microsoft Azure being the two providers currently with the largest market share. To consider the proposed framework in the context of cloud computing, we compare DSE throughput using a reference multi-core workstation. Performance-related attributes of this reference workstation was earlier reported in this thesis in Table 3.2). The reference workstation was powered by a 2010 CPU which was used to conduct research for this thesis because it was the principle workstation made available for research with the Western Canadian Research Grid. This cluster has since been discontinued as of September 2019. This reference workstation was compared against a modern multi-core workstation currently available in 2019 and multi-processor cloud-computing instances. Because cloud computing was not available for this research project, the performance results reported for various computing instances has been estimated by utilizing results using the PassMark CPU benchmark [135]. PassMark software and hardware benchmarks are widely used
(e.g. [136–138]) for evaluating the runtime performance of thousands of computing platforms. When using benchmarking data, we make the following set of assumptions and approximations:

- An evaluation of a single point in the design space (e.g. computing the PEP or estimating channel current) is executed as a single thread occupying a single core.
- An exhaustive search of a design space involves running multiple simulations in parallel as separate threads (see Section 7.5).
- The maximum number of simulations running on a computing instance is equal to the number of cores.
- The level of parallelism for design point evaluation is 100%. This implies that $n$ evaluations can run on an $n$ core computing instance in $1/n$ run time compared to a single core computing instance.

Table 7.3 lists simulation runtime results for the reference workstation in addition to estimated runtime results for a 2019 workstation and four example cloud computing instances currently available on Amazon Elastic Computing (EC2) [6]. For each computing instance, the table includes the CPU-type, number of cores, memory, the single-thread benchmark performance metric according to PassMark CPU benchmarks [135], the single-thread runtime of 100k simulation iterations, the multi-thread runtime of 100k simulation iterations, and the multi-thread speedup over the 2010 workstation. The following equations are used to estimate the figures reported for the runtime performance in table 7.3 and should be considered as rough estimates. An accurate runtime requires running the 100k iterations directly on the computing instances in Table 7.3.

The single-thread runtime of 100k simulation iterations is estimated using:

$$ST_{\text{runtime}}_i \approx ST_{\text{runtime}}_{\text{ref}} \times \frac{Benchmark_{\text{ref}}}{Benchmark_i}$$

The multi-thread (multi-core) runtime of 100k simulation iterations:

$$MT_{\text{runtime}}_i = \frac{ST_{\text{runtime}}_i}{\text{(number of cores)}}.$$
<table>
<thead>
<tr>
<th>Instance Type</th>
<th>CPU type</th>
<th># cores</th>
<th>Memory (GB)</th>
<th>Single Thread Benchmark</th>
<th>ST100k iters (s)</th>
<th>MT100k iters (s)</th>
<th>MT speedup (over 2010 workstation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference workstation</td>
<td>Intel Xeon E5430</td>
<td>4</td>
<td>4</td>
<td>1132</td>
<td>18</td>
<td>4.5</td>
<td>1.0</td>
</tr>
<tr>
<td>2019 workstation</td>
<td>Intel Xeon E5440</td>
<td>4</td>
<td>4</td>
<td>1266</td>
<td>16.1</td>
<td>4.02</td>
<td>1.1</td>
</tr>
<tr>
<td>EC2 c5.2large</td>
<td>Intel Xeon Platinum 8010</td>
<td>4</td>
<td>16</td>
<td>2063</td>
<td>9.9</td>
<td>2.47</td>
<td>1.8</td>
</tr>
<tr>
<td>EC2 c5.9xlarge</td>
<td>Intel Xeon Platinum 8010</td>
<td>18</td>
<td>72</td>
<td>2063</td>
<td>9.9</td>
<td>0.55</td>
<td>8.2</td>
</tr>
<tr>
<td>EC2 c5.12xlarge</td>
<td>Intel Xeon Platinum 8010</td>
<td>24</td>
<td>96</td>
<td>2063</td>
<td>9.9</td>
<td>0.41</td>
<td>10.9</td>
</tr>
<tr>
<td>EC2 c5.18xlarge</td>
<td>Intel Xeon Platinum 8010</td>
<td>36</td>
<td>144</td>
<td>2063</td>
<td>9.9</td>
<td>0.27</td>
<td>16.4</td>
</tr>
<tr>
<td>EC2 c5.24xlarge</td>
<td>Intel Xeon Platinum 8010</td>
<td>48</td>
<td>192</td>
<td>2063</td>
<td>9.9</td>
<td>0.21</td>
<td>21.9</td>
</tr>
</tbody>
</table>

Table 7.3: Estimates of runtime for 100k iterations of a single design point evaluation for the reference workstation (used in the computational cost table in Section 3.3.2), a 2019 workstation, and four example cloud computing instances currently available on Amazon Elastic Computing (EC2) [6]
It is important to note that, unlike general purpose workloads, evaluating a single point in the design space (i.e. evaluating channel behavior using FGBD simulations) may not benefit substantially from per core hardware multi-threading (e.g. Hyper-Threading by Intel). Like many scientific applications, FGBD simulations make significant use of the floating-point unit in the processor and are demanding with respect to memory access. With two threads running on a core that are both taxing on the CPU with respect to memory access and the floating-point unit, the gain from hardware multi-threading is likely not significant. It is for this reason that we enforce a maximum of one thread per core in this thesis.

Several important results can be identified from Table 7.3. For the CPUs listed in the table, improvements in single-thread performance from 2010 to 2019 workstations was about 10%. This performance increase is likely a consequence of architectural innovations as opposed to the improvement in clock-speed because clock-speed has somewhat saturated in this time frame. Further, c5.24xlarge, which is the largest of the Amazon Elastic Computing instances, is about 22x faster than the reference workstation. This increase does not match the performance improvement one might predict using Moore’s law which would estimate a gain of 2x every 18 months translating to 64x over 9 years.

It is difficult to draw too many conclusions from the comparison in Table 7.3 because of the many factors that contribute to workstation performance, however, this data demonstrates a trend towards the increased parallelism and its effectiveness for improving DSE throughput.
Chapter 8

Conclusion

8.1 Problem Addressed

Modeling the way an ion channel works is critical in many applications such as drug development, nano-medicine, the design of artificial nano-pores, and others. And an iterative process through which a particular model constantly gets modified and tested until it is refined and its validity is determined would be very beneficial in these applications. Such an iterative process is a form of Design Space Exploration.

The challenge here is related to the spatial scale of the ion channel system and the temporal resolution at which we need to observe the channel dynamics. The ion channel system is a nanometer in size and permeation has to be resolved at an Angstrom scale resolution. On a temporal scale, conduction (which is the only experimentally observable characteristic that a model can be tested against for validation) occurs at a micro-second time-scale and femtosecond time resolution is required for proper model accuracy. The evaluation of each design point requires billions of iterations. The problem complexity is further exacerbated by the complexity of the ion channel system. An ion channel model must be sufficiently complex to accurately model the membrane bi-layer, protein atoms that form the channel, water molecules inside the channel, and ions that traverse the channel. This multi-aspect complexity of the modeling problem along with the computational cost of each
DSE time-iteration of simulation and the large number of time-iterations of simulation needed to evaluate each point in the design space imply that evaluating every point in the design space using a brute-force approach can be prohibitive depending on the cost of the approach used to evaluate each point in the design space.

There are several common modeling approaches that can be used for evaluating each point in the design space. These approaches range from very detailed and accurate but computationally intensive (such as MD), to more abstract, less costly, and less accurate (such as continuum models). DSE using MD is computationally prohibitive due to MD’s large sensitivity to any variation in the channel design parameters (required for DSE) and the substantial computational demand required for such DSE. Using a computationally intensive modeling technique for each DSE iteration would only compound the problem further. BD is another common modeling approach which improves on cost at the expense of accuracy. In classic BD, an ion channel is represented using a homogeneous dielectric model in which the dielectric value can be “guessed”. This implies that the model ignores important polarization effects which results in a loss of modeling accuracy and the inability to model complex ion channel modeling scenarios.

This research work addresses these issues by proposing a novel BD-based approach, called FBGD, as well as a DSE methodology. FBGD significantly improves on accuracy by capturing local effects by allowing geometric models with a spatially-varying dielectric. The DSE methodology cuts the cost of DSE by reducing the size of the design space and cutting the number of the required DSE iterations.

8.2 Research Summary

This research work explores the use of Design Space Exploration techniques for modeling ion channels. DSE is used to determine a set of design parameters for the system being modeled such that a particular metric is optimized. Due to the computational cost of evaluating the metric for each combination of parameters (i.e. each point in the design space) and the sheer number of possible
parameter combinations (i.e. size of the design space), the problem has a significant computational complexity. For most practical models, a brute-force exhaustive approach is computationally prohibitive. To improve the performance of DSE for ion channels, this thesis proposes techniques to improve the speed by which each parameter combination for a model is evaluated and by reducing the number of evaluations needed to find an optimal parameter combination.

To improve the performance of evaluating each point in the design space, this thesis proposes a new approach to modeling the physical behavior of the ion channel which we call FPGD. This approach does not suffer from the computational complexity of MD-based approaches and has improved accuracy over BD-based approaches. The proposed FPGD approach provides a speed-accuracy trade-off well-suited for DSE while providing a level of accuracy sufficient for exploring function/structure relationships. FPGD also has the additional benefit that it allows one to explore channels that are heterogeneous with respect to their dielectric structure.

To improve the efficiency by which a design space is explored, this thesis proposes a number of techniques for minimizing the design space through statistical inference techniques. Structure in the design space can be used to estimate the values of large sets of design points without the need to simulate them. This allows us to find near-optimal combinations of design parameters without having to perform an exhaustive search. DSE speed is improved by orders of magnitude while finding near optimal solutions. The approach is well-suited for finding a rough estimate of the optimal design parameters.

To demonstrate the validity and efficiency of the proposed DSE methodology, two DSE studies were conducted. The first study investigated the proposed framework’s ability to estimate the correct design parameters. In particular, the ability to estimate the correct dielectric map for the KcsA biological ion channel. There were two significant findings from the study. First, the channel dielectric map computed using the FGBD proposed framework is consistent with that computed directly using Molecular Dynamics with a 7% error margin. Second, a homogeneous dielectric combination leads to a dielectric map consistent with that computed using MD. This raised a question regarding the relevance of modeling the channel as a heterogeneous dielectric and whether
that improves modeling accuracy. The DSE study also demonstrated the effectiveness of the DSE methodology in the case of the blocker-free KcsA channel by showing that the design space was reduced from $R = \prod_{i=1}^{6} R_i$ to $R = 2R_{12} + R_1$, where $R_{12} = \max\{R_1, R_2\}$.

The second DSE study investigated the influence of extracellular blockade of the KcsA channel by the TEA molecule on the dielectric map of the channel. This study is a continuation of an investigation led by Dr. Shin-Ho Chung’s research group at the Australian National University to study TEA binding in KcsA [2]. This work [2] identified the PEP that reproduces the experimentally determined inhibitory constant for TEA. To do this, the MD-computed free energy profile was used to estimate the PEP which was used to compute the conductance. The PEP estimated using the MD free energy profile was altered by systematically dropping the well until the PEP that leads to inhibitory constant that is consistent with experimental data was identified. The DSE study conducted two experiments: in the first experiment, the MD-computed PEP reported in [4] was used as the golden reference (target) PEP to estimate the dielectric map for the TEA-occupied KcsA ion channel. The second experiment used the PEP that reproduces the experimentally determined inhibitory constant for TEA as the golden reference. There were four important findings of this study: First, using a particular dielectric map, the PEP computed using the framework is consistent with PEP computed using MD. In other words, FGBD computations are consistent with MD at a much lower cost. Second, the PEP computed using BD that is based on homogeneous dielectric significantly deviates from the correct PEP indicating that a spatially varying dielectric significantly improves accuracy ($R^2$ drops from %82 to %65 when a homogenous model is used). Third, using a particular dielectric map, FGBD can compute a PEP consistent with the PEP that reproduces experimental results (in this case even the MD estimated PEP did not reproduce experimental data). Finally, the impact of the presence of the TEA molecule on dampening the dielectric was shown: the dielectric map for the TEA-occupied channel is identical to that of the empty channel except in and around the TEA molecule’s location where the dielectric significantly drops (see Table 6.2). Similar to the first DSE study, The DSE study also demonstrated the effectiveness of the DSE methodology in the case of the blocker-occupied KcsA channel by showing that the design space was successfully reduced from $R = \prod_{i=1}^{9} R_i$ to $R = R_1(5 + R_2)$. 

120
One of the most important outcomes of the DSE studies is showing the significance of modeling the channel as a heterogeneous medium for more complex scenarios. The DSE study provided in Chapter 6 confirmed that in more complex scenarios (when the channel is blocked by a drug molecule), the homogeneous BD model is likely to fail. In this study, it was not possible to drop the well in the PEP using a homogeneous dielectric channel model to achieve the well depth required to reproduce the experimentally determined inhibitory constant. The significance of dielectric heterogeneity was further emphasized by showing that it is only possible to drop the depth of the well in the PEP by dividing the upper part of the channel into two different dielectric regions and adjusting their dielectric values separately. The dielectric heterogeneity allowed the ability to estimate design parameters for FGBD that result in improved modeling accuracy that is consistent with MD computations. This is significant because once these design parameters have been estimated, an FGBD channel model that is configured using these parameters can be used to investigate other channel scenarios such as the impact of different conditions on channel behavior. For example, an increase/decrease of the number of ions on the extracellular/intracellular spaces, or an increase/decrease in the membrane potential. An additional use for estimated parameters is to use them as seed values (i.e. a first guess starting point) for conducting DSE studies for similar channels or the same channel but with a different blocker etc.

The scalability and resolution of the proposed framework were also explored in the computational study provided in Chapter 7. It was shown that for the KcsA problem, the scalability of FGBD which uses FEM is significantly better than classic BD for both homogeneous and heterogeneous models. In fact, the memory requirements for classic BD scale exponentially as the number of parameters (degree of heterogeneity) increases but they remain constant for FGBD. Results predicted that the number of iterations possible for a DSE experiment will scale exponentially with time. Further, the computational cost of DSE experiments using various computing platforms including cloud computing instances was discussed. One of the benefits of using the proposed DSE methodology is the ability to completely parallelize the computations of all design points.

These studies provided validation of the FGBD computations and DSE methodology estimations as
well as the computational feasibility of FGBD and DSE methodology. It is important to note that although a significant reduction of the Design Space by utilizing knowledge of channel structure along with the use of statistical inference techniques was possible in the provided studies, this reduction is case dependent. It is possible that there is a large correlation between design parameters such that the design space size reduction is less. As a result, the techniques presented in this thesis should be considered to be a set of tools that can be used on a case-by-case basis for other channel-related DSE problems. For some problems, all tools will provide value and other cases only a subset will work.

The success of these DSE experiments is significant because such DSE would be beneficial in applications such as drug design. For example, using DSE, one can deduce the dielectric map that would reproduce a particular channel behavior. The dielectric dampening effect can then be determined by comparing this dielectric map to the dielectric map of the channel when the drug is not present. This information is valuable in deciding on the drug charge strength and size as that determines the strength of the field around the blocker drug and as a result the dielectric strength.

### 8.3 Limitations

Following are limitations of the approach proposed in this thesis:

1. One limitation that has to be considered is that the FGBD model is only a partial capture of the reality of the ion channel system. This is the nature of modeling, and all models will have this limitation to a varying degree. Sections 1.7 and 2.4 described the trade offs various models make and their potential effectiveness in DSE of ion channels. In this thesis, it was shown that FGBD is an effective approach for DSE and achieves accuracy consistent with MD's. MD is considered to be a high-resolution modeling approach that is significantly more computationally intensive in comparison to FGBD.

2. One of the main challenges of this research work was handling the increased complexity in the ion channel system. This includes increased dielectric heterogeneity resolution, and the
use of channel structures that do not assume symmetric geometry (i.e. geometries generated using the asymmetric geometry generator tool). Although a mechanism was proposed for constructing an asymmetric channel, meshing limitations in FEM software made it not possible to use such a construct for FEM.

3. The DSE methodology used to reduce the size of the design space resulted in only semi-optimal design points. Given that the size of the problem was reduced by orders of magnitude, the loss in accuracy was justified by the significant reduction in run-time. As a consequence, the approach is well suited for fast exploration of the design space. Once a design point has been selected via the proposed approach, this solution could be used as a seed to find further improvements using a more exhaustive approach.

4. There is a significant overhead involved when using the Langevin Simulator developed by Dr. Shin-Ho Chung’s group. A new, more flexible, version of the simulator should be developed in order to conduct future studies.

8.4 Future Work

8.4.1 Introduction of Machine Learning

Many machine learning (ML) based computational methods have recently been developed for predicting ion channel and cellular models (e.g. [139], [140]). In these models, the system is treated as a "black box" where there is no visibility into the mechanisms behind the final output. So, a channel function can be “predicted” given atoms location and charges. Using ML-based algorithms in this manner does not provide visibility into the mechanisms behind channel function. An alternative approach would be utilizing ML-based algorithms with FGBD as well as wetlab testing (physiological experiments) in order to understand the impact of the PEP on channel function:
1. Using simulations along with an ML-based algorithm, predict what channel parameters (including molecular structure) would result in a particular PEP. This gives the Channel Parameters - PEP relationship.

2. Using physiological experiments along with an ML-based algorithm, predict what channel parameters would result in a particular channel function. This gives the Channel Parameters - Channel Function relationship.

3. Using the Channel Parameters - Channel Function and Channel Parameters - PEP relationships as a dataset, train the ML-based algorithm to predict channel function based on the given PEP.

In this way, the ion channel system is no longer a black box and it is possible to examine the PEP characteristics along with channel parameters to examine how parameters impact the mechanisms behind channel function.

The choice of the ML-algorithm and channel structure-function (obtained through physiological experiments) datasets is important. This is because these issues can make using ML-based approaches for modeling ion channels challenging due to the following:

1. Parallelizing combinatorial optimization algorithms is challenging. Very recent advances such as Toshiba’s highly parallelizable combinatorial optimization Simulated Bifurcation algorithm address this issue and hold a great promise for accelerating ion channel modeling. The algorithm’s performance was demonstrated in 2019 [141] showing that a good solution to an optimization problem with 2,000 connected variables (≈ 2 million connections) can be obtained in 0.5 milliseconds (≈ 10 times faster than the laser-based quantum computer recognized as the world’s fastest can solve the same problem). And using a cluster of eight GPUs, a good solution for a large-scale problem with 100,000 connected variables (≈ 5 billion connections) can be obtained in a few seconds.

2. The quality and resolution of datasets as well as their limited number can make the ML-algorithm training challenging. This issue will gradually get addressed as the quality and
resolution of the datasets obtained from the protein databank and through homology modeling continue to be improved, and as more datasets become available.

8.4.2 Correlating Structural Changes and Dielectric Value

It was shown in this thesis that it is possible to manipulate the PEP by adjusting design parameters. In the DSE studies provided in Chapters 5 and 6, these parameters were the dielectric constants corresponding to the various channel regions. It would be valuable to investigate what reducing or increasing the dielectric constant in a particular region implies in terms of structural changes in the channel.

8.4.3 Improving the Resolution of Dielectric Heterogeneity

As FEM solvers improve, it would be interesting to experiment with the resolution of dielectric heterogeneity, and the possibility of “engineering” the PEP curve - that is, can we manipulate smaller parts of the PEP by manipulating certain small dielectric regions? Is it possible to create smaller (higher resolution) wells and barriers within the PEP. This would have a substantial benefit in drug design and development.

8.4.4 Incorporating Asymmetric Channels Models

The impact of channel structure asymmetry on channel behavior can be investigated. Such an investigation can be done once FEM solvers and meshing algorithms are better able to deal with more complex channel structures.
8.4.5 Investigating Drugs

The framework can be used to: investigate the effect of other drugs on the KcsA channel, investigate the effect of drugs on other channels, predict the structure of a channel for which the structure is currently unknown, etc.
Bibliography


[52] Wonpil Im and Charles L. Brooks. Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proceedings of the National Academy


D. A. Doyle, J. Morais Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. The structure of the potassium channel: molecular basis of k+ conduction and selectivity. *Science (New York, N.Y.)*, 280(5360):69–77, Apr 3 1998. LR: 20081121; JID: 0404511; 0 (Bacterial Proteins); 0 (Lipid Bilayers); 0 (Potassium Channel Blockers); 0 (Potassium Channels); 0 (Scorpion Venoms); 0 (prokaryotic potassium channel); 66-40-0 (Tetraethylammonium); 7440-09-7 (Potassium); 7440-17-7 (Rubidium); 7440-23-5 (Sodium); 7440-46-2 (Cesium); 7732-18-5 (Water); CIN: Science. 1998 Apr 3;280(5360):56-7. PMID: 9556453; CIN: Science. 1998 Aug 14;281(5379):883; CIN: Science. 1998 Aug 21;281(5380):1146. PMID: 9735029; ppublish.

Y. Zhou, J. H. Morais-Cabral, A. Kaufman, and R. MacKinnon. Chemistry of ion coordination and hydration revealed by a k+ channel-fab complex at 2.0 a resolution. *Nature*, 414(6859):43–48, Nov 1 2001. LR: 20061115; PDB/1K4C; PDB/1K4D; JID: 0410462; 0 (Antibodies, Monoclonal); 0 (Bacterial Proteins); 0 (Immunoglobulin Fab Fragments); 0 (Potassium Channels); 0 (Solutions); 0 (prokaryotic potassium channel); 7440-09-7 (Potassium); 7732-18-5 (Water); CIN: Nature. 2001 Nov 1;414(6859):23-4. PMID: 11689922; ppublish.


New approaches for bridging computation and experiment on membrane proteins.


[118] Comsol [online].


[124] S. H. Chung, T. W. Allen, and S. Kuyucak. Conducting-state properties of the kcsa potassium channel from molecular and brownian dynamics simulations. *Biophysical journal*, 82 (2):628–645, Feb 2002. LR: 20081121; JID: 0370626; 0 (Bacterial Proteins); 0 (Ions); 0 (Potassium Channels); 0 (prokaryotic potassium channel); 7440-09-7 (Potassium); OID: NLM: PMC1301874; ppublish.


