A STUDY ON COMPUTATIONAL ELECTROMAGNETICS PROBLEMS WITH APPLICATIONS TO CASIMIR FORCE CALCULATIONS

BY

PHILLIP R. ATKINS

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:
Professor Weng Cho Chew, Chair
Professor Narayana Aluru
Professor Jianming Jin
Professor Erhan Kudeki
ABSTRACT

This work covers two main subjects, the efficient calculation of the dyadic Green’s function for layered media (DGLM) and the calculation of the Casimir force using computational electromagnetics (CEM). The calculation of the DGLM is desirable for many problems that involve stratified media, like the cases of the subsurface of the earth and seabed, or circuits involving layers of substrates. The DGLM helps make the solving of these problems more efficient by allowing for the layered inhomogeneity to be incorporated into the Green’s function. In this manner, we no longer need to explicitly model the background inhomogeneity. However, the calculation of the DGLM requires the evaluation of semi-infinite Sommerfeld integrals which prove to be a costly venture. An interpolation scheme is proposed here to allow for more efficient calculations of the Sommerfeld integrals. The interpolation is unique by its reuse of the same interpolation matrix to estimate the function and its derivatives that make up the DGLM.

The Casimir force is an intermolecular force of recent interest due to up and coming micro-electromechanical systems (MEMS) devices. The Casimir force becomes relevant in the behavior of MEMS devices and efficient calculation of the forces would be desirable to aid in their design. Recent methods have been developed that incorporate CEM techniques, notably the electric field integral equation (EFIE). Herein, we present three methods, one of a new devising, that make use of various CEM techniques. Our new method allows for the easy incorporation of different techniques to address problems like low frequency breakdown, arbitrary materials, and domain decomposition methods (DDM). In addition, work on incorporating large scale techniques to increase the size of the problems is discussed.
To my parents and Qi
ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Dr. Weng Cho Chew, for his assistance, insight, and patience over the course of my studies. Thanks are also extended to Jie Xiong, Yuan Liu, Maokun Li, Lin Sun, and Zhiguo Qian for their development and contributions to the Object-Oriented Method of Moments code project and Yat-Hei Lo for his NewMoM codes that helped form the basis for much of our later codes. The EPA and A-EPA codes were developed with help and guidance from Maokun Li, Lin Sun, Zuhui Ma, and Shaoying Huang. I would also like to acknowledge the hospitality and support of the University of Hong Kong where part of this work was conducted and funded. Finally, I would like to thank my parents, Robert and Andrea, and Qi for their encouragement and support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF ABBREVIATIONS</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 INTERPOLATION OF THE DYADIC GREEN’S FUNCTION FOR LAYERED MEDIA</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2.2 The Freespace Problem</td>
<td>5</td>
</tr>
<tr>
<td>2.3 Dyadic Green’s Function for Layered Media</td>
<td>7</td>
</tr>
<tr>
<td>2.4 Tabulation and Interpolation Scheme</td>
<td>24</td>
</tr>
<tr>
<td>2.5 Validation and Results</td>
<td>46</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>49</td>
</tr>
<tr>
<td>CHAPTER 3 CALCULATION OF THE CASIMIR FORCE USING COMPUTATIONAL ELECTROMAGNETICS</td>
<td>51</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>51</td>
</tr>
<tr>
<td>3.2 Stress Tensor Approach</td>
<td>54</td>
</tr>
<tr>
<td>3.3 RRWJ Method</td>
<td>59</td>
</tr>
<tr>
<td>3.4 New Formulation</td>
<td>78</td>
</tr>
<tr>
<td>3.5 Low Frequency Methods</td>
<td>87</td>
</tr>
<tr>
<td>3.6 Large-Scale Methods</td>
<td>95</td>
</tr>
<tr>
<td>3.7 Large-Scale Method Findings</td>
<td>109</td>
</tr>
<tr>
<td>CHAPTER 4 EQUIVALENCE PRINCIPLE ALGORITHM</td>
<td>121</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>121</td>
</tr>
<tr>
<td>4.2 EPA Derivation</td>
<td>124</td>
</tr>
<tr>
<td>4.3 Incorporation into Casimir Force</td>
<td>131</td>
</tr>
<tr>
<td>4.4 Results</td>
<td>135</td>
</tr>
<tr>
<td>4.5 Augmented-Equivalence Principle Algorithm</td>
<td>136</td>
</tr>
<tr>
<td>4.6 Casimir Results Using A-EPA</td>
<td>143</td>
</tr>
<tr>
<td>CHAPTER 5 FUTURE WORK</td>
<td>149</td>
</tr>
<tr>
<td>CHAPTER 6 CONCLUSION</td>
<td>152</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-EFIE</td>
<td>Augmented-Electric Field Integral Equation</td>
</tr>
<tr>
<td>A-EPA</td>
<td>Augmented-Equivalence Principle Algorithm</td>
</tr>
<tr>
<td>CEM</td>
<td>Computational Electromagnetics</td>
</tr>
<tr>
<td>DDM</td>
<td>Domain Decomposition Method</td>
</tr>
<tr>
<td>DGLM</td>
<td>Dyadic Green’s Function for Layered Medium</td>
</tr>
<tr>
<td>ES</td>
<td>Equivalent Surface</td>
</tr>
<tr>
<td>EFIE</td>
<td>Electric Field Integral Equation</td>
</tr>
<tr>
<td>FMA</td>
<td>Fast Multipole Algorithm</td>
</tr>
<tr>
<td>MEMS</td>
<td>Micro-Electromechanical Systems</td>
</tr>
<tr>
<td>LF-FMA</td>
<td>Low Frequency-Fast Multipole Algorithm</td>
</tr>
<tr>
<td>MF-FMA</td>
<td>Mixed Form-Fast Multipole Algorithm</td>
</tr>
<tr>
<td>MLFMA</td>
<td>Multilevel Fast Multipole Algorithm</td>
</tr>
<tr>
<td>MOM</td>
<td>Method of Moments</td>
</tr>
<tr>
<td>MPI</td>
<td>Message Passing Interface</td>
</tr>
<tr>
<td>MVP</td>
<td>Matrix-Vector Product</td>
</tr>
<tr>
<td>PEC</td>
<td>Perfect Electrical Conductor</td>
</tr>
<tr>
<td>PFA</td>
<td>Proximity Force Approximation</td>
</tr>
<tr>
<td>QED</td>
<td>Quantum Electrodynamics</td>
</tr>
<tr>
<td>REGJK</td>
<td>Rahi, Emig, Graham, Jaffe, and Kardar</td>
</tr>
<tr>
<td>RRWJ</td>
<td>Reid, Rodriguez, White, and Johnson</td>
</tr>
<tr>
<td>RWG</td>
<td>Rao-Wilton-Glisson</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------</td>
</tr>
<tr>
<td>TE</td>
<td>Transverse Electric</td>
</tr>
<tr>
<td>TM</td>
<td>Transverse Magnetic</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

This body of work will focus on various computational electromagnetic (CEM) problems and their applications to the dyadic Green’s function for layered media (DGLM) and the Casimir force. The techniques involved in the DGLM problem are the Sommerfeld integrals for layered media and interpolation methods. These are applied to produce an efficient means for modeling the behavior of scatterers on and inside a stratified media. This is an important problem for modeling the response of antennas above the Earth’s surface or for the simulation of circuits built on layers of substrate. Sommerfeld integrals are commonly used in CEM problems due to the fact that they represent the expansion of a spherical wave into the summation of directed plane waves multiplied by directed cylindrical waves. The ability to decompose a spherical wave into its directed wave allows for easy calculation of the source’s response due to layered media. This allows problems with a planarly layered background medium to be reduced in complexity by absorbing the layering behavior into the Green’s function in the electric field integral equation (EFIE).

The EFIE can then be solved using a variety of computational methods like the method of moments (MOM). The downside to the layered media Green’s function is that the Green’s function is costly to calculate due to the fact that it can only be expressed as an infinite integral. The solution for the MOM problem requires the volume or surface integration of the Green’s function, resulting in costly calls to solve the Green’s function at each quadrature point in the volume or surface integration. A proposed solution to decrease the computational time is to pre-tabulate the possible values of the Green’s function for a given problem and use interpolation at each quadrature point in the integrations.

These Sommerfeld integrals are used in the EFIE to relate currents that flow on the surface of conductors with the electric fields that they excite. The unknown currents can
be found given an initial electric field source through the use of integral equation solvers like the Galerkin method, also known as the method of moments. MOM is a specific boundary element method that was first used by Harrington for electromagnetic problems [1]. Part of this thesis will present the interpolation of the Green’s function and its implementation into an EFIE MOM code. The interpolation will be incorporated into a matrix friendly formulation and discussed in detail in Chapter 2. One of the salient features of this implementation is that the derivatives of the Sommerfeld integral are also interpolated from the same tabulated data. This allows for an overall reduction in memory usage and the time used to fill the tabulation matrices.

The second half of the thesis will focus on a different topic, the Casimir force. The Casimir force is a quantum electrodynamic intermolecular force that is quickly becoming of increased interest to theorists and experimentalists. The calculation of the Casimir force is often a difficult process. Early on it was limited to a small class of problems like the attraction between infinite plates [2] and the attraction between dielectric slabs [3, 4]. Extending the calculations to other classes of objects was achieved using the proximity force approximation (PFA) [5]. The PFA, however, is only useful for specific objects like spheres, plates, and cylinders. Two-dimensional problems were solved using methods like the finite-difference [6] but it was not until recently that three-dimensional solvers for arbitrary objects were developed [7–9].

Two methods for calculating the Casimir force using a stress tensor approach [8] and a path integral approach [10] will be reviewed while a novel third method using the argument principle will be presented. This new derivation using the argument principle will be shown to be equivalent to the path integral approach while incorporating new flexibility that the previous path integral approach developed by Reid, Rodriguez, White, and Johnson (RRWJ) lacks. The flexibility will allow for the use of different computational techniques to address various limitations. For example, solving the Casimir force requires an integration over the frequency where the integrand is concentrated in the low frequency spectrum. Numerically solving the problem using traditional CEM techniques is difficult at these low frequencies. The new derivation, however, allows for the implementation of a low frequency method called the augmented-electric field integral equation (A-EFIE) [11]. The
A-EFIE improves the stability of the resulting code at low frequencies.

Another extension to the current capabilities of the Casimir solvers will be large scale techniques. Currently, these algorithms make use of direct methods that have limitations on the number of unknowns that can be solved and thus limit the size and complexity of the objects involved. Using a technique called the Fast Multipole Algorithm (FMA) \[12–14\] allows for problem sizes to go into the order of 1 million unknowns as opposed to 10,000 unknowns using dense matrices. However, the implementation of FMA into these algorithms is not trivial due to various numerical problems that arise and a full demonstration of its use was not achieved.

Finally, we will discuss the use of the domain decomposition method (DDM) for integral equations called the Equivalence Principle Algorithm (EPA) \[15–17\] for the calculation of the Casimir force. The EPA allows for a problem to be divided up into multiple subproblems. The solving of these subproblems and the tying of the results together to solve the original problem can be cheaper in CPU time and memory than the original EFIE or A-EFIE problem for complex or multi-scale objects. This makes it a strong candidate for the simulation of detailed geometries like corrugated plates.

The overarching purpose of this work is to extend the capabilities of current CEM techniques and its applications. The use of the DGLM has been broadened by reducing the computation time required to include it in an algorithm through the use of our interpolation technique. The robustness and functionality of the Casimir force calculations have been extended by our new formulations and incorporation of various CEM techniques.
CHAPTER 2

INTERPOLATION OF THE DYADIC GREEN’S FUNCTION FOR LAYERED MEDIA

2.1 Introduction

Computational electromagnetics (CEM) has evolved a myriad of techniques to be used in order to model the behavior of complex structures. One of the more widely used is the method of moments (MOM) [1]. The formulation for the integral equations used in the MOM varies but centers around the evaluation of the Green’s function. When the modeling problem includes an infinite layered medium, the Green’s function can account for the layering in the form of a dyadic Green’s function for layered media (DGLM). Normally when modelling a system of scatterers, each scatterer is represented by a set of surface or volume basis functions. For a system that has planar layers, like a PCB or a simple subsurface model, then each layer would need to be represented by a set of basis functions in the MOM. This vastly increases the number of unknowns in the resulting matrix problem. However, the DGLM allows for the scattering effects of the layered material to be accounted for in the Green’s function used in the integro-differential operator in the MOM problem. This allows the layering to be effectively removed from the problem so that we no longer need to represent it in the matrix.

However, the evaluation of the DGLM is costly as it involves the computation of Sommerfeld integrals, which are semi-infinite integrations. Various techniques have been used in the past for decreasing the computation time of these integrals. We propose to evaluate the DGLM using interpolation techniques in order to decrease the overall matrix filling time for the MOM. While this technique has been used in many previous multi-layered problems [18–24], previous works have not presented a comprehensive
discussion on the implementation, considerations, and results of interpolation.

This chapter will discuss the specific interpolation scheme used for a matrix-friendly formulation [25], focusing on various considerations needed for obtaining accurate results. Our implementation differs from previous ones in that we have reduced the inner products for the impedance matrix into integrals dependent upon two Sommerfeld integrals and their derivatives. Instead of calculating, storing, and interpolating the terms of the DGLM itself, we will only store these two basic Sommerfeld integrals and interpolate their values and their derivatives and use them to evaluate the DGLM. In this way, we can save on memory costs and reduce the CPU time devoted to generating the tabulation grids for the interpolations. We will also discuss a new singularity-subtraction technique and its application for interpolating the DGLM. In addition, this work will address methods in achieving desired error bounds to achieve error control similar to that when performing quadrature integration.

2.2 The Freespace Problem

The formulation of the MOM is dependent upon the specific problem being considered as the constituent materials of the scatterers affect the derivation and the resulting formulation. Regardless, the formulations still incorporate a Green’s function and as a simple demonstration we will consider the case of perfect electrically conducting (PEC) scatterers. On the surface of a PEC, the total electric field must be zero according to the boundary conditions of Maxwell’s equations. In a homogeneous, isotropic media, the electric field must satisfy [26]

\[ \nabla^2 \mathbf{E}(\mathbf{r}) + k^2 \mathbf{E}(\mathbf{r}) = -i\omega \mu \left[ \mathbf{I} + \nabla \nabla \frac{k^2}{k^2} \right] \cdot \mathbf{J}(\mathbf{r}) \]  

(2.1)

where \( \mathbf{J}(\mathbf{r}) \) is a source current. Taking each vector component on the left-hand side individually, the left-hand is composed of a set of Helmholtz scalar wave equations. Given
the scalar wave equation

\[(\nabla^2 + k^2) \psi(\mathbf{r}) = s(\mathbf{r})\]  \hspace{1cm} (2.2)

the solution can be described by a Green’s function,

\[\psi(\mathbf{r}) = -\int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') s(\mathbf{r}')\]  \hspace{1cm} (2.3)

where the Green’s function is

\[g(\mathbf{r}, \mathbf{r}') = \frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{4\pi |\mathbf{r} - \mathbf{r}'|}\]  \hspace{1cm} (2.4)

Similarly, the electric field can be represented using the same homogeneous Green’s function as

\[\mathbf{E}(\mathbf{r}) = i\omega\mu \int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \left[ \mathbf{I} + \nabla' \nabla' \right] \cdot \mathbf{J}(\mathbf{r}')\]  \hspace{1cm} (2.5)  
\[= i\omega\mu \int d\mathbf{r}' \mathbf{G}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}')\]  \hspace{1cm} (2.6)

where we have simplified the right-hand side by use of the dyadic Green’s function defined as

\[\mathbf{G}(\mathbf{r}, \mathbf{r}') = \left[ \mathbf{I} + \nabla' \nabla' \right] g(\mathbf{r}, \mathbf{r}')\]  \hspace{1cm} (2.7)

In this manner, the electric field can be described using the dyadic Green’s function and the source currents. Assuming PEC scatterers, then on the surface of the scatterers we have

\[-\hat{n} \times \mathbf{E}^s(\mathbf{r}) = \hat{n} \times \mathbf{E}^i(\mathbf{r}) - \hat{n} \times i\omega\mu \int d\mathbf{r}' \mathbf{G}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}')\]  \hspace{1cm} (2.8)

Using the method of moments, this results in a matrix equation:

\[\mathbf{Z} \cdot \mathbf{J} = \mathbf{V}\]  \hspace{1cm} (2.9)
where $\mathbf{Z}$ is the matrix representation of the above dyadic Green’s function and $\mathbf{J}$ and $\mathbf{V}$ are the vector representations of the current $\mathbf{J}(\mathbf{r})$ and exciting fields respectively. The above dyadic Green’s function represents an isotropic, homogeneous background. We can replace this closed form dyadic Green’s function with one that is formulated to represent a layered background material. The construction of the matrix representing the layered problem only differs from the preceeding freespace case in the dyadic Green’s function. The rest of this chapter is focused upon how we can calculate the DGLM quickly so that we can efficiently construct the $\mathbf{Z}$ matrix to solve the scattering problem.

### 2.3 Dyadic Green’s Function for Layered Media

The proposed interpolation technique is specific to the formulation chosen for the DGLM. The DGLM has been derived in a variety of ways but more recently a matrix-friendly formulation was proposed that will be used in our interpolation scheme [25]. The formulation is derived in terms of vector wave functions, starting from the $z$-directed pilot vector potential approach [26]. The basic form of the dyadic Green’s function is

\[
\mathbf{G}(\mathbf{r}, \mathbf{r}') = \frac{i}{8\pi^2} \int \frac{d\mathbf{k}_\rho}{2\mathbf{k}_m^2 k^2_m} \left[ \mathbf{M}(\mathbf{k}_\rho, \mathbf{r}) \mathbf{M}(\mathbf{-k}_\rho, \mathbf{r}') + \mathbf{N}(\mathbf{k}_\rho, \mathbf{r}) \mathbf{N}(\mathbf{-k}_\rho, \mathbf{r}') \right] - \frac{\hat{\mathbf{z}} \hat{\mathbf{z}}}{k^2_m} \delta(\mathbf{r} - \mathbf{r}')
\] (2.10)

For a layered medium whose boundaries lie in the $x - y$ plane, that is, the medium is homogeneous in the $x$ and $y$ dimensions and inhomogeneous in the $z$ dimension, the dyadics $\mathbf{M}(\mathbf{k}_\rho, \mathbf{r}) \mathbf{M}(\mathbf{-k}_\rho, \mathbf{r}')$ and $\mathbf{N}(\mathbf{k}_\rho, \mathbf{r}) \mathbf{N}(\mathbf{-k}_\rho, \mathbf{r}')$ are defined as

\[
\mathbf{M}(\mathbf{k}_\rho, \mathbf{r}) \mathbf{M}(\mathbf{-k}_\rho, \mathbf{r}') = (\nabla \times \hat{\mathbf{z}})(\nabla' \times \hat{\mathbf{z}}) e^{i\mathbf{k}_\rho \cdot (\mathbf{r}_s - \mathbf{r}_s')} F^{TE}(k_m, z, z')
\] (2.11)

\[
\mathbf{N}(\mathbf{k}_\rho, \mathbf{r}) \mathbf{N}(\mathbf{-k}_\rho, \mathbf{r}') = -\left( \frac{\nabla \times \nabla' \times \hat{\mathbf{z}}}{i\omega \varepsilon_n} \right) \left( \frac{\nabla' \times \nabla'' \times \hat{\mathbf{z}}}{i\omega \mu_m} \right) e^{i\mathbf{k}_\rho \cdot (\mathbf{r}_s - \mathbf{r}_s')} F^{TM}(k_m, z, z')
\] (2.12)

where $m$ denotes the layer where the source point lies and $n$ is the layer where the observation point lies. Here, $k_{mz} = \sqrt{k^2_m - k^2_\rho}$ where $\mathbf{k}_\rho = k_x \hat{x} + k_y \hat{y}$ and $\mathbf{r}_s = x\hat{x} + y\hat{y}$. The
vector $\mathbf{r}$ is the position vector of the observation point and the vector $\mathbf{r}'$ is the position vector of the source point. The function $F$ describes the response of a point source potential embedded in the layered medium and will be defined in more detail shortly. The resulting DGLM is formulated as

$$
\mathbf{G}(\mathbf{r}, \mathbf{r}') = (\nabla \times \hat{z}) (\nabla' \times \hat{z}) g^{TE}(\mathbf{r}, \mathbf{r}') + \frac{1}{k_{nm}^2} (\nabla \times \nabla \times \hat{z}) (\nabla' \times \nabla' \times \hat{z}) g^{TM}(\mathbf{r}, \mathbf{r}'),
$$

where $k_{nm}^2 = \omega^2 \epsilon_n \mu_m$. The Green’s function terms, $g^{TE}(\mathbf{r}, \mathbf{r}')$ and $g^{TM}(\mathbf{r}, \mathbf{r}')$, can be expressed in cylindrical coordinates and Bessel functions as

$$
g^{TE}(\mathbf{r}, \mathbf{r}') = \frac{i}{4\pi} \int_0^\infty \frac{dk_\rho}{k_{mz} k_\rho} J_0(k_\rho r_s) F^{TE}(k_\rho, z, z') \quad (2.14)$$

$$
g^{TM}(\mathbf{r}, \mathbf{r}') = \frac{i}{4\pi} \int_0^\infty \frac{dk_\rho}{k_{mz} k_\rho} J_0(k_\rho r_s) F^{TM}(k_\rho, z, z') \quad (2.15)
$$

where $r_s = \sqrt{(x - x')^2 + (y - y')^2}$. In solving a MOM problem, the above DGLM in Equation (2.13) will be applied to the inner product between the testing and basis currents to populate the impedance matrix for the electric field integral equation (EFIE). The resulting impedance matrix problem becomes

$$
Z_{ij} = i \omega \mu_m \langle \mathbf{J}_{Ti}(\mathbf{r}), \mathbf{G}(\mathbf{r}, \mathbf{r}'), \mathbf{J}_j(\mathbf{r}') \rangle = i \omega \mu_m \int d\mathbf{r} \mathbf{J}_{Ti}(\mathbf{r}) \cdot \int d\mathbf{r}' \mathbf{G}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_j(\mathbf{r}') \quad (2.16)
$$

Expanding out the inner product and shifting around the derivatives,

$$
\langle \mathbf{J}_{Ti}(\mathbf{r}), \mathbf{G}(\mathbf{r}, \mathbf{r}'), \mathbf{J}_j(\mathbf{r}') \rangle = \langle \nabla \cdot \mathbf{J}_{Ti}(\mathbf{r}), g_\phi(\mathbf{r}, \mathbf{r}'), \nabla' \cdot \mathbf{J}_j(\mathbf{r}') \rangle
- \langle \mathbf{J}_{Ti}(\mathbf{r}) \cdot \hat{z}, g_{z1}(\mathbf{r}, \mathbf{r}'), \nabla' \cdot \mathbf{J}_j(\mathbf{r}') \rangle - \langle \nabla \cdot \mathbf{J}_{Ti}(\mathbf{r}), g_{z2}(\mathbf{r}, \mathbf{r}'), \hat{z} \cdot \mathbf{J}_j(\mathbf{r}') \rangle
+ \langle \hat{z} \cdot \mathbf{J}_{Ti}(\mathbf{r}), g_{z2}(\mathbf{r}, \mathbf{r}'), \hat{z} \cdot \mathbf{J}_j(\mathbf{r}') \rangle + \langle \mathbf{J}_{TSi}(\mathbf{r}), g_{ss}(\mathbf{r}, \mathbf{r}'), \mathbf{J}_{Sj}(\mathbf{r}') \rangle \quad (2.17)
$$
Thus, the DGLM that is used for the calculation of the impedance matrix for the EFIE can be decomposed into the summation of $g^{TE}$, $g^{TM}$, and their derivatives with respect to $z$ and $z'$. It is therefore possible to only calculate $g^{TE}$ and $g^{TM}$ over a prescribed gridspace and then interpolate the required functions and their derivatives at a given $\mathbf{r}$ and $\mathbf{r}'$.

Normally, the Green’s function terms $g^{TE}$ and $g^{TM}$ are dependent upon three spatial variables, $r_s$, $z$, and $z'$. However, by taking into consideration special cases, $g^{TE}$ and $g^{TM}$ can be made dependent upon one or two variables, allowing for faster and more accurate

![Figure 2.1: Characterization of a source embedded in N layered medium.](image)

...
interpolation. As of now, the function \( F \) has not been defined, but its solution can be determined by examining the response of a point source embedded in a layered media [27]. We will assume that the layered medium is defined as shown in Figure 2.1.

The resulting potential \( F \) is described as

\[
F^+ (z, z') = \left[ e^{iknz} + \tilde{R}_{m,m-1} e^{-ikn(z+2d_{n-1})} \right] A^+_n
\]

\[
A^+_n = \prod_{j=n+1}^{m-1} S^+_{j,j-1} e^{i(k_{jz}(d_{j-1}-d_{j-1}))} e^{i(k_{nz}d_n-k_{nz}d_{m-1})} A^+_m S^+_{m,m-1}
\]

\[
A^+_m = \left[ e^{-ik_{mz}z'} + e^{ik_{mz}(z'+2dm)} \tilde{R}_{m,m+1} \right] \tilde{M}_m
\]

\[
S^+_{i,i-1} = \frac{1 - R_{i,i-1}}{1 - R_{i,i-1} \tilde{R}_{i-1,i-2} e^{2i(k_{i,z}(d_{i-1}-d_{i-2}))}}
\]

\[
F^- (z, z') = \left[ e^{-iknz} + \tilde{R}_{n,m+1} e^{ikn(z+2dn)} \right] A^-_n
\]

\[
A^-_n = \prod_{j=m}^{n-2} S^-_{j,j+1} e^{i(k_{j,z}(d_{j+1}-d_{j}))} e^{i(k_{nz}d_n-k_{nz}d_{n-1})} A^-_m S^-_{m,m+1}
\]

\[
A^-_m = \left[ e^{ik_{mz}z'} + e^{-ik_{mz}(z'+2dm)} \tilde{R}_{m,m-1} \right] \tilde{M}_m
\]

\[
S^-_{i,i+1} = \frac{1 + R_{i,i+1}}{1 - R_{i,i+1} \tilde{R}_{i+1,i+2} e^{2i(k_{i,z}(d_{i+1}-d_{i}))}}
\]

\[
\tilde{M}_m = \left[ 1 - \tilde{R}_{m,m+1} \tilde{R}_{m,m-1} e^{2i(k_{nz}(d_{m}-d_{m-1}))} \right]^{-1}
\]

\[
\tilde{R}_{i,i+1} = \frac{R_{i,i+1} + \tilde{R}_{i+1,i+2} e^{2i(k_{i,z}(d_{i+1}-d_{i}))}}{1 + R_{i,i+1} \tilde{R}_{i+1,i+2} e^{2i(k_{i,z}(d_{i+1}-d_{i}))}}
\]

\[
R^T_{i,i+1} = \frac{\mu_{i+1}k_{i,z} - \mu_i k_{i+1,z}}{\mu_{i+1}k_{i,z} + \mu_i k_{i+1,z}}
\]

\[
R^M_{i,i+1} = \frac{\epsilon_{i+1}k_{i,z} - \epsilon_i k_{i+1,z}}{\epsilon_{i+1}k_{i,z} + \epsilon_i k_{i+1,z}}
\]

Here, \( \tilde{R} \) represents the generalized reflection coefficients while \( R^{TE} \) and \( R^{TM} \) are the Fresnel reflection coefficients for the transverse electric (TE) and transverse magnetic (TM) modes respectively. The function \( F^+ \) is used when the observation point lies in a layer above the source layer, that is \( n < m \), and \( F^- \) is used when \( n > m \). However, when the source and observation points lie in the same layer, \( n = m \), the function \( F \) can be decomposed into the summation of two functions, each dependent upon two independent
variables. In the case where the source and observation points lie in the same plane parallel to the boundaries, when \( z = z' \), the function \( F \) is dependent only upon \( r_s \). When \( n = m \), \( F \) can be decomposed as

\[
F(r_s, z, z') = F_1(r_s, |z - z'|) + F_2(r_s, z + z') \tag{2.35}
\]

\[
F_1(r_s, |z - z'|) = e^{ik_mz|z-z'|} + 2\cos(k_mz - z')e^{2ik_mz(d_m-d_m-1)}\tilde{R}_{m,m-1}\tilde{R}_{m,m+1}M_m \tag{2.36}
\]

\[
F_2(r_s, z + z') = \left[ e^{-ik_mz(z+z'+2d_m-1)}\tilde{R}_{m,m-1} + e^{ik_mz(z+z'+2d_m)}\tilde{R}_{m,m+1} \right] \tilde{M}_m \tag{2.37}
\]

It can be seen that \( F_1 \) contains the direct contribution from the source and this can be ignored since it can be calculated in closed form. We will henceforth refer to \( g \), whether TE or TM, as \( g = g_1 + g_2 \), where \( g_1 \) is the integral containing \( F_1 \) and \( g_2 \) is the integral containing \( F_2 \). Specifically,

\[
g_1(r, r') = \frac{i}{2\pi} \int_0^\infty \frac{dk_{\rho}}{k_mz k_{\rho}} J_0(k_{\rho}r_s) \cos(k_mz|z-z'|) e^{2ik_mz(d_m-d_m-1)}\tilde{R}_{m,m-1}\tilde{R}_{m,m+1}M_m \tag{2.38}
\]

\[
g_2(r, r') = \frac{i}{4\pi} \int_0^\infty \frac{dk_{\rho}}{k_mz k_{\rho}} J_0(k_{\rho}r_s) \left[ e^{-ik_mz(z+z'+2d_m-1)}\tilde{R}_{m,m-1} + e^{ik_mz(z+z'+2d_m)}\tilde{R}_{m,m+1} \right] \tilde{M}_m \tag{2.39}
\]

**Singularity subtraction**

The first difficulty in evaluating the Sommerfeld integrals contained in Equations (2.38) and (2.39) involves the singularity that exists when \( k_{\rho} = 0 \). In each integration, it is desirable to numerically integrate an equation of the form

\[
H(r, r') = \int_0^\infty \frac{dk_{\rho}}{k_mz k_{\rho}} J_0(k_{\rho}r_s) e^{ik_mz f(z,z')} \tilde{R} \tag{2.40}
\]
This can be done using singularity subtraction to remove the pole at the origin by using a function \( h(k_ρ, r_s, f(z, z')) \) which will be evaluated at \( k_ρ = 0 \).

\[
H(r, r') = \int_0^\infty \frac{dk_ρ}{k_ρ} \left[ \frac{1}{k_m} J_0(k_ρ r_s) e^{ik_m z' f(z, z')} \widetilde{R} - h(0, r_s, f(z, z')) \frac{a^ℓ}{k_ρ^ℓ + a^ℓ} \right] \\
+ h(0, r_s, f(z, z')) \int_0^\infty \frac{dk_ρ}{k_ρ} \frac{a^ℓ}{k_ρ^ℓ + a^ℓ} \\
h(0, r_s, f(z, z')) = \frac{1}{k_m} e^{ik_m \cdot f(z, z')} \left. \widetilde{R} \right|_{k_ρ=0}
\]  

(2.41)

When evaluated at \( a = 1 \), the second integral that represents the singularity subtraction contribution becomes

\[
h(0, r_s, f(z, z')) \int_0^\infty \frac{dk_ρ}{k_ρ} \frac{a^ℓ}{k_ρ^ℓ + a^ℓ} = h(0, r_s, f(z, z')) \left[ \frac{1}{ℓ} \ln (k_ρ^ℓ) - \frac{1}{ℓ} \ln (k_ρ^ℓ + a^ℓ) \right] \bigg|_{k_ρ=0} \\
= h(0, r_s, f(z, z')) \left[ \lim_{k_ρ \to 0} \frac{1}{ℓ} \ln \left( \frac{k_ρ^ℓ}{k_ρ^ℓ + a^ℓ} \right) - \ln (a) \right] \\
= -h(0, r_s, f(z, z')) \lim_{k_ρ \to 0} \ln (k_ρ) = +\infty
\]  

(2.42)

The above is true when the factor \( ℓ \) is greater or equal to 1. This factor can be used to control the convergence of the singularity subtraction term in the integrand. A low \( ℓ \) will allow for slow convergence, but since it converges slower it will need fewer quadrature points over a given period to satisfy a specified error bound. Thus, increasing \( ℓ \) will decrease the computation time of the singularity subtraction term’s contribution to the integrand, but at the same time it can decrease accuracy. The ideal choice of \( ℓ \) would be such that the convergence of the numerical integral will no longer improve by increasing \( ℓ \), thus making the original integrand the limiting factor in the convergence. We have found \( ℓ = 4 \) to be a sufficient choice. Although there remains an infinite contribution from the singularity subtraction term, this will be cancelled out in each of the five Green’s function terms in Equations (2.18) through (2.22). This is to be expected since the singularity does not exist in the original integrals in Equation (2.17). The cancellation can be seen...
explicitly in the case of $g_{ss}$.

$$g_{ss}(\mathbf{r}, \mathbf{r}') = (k_n^2 + \partial_z^2) g^{TE}(\mathbf{r}, \mathbf{r}') = (k_n^2 + \partial_z^2) \left( g_s^{TE} - h^{TE} \lim_{\epsilon \to 0} \ln(\epsilon) \right)$$

$$= (k_n^2 + \partial_z^2) g_s^{TE} - k_n h^{TE} \lim_{\epsilon \to 0} \ln(\epsilon) + k_m h^{TE} \lim_{\epsilon \to 0} \ln(\epsilon)$$

$$= (k_n^2 + \partial_z^2) g_s^{TE}$$

(2.44)

where $g_s^{TE}$ is the singularity subtracted version of $g^{TE}$. However, the following relationships can be shown:

$$\tilde{R}_{i,i+1}^{TM}(k_\rho = 0) = -\tilde{R}_{i,i+1}^{TE}(k_\rho = 0)$$

(2.45)

$$\tilde{R}_{i,i-1}^{TM}(k_\rho = 0) = -\tilde{R}_{i,i-1}^{TE}(k_\rho = 0)$$

(2.46)

$$\tilde{M}_m^{TM}(k_\rho = 0) = \tilde{M}_m^{TE}(k_\rho = 0)$$

(2.47)

This means that for the general case,

$$h_1^{TM}(0, r_s, |z - z'|) = h_1^{TE}(0, r_s, |z - z'|)$$

(2.48)

$$h_2^{TM}(0, r_s, z + z') = -h_2^{TE}(0, r_s, z + z')$$

(2.49)

where $h_1$ is the singularity contribution associated with $g_1$ and $h_2$ is associated with $g_2$.

Using the above relations, it can be shown that the closed form singularity subtraction terms will all cancel out in Equations (2.19) through (2.22).

**Quasi-static subtraction**

When evaluating the function $g_2(\mathbf{r}, \mathbf{r}')$, two problems occur when a source and observation point are both located near a boundary. The close proximity of the source point to the boundary gives rise to a quasi-static image term in the neighboring layer which causes poor accuracy in the interpolation and slow convergence of the numeric integral. The poor accuracy in the interpolation is due to the fact that the main contribution of the reflected
fields will come from the quasi-static image, which will behave like a weak singularity. Thus, the Green’s function terms for source and observation points near and along a boundary will be non-analytic, decreasing the accuracy of a polynomial matching interpolation algorithm. Likewise, the convergence of the integral will be slow due to the contribution of a weak singularity in the integrand from the quasi-static image. The function \( g_1 (r, r’) \) does not suffer greatly from these problems because \( g_1 (r, r’) \) accounts for the second order reflections that occur when the source and observation points are embedded between layers. In this case, the quasi-static image is always displaced away from the source and observation points, even when they lie on a boundary. This can be thought of as calculating the resulting fields from the image of an image.

Mathematically, the quasi-static image plaguing \( g_2 \) can be seen from the fact that for a small \( r_s \), and when \( z \) and \( z’ \) are on or near a boundary, the convergence of the integrand largely depends upon \( \frac{1}{k \rho} \tilde{R} \). As \( k \rho \to \infty \), \( \Im \{k \rho \} \to +\infty \) causing the generalized reflection coefficients to approach the constants

\[
\lim_{k \rho \to \infty} \tilde{R}_{m,m+1}^{TE} = R_{m,m+1}^{TE,S} = \frac{\mu_{m+1} - \mu_m}{\mu_{m+1} + \mu_m} \\
\lim_{k \rho \to \infty} \tilde{R}_{m,m+1}^{TM} = R_{m,m+1}^{TM,S} = \frac{\epsilon_{m+1} - \epsilon_m}{\epsilon_{m+1} + \epsilon_m}
\]

Thus, the worst-case convergence of the integrand occurs for source and observation points along the boundary and in close proximity of each other. The convergence rate goes as \( \frac{1}{k \rho} \tilde{R} \) which is approximately \( k^{-2} \rho \) for large \( k \rho \). This makes it desirable to subtract out the quasi-static contribution from the integrand. This problem has been dealt with previously [28]; however, a simple closed form solution for the quasi-static subtraction only arises when we are subtracting out the Sommerfeld integral for a free-space point source. Instead, the typical integral that we wish to evaluate is

\[
g_2 (r_s, z + z’) = \int_0^\infty \frac{dk \rho}{k \rho} \left[ \frac{1}{k \rho} J_0 (k \rho r_s) e^{ik \rho Z} \tilde{R}_{m,m+1} \tilde{M} - h_2 (0, r_s, z + z’) \frac{1}{k^4 \rho + 1} \right]
\]

where \( Z = z + z’ + 2d_m \). Taking the limit of \( k \rho \to +\infty \) gives the ideal quasi-static
The proposed quasi-static subtraction term in Equation (2.54) has the added benefit that

\[
\frac{\partial^2 g_2^{QS}}{\partial Z^2} = i R_{m,m+1}^S \frac{e^{ik_m R}}{R}
\]  

(2.55)

where \( R = \sqrt{r_s^2 + Z^2} \). We will subtract the integrand from Equation (2.54) from the integrand in Equation (2.52) and integrate the resulting equation. The results will be stored for interpolation and after interpolating, we need to add back in the equivalent of Equation (2.54). For the proposed quasi-static term to be useful, it must have a closed form solution that can be easily calculated and added back in after the interpolation. Equation (2.54) has an explicit solution only when \( Z = 0 \) [29].

\[
g_2^{QS} (r_s, 0) = -R_{m,m+1}^S \frac{e^{ik_m r_s}}{k_m}
\]  

(2.56)

The general solution of \( g_2^{QS} \) must be obtained from solving the ordinary differential equation in Equation (2.55). However, there is no closed form solution to the right-hand side of the equation when it is integrated with respect to \( Z \). Still, the differential equation
can be reduced to a first order ordinary differential equation. We will define as follows:

\[ F(Z) = \int_0^\infty dk \rho(k) \rho(k) J_0(k \rho r_s) e^{ikmZ} \]  
\[ F''(Z) = i \frac{e^{ikmR}}{R} \]  
\[ F(0) = -\frac{e^{ikm}}{k_m} \]  

Here, the primes indicate partial derivatives with respect to \( Z \). Equation (2.58) can be reduced to first order by multiplying both sides by \( Z \) and integrating.

\[ \intZF''(Z) = ZF'(Z) - \int F'(Z) dZ \]  
\[ = i \int \frac{Ze^{ikmR}}{R} dZ = \frac{e^{ikmR}}{k_m} + C_1 \]  

Thus, we arrive at the ordinary differential equation

\[ ZF'(Z) - F(Z) = \frac{e^{ikmR}}{k_m} + C_1 \]  

To solve the desired ordinary differential equation

\[ F'(Z) - \frac{1}{Z} F(Z) = \frac{e^{ikmR}}{k_mZ} + \frac{C_1}{Z} \]  

the integrating factor \( \rho(Z) = e^{-\int \frac{1}{Z} dZ} = \frac{1}{Z} \) is used. The integration will be taken with the limits \( a \) and \( Z \). We will later set \( a = 0 \) but for now it is enough to note that \( a \) is an arbitrary constant. The resulting solution is

\[ F(Z) = Z \int_a^Z \frac{e^{ikmR}}{k_mZ^2} dZ + Z \left[ C_1 \left( \frac{1}{a} - \frac{1}{Z} \right) + C_2 \right] \]  

where \( C_2 = \frac{1}{Z} F(Z) |_{Z=a} \). The integral above can be evaluated using integration by parts
resulting in

\[
F(Z) = -\frac{e^{ik_m R}}{k_m} + Z \left[ C_1 \left( \frac{1}{a} - \frac{1}{Z} \right) + i \int_a^Z \frac{e^{ik_m R}}{R} dZ + C_3 \right]
\] (2.65)

where \( C_3 = \left[ \frac{1}{Z} F(Z) + \frac{e^{ik_m R}}{k_m Z} \right]_{Z=a} \). The first order derivative of Equation (2.65) can be used to elucidate the value of the unknown constants.

\[
F'(Z) = C_1 + C_3 + i \int_a^Z \frac{e^{ik_m R}}{R} dZ
\] (2.66)

Evaluating \( F'(a) \) removes the integral in Equation (2.66). It also should be noted that \( F'(0) \) is finite in value due to the fact that the original integral has a closed form solution under these conditions. This requires that \( C_1 = 0 \) for \( F'(0) \) to be finite for arbitrary values of \( a \) and \( r_s \). It should also be noted that \( C_1 = 0 \) is necessary for Equation (2.62) to satisfy the boundary condition set forth in Equation (2.59). Thus, if we allow \( a = 0 \) and evaluate \( F'(0) \), then the result would be \( C_3 \). If the derivative of \( F(Z) \) is taken from its integral form, shown in Equation (2.57), and evaluated at \( Z = 0 \), then a second closed form boundary condition arises [29].

\[
F'(0) = i \int_0^\infty dk_\rho \frac{k_\rho}{k_{mz}} J_0(k_\rho r_s) = \frac{\pi}{2} H_0^{(1)}(k_m r_s) = C_3
\] (2.67)

The solution to Equation (2.62), when \( a = 0 \), is

\[
F(Z) = -\frac{e^{ik_m R}}{k_m} + Z \left[ \frac{\pi}{2} H_0^{(1)}(k_m r_s) + i \int_a^Z \frac{e^{ik_m R}}{R} dZ \right]
\] (2.68)
The final solutions for the quasi-static term are

\[
g_2^{QS}(r_s, Z) = R_{m,m+1}^{s} \left[ -\frac{e^{ik_mR}}{k_m} + Z \left\{ \frac{\pi}{2} H_0^{(1)}(k_mr_s) + i \left( \ln \left( \frac{Z + R}{r_s} \right) + \int_0^Z \frac{e^{ik_mR} - 1}{R} dZ \right) \right\} \right] \tag{2.69}
\]

\[
\frac{\partial}{\partial Z}g_2^{QS}(r_s, Z) = R_{m,m+1}^{s} \left\{ \frac{\pi}{2} H_0^{(1)}(k_mr_s) + i \left[ \ln \left( \frac{Z + R}{r_s} \right) + \int_0^Z \frac{e^{ik_mR} - 1}{R} dZ \right] \right\} \tag{2.70}
\]

\[
\frac{\partial^2}{\partial Z^2}g_2^{QS}(r_s, Z) = iR_{m,m+1}^{s} \frac{e^{ik_mR}}{R} \tag{2.71}
\]

\[
g_2^{QS}(0, Z) = -R_{m,m+1}^{s} \left[ \frac{e^{ik_mR}}{k_m} + iZE_1(-ik_mZ) \right] \tag{2.72}
\]

\[
\frac{\partial}{\partial Z}g_2^{QS}(0, Z) = -iR_{m,m+1}^{s}E_1(-ik_mZ) \tag{2.73}
\]

\[
g_2^{QS}(0, 0) = -\frac{R_{m,m+1}^{s}}{k_m} \tag{2.74}
\]

\[
E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt \tag{2.75}
\]

where \(E_1\) is the exponential integral. Of particular note is the fact that we have manipulated the integral of \(\frac{e^{ik_mR}}{R}\) into the integral of \(\frac{e^{ik_mR} - 1}{R}\). This was done to allow the integral to be computed using numeric integration via Gaussian quadrature quickly and accurately.

When \(r_s = 0\), a singularity arises in the integration of \(\int_0^Z \frac{e^{ik_mR}}{R} dZ\). However, this singularity is cancelled from the singularity arising from the Hankel function. That is,
when \( r_s = 0 \) [30],

\[
g_2^{QS} (0, Z) = R_{m,m+1}^S \left[ -\frac{e^{ik_mZ}}{k_m} + Z \left\{ \frac{\pi}{2} \lim_{r_s \to 0} H^{(1)}_0 (k_mr_s) + i \lim_{r_s \to 0} \left( \ln \left( \frac{Z + R}{r_s} \right) \right) \right\} + \int_0^Z \frac{e^{ik_mR} - 1}{R} dZ \right] \\
= R_{m,m+1}^S \left[ -\frac{e^{ik_mZ}}{k_m} + \left\{ \frac{\pi}{2} \lim_{r_s \to 0} \left( 1 + \frac{2i}{\pi} \left[ \ln \left( \frac{k_mr_s}{2} \right) + \gamma \right] \right) + i \ln (2Z) \right\} \right] \\
= R_{m,m+1}^S \left[ -\frac{e^{ik_mZ}}{k_m} + iZ \left\{ -\frac{i\pi}{2} + \lim_{r_s \to 0} \ln (k_mr_s) - \ln (2) + \gamma + \ln (2) + \ln (Z) \right\} \right] \\
= R_{m,m+1}^S \left[ -\frac{e^{ik_mZ}}{k_m} + iZ \left\{ -\frac{i\pi}{2} + \lim_{r_s \to 0} \ln (k_mr_s) + \gamma - E_1 (-ik_mZ) \right\} \right] \\
= R_{m,m+1}^S \left[ -\frac{e^{ik_mZ}}{k_m} + iZ \left\{ -\frac{i\pi}{2} + \lim_{r_s \to 0} (\ln |k_mr_s| + i \arg (k_mr_s)) - E_1 (-ik_mZ) \right\} \right] \\
= R_{m,m+1}^S \left[ -\frac{e^{ik_mZ}}{k_m} + iZ \left\{ -\frac{i\pi}{2} + \lim_{r_s \to 0} (\ln |k_mr_s| + i \arg (k_mr_s)) - E_1 (-ik_mZ) \right\} \right] \\
= -R_{m,m+1}^S \left[ \frac{e^{ik_mR}}{k_m} + iZE_1 (-ik_mZ) \right]
\]

As such, the special cases in Equations (2.72) through (2.74) need to be used when appropriate. However, under the condition where the source and observation points coincide on a boundary, i.e. \( r_s = Z = 0 \), there will arise a nonessential singularity. This nonessential singularity exists in the first-order and second-order derivatives, namely Equations (2.70) and (2.71). A more technical discussion regarding this can be found in Appendix A.1.

Figures 2.2 through 2.5b demonstrate the improvement in interpolation due to
Figure 2.2: The relative error between the integrated result and derived closed form for $g_2^{QS}$ using Matlab.
Figure 2.3: The absolute error between $g^{TM}$ integrated without the quasi-static subtraction term and with the subtraction taken place and corrected using the derived closed form.
Figure 2.4: The magnitude of $g^{TM}$ calculated by integrating without the quasi-static subtraction for comparison with Figure 2.3.

Figure 2.5: The relative error in interpolation (a) without quasi-static subtraction and (b) with quasi-static subtraction.
Figure 2.6: The behavior of $g^{TM}$ (a) with the quasi-static contribution and (b) without the quasi-static contribution.

quasi-static subtraction and verify the accuracy of the proposed closed form correction.

The relative error bound used for the integration routines was $10^{-8}$ with the integration path being truncated when the relative change to the result was less than $10^{-8}$. Because of the truncation in the integration path, there can be an error floor at $10^{-8}$. The upper and lower layers had a relative permittivity of unity and 16 respectively. The conductivity of the upper and lower layers was 0.0001 S/m and 0.1 S/m respectively. The permeability of the medium was homogeneous. Figure 2.2 compares the results, produced in Matlab, of integrating Equation (2.54) with the results derived from Equations (2.69) and (2.74). The low error confirms the validity of the proposed correction equations. Figure 2.3, with Figure 2.4 for comparison, compares the integration of $g^{TM}$ without the quasi-static subtraction with the integration of $g^{TM}$ with the quasi-static subtraction and using Equations (2.69) and (2.74) to correct for the subtraction. Figures 2.5a and 2.5b show the relative error for $g_2^{TM}$ taken for source and observation points along the boundary of a half-space.

Of particular note is the behavior of $g_2$ near the origin of our tabulation grids, where $r_s = Z = 0$. This is the case where the source and observation points coincide on a boundary. Without any quasi-static subtraction, that is when the quasi-static subtraction term is added back into $g_2$, the function appears decidedly non-analytic when viewed with
varying \( r_s \) along the boundary \((Z = 0)\). This can be seen in Figure 2.6a. The results shown in Figure 2.6b are the values that do not have the quasi-static correction added back in and are the values used for tabulation and interpolation. The behavior of the imaginary part of the \( g^{TM} \) in Figure 2.6a explicitly shows the non-analytic property by the fact that the plot of the imaginary part creates a cusp at \( r_s = 0 \). The function \( g^{TM}_2 \) is symmetric about \( r_s = 0 \) since it is only dependent, in terms of \( r_s \), upon a Bessel function. Thus, the cusp results from the symmetry of the function. However, the cusp is not present in Figure 2.6b suggesting that the resulting interpolation errors will be reduced. Figure 2.5a shows the relative error between the interpolated \( g^{TM}_2 \) along a boundary and the directly integrated result when we do not remove the quasi-static subtraction term from the values of \( g^{TM}_2 \) in the tabulation grid. As expected, there is a large error spike near \( r_s = 0 \) that results from the non-analytic behavior of \( g^{TM}_2 \) in the tabulation grid. Upon using quasi-static subtraction, the adjusted \( g^{TM}_2 \) is compared between the interpolated and directly integrated results in Figure 2.5b. The successful quasi-static subtraction removes the large error spike around \( r_s = 0 \).

### 2.4 Tabulation and Interpolation Scheme

The Green’s function terms, \( g^{TE} \) and \( g^{TM} \), are dependent upon the spatial position of the source and observation points, that is, \( r_s, z \) and \( z' \). Normally, the two functions will require a three-dimensional interpolation and tabulation due to the separation of the three variables. This separation is seen when we take into consideration the function \( F(z, z') \) from Equations (2.23) and (2.27) when used in Equations (2.14) and (2.15). As shown above, however, when the source and observation points are located in the same layer, the interpolation and tabulation can be defined in one or two dimensions by splitting \( g^{TE,TM} \) into the summation of the functions \( g^{TE,TM}_1 \) and \( g^{TE,TM}_2 \) who are dependent upon \((r_s, |z - z'|)\) and \((r_s, z + z')\) respectively.

The proposed tabulation scheme is to precompute and tabulate over a gridspace of all combinations of possible source and observation points. This gridspace will therefore
encompass the physical volume of the scatterer. The algorithm systematically steps through the volume enclosing the scatterer and notes the possible source and observation points. When the source and observation points lie across different layers of inhomegeneity, the values of \( g^{TE,TM} \) will need to be calculated using a three-dimensional interpolation and thus the source and observation points will be marked on a three-dimensional grid. But when the source and observation points lie in the same layer, then they are mapped onto two-dimensional grids for the interpolation of \( g_1^{TE,TM} \) and \( g_2^{TE,TM} \).

It is necessary to keep track of the locations on the two-dimensional grid that have already been marked for tabulation. This is because multiple combinations of source and observation points in the three-dimensional space will map to the same point on the two-dimensional grid. For example, \( g_1 \) is dependent upon \( r_s \) and \(|z - z'|\). As such, the source and observation points that lie in the planes \( z' = a \) and \( z = b \) will result in the same values for \( g_1^{TE,TM} \) as if the source and observation points lay in the planes \( z' = -a \) and \( z = -b \). A simple way to divide up the tabulation grid for the two-dimensional interpolations is to find the minimum and maximum values in the two-dimensional grids and then find the appropriate grid points between these extrema. An example of various extrema can be seen in Figure 2.7. Here, \( r_{s,\text{max}} \) is taken as the largest horizontal displacement between two points on the mesh. When calculating the tabulation grid for the interactions between points only within layer 2, the limits of \( z \) and \( z' \) would be taken from \(-d_2\) to \(-d_1\). After the volume of the scatterer has been traversed, the resulting three-dimensional and two-dimensional grids of the possible source and observation points are stored and the Green’s function terms are calculated at the desired points.

Additionally, there are some problems where the scatterer lies solely in one or more planes parallel to the boundaries of the layers. In these circumstances, one-dimensional grids and interpolation can be used. Each one-dimensional grid will handle the interaction between two possible planes. For example, if one were to model a microstrip line and a stripline, then there will need to be three one-dimensional grids. Shown in Figure 2.8, we assume that we have a three layered medium of air, FR4 and copper cladding. The microstrip and stripline are assumed to be two-dimensional strips and thus are confined to some vertical displacement of \( z_1 \) and \( z_2 \). Thus, one grid will contain all the source and
Figure 2.7: The extents for the values of $r_s$, $z$ and $z'$ used for specifying the tabulation grids.

Figure 2.8: The three one-dimensional grids that contain the interactions of the scatterers in a typical layered medium problem.
observation points for the interaction of currents on the microstrip line itself, another grid will be the interactions within the stripline and the final grid will contain the interactions between the microstrip and stripline. These correspond to the possible combinations of \( z + z' \) and \( |z - z'| \).

It should be noted that since the microstrip lies on the boundary between the air and FR4 layers, it could be considered to lie inside either the air or the FR4 layers. Considering it to reside in the FR4 layer allows us to have three one-dimensional grids due to the dependence upon \( z + z' \) and \( |z - z'| \). However, if the microstrip is considered to be in the air layer, then we have four one-dimensional grids. We will still have grids 1 and 2 as shown in Figure 2.8, but we will need two grids to relate the information of the currents on the microstrip interacting at the stripline and the currents on the stripline interacting at the microstrip. This is because the interactions are no longer contained within the same layer and thus are dependent upon \( z \) and \( z' \) instead of \( z + z' \) and \( |z - z'| \). Further discussion about the one-dimensional interpolation can be found in Appendix A.2.

To the purposes of having a flexible interpolation scheme capable of performing across multiple dimensions, the simple Lagrange interpolation will be sufficient [31]. If \( N \) is the number of data points along a dimension that is passed into the interpolation algorithm, then to interpolate the value \( f(x) \),

\[
\ell_j(x) = \frac{\prod_{k=1, k \neq j}^{N} (x - x_k)}{\prod_{k=1}^{N} (x_j - x_k)}, \quad j = 1, ..., N \tag{2.76}
\]

\[
y_i = f(x_i) \tag{2.77}
\]

\[
f_{N-1}(x) = y_1 \ell_1(x) + y_2 \ell_2(x) + ... + y_N \ell_N(x) \tag{2.78}
\]

The coefficients, calculated using Equation (2.76), can be easily determined for Lagrangian interpolation. This is desirable for our algorithm because the coefficients for the interpolating polynomial need to be recalculated for each new Green’s function call. Extending a one-dimensional interpolation into higher dimensions will be accomplished by performing the one-dimensional interpolation along each successive dimension [32]. Thus, the interpolation coefficients of the higher dimensions are dependent upon the results from
the interpolation of the previous dimensions. For example, we want to interpolate \( f(x, y) \) using the tabulation points \((x_i, y_j)\) where \(i, j = 1, 2, ..., N\). This is done by first interpolating along the \(x\)-dimension to find the points \(f(x, y_i)\) and then interpolating along the \(y\)-dimension using the values of \(f(x, y_i)\) to find \(f(x, y)\). It may be possible to calculate and store ahead of time the interpolation coefficients for finding \(f(x, y_i)\), but the interpolation coefficients for interpolating along the \(y\)-dimension will always be dependent upon the given \(x\). As such, having an interpolating algorithm that has a fast calculation time for the interpolation coefficients is very advantageous.

Langrangian interpolation is also useful because if it is desirable to interpolate the derivatives of the function \(f(x)\), then we can simply calculate the derivative of the interpolation coefficients equation. Doing so gives us

\[
\ell'_j(x) = \sum_{m=1, m \neq j}^N \frac{\prod_{k=1, k \neq j, m}^N (x - x_k)}{\prod_{k=1, k \neq j}^n (x_j - x_k)}, \quad j = 1, ..., N \tag{2.79}
\]

\[
\ell''_j(x) = \sum_{i=1, i \neq j}^N \sum_{m=1, m \neq i, j}^N \frac{\prod_{k=1, k \neq i, j, m}^N (x - x_k)}{\prod_{k=1, k \neq j}^n (x_j - x_k)}, \quad j = 1, ..., N \tag{2.80}
\]

\[
f'_{N-1}(x) = y_1 \ell'_1(x) + y_2 \ell'_2(x) + ... + y_N \ell'_N(x) \tag{2.81}
\]

\[
f''_{N-1}(x) = y_1 \ell''_1(x) + y_2 \ell''_2(x) + ... + y_N \ell''_N(x) \tag{2.82}
\]

Equations (2.79) and (2.80) show that for each order of a derivative, the number of summations in the calculation of the interpolation coefficients increases by an order of \(N - 1\). Thus, it is more expensive to interpolate the higher order derivatives. In addition, each derivative effectively reduces the order of the interpolating polynomial by 1. This can be seen conceptually by the fact that we are taking the derivative of the original interpolating polynomial, which will reduce the order of the polynomial by one. This forces a minimum number of points that need to be passed into our interpolation algorithm to achieve sufficient accuracy for \(f(x)\), \(f'(x)\), and \(f''(x)\). The number of points should be at least four points to allow for a quadratic polynomial being matched for the second-order derivative interpolation. The error for Lagrangian interpolation can be computed using the
Taylor series expansion [33],

\[
\text{Error} = \frac{1}{(N+1)!} f^{(N+1)}(\xi) w(x) \tag{2.83}
\]

\[
w(x) = \prod_{k=1,k\neq j}^{N} (x - x_k) \tag{2.84}
\]

\[
\xi \in [x_1, x_N] \tag{2.85}
\]

The error for the interpolations of the derivatives of \( f \) are

\[
\text{Error'} = \frac{1}{(N+1)!} \left[ f^{(N+2)}(\xi) w(x) + f^{(N+1)}(\xi) w'(x) \right] \tag{2.86}
\]

\[
\text{Error''} = \frac{1}{(N+1)!} \left[ f^{(N+3)}(\xi) w(x) + 2f^{(N+2)}(\xi) w'(x) + f^{(N+1)}(\xi) w''(x) \right] \tag{2.87}
\]

Since we are extending a one-dimensional interpolation into higher dimensions by performing the one-dimensional interpolation along each successive dimension, it is apparent that the time to interpolate the two-dimensional and three-dimensional grids will be \( O(N^2) \) and \( O(N^3) \), respectively. However, the error bounds for the multidimensional interpolations are much more difficult to calculate, but the original error bounds for the one-dimensional interpolation can be indicative of the overall error.

One final consideration needs to be made for the interpolation algorithm. Because we cannot interpolate across the boundaries of the tabulation grids, we run into problems interpolating near the boundaries. This occurs because the data points used for interpolation are no longer spatially centered about the desired interpolated data point. Instead, the data points are shifted due to the impassable boundaries. This will increase the error in the interpolation [33]. To counteract this, we can create a finer mesh for the tabulation grid as the smaller step size will decrease the error in the interpolation.

However, since the smaller step size will also increase the number of points that must be precalculated, we will only provide a finer mesh along the boundaries. Specifically, we will divide the tabulation grids as regular grids according to a desired step size in each dimension. Then we divide up the region between the points along the boundary such that
the number of points between the boundary and the first coarse grid point is equal to half the number of points that are passed into the interpolation algorithm. This improved the error near the boundaries, as shown in Figure 2.9b in comparison to Figure 2.9a.

Figure 2.9: Relative error for $g^{TM}$ for a (a) regular coarse grid and a (b) grid with fine edges.

Figure 2.10: Relative error for $g^{TM}$ for the final tabulation grid.

However, in the region between the second and third coarse grid points, the error was
still much larger than the general error floor. We then reduced the step size by a factor of two for the points between the second and fourth coarse grid points, yielding the results seen in Figure 2.10. In all of the aforementioned error plots, the interpolating polynomial used 6 points in each dimension and the coarse grid spacing was \( \frac{\lambda}{25} \), where \( \lambda \) was the wavelength in the sampling path’s layer. The limits of the tabulation grid were taken as the limits of the sampled path, that is \( 0 \leq r_s, z, z' \leq \lambda \). The geometry is still the half-space used earlier; however, the error is calculated along the line \( r_s = z = z' \) from 0 to \( \lambda \).

![Diagram](image)

**Figure 2.11:** Example of the sampling of the tabulation grid along one dimension. Here, \( N = 6 \).

An example of the sampling of the tabulation grid along a single dimension can be seen in Figure 2.11 where we are assuming that the number of points used for the interpolating polynomial is \( N = 6 \). The vertical crosses represent the coarse grid points, used in Figure 2.9a. The circles were added in the results for Figure 2.9b. The crosses were used in the final relative error plot, which can be seen in Figure 2.10. The resulting error plots for the Green’s function of Equations (2.18) through (2.22) are shown in Figures 2.13a through 2.14b. It should be noted that under the conditions of these simulations, \( g_{z1} = g_{z2} \). It is necessary that the error bars for our interpolation scheme be consistent across various geometries. As such, Figures 2.15a through 2.16a show the error in the interpolation when we go over the same path, \( r_s = z + z' \) from 0 to \( \lambda \), above three layers instead of one. The boundaries are located at 0, \(-\frac{\lambda}{2}\), and \(-\frac{3\lambda}{4}\). The relative permittivities were 5 + 0.001i, 16 + 0.1i, 1, and 3. In comparison to Figures 2.10 through 2.12b, the errors are about the same. The embedded case is shown in Figures 2.16b through 2.17b. The embedded geometry was three layers with relative permittivities of 5.0 + 0.001i, 16 + 0.1i, and 1. The boundaries were located at \( \lambda \) and 0. Again, the errors are consistent with the previous results.

Since the error due to interpolation is fairly consistent across different geometries, it
Figure 2.12: Relative error for (a) $\frac{\partial g^{TM}}{\partial z}$ and (b) $\frac{\partial^2 g^{TM}}{\partial z^2}$ for the final tabulation grid.

Figure 2.13: Relative error for (a) $g_{ss}$ and (b) $g_{zz}$ for the final tabulation grid.
Figure 2.14: Relative error for (a) $g_{z1}$ and $g_{z2}$ and (b) $g_{\phi}$ for the final tabulation grid.

Figure 2.15: Relative error for (a) $g^{TM}$ and (b) $\frac{\partial g^{TM}}{\partial z}$ above multiple layers.
Figure 2.16: Relative error for (a) $\frac{\partial^2 g^{TM}}{\partial z^2}$ above multiple layers and (b) $g^{TM}$ embedded in three layers.

Figure 2.17: Relative error for (a) $\frac{\partial g^{TM}}{\partial z}$ and (b) $\frac{\partial^2 g^{TM}}{\partial z^2}$ embedded in three layers.
would be instructive to be able to quantify the error in terms of the coarse spacing of the tabulation grid and the number of points used in the interpolating polynomial. To this effect we generated error results for the same half-space and path as in Figures 2.9a through 2.10, the only variation being the number of points used in the interpolating polynomial varied from four to twelve inclusive and the step size for the coarse spacing of the tabulation grid varied from $\frac{\lambda}{10}$ to $\frac{\lambda}{40}$ inclusive. We compared various metrics that were used to quantify the error over the entire path into a single value. Mean error, root mean squared error, and the maximum error were used and it was seen that the results generally followed a linear relationship when viewed on a logarithmic scale. A comparison of these three for the value of $g^{TM}$ can be seen in Figures 2.18a through 2.19. The maximum error used was the largest relative error that was at a distance greater than or equal to $\frac{\lambda}{15}$ from the origin. This is done because the values of the self term are always calculated exactly and it is assumed that the finite size of the mesh used to represent the scatterer ensures that there is a minimum distance between adjacent source points.

![Figure 2.18: Comparison of the (a) mean error and the (b) root mean squared error for $g^{TM}$ for varying step sizes and number of points used for interpolation.](image)

It can be seen in the these figures that there exists an error floor at around $10^{-16}$ in Figure 2.18a. This error floor is due to the use of double precision floating point numbers by the code. However, the error floor can also arise due to the error bounds used in the
numeric integration of the reference solution and the tabulation grid points. In these simulations, an adaptive Gaussian quadrature was used with a relative error bound of $10^{-13}$ which truncated the semi-infinite integration after the relative change in the result fell below $10^{-12}$. The linearity seen in the above figures is reduced when we interpolate the derivatives. In addition, since the medium only has an inhomogeneous permittivity, the TE reflections do not have a quasi-static image while the TM reflections do. This means that the TM interpolation will have to deal with the existence of weak singularities near the boundary. The rest of the results for $g^{TM}$ and $g^{TE}$ and their derivatives can be found in Appendix A.2.

It is possible to fit a regression against each of these plots to derive an equation relating the logarithm of the relative error to the logarithm of the step size and the number of points passed in. This regression was applied to only the first six sets of data in varying numbers of points passed in for interpolation. That is, only the results for the number of interpolation points from four to nine were used due to the fact that results for larger number of points were corrupted by the error floor. The resulting coefficients are shown in

Figure 2.19: Comparison of the maximum error for $g^{TM}$ for varying step sizes and number of points used for interpolation.
Table 2.1 for the base 10 logarithm of the mean error. For example,

$$\log_{10} (\text{Mean error of } g^{TM}) = 9.0664 + 8.1208 \log_{10} (h) + 0.5037 \left( \log_{10} (h) \right)^2$$

$$- 1.9385N + 0.0807N^2 + 0.0575N \log_{10} (h) \quad (2.88)$$

where $N$ is the number of points used in the interpolating polynomial and $h$ is the step size of the coarse grid in wavelengths. Examples of the regression surfaces in comparison with the actual results can be seen in Figures 2.20 through Figures 2.22. The equations governing the mean and maximum errors, at a distance greater than $\frac{\lambda}{15}$ from the point $r_s = Z = 0$, can be seen in Tables 2.1 and 2.2 respectively.

### Table 2.1: Coefficients for mean relative error equations.

<table>
<thead>
<tr>
<th>Result</th>
<th>Constant</th>
<th>$\log_{10} (h)$</th>
<th>$\left( \log_{10} (h) \right)^2$</th>
<th>$N$</th>
<th>$N^2$</th>
<th>$N \log_{10} (h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g^{TM}$</td>
<td>9.0664</td>
<td>8.1208</td>
<td>0.5037</td>
<td>-1.9385</td>
<td>0.0807</td>
<td>0.0575</td>
</tr>
<tr>
<td>$\partial g^{TM} / \partial z$</td>
<td>7.6717</td>
<td>6.5442</td>
<td>0.5739</td>
<td>-1.8012</td>
<td>0.0805</td>
<td>-0.0642</td>
</tr>
<tr>
<td>$\partial^2 g^{TM} / \partial z^2$</td>
<td>5.2180</td>
<td>4.1092</td>
<td>0.3956</td>
<td>-1.2858</td>
<td>0.0598</td>
<td>-0.0667</td>
</tr>
<tr>
<td>$g^{TE}$</td>
<td>21.8318</td>
<td>12.6452</td>
<td>1.1776</td>
<td>-4.2585</td>
<td>0.2078</td>
<td>-0.1301</td>
</tr>
<tr>
<td>$\partial g^{TE} / \partial z$</td>
<td>19.7841</td>
<td>13.0643</td>
<td>1.9575</td>
<td>-3.9156</td>
<td>0.1986</td>
<td>-0.1520</td>
</tr>
<tr>
<td>$\partial^2 g^{TE} / \partial z^2$</td>
<td>25.3081</td>
<td>16.1827</td>
<td>3.0583</td>
<td>-4.9609</td>
<td>0.2395</td>
<td>-0.5986</td>
</tr>
</tbody>
</table>

### Table 2.2: Coefficients for maximum relative error equations.

<table>
<thead>
<tr>
<th>Result</th>
<th>Constant</th>
<th>$\log_{10} (h)$</th>
<th>$\left( \log_{10} (h) \right)^2$</th>
<th>$N$</th>
<th>$N^2$</th>
<th>$N \log_{10} (h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g^{TM}$</td>
<td>9.7750</td>
<td>11.0969</td>
<td>2.5711</td>
<td>-0.7652</td>
<td>0.0151</td>
<td>0.0976</td>
</tr>
<tr>
<td>$\partial g^{TM} / \partial z$</td>
<td>7.3597</td>
<td>6.8865</td>
<td>1.5875</td>
<td>-0.7194</td>
<td>0.0265</td>
<td>0.2191</td>
</tr>
<tr>
<td>$\partial^2 g^{TM} / \partial z^2$</td>
<td>3.6546</td>
<td>2.1369</td>
<td>0.7370</td>
<td>0.2926</td>
<td>0.0247</td>
<td>0.5193</td>
</tr>
<tr>
<td>$g^{TE}$</td>
<td>13.2294</td>
<td>6.7109</td>
<td>0.3312</td>
<td>-2.6490</td>
<td>0.1258</td>
<td>-0.2562</td>
</tr>
<tr>
<td>$\partial g^{TE} / \partial z$</td>
<td>11.1359</td>
<td>7.2610</td>
<td>1.1841</td>
<td>-2.0356</td>
<td>0.1027</td>
<td>-0.0803</td>
</tr>
<tr>
<td>$\partial^2 g^{TE} / \partial z^2$</td>
<td>8.4114</td>
<td>5.3642</td>
<td>1.9309</td>
<td>-0.9942</td>
<td>0.0620</td>
<td>0.5526</td>
</tr>
</tbody>
</table>
Figure 2.20: Comparison of the regression error equations from Table 2.1 in comparison with the actual results for $g^{TM}$. 
Figure 2.21: Comparison of the regression error equations from Table 2.1 in comparison with the actual results for $\frac{\partial g^{TE}}{\partial z}$.
Figure 2.22: Comparison of the regression error equations from Table 2.1 in comparison with the actual results for $\frac{\partial^2 \phi^{r.e}}{\partial z^2}$.
The above results are useful in characterizing the error that is introduced by the interpolation and the various techniques used for proper error control. However, we are mainly concerned with Equations (2.18) through (2.22). The errors here are expected to increase due to the effects of cancellation. The error plots can be found in Appendix A.2. Tables 2.3 and 2.4 contain the coefficients for the error equations while Figures 2.23 through 2.26 demonstrate examples of the regression surfaces.

Table 2.3: Coefficients for mean relative error equations.

<table>
<thead>
<tr>
<th>Result</th>
<th>Constant</th>
<th>$\log_{10} (h)$</th>
<th>$(\log_{10} (h))^2$</th>
<th>$N$</th>
<th>$N^2$</th>
<th>$N \log_{10} (h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{ss}$</td>
<td>11.1303</td>
<td>6.6743</td>
<td>0.8040</td>
<td>-2.2849</td>
<td>0.1099</td>
<td>-0.2492</td>
</tr>
<tr>
<td>$g_{z1}$</td>
<td>13.5300</td>
<td>8.7085</td>
<td>0.8417</td>
<td>-2.6576</td>
<td>0.1293</td>
<td>-0.0198</td>
</tr>
<tr>
<td>$g_{z2}$</td>
<td>13.5300</td>
<td>8.7085</td>
<td>0.8417</td>
<td>-2.6576</td>
<td>0.1293</td>
<td>-0.0198</td>
</tr>
<tr>
<td>$g_{zz}$</td>
<td>14.0925</td>
<td>8.2678</td>
<td>1.1376</td>
<td>-2.6985</td>
<td>0.1295</td>
<td>-0.3326</td>
</tr>
<tr>
<td>$g_{\phi}$</td>
<td>18.1498</td>
<td>9.6382</td>
<td>1.2828</td>
<td>-3.8057</td>
<td>1.2828</td>
<td>-0.5117</td>
</tr>
</tbody>
</table>

Table 2.4: Coefficients for maximum relative error equations.

<table>
<thead>
<tr>
<th>Result</th>
<th>Constant</th>
<th>$\log_{10} (h)$</th>
<th>$(\log_{10} (h))^2$</th>
<th>$N$</th>
<th>$N^2$</th>
<th>$N \log_{10} (h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{ss}$</td>
<td>6.4190</td>
<td>3.1710</td>
<td>1.1009</td>
<td>-0.7637</td>
<td>0.0529</td>
<td>0.5028</td>
</tr>
<tr>
<td>$g_{z1}$</td>
<td>10.6975</td>
<td>9.2530</td>
<td>2.1805</td>
<td>-1.3074</td>
<td>0.0548</td>
<td>0.0001</td>
</tr>
<tr>
<td>$g_{z2}$</td>
<td>10.6975</td>
<td>9.2530</td>
<td>2.1805</td>
<td>-1.3074</td>
<td>0.0548</td>
<td>0.0001</td>
</tr>
<tr>
<td>$g_{zz}$</td>
<td>7.2949</td>
<td>3.4817</td>
<td>1.1916</td>
<td>-0.8649</td>
<td>0.0584</td>
<td>0.5019</td>
</tr>
<tr>
<td>$g_{\phi}$</td>
<td>3.5682</td>
<td>0.4251</td>
<td>0.6388</td>
<td>-0.2126</td>
<td>0.0317</td>
<td>0.8551</td>
</tr>
</tbody>
</table>
Figure 2.23: Comparison of the regression error equations from Table 2.3 in comparison with the actual results for $g_{ss}$. 
Figure 2.24: Comparison of the regression error equations from Table 2.3 in comparison with the actual results for $g_{zz}$. 
Figure 2.25: Comparison of the regression error equations from Table 2.3 in comparison with the actual results for $g_{z1}$ and $g_{z2}$. 
Figure 2.26: Comparison of the regression error equations from Table 2.3 in comparison with the actual results for $g_\phi$. 
2.5 Validation and Results

The validity of the original DGLM has been previously presented in [25] and the interpolation and tabulation method was checked against the results of directly integrating the DGLM as in [25]. A comparison was made in the bistatic radar cross section (RCS) produced by the scattering off a PEC sphere. The incident wave is a plane wave operating at 60 MHz incident at an elevation angle of $30^\circ$ and an azimuth angle of $0^\circ$. The layers all have a permeability of free space. Three sets of data were compiled and compared with the PEC sphere being located in different situations. The first situation was with the sphere located 2.5 m above a half space with relative permittivities of $\epsilon_1 = 1.0 + i0.0001$ and $\epsilon_2 = 16.0 + i0.1$. The second case has the sphere embedded in the middle of a three layer medium with relative permittivities of $\epsilon_1 = 4.0 + i0.1$, $\epsilon_2 = 16.0 + i0.01$, and $\epsilon_3 = 1.0 + i0.0001$. The top layer’s boundary lies 5 m above the sphere, while the boundary between the second and third layers lies 2.5 m below the sphere. Finally, the sphere is placed in the bottom layer of a three layer geometry. The relative permittivities are the same for the embedded sphere case but the boundary between the second and third layers lies 2.5 m above the sphere.

In all the following simulations, the direct integration results were calculated using an adaptive Simpson’s integration routine with an absolute error bound of $10^{-4}$ and a threshold of $10^{-5}$. The threshold is the relative change in the integral at which the semi-infinite integration ceases. The interpolation results used an adaptive quadrature routine from QUADPACK [34] with a relative and absolute error bound of $10^{-6}$ and a threshold of $10^{-8}$. The tabulation grids used 6 points for the interpolating polynomials and had a coarse grid spacing of $\frac{\lambda}{25}$.

The results can be seen in Tables 2.5 and 2.6. The reported times in Table 2.6 are the times taken to fill the impedance matrix. The results show that for the cases where the scatterer is confined to the top and bottom layer, the speedup is appreciable, on the order of a factor of 10, and the time needed to account for an added layer is low. Placing the sphere in an embedded layer greatly increases the interpolation and tabulation time. This is because an additional set of integrals, $g_1$, needed to be tabulated and interpolated.
Table 2.5: RCS results of various simulations.

<table>
<thead>
<tr>
<th>Case</th>
<th>Direct Result ($\phi, \theta$)</th>
<th>Interpolation Result ($\phi, \theta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above</td>
<td>(-0.279545, -53.4252) dBsm</td>
<td>(-0.278645, -53.4188) dBsm</td>
</tr>
<tr>
<td>Embedded</td>
<td>(0.121033, -49.5591) dBsm</td>
<td>(0.119443, -49.5719) dBsm</td>
</tr>
<tr>
<td>Below</td>
<td>(-9.30028, -74.4569) dBsm</td>
<td>(-9.30134, -74.4824) dBsm</td>
</tr>
</tbody>
</table>

Table 2.6: Comparison of runtimes for various simulations.

<table>
<thead>
<tr>
<th>Case</th>
<th>Direct Time</th>
<th>Interpolation Time</th>
<th>Tabulation Fill Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above</td>
<td>22min 14.07sec</td>
<td>2min 5.96sec</td>
<td>0min 18.83sec</td>
</tr>
<tr>
<td>Embedded</td>
<td>70min 53.72sec</td>
<td>4min 50.28sec</td>
<td>5min 52.14sec</td>
</tr>
<tr>
<td>Below</td>
<td>30min 1.40sec</td>
<td>2min 23.34sec</td>
<td>0min 22.22sec</td>
</tr>
</tbody>
</table>

Another set of tests were run to verify the results for planar structures. These planar structures will allow the code to use one-dimensional interpolations. The scatterer is a circular patch with a radius of 0.65 cm. The patch is placed at the upper boundary of a three layer system. The top layer is air, followed by a dielectric with a relative permittivity of 2.33, finally followed by a PEC ground plate. The thickness of the dielectric layer is 0.07874 cm. The resulting RCS bistatic measurements, using the same parameters as in the previous set of results, compared between the results from direct integration and interpolation can be seen in Figure 2.27a. The relative error, normalized to the direct integration results, can be seen in Figure 2.27b. The errors are consistently on the order of $10^{-3}$ or lower. The total runtime to generate the eleven RCS calculations using the interpolation was 2 minutes and 24.75 seconds. Of this, 96.63 seconds were spent interpolating and 26.39 seconds were used to generate the tabulation grids. This is compared to a total time of 666 minutes and 33.75 seconds using the direct integration.

Thus, we have a speedup of 276.3 or 831.1, including or excluding the time used to generate the tabulation grids respectively. The advantages of the interpolation are apparent even when we need to generate the tabulation grids for each simulation. However, if one were to generate the grids and store them ahead of time, then the advantages improve greatly. This would be useful, for example, in optimizing a design. If one were running a parametric optimization of the circular patch antenna, then being able to reuse the same tabulation grid would allow a speedup on the order of 100 instead of 10. In addition, the
error results show that the interpolation is able to control the error independent of the frequency due to the fact that the tabulation grids are spaced in terms of wavelengths.

A second test was run using two circular patches of radius of 0.65 cm in the same three layered geometry as before. Here, the second patch was translated horizontally by 0.70 cm and vertically by 0.07874 cm. The two patches overlap to some degree, lying in the same layer, but having different vertical displacements. Thus, the interpolation code will create three sets of tabulation grids and use one-dimensional interpolations. The results are shown in Tables 2.7 and 2.8. Once again the interpolation results compare very favorably, with the direct integration but have an appreciable speedup. The total runtime for the two simulations was 337 minutes and 23.63 seconds for the direct integration and 4 minutes and 53.81 seconds for the interpolation. This results in a speedup of 68.9 but when only comparing the impedance matrix fill time the speedup is 266.1.

Table 2.7: RCS results of the two circular patches.

<table>
<thead>
<tr>
<th>Frequency (GHz)</th>
<th>Direct Result $(\phi, \theta)$</th>
<th>Interpolation Result $(\phi, \theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>(-43.0374, -111.600) dBsm</td>
<td>(-43.0621, -111.726) dBsm</td>
</tr>
<tr>
<td>15.0</td>
<td>(-46.1329, -89.2315) dBsm</td>
<td>(-46.226, -89.0889) dBsm</td>
</tr>
</tbody>
</table>
Table 2.8: Comparison of runtimes for the two circular patches.

<table>
<thead>
<tr>
<th>Frequency (GHz)</th>
<th>Direct Time</th>
<th>Interpolation Time</th>
<th>Tabulation Fill Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>191min 32.63sec</td>
<td>34.93sec</td>
<td>1min 42.17sec</td>
</tr>
<tr>
<td>15.0</td>
<td>145min 19.64sec</td>
<td>41.03sec</td>
<td>1min 40.52sec</td>
</tr>
</tbody>
</table>

2.6 Conclusion

A new interpolation technique for the DGLM has been presented in this work. While interpolation has been used previously in many layered medium MOM problems, we have given a detailed report on the limitations of interpolation and techniques to address these limitations and improve error control. The DGLM used has five Green’s function terms, shown in (2.18) through (2.22), that need to be evaluated for each inner product for each impedance matrix entry. We were able, at the cost of only one additional interpolation, to only require the storage and evaluation of two integrals instead of five. This comes at the cost of having to increase the density of the tabulation grids in order to account for the decreased accuracy of interpolating the derivatives and the effects of cancellation. In addition, this required us to formulate a new quasi-static subtraction technique. Despite this, our formulation allows us to have a predictable matrix filling time that is much faster than traditional integration. The ability to predict and control the error in the five Green’s function terms allows the user to control the accuracy of the DGLM in a manner similar to the error controls of a traditional adaptive integration routine.

The speedup obtained by the interpolation varies due to the fact that the integration time is dependent upon the geometry of the background layers and the scatterer. However, we easily see a vast improvement for using one-dimensional interpolation in planar structures and a modest improvement using two-dimensional interpolation for structures confined to a single layer. Because of the fact that the CPU time for interpolation increases as a factor of the dimensionality, the improvement for using three-dimensional interpolation will be much less. Future work may explore ways of incorporating three-dimensional interpolation for problems spanning multiple layers. In this regard, it may be more feasible to forgo memory concerns in favor of speed by storing and interpolating the five Green’s function terms directly. This would increase the memory.
needed for the tabulation grids but reduce the number of interpolations from six to five and allow for a sparser grid and fewer points needed for the interpolating polynomial. This can be done independently of the impedance matrix entries for interactions within the same layer which can be handled as discussed in this chapter.
CHAPTER 3

CALCULATION OF THE CASIMIR FORCE USING COMPUTATIONAL ELECTROMAGNETICS

3.1 Introduction

The Casimir force is an extension to the traditional intermolecular London-van der Waals forces which describe the intermolecular attraction between molecules due to their electrical moments and polarizabilities. The two classical van der Waals forces are the orientation effect and the induction effect. The orientation effect, derived by Keesom [35,36], is the force that arises due to the dependence of a molecule’s permanent electric dipole field upon its orientation with respect to the permanent moments of its neighbors. With two molecules both having dipole moments, it is easy to see that, due to the orientation of the dipoles, the energy of the dipole system is dependent upon the relative positions of the dipoles. Normally this would result in no net force since one would average the force over all positions. However, Boltzmann statistics give statistical preference to positions that represent lower energies at finite temperatures. Averaging the force or energy over all positions weighted by Boltzmann statistics gives rise to a net force that disappears as the temperature increases.

The second classical force is called the induction effect, developed by Debye [37] and Falkenhagen [38]. The induction effect comes into play when one accounts for the fact that an electric field can induce a dipole moment in another molecule. Calculating the forces between two molecules due to the induced dipole moments from the other’s dipole field gives the induction effect. These classical intermolecular forces require that some of the constituent molecules have permanent electric moments. Quantum mechanical theory gave physicists the ability to estimate the permanent moments in molecules and found that
some molecules do not have strong enough moments to account for their intermolecular forces. London solved this problem with a third force called the dispersion effect, named for its similarity to the quantum mechanical theory of light dispersion, or the London force [39]. London, using the newly developed quantum mechanics, proposed that molecules had fluctuating dipole moments modeled after the quantum harmonic oscillator. This oscillating dipole moment still has an average moment of zero but non-zero fluctuations even at the ground state. Calculating the forces between molecules due to the presence and inductive effects of these fluctuating dipole moments gave a result that agreed well with contemporary experimental results.

However, later experiments showed that the London force did not accurately predict the intermolecular forces for molecules at large separations. It was at this point that Casimir and Polder looked into using the new ideas of quantum electrodynamics (QED) to solve the problem. In 1948, Casimir and Polder published a paper detailing a correction to the London-van der Waals force [2] in which they rederived the London-van der Waals forces by taking into account the finite speed of light and the quantization of the electromagnetic field. This differed from previous derivations as the London force did not consider the effects of retardation, allowing for perfect correlation between a fluctuating dipole and its induction. In doing so, Casimir and Polder reconciled the theoretical force calculation with observed results at greater distances. One of the consequences of the new QED theory was that Casimir found that molecules placed near a perfect electrically conducting (PEC) sheet would be attracted to the sheet. Out of this paper grew a short paper published by Casimir that same year that showed that this corrected force could also be derived by looking at the perturbation of the quantum vacuum energy [40]. Renewed interest in the Casimir force came about in the 1990s when it became possible to measure the force accurately in the laboratory [41]. Previously a phenomenon relegated to scientific curiosity, the Casimir force has recently become a factor in the design of micro-electromechanical systems (MEMS) devices [42]. As such, there has been an increased desire to model the Casimir force of arbitrary systems in order to facilitate the design and modeling of MEMS devices.

Those who are unfamiliar with the Casimir force can refer to the Appendix B for more
introductory details. The purpose of the appendices is to demonstrate that the Casimir force can be interpreted as resulting from the fluctuating fields of the photon vacuum state. These fluctuating fields can be thought of as conforming to the macroscopic boundary conditions. As such, the presence of scatterers in the vacuum state means that the energy density of the vacuum state must change due to the boundary conditions. Modes that existed before may no longer be allowed and the spatial dependence upon the relative position of the scatterers in the energy density gives rise to the Casimir force.

There have been a number of methods for modeling the Casimir force. The most straightforward manner, theoretically, is to find the allowed modes that are a function of the relative separation of the scatterers in question \[40,43–47\]. This was done by finding the eigenfrequencies of the problem and calculating the energy and force due to the zero-point energies of each mode. This method is useful for closed form calculations, like in the van Kampen method \[4\], but it is difficult to solve for numerically though a few finite-difference methods have been developed \[6\]. Another approach is to integrate over the mean energy density or force density by finding the field correlations for the stress tensor. This is done using the classical electromagnetic Green’s functions in a relation with the field correlations derived using the fluctuation-dissipation theorem \[3\]. Our research group has previously developed a numerical method using this approach that we will discuss briefly \[8\]. Finally, there is the approach using path integrals to express the energy of the system as constrained by the boundary conditions. This was recently the focus of one group and inspired a new method using traditional method of moments (MOM) electromagnetic solvers \[9,10\] which will be referred to as the Reid, Rodriguez, White, and Johnson (RRWJ) method.

The overall focus of this work is to find ways of applying our knowledge of computational electromagnetic (CEM) solvers to extend the performance and problem spectrum of existing methods. The following pages will cover the stress tensor method developed by our group \[8\] and the path integral formulation developed by RRWJ \[9,10\]. The functionality and robustness of the stress tensor code was improved and means of incorporating large-scale methods \[12\] into the algorithm were investigated. In the RRWJ formulation, a new way of deriving the algorithm has been found that allows for the inclusion of a greater
range of MOM techniques and work continues to incorporate large-scale methods. The new derivation allows for the use of low frequency methods [11] and easily solves different classes of materials like dielectrics [48–51]. It is our hope to extend the recently developed techniques for dielectrics to use the Equivalent Principle Algorithm (EPA) [52].

3.2 Stress Tensor Approach

Previous work involving traditional computational electromagnetic techniques was done by Xiong, Tong, Atkins, and Chew [8] that focused around the use of the stress tensor approach. The method is based on the use of the fluctuation-dissipation theorem. The fluctuation-dissipation theorem grew out of the study of topics like Brownian motion [53] and the Johnson-Nyquist noise [54,55]. The Johnson-Nyquist noise describes the phenomenon that occurs when a resistor, or any dissipative circuit element, is in thermal equilibrium. The resistor experiences a noise voltage with a mean square spectral density of $4k_B T R V^2/\text{Hz}$. This noise arises due to the thermal oscillations of the electrons within the resistor. The oscillating electrons induce a voltage across the resistor and give up their energy back to the environment’s heat bath. In this way, a lossy system in equilibrium gives rise to a fluctuating quantity while conserving energy. The magnitude of these fluctuations is related to the loss of the system and this relationship is described in general by the fluctuation-dissipation theorem. The generalized fluctuation-dissipation theorem arose by deriving a theorem of the Nyquist noise for generalized forces in linear dissipative systems [56]. A more in-depth discussion of the fluctuation-dissipation theorem can be found in a review paper by Kubo [53]. For our purposes, the fluctuation-dissipation theorem can be applied to the electromagnetic fields. The correlation of the fluctuating fields can be related to the loss of the system via the dyadic Green’s function [57]. The resulting Casimir method starts by noting that the net force acting on a surface $S$ is given by the surface integral of the mean stress tensor [58].

$$\mathbf{F} = \oint_S \langle \mathbf{T}(\mathbf{r}') \rangle \cdot d\mathbf{s}'$$  \hspace{1cm} (3.1)
where $\mathbf{T}$ is the Maxwell stress tensor. Two tangent vectors, $\hat{u}$ and $\hat{v}$, are defined on an arbitrary surface and a normal vector, $\hat{n}$, is related as $\hat{u} \times \hat{v} = \hat{n}$. Assuming a PEC object, $E_u, E_v$, and $B_n$ must vanish along the surface of the object according to boundary conditions. Thus the elements of the stress tensor needed to calculate for the surface integral become

$$T_{nn}(\mathbf{r}) = \frac{\epsilon_0}{2} E_n^2(\mathbf{r}) - \frac{1}{2\mu_0} B_u^2(\mathbf{r}) - \frac{1}{2\mu_0} B_v^2(\mathbf{r}) \quad (3.2)$$

$$T_{un}(\mathbf{r}) = T_{vn}(\mathbf{r}) = 0 \quad (3.3)$$

When the temperature is at absolute zero, that is $T = 0$, it can be assumed that the electromagnetic field is in the ground state, represented by the zero photon number state, otherwise known as the vacuum state. The ground state of the photon number state, while an eigenstate of the Hamiltonian operator, is not an eigenstate of the electric or magnetic field operators. As such, the electromagnetic fields in the vacuum exhibit fluctuations about their mean value of zero at a given vacuum energy. It is these fluctuating fields that give rise to a pressure on the objects in question that is the Casimir force. For use in the above stress tensor, the average of the fluctuating electric and magnetic fields in the ground state is found from the fluctuation-dissipation theorem

$$\langle 0 | \hat{E}_i(\mathbf{r},t) \hat{E}_j(\mathbf{r}',t) | 0 \rangle = \frac{\hbar}{\pi} \text{Im} \int_0^\infty \omega^2 G_{ij}(\mathbf{r},\mathbf{r}',\omega) d\omega \quad (3.4)$$

$$\langle 0 | \hat{B}_i(\mathbf{r},t) \hat{B}_j(\mathbf{r}',t) | 0 \rangle = \frac{\hbar}{\pi} \text{Im} \int_0^\infty (\nabla \times \hat{u} \cdot \nabla)_{il}(\nabla \times \hat{v})_{jm} G_{lm}(\mathbf{r},\mathbf{r}',\omega) d\omega \quad (3.5)$$

The dyadic Green’s function above is the same used in computational electromagnetics. Specifically, the Green’s function is related to the fields by

$$E_n(r) = i\omega \int_S d\mathbf{r}' G_{n\beta}(k,\mathbf{r},\mathbf{r}').J_\beta(\mathbf{r}) \quad (3.6)$$

$$B_u(r) = i\omega \int_S d\mathbf{r}' \left[ \frac{\partial}{\partial v} G_{n\beta}(k,\mathbf{r},\mathbf{r}') - \frac{\partial}{\partial n} G_{v\beta}(k,\mathbf{r},\mathbf{r}') \right] J_\beta(\mathbf{r}') \quad (3.7)$$
Table 3.1: Source and Field Operators for the Stress Tensor

<table>
<thead>
<tr>
<th>Field Component</th>
<th>Source $s_k \ (r_i = r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $E_n(r, s_1)$</td>
<td>$\hat{n}\delta(r' - r_i)$</td>
</tr>
<tr>
<td>(2) $H_u(r, s_2)$</td>
<td>$\hat{v}\frac{\partial}{\partial n}\delta(r' - r_i)$</td>
</tr>
<tr>
<td>(3) $H_u(r, s_3)$</td>
<td>$\hat{n}\frac{\partial}{\partial v}\delta(r' - r_i)$</td>
</tr>
<tr>
<td>(4) $H_v(r, s_4)$</td>
<td>$\hat{u}\frac{\partial}{\partial u}\delta(r' - r_i)$</td>
</tr>
<tr>
<td>(5) $H_v(r, s_5)$</td>
<td>$\hat{u}\frac{\partial}{\partial n}\delta(r' - r_i)$</td>
</tr>
</tbody>
</table>

Notice that to find the value of the dyadic Green’s function for a desired source and observation point, the excited electric field at the observation point due to a Dirac delta current excitation at the source point must be found. Likewise, to find the elements of the magnetic dyadic Green’s function, one can find the magnetic field excited by derivatives of the Dirac delta current excitations. In other words, using the above equations that relate the fields to the volume integral of the sources and dyadic Green’s functions, the terms for the stress tensor become

$$\frac{\epsilon_0}{2}\langle E_n(r)E_n(r')\rangle_{r'=r} = \frac{\hbar c_0}{2\pi \mu_0} Im \int_0^\infty \frac{k_0^2}{i\omega} E_n(k_0, r, s_1) dk_0 \quad (3.8)$$

$$\frac{1}{2\mu_0}\langle B_u(r)B_u(r')\rangle_{r'=r} = \frac{\hbar c_0}{2\pi} Im \int_0^\infty [H_u(k_0, r, s_3) - H_u(k_0, r, s_2)] dk_0 \quad (3.9)$$

$$\frac{1}{2\mu_0}\langle B_v(r)B_v(r')\rangle_{r'=r} = \frac{\hbar c_0}{2\pi} Im \int_0^\infty [H_v(k_0, r, s_5) - H_v(k_0, r, s_4)] dk_0 \quad (3.10)$$

Now the problem can be changed from one involving the dyadic Green’s functions to that of one involving the electric and magnetic fields excited by a set of point sources and their derivatives. This requires one to find the observed electric fields when they are excited at the observation point by the sources $s_i$ which are given in Table 3.1. To do this, it is assumed that computational code first uses an electric field excitation source $s_i$ and solves for the induced currents. With these induced currents, the scattered field at the observation point is found.

The necessary fields can be solved using EFIE by using the sources in Table 3.1 as the excitations on the right hand side. The impedance matrix remains the same with the exception that a Wick rotation is performed to expedite convergence of the integration.
over frequency. In essence, the Wick rotation transforms the problem from real frequencies to purely imaginary frequencies. Working in the imaginary frequencies gives rise to an integrand in the frequency integration that is smoothly varying and highly convergent. In the real frequencies, the integrand becomes highly oscillatory and as such is difficult to numerically integrate.

Implementation

To find the Casimir force, the resulting equation from the previous section is used.

\[
F = \oint_S \langle T(r') \rangle \cdot ds' = \frac{\hbar c_0}{2\pi} Im \int_0^\infty \oint_S \left[ \frac{k_0^2}{i\omega \mu_0} E_n(k_0, r, s_1) - H_u(k_0, r, s_3) + H_u(k_0, r, s_2)
\right.
\]
\[
- H_v(k_0, r, s_5) + H_v(k_0, r, s_4) \bigg] \cdot ds' dk_0
\]

Gauss-Laguerre quadrature is used to evaluate the frequency integral, over \( k_0 \), allowing the use of Gaussian quadrature over the domain of \([0, \infty)\). The evaluation of the surface integral is done using a Gaussian quadrature rule appropriate for the object’s surface. Thus, the final equations are

\[
F \approx \frac{\hbar c_0}{2\pi} Im \sum_{i=1}^{N} \sum_{j=1}^{M} \left[ \frac{k_0^2}{i\omega \mu_0} E_n(k_{0i}, r_j, s_1) + H_u(k_{0i}, r_j, s_2) - H_u(k_{0i}, r_j, s_3)
\right.
\]
\[
+ H_v(k_{0i}, r_j, s_4) - H_v(k_{0i}, r_j, s_5) \bigg]
\]

So, to find the Casimir force one must calculate the desired electric and magnetic field components that arise due to the current excitations given in Table 3.1 at specific locations in space and specific frequencies. This is done by using the method of moments (MOM) to solve for the induced currents on the PEC objects due to our excitations. Then the solved currents are used to find the scattered fields in the above equation. The right-hand side
vector becomes
\[ b_m = \int_{T_m} f_m(r) \cdot E^{inc}(r) dr = \frac{i}{\omega \mu_0} \int_{T_m} f_m(r) \cdot \int G(r, r') (k_0, r, r') s_k(r') dr' dr \]  
(3.13)

where \( f_m(r) \) is the \( m \)-th basis function (assumed to be Rao-Wilton-Glisson (RWG) basis [59]) and \( T_m \) is the surface over the \( m \)-th basis function (two adjacent triangular patches in the case of RWG). This is the first source of singularities because some of the sources \( s_k \) are super-hyper singular due to the derivative of the Dirac delta function. Xiong et al. use Tong and Chew’s work to evaluate these cases [60]. Xiong et al. also stated that there is another source of singularity. Since it is necessary to find the system’s dyadic Green’s function on the surface of the object to integrate the Maxwell stress tensor over the surface, the scattered fields due to Dirac delta sources on the surface of the object are needed. However, the PEC surface, observation point, and source point all coincide and using a point source results in a scattered field of infinite magnitude. To avoid this, Xiong et al. argues that the location of the source and observation points from the PEC surface need to be offset. A small distance offset is applied along the normal direction to the locations where the sources and observation points are located. That is, given the quadrature points that we wish to integrate along the surface \( \mathbf{r}_j \), we shift them via
\[ \mathbf{r}_j = \mathbf{r}_j + \delta \hat{n} \]  
(3.14)

With the above considerations, the resulting implemented code can produce results comparable to other published simulations. There is the added benefit that by using the stress tensor, a pressure map can be calculated over a given surface in addition to the net force. However, the disadvantages lie in having to integrate across a surface and carefully dealing with the singularities, in point of which is a short monograph located in Appendix C that discusses a new way of treating the normalization of the Casimir force. Work has also progressed on the stability of the code and expanding the types of problems the code can handle. The ability for arbitrary objects has been added in addition to improvements in the numerical treatment of singularity subtraction. Further work has been done on
implementing large scale techniques using the fast multipole algorithm (FMA) which is discussed in 3.6. To address the added computation time from the surface integrations, basic parallel processing techniques can be used to make use of the embarrassingly parallel nature of the algorithm.

3.3 RRWJ Method

The path integral method introduced by Reid, Rodriguez, White, and Johnson (RRWJ) method is a complicated procedure but produces a rather simple result. The process starts by relating the ground state energy of a quantum system to the partition function. The partition function itself can be calculated using a path integral. The difficulty lies in calculating the path integral as the path integral is a functional integration over all possible values of the fields in the system that conform to the system’s boundary conditions. The problem of finding the possible values of the fields can be circumvented by using what is known as an effective action. In perturbation quantum mechanics, the effective action is useful because it allows one to go from a constrained path integral to an unconstrained path integral. Normalization allows for many of the factors in this process to be ignored. With the effective action, various techniques are used to simplify the expression of the action from involving an inverse operator to a single operator. The path integral involving this simplified effective action can be evaluated by using a basis expansion thus changing the operator acting on continuous functions to a matrix-vector operation.

Partition Function

The process to find the Casimir energy starts by relating the ground state energy to what is known as the partition function in order to relate the Casimir energy to a path integral. If we have the operator \( \hat{\Omega} \), then the trace of the operator is defined as

\[
\text{Tr} \left[ \hat{\Omega} \right] = \sum_n \langle n | \hat{\Omega} | n \rangle = \Omega_{nn}, \]

where \( n \) represents the eigenstate of the operator [61]. Assuming that we will be working with the Hamiltonian operator \( \hat{H} \) for a system that is
time independent, we can thus show

\[
\text{Tr} \left[ e^{-\frac{\hat{H}t}{\hbar}} \right] = \sum_n \langle n | e^{-\frac{\hat{H}t}{\hbar}} | n \rangle \\
= \sum_n \langle n | 1 - \frac{it}{\hbar} \hat{H} + \frac{t^2}{2\hbar^2} \hat{H}^2 + \ldots | n \rangle \\
= \sum_n \langle n | 1 - \frac{it}{\hbar} E_n + \frac{t^2}{2\hbar^2} E_n^2 + \ldots | n \rangle \\
= \sum_n e^{-\frac{t}{\hbar} E_n t} \langle n | n \rangle = \sum_n e^{-\frac{t}{\hbar} E_n t}
\]  

(3.15)

If there exists a system with a ground state \( E_0 \) that is non-degenerate, then we can show that

\[
- \lim_{\Lambda \to \infty} \frac{1}{\Lambda} \ln \left( \sum_n e^{-E_n \Lambda} \right) = - \lim_{\Lambda \to \infty} \frac{1}{\Lambda} \ln \left[ e^{-E_0 \Lambda} \left( 1 + \sum_n e^{-(E_n - E_0) \Lambda} \right) \right] \\
= - \lim_{\Lambda \to \infty} \frac{1}{\Lambda} \left[ \ln \left( e^{-E_0 \Lambda} \right) + \ln \left( 1 + \sum_n e^{-(E_n - E_0) \Lambda} \right) \right] \\
= E_0 - \lim_{\Lambda \to \infty} \frac{1}{\Lambda} \ln \left( 1 + \sum_n e^{-(E_n - E_0) \Lambda} \right) \\
= E_0
\]  

(3.16)

Equation (3.16) above is almost in the desired form. The last step is to let the systems have Euclidean time by defining \( t_b - t_a = -i\beta \). Here, \( \beta \) will be considered as the inverse temperature as taking the limit of \( 1/\beta \) to zero will give back the ground state, or, equivalently, the energy at zero temperature. This allows for partition function to be defined as \( Z \) where \( Z = \text{Tr} \left[ e^{-\frac{\hat{H}t}{\hbar}} \right] \). Using the results from Equations (3.15) and (3.16), the ground state of a system defined by the partition function \( Z \) becomes

\[
E_0 = - \lim_{\beta \to \infty} \frac{\hbar}{\beta} \ln Z
\]  

(3.17)

The Casimir energy requires normalization which will be taken as the difference between the ground state energy of the desired system, \( Z \), and the ground state energy when the
objects are at infinite separation, \( Z_\infty \). Thus, the Casimir energy, \( \mathcal{E} \), becomes

\[
\mathcal{E} = - \lim_{\beta \to \infty} \frac{\hbar}{\beta} \ln \left( \frac{Z}{Z_\infty} \right)
\]  

(3.18)

The Propagator and the Path Integral

It was shown above that the ground state energy can be expressed using the partition function \( Z \) defined as

\[
Z = \sum_n \langle n | e^{-\frac{\hbar}{\beta} \hat{H}} | n \rangle
\]

(3.19)

This can also be equivalently expressed using the quantum mechanical propagator. Assuming that the Hamiltonian operator of interest is time independent, there will be a set of eigenstates associated with position \( r_a \) at time \( t = t_a \) and position \( r_b \) at time \( t = t_b \). The propagation from the initial state to the final state can be mediated using the time evolution operator, defined as [62]

\[
\hat{U} = e^{-\frac{\hbar}{\beta} \hat{H} t}
\]

(3.20)

such that

\[
|\Psi(r, t)\rangle = \hat{U} |\Psi(r, 0)\rangle
\]

(3.21)

Thus, the probability amplitude that the state at position \( r_a \) and time \( t_a \) will evolve to the state at position \( r_b \) and time \( t_b \) can be found using the time evolution operator. The probability amplitude of this occurring is known as the propagator and is defined as

\[
K(r_b, t_b; r_a, t_a) = \langle \Psi(r_b, t_b) | \Psi(r_a, t_a) \rangle = \langle \psi(r_b) | e^{-\frac{\hbar}{\beta}(t_b-t_a) \hat{H}} | \psi(r_a) \rangle
\]

(3.22)
The time evolution operator can be expanded using $e^{-i/\hbar(t_b-t_a)}\hat{H} = e^{-i/\hbar(t_b-t_1)}e^{-i/\hbar(t_1-t_a)}\hat{H}$ and a factor of 1 can be inserted from the fact that $\int \, dr_1 |\psi(r_1,0)\rangle \langle \psi(r_1,0)| = 1$ to give

$$K(r_b, t_b; r_a, t_a) = \langle \psi(r_b) | e^{-\hat{\mathcal{H}}(t_b-t_1)} | \psi(r_a) \rangle \langle \psi(r_1) | e^{-i/\hbar(t_1-t_a)}\hat{H} | \psi(r_a) \rangle$$

$$= \int \, dr_1 K(r_b, t_b; r_1, t_1) K(r_1, t_1; r_a, t_a) \quad (3.23)$$

The above is another way of stating the rule for combining amplitudes, which states that if a process can occur so many ways, then the amplitudes for each of these ways add. In other words, an intermediate state at position $r_1$ and time $t_1$ can be chosen and the probability amplitude from $r_a$ at $t_a$ to $r_b$ at $t_b$ becomes the sum over the probability amplitudes from $r_a$ at $t_a$ to all possible positions $r_1$ at $t_1$ to $r_b$ at $t_b$. Feynman likened the probability amplitude in Equation (3.22) to having a source at $r_a$ and a detector at $r_b$ and measuring the probability that a particle sourced at $t_a$ would be detected at $t_b$ [63]. The probability amplitude in Equation (3.23) is equivalent but can be thought of as being constructed by placing a screen between the source and detector with a hole at some position $r_1$ that opens at $t_1$. The total probability amplitude is the sum of the probability amplitudes from every possible position of the hole with the amplitudes interfering with each other in a manner that is not unlike a double slit experiment. Figure 3.1 shows this visually with each arrow representing a propagator from one state to another with the position on the vertical axis indicating different states and the horizontal axis representing time.

So the propagator can be divided over intervals of time using the time evolution operator in addition to being divided up over a set of states. If the time is divided up into $N$ intervals, then each time step will be $\delta = \frac{t_b-t_a}{N}$ and the probability amplitude becomes

$$K(r_b, t_b; r_a, t_a) = \langle \psi(r_b) | \prod_{n=1}^{N} e^{-\frac{i}{\hbar}\delta \hat{H}} | \psi(r_a) \rangle \quad (3.24)$$

In between each exponential, another intermediate state can be inserted and the
Figure 3.1: Visual interpretation of the propagator with an intermediate state at $t = t_1$ defined at some position $r_1$.

The propagator thus becomes

$$K(r_b, t_b; r_a, t_a) = \int dr_1 \ldots dr_{N-1} K(r_b, t_b; r_{N-1}, (N-1)\delta + t_a) \cdot K(r_{N-1}, (N-1)\delta + t_a; r_{N-2}, (N-2)\delta + t_a) \ldots K(r_2, 2\delta + t_a; r_1, \delta + t_a) \cdot K(r_1, \delta + t_a; r_a, t_a)$$  \hspace{1cm} (3.25)

Now the amplitude has become the contribution of all possible $N$-part paths. One could take the limit of $N \to \infty$ and conceptually the propagator becomes the contribution from all possible paths from the initial state to the final state. This forms the basis of the path integral and in our case the path integral becomes

$$K(r_b, t_b; r_a, t_a) = \int_{r_a}^{r_b} D\{r\} e^{iS[r_b, r_a]}$$  \hspace{1cm} (3.26)

The function $S$ in the exponential is the quantum mechanical analog of the Lagrangian action while $D\{r\}$ indicates that we are taking an integral over all possible values of $r$. In
the above discussion, the path integral was discussed in terms of position and time coordinates, perhaps relating to a particle like an electron. However, the path integral can be extended from a phase-space integral to a field integral as well.

Recall from above that the partition function is defined in Equation (3.19) and that the propagator is defined, using \( q \) to define our initial state and \( q' \) our final state, as

\[
K(q', \tau; q, 0) = \langle q' | e^{-i \hat{H} \tau} | q \rangle \tag{3.27}
\]

Then the time period \( \tau \) can be set to the inverse temperature parameter as \( \tau = -i \beta \).

Then, letting our initial and final states be the same,

\[
Z = \sum_n K(n, -i \beta; n, 0) \tag{3.28}
\]

Thus, the partition function can be related to the propagator and calculated using a path integral. Essentially, while the original propagator was limited to the paths between the two different states, the propagator in the partition function consists of all paths starting and ending at state \( n \). The partition function is then the summation of all these propagators which is equivalent to a path integral that integrates over all possible closed paths in the state space.

**Path Integral Formulation**

It can be shown that our partition function \( Z \) can be related to a quantum field path integral in Euclidean space as [61,62,64]

\[
Z = \int D \{ \phi (r, \tau) \}_C e^{-1/\hbar S_{\beta} [\phi]} \tag{3.29}
\]

where \( \tau = it \), \( \phi \) is a scalar field, the subscript \( C \) denotes that the fields over which the functional integration is constrained and we have taken the liberty of using the imaginary time transformation. Since we are using Euclidean (imaginary) time, it should be noted
that any purely real frequencies and wavenumbers are imaginary in the normal Minkowski space. The action $S_\beta$ is defined as

$$S_\beta [\phi] = \int_0^{h \beta} d\tau \int d\mathbf{r} \mathcal{L}_E \{\phi (\mathbf{r}, \tau)\}$$

(3.30)

where $\mathcal{L}_E$ is the Euclidean Lagrangian density. Assuming that the boundary conditions are time independent and the Lagrangian density contains no terms higher than quadratic order in $\phi$ and its derivatives, field $\phi$ can be expanded using a Fourier series.

$$\phi(\tau, \mathbf{r}) = \sum_{n=0}^{\infty} \phi_n(\mathbf{r}) e^{-i \Omega_n \tau}, \quad \Omega_n = \frac{2 \pi n}{c \beta}$$

(3.31)

The frequencies $\Omega_n$ are known as the Matsubara frequencies. Likewise, the path integral now becomes the product of integrals over the Fourier components.

$$\int \mathcal{D} \{\phi(\tau, \mathbf{r})\} = \prod_{n=0}^{\infty} \int \mathcal{D} \{\phi_n(\mathbf{r})\}$$

(3.32)

The partition function becomes

$$Z = \prod_{n=0}^{\infty} Z(\Omega_n) = \prod_{n=-\infty}^{\infty} \int \mathcal{D} \{\phi_n(\mathbf{r})\} e^{-S[\phi_n; \Omega_n]}$$

(3.33)

where

$$S[\phi_n; \Omega_n] = \beta \int d\mathbf{r} \mathcal{L}_E \{\phi(\mathbf{r})e^{-i \Omega_n \tau}\}$$

(3.34)

Using the partition functions from Equation (3.33) in (3.18) gives a new expression for the Casimir energy.

$$\mathcal{E} = - \lim_{\beta \to \infty} \frac{\hbar}{\beta} \ln \left( \frac{Z}{Z_\infty} \right) = - \lim_{\beta \to \infty} \frac{\hbar}{\beta} \sum_{n=0}^{\infty} \ln \left( \frac{Z(\Omega_n)}{Z_\infty(\Omega_n)} \right)$$

(3.35)
In the limit, we have

\[- \lim_{\beta \to \infty} \beta \frac{\hbar}{2\pi} \sum_{n=1}^{\infty} F(\Omega_n) = - \frac{2\hbar}{\beta} \int_{0}^{\infty} \frac{\beta d\Omega}{2\pi} F(\Omega) = - \frac{\hbar}{\pi} \int_{0}^{\infty} d\Omega F(\Omega) \quad (3.36)\]

which finally gives

\[ \mathcal{E} = - \frac{\hbar}{\pi} \int_{0}^{\infty} d\Omega \ln \frac{\mathcal{Z}(\Omega)}{\mathcal{Z}_{\infty}(\Omega)} = - \frac{\hbar c}{\pi} \int_{0}^{\infty} d\kappa \ln \frac{\mathcal{Z}(c\kappa)}{\mathcal{Z}_{\infty}(c\kappa)} \quad (3.37)\]

where \( \Omega = c\kappa \). The wavenumber \( \kappa \) is used to clarify that the partition functions are in Euclidean space. Evaluating the partition function \( \mathcal{Z}(\Omega) \) requires integrating a path integral over a constrained path that is dictated by the boundary conditions of the system to which the fields \( \phi \) must conform. This presents a great obstacle because determining the constrained path is essentially the same as determining the eigenfrequencies of the system. At first glance then, moving to the path integral does not seem to make the task easier. However, it is possible to allow for a path integral that integrates over the free fields as opposed to the constrained fields. This is done by using a path integral taken over all possible configurations of the field \( \phi \) where the integrand is multiplied by Dirac delta functions that evaluate at the configurations of \( \phi \) that conform to the boundary conditions [64–66]. Assuming that the boundary conditions are represented by a set of linear operators \( L_\alpha \), then the constrained path integral can be expressed as

\[ \mathcal{Z}(\Omega_n) = \int \mathcal{D} \{ \phi_n(r) \} e^{-S[\phi_n;\Omega_n]} = \int \mathcal{D} \{ \phi_n(r) \} \prod_\alpha \delta (L_\alpha \phi) e^{-S[\phi_n;\Omega_n]} \quad (3.38)\]

so that now the functional integration is taken over the unconstrained path and the Dirac delta functions can be expressed as Gaussian integrals. A one-dimensional Dirac delta can be represented using the Fourier transform as

\[ \delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-2\pi ikx} dk \quad (3.39)\]
The Fourier transform can be used to redefine the Dirac delta distribution used in Equation (3.38).

\[ Z(\Omega_n) = \int \mathcal{D}\{\phi_n(r)\} \int \prod_{\alpha} \frac{dk_\alpha}{2\pi} e^{-S[\phi_n;\Omega_n] + i \sum_{\alpha} k_\alpha L_\alpha} \phi \] (3.40)

With the above exponent being quadratic in \( \phi \), the integration over the unconstrained \( \phi \) can be done in the same manner as other Gaussian integrals. The result becomes

\[ Z(\Omega_n) = W_0 \int \prod_{\alpha} dk_\alpha e^{-S_{\text{eff}}[k_\alpha]} \] (3.41)

where \( W_0 \) is the partition function of the free field and cancels out in the normalization in Equation (3.37) and \( S_{\text{eff}} \) is known as the effective action.

### Euclidean Lagrangian Density

The formulation in the previous sections has been done without any reference to the actual system at hand. The derivation so far pertains to any number of equivalent quantum field systems. To define the path integral for the Casimir force, the effective action needs to be derived using the appropriate Euclidean Lagrangian density and boundary conditions of the problem. The Lagrangian density of the electromagnetic field in vacuum in Minkowski space is [67]

\[ \mathcal{L} = -\frac{1}{4\mu_0} F_{\alpha\beta} F^{\alpha\beta} = -\frac{2}{\mu_0} \left( \partial_\alpha A_\beta \partial^\alpha A^\beta - \partial_\beta A_\alpha \partial^\alpha A^\beta \right) \] (3.42)

where \( A^\alpha \) is the electromagnetic four-potential (where we have used the script font to differentiate the four-vector from the vector potential) defined as

\[ A^\alpha = \left( \frac{\Phi}{c}, \mathbf{A} \right) \] (3.43)
and

\[ \partial^o = \left( \frac{\partial}{\partial x_0}, -\nabla \right) \] (3.44)

The above Lagrangian can be shown to be equivalent to the expected Lagrangian density using the electromagnetic field vectors,

\[ \mathcal{L} = \frac{1}{2} \left( \epsilon_0 |E|^2 - \frac{1}{\mu_0} |H|^2 \right) \] (3.45)

where

\[ E = -\frac{\partial A}{\partial t} - \nabla \Phi \] (3.46)

\[ B = \nabla \times A \] (3.47)

From either form, we can generate the Lagrangian density using the potentials and assuming time-harmonic solutions,

\[ \mathcal{L}_E = \frac{1}{2} \left[ \epsilon_0 \left( -i\omega A^* - \nabla \Phi^* \right) \cdot (i\omega A - \nabla \Phi) + \frac{1}{\mu_0} \left( \nabla \times A^* \right) \cdot \left( \nabla \times A \right) \right] \]

\[ = \frac{1}{2} \left[ \epsilon_0 \left( \omega^2 A^* \cdot A + \nabla \Phi^* \nabla \Phi + i\omega A^* \cdot \nabla \Phi - i\omega \nabla \Phi^* \cdot A \right) + \frac{1}{\mu_0} \left( \nabla \times A^* \right) \cdot \left( \nabla \times A \right) \right] \] (3.48)

It is desired to transform the above into a vector-matrix-vector like product. This is done by moving the derivatives from the complex conjugate potentials onto the unconjugated potentials using integration by parts with the understanding that the action takes the integration of the Lagrangian density over all space. For example,

\[ \int d\mathbf{r} \nabla \Phi^* \nabla \Phi = \int d\mathbf{r} \nabla [\Phi^* \cdot \nabla \Phi] - \int d\mathbf{r} \Phi^* \nabla \cdot \nabla \Phi = \int \Phi^* \cdot \nabla \Phi \hat{n} dS - \int d\mathbf{r} \Phi^* \nabla \cdot \nabla \Phi \]

\[ = -\int d\mathbf{r} \Phi^* \nabla \cdot \nabla \Phi \] (3.49)
where it is assumed that the potentials go to zero at infinity. The other integrations fall out as

\[
\int dr \nabla \Phi^* \cdot A = - \int dr \Phi^* \nabla \cdot A \tag{3.50}
\]

\[
\int dr \nabla \times A^* \cdot \nabla \times A = \int dr A^* \cdot \nabla \times \nabla \times A + \int dr \nabla \cdot (A^* \times \nabla \times A) = \int dr A^* \cdot \nabla \times \nabla \times A + \int A^* \times \nabla \times A \cdot \hat{n} dS = \int dr A^* \cdot \nabla \times \nabla \times A \tag{3.51}
\]

Thus,

\[
\mathcal{L}_E = \frac{1}{2} \left[ \epsilon_0 \left( -\Omega^2 A^* \cdot A - \Phi^* \nabla \cdot \nabla \Phi + i\omega A^* \cdot \nabla \Phi + i\omega \Phi^* \nabla \cdot A \right) + \frac{1}{\mu_0} A^* \cdot \nabla \times \nabla \times A \right] \tag{3.52}
\]

The derivations for the path integral call for a Lagrangian in the Euclidean space. This can be done using the following transformations

\[
\omega \rightarrow i\Omega, \quad \Phi \rightarrow i\Phi, \quad \Phi^* \rightarrow i\Phi^* \tag{3.53}
\]

The resulting Euclidean Langrangian becomes

\[
\mathcal{L}_E = \frac{1}{2} \left[ \epsilon_0 \left( -\Omega^2 A^* \cdot A + \Phi^* \nabla \cdot \nabla \Phi - i\Omega A^* \cdot \nabla \Phi - i\Omega \Phi^* \nabla \cdot A \right) + \frac{1}{\mu_0} A^* \cdot \nabla \times \nabla \times A \right] \tag{3.54}
\]

This can now be written in a vector-matrix-vector product form using the previously defined four-potential.

\[
\mathcal{L}_E = \frac{1}{2} A^\dagger \cdot \mathbf{D}_{EM} \cdot A \tag{3.55}
\]
where

\[
\mathbf{A} = \begin{bmatrix}
\Phi/
\frac{c}
A_x \\
A_y \\
A_z
\end{bmatrix}
\]

(3.56)

\[
\mathbf{D}_{EM} = \begin{bmatrix}
\frac{1}{\mu_0} \nabla^2 & -i \frac{1}{\mu_0} \Omega \partial_x & -i \frac{1}{\eta_0} \Omega \partial_y & -i \frac{1}{\mu_0} \Omega \partial_z \\
-i \frac{1}{\eta_0} \Omega \partial_x & -\epsilon_0 \Omega^2 + \frac{1}{\mu_0} (\partial_y^2 + \partial_z^2) & -\frac{1}{\mu_0} \partial_x \partial_y & -\frac{1}{\mu_0} \partial_x \partial_z \\
-i \frac{1}{\eta_0} \Omega \partial_y & -\frac{1}{\mu_0} \partial_y \partial_x & -\epsilon_0 \Omega^2 + \frac{1}{\mu_0} (\partial_x^2 + \partial_z^2) & -\frac{1}{\mu_0} \partial_y \partial_z \\
-i \frac{1}{\eta_0} \Omega \partial_z & -\frac{1}{\mu_0} \partial_z \partial_x & -\frac{1}{\mu_0} \partial_z \partial_y & -\epsilon_0 \Omega^2 + \frac{1}{\mu_0} (\partial_x^2 + \partial_y^2)
\end{bmatrix}
\]

(3.57)

While the above has not been generalized for arbitrary permittivity and permeability, one could replace the vacuum permittivity and permeability directly and regain the more general result [64]. Finally, there is the necessity of symmetry breaking [68]. In a more general sense, breaking symmetry in quantum field theory can become necessary when applying a symmetric operation to the Hamiltonian (or in this case the Lagrangian) of the system. Under such circumstances, the ground state is transformed to a different but equivalent ground state. This ground state, however, is now degenerate due to the symmetry. When performing the path integral, it is possible to include paths that belong to these degenerate states causing one to overcount the actual paths of the ground state. The solution then is to break the symmetry of the transformed Hamiltonian in such a way that the ground state is no longer degenerate without affecting the ground state itself. Conceptually, this is similar to the classical gauge fixing of the electromagnetic potentials. Without fixing the gauge, the path integral will integrate over the paths that are invariant under the various gauge transformations. To solve this problem, RRWJ chose to use a Faddeev-Popov gauge. This gauge is commonly taken as writing the tensor operator \( \mathbf{D} \) as the sum of a diagonal operator and a rank-one update. Then the gauge fixing is
accomplished by adding back the scaled rank-one update.

\[ D_{EM} = U_1 - U_2 \]  \hspace{1cm} (3.58)

\[ U_1 = \left( -\epsilon_0 \Omega^2 + \frac{1}{\mu_0} \nabla^2 \right) I \]  \hspace{1cm} (3.59)

\[ U_2 = \left[ \begin{array}{cccc}
-\epsilon_0 \Omega^2 & -i \frac{1}{\eta_0} \Omega \partial_x & -i \frac{1}{\eta_0} \Omega \partial_y & -i \frac{1}{\eta_0} \Omega \partial_z \\
-i \frac{1}{\eta_0} \Omega \partial_x & \frac{1}{\mu_0} \partial_x^2 & \frac{1}{\mu_0} \partial_x \partial_y & \frac{1}{\mu_0} \partial_x \partial_z \\
-i \frac{1}{\eta_0} \Omega \partial_y & \frac{1}{\mu_0} \partial_y \partial_x & \frac{1}{\mu_0} \partial_y^2 & \frac{1}{\mu_0} \partial_y \partial_z \\
-i \frac{1}{\eta_0} \Omega \partial_z & \frac{1}{\mu_0} \partial_z \partial_x & \frac{1}{\mu_0} \partial_z \partial_y & \frac{1}{\mu_0} \partial_z^2
\end{array} \right] \]  \hspace{1cm} (3.60)

where \( \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \) is the impedance of freespace. The Faddeev-Popov gauge fixing term is then

\[ \overline{D}_{FP} = \frac{1}{\alpha_{FP}} U_2 \]  \hspace{1cm} (3.61)

where \( \alpha_{FP} \) is the gauge parameter and can be arbitrarily chosen as it will have no physical impact. The final Euclidean Lagrangian density for the PEC case becomes

\[ \mathcal{L}_E = \frac{1}{2} A^\dagger \cdot [D_{EM} + D_{FP}] \cdot A = \frac{1}{2} A^\dagger \cdot D \cdot A \]  \hspace{1cm} (3.62)

**Effective Action**

For the effective action to be derived, the boundary conditions must be expressed as a Dirac delta in the path integral. For the PEC case, it is normally considered that the EFIE provides adequate information to solve for the electromagnetic fields. The boundary conditions for the EFIE are that the total tangential electric fields along the surface of the objects must be zero. In other words,

\[ \hat{n} \times E(\mathbf{r}) = 0 \]  \hspace{1cm} (3.63)
There is no distinction of incident and scattered fields in this case because the path integral is looking at total field configurations. Equivalently, the boundary conditions can be described using the four-potential via the differential operator $\underline{L}$.

$$\hat{n} \times \underline{L} \cdot \mathbf{A} = 0 \tag{3.64}$$

where $\underline{L}$ can be easily inferred from Equation (3.46).

$$\underline{L} = \begin{bmatrix}
-\partial_x & i\Omega & 0 & 0 \\
-\partial_y & 0 & i\Omega & 0 \\
-\partial_z & 0 & 0 & i\Omega
\end{bmatrix} \tag{3.65}$$

The Dirac delta functions for the conditions at a given location $\mathbf{r}$ on a PEC surface can now be defined using what can be thought of as a surface current $\mathbf{J}(\mathbf{r})$, although the label is not to be taken too literally.

$$\delta(\hat{n} \times \mathbf{E}(\mathbf{r})) = \int \frac{d\mathbf{J}(\mathbf{r})}{(2\pi)^2} e^{i \int T(\mathbf{r}) \cdot \underline{L} \cdot \mathbf{A}(\mathbf{r})} \tag{3.66}$$

Taking the superposition of the boundary conditions across a surface $S$ gives us the desired functional Dirac deltas.

$$\int \mathcal{D}\{\mathbf{J}(\mathbf{r})\} e^{i \int_S \mathbf{J}^T(\mathbf{r}) \cdot \underline{L} \cdot \mathbf{A}(\mathbf{r})} d\mathbf{r} \tag{3.67}$$

The partition function for the electromagnetic fields involved in the Casimir force can now be defined over the unconstrained path.

$$\mathcal{Z}(\Omega) = \int \mathcal{D}\{\mathbf{A}\} e^{-\frac{\beta}{2} \int \mathbf{A}^\dagger \cdot \underline{D} \cdot \mathbf{A} d\mathbf{r}}$$

$$= \int \mathcal{D} \prod_i \{\mathbf{J}_i(\mathbf{r})\} \int \mathcal{D}\{\mathbf{A}\} e^{-\frac{\beta}{2} \int \mathbf{A}^\dagger \cdot \underline{D} \cdot \mathbf{A} d\mathbf{r}} \prod_i e^{i \int S_i \mathbf{J}_i^T \cdot \underline{L} \cdot \mathbf{A} d\mathbf{r}} \tag{3.68}$$

where the product is taken over $i$ independent surfaces. Taking note of the fact that $\underline{D}$ is a tensor operator, then the above path integral can be taken as a series of Gaussian integrals.
Specifically, it can be shown that \[69\]

\[
\int \mathcal{D}\{\phi\} e^{\int \left[-\frac{1}{2} \phi \cdot \hat{A} \cdot \phi + i J \cdot \phi \right] dr} \propto e^{-\frac{1}{2} \int \int \left[ J(r) \hat{A}^{-1}(r,r') J(r') \right] dr' dr}
\]

where \(\hat{A}\) is a differential operator and this can be generalized to the tensor case. Now \(\mathcal{A}\) is the four-potential over all space while the “currents” are confined to their associated surfaces. One can adapt the path integral over \(\mathcal{A}\) taken over all space with Dirac deltas showing the explicit confinement of the “currents” to the surfaces.

\[
\int \mathcal{D}\{A\} e^{-\frac{\beta}{2} \int A^\dagger \cdot \mathcal{D} \cdot A d\mathbf{r}} \prod_i e^{i \int S_i J^T_i L \cdot A d\mathbf{r}} = \int \mathcal{D}\{A\} e^{\beta \int d\mathbf{r} \left[-\frac{1}{2} A^\dagger \cdot \mathcal{D} \cdot A + i \frac{1}{\beta} \sum_i J^T_i \delta(S_i) \cdot L \cdot A \right]} \] (3.70)

Evaluating the Gaussian integrals,

\[
\int \mathcal{D}\{A\} e^{\beta \int d\mathbf{r} \left[-\frac{1}{2} A^\dagger \cdot \mathcal{D} \cdot A + i \frac{1}{\beta} \sum_i J^T_i \delta(S_i) \cdot L \cdot A \right]} = W_0 \left[ \int S_{1d} d\mathbf{r} \int S_{1d} d\mathbf{r'} J^T_1(r) \mathbf{L} \hat{A}^{-1} \mathbf{L}^T \cdot J_1(r') + \int S_{2d} d\mathbf{r} \int S_{2d} d\mathbf{r'} J^T_2(r) \mathbf{L} \hat{A}^{-1} \mathbf{L}^T \cdot J_2(r') + \ldots \right] \] (3.71)

To find an equivalent operator to replace \(\mathbf{L} \cdot \mathbf{D}^{-1} \cdot \mathbf{L}^T\), the problem will be momentarily taken to the momentum space which takes the differential operators to numerical values of the components of a wavevector. For example, \(\partial_x \rightarrow i k_x\) and in the momentum space, the
tensors involved become

\[
\mathbf{L} = \begin{bmatrix}
-\imath c k_x & \imath \Omega & 0 & 0 \\
-\imath c k_y & 0 & \imath \Omega & 0 \\
-\imath c k_z & 0 & 0 & \imath \Omega
\end{bmatrix}
\]  

(3.72)

\[
\mathbf{U}_1 = \left( -\epsilon_0 \Omega^2 - \frac{1}{\mu_0} k^2 \right) \mathbf{I} = -\frac{1}{\mu_0} \left( \kappa^2 + k^2 \right) \mathbf{I}
\]  

(3.73)

\[
\mathbf{U}_2 = \begin{bmatrix}
-\epsilon_0 \Omega^2 & -\frac{1}{\eta_0} \Omega k_x & -\frac{1}{\eta_0} \Omega k_y & -\frac{1}{\eta_0} \Omega k_z \\
-\frac{1}{\eta_0} \Omega k_x & -\frac{1}{\mu_0} k_x^2 & -\frac{1}{\mu_0} k_x k_y & -\frac{1}{\mu_0} k_x k_z \\
-\frac{1}{\eta_0} \Omega k_y & -\frac{1}{\mu_0} k_y k_x & -\frac{1}{\mu_0} k_y^2 & -\frac{1}{\mu_0} k_y k_z \\
-\frac{1}{\eta_0} \Omega k_z & -\frac{1}{\mu_0} k_z k_x & -\frac{1}{\mu_0} k_z k_y & -\frac{1}{\mu_0} k_z^2
\end{bmatrix}
\]  

(3.74)

\[
\mathbf{D}^{-1} = \begin{bmatrix}
\mathbf{U}_1 - \mathbf{U}_2 + \frac{1}{\alpha_{FP}} \mathbf{U}_2
\end{bmatrix}^{-1}
\]  

(3.76)

What is of interest now is the tensor product \( \mathbf{L} \cdot \mathbf{D}^{-1} \mathbf{L}^T \).

Let

\[
\mathbf{P} = -\frac{\mu_0 (\alpha_{FP} - 1)}{\alpha_{FP} (\kappa^2 + k^2)} \mathbf{U}_2
\]  

(3.77)

then via the Neumann series,

\[
[\mathbf{I} - \mathbf{P}]^{-1} = \mathbf{I} + \frac{1}{1 - \text{Tr}[\mathbf{P}]}
\]  

(3.78)
which gives
\[ \mathbf{D}^{-1} = -\frac{\mu_0}{\kappa^2 + k^2} \left[ \mathbf{I} - \frac{\mu_0}{\kappa^2 + k^2} (\alpha_{FP} - 1) \mathbf{U}_2 \right] \] (3.79)

The second term is scaled by the gauge fixing parameter, but it can be shown that
\[ \mathbf{L} \cdot \mathbf{U}_2 \cdot \mathbf{L}^T = \frac{(\Omega - c\kappa)^2}{\mu_0} \begin{bmatrix} k_x^2 & k_x k_y & k_x k_z \\ k_x k_y & k_y^2 & k_y k_z \\ k_x k_z & k_y k_z & k_z^2 \end{bmatrix} = 0 \] (3.80)
by virtue of the fact that we have chosen \( \Omega = c\kappa \) and thus, as expected, the gauge fixing does not affect the final result. Finally,
\[ \mathbf{L} \cdot \mathbf{D}^{-1} \mathbf{L}^T = \mathbf{L} \cdot -\frac{\mu_0}{\kappa^2 + k^2} \mathbf{I} \cdot \mathbf{L}^T = \mathbf{L} \cdot \frac{\mu_0}{\kappa^2 + k^2} \begin{bmatrix} \Omega^2 + c^2 k_x^2 & c^2 k_x k_y & c^2 k_x k_z \\ c^2 k_x k_y & \Omega^2 + c^2 k_y^2 & c^2 k_y k_z \\ c^2 k_x k_z & c^2 k_y k_z & \Omega^2 + c^2 k_z^2 \end{bmatrix} \]
\[ = \frac{\mu_0}{\kappa^2 c^2 (\kappa^2 + k^2)} \left( \mathbf{I} - \frac{1}{\kappa^2} \nabla \nabla \right) \] (3.81)
where the momentum space tensor can be reduced to the tensor operator \( \nabla \nabla \). To convert from the momentum space back to the spatial space,
\[ \frac{1}{(2\pi)^3} \int d\mathbf{k} \mathbf{L} \cdot \mathbf{D}^{-1} \mathbf{L}^T e^{i\mathbf{k} \cdot \mathbf{r}} = \frac{\mu_0}{(2\pi)^3 \kappa^2 c^2} \int d\mathbf{k} \frac{1}{\kappa^2 + k^2} e^{i\mathbf{k} \cdot \mathbf{r}}
= \frac{\mu_0}{\kappa^2 c^2} \left( \mathbf{I} - \frac{1}{\kappa^2} \nabla \nabla \right) \frac{e^{-\kappa |\mathbf{r}|}}{4\pi |\mathbf{r}|}
= \frac{\mu_0}{\kappa^2 c^2} \mathbf{G}_\kappa (\mathbf{r}, \mathbf{r}') \] (3.82)
where $\overline{G}_\kappa (r, r')$ is the homogeneous dyadic Green’s function for imaginary frequencies.

The partition function thus becomes

$$Z(\kappa) = W_0 \int \mathcal{D} \prod_i \{J_i(r)\} e^{-\frac{1}{2\beta} \frac{\mu_0}{\kappa^2 c^2} \int dr \int dr' J^T(r) \overline{G}_\kappa (r, r') \cdot J(r')} \tag{3.83}$$

where $J^T = \begin{bmatrix} J^T_1 & J^T_2 & \cdots \end{bmatrix}$. The above can still be evaluated using the same Gaussian integration methods as before. To do this, the currents are expanded over a set of basis functions.

$$J(r) = \sum_{n=1}^{N} J_n f_n(r) \tag{3.84}$$

The functional integration over the continuous surface currents becomes an $N$-dimensional integration over the possible values of the $N$ coefficients of the basis expansion. The change of variables in the integration is

$$\int \mathcal{D} \{J\} \rightarrow J \int \prod_n d J_n \tag{3.85}$$

where $J$ is the Jacobian of the transformation which once again will be removed via normalization. The partition function becomes

$$Z(\kappa) = W_0 J \int \prod_n d J_n e^{-\frac{1}{2\beta} \frac{\mu_0}{\kappa^2 c^2} J^T \cdot \left[ \int dr \int dr' f(r) \cdot \overline{G}_\kappa (r, r') \cdot f(r') \right] \cdot M \cdot J} \tag{3.86}$$

where $J$ now holds the coefficients $J_n$ and the elements of $\overline{M}$ are defined as

$$\overline{M}_{i,j} = \int dr \int dr' f_i(r) \cdot \overline{G}_\kappa (r, r') \cdot f_j(r') \tag{3.87}$$
Via Gaussian integration [70], the above becomes

\[ Z(\kappa) = W_0 \mathcal{J} \left[ \frac{\mu_0 \det \overline{M}}{\beta^2 \kappa^2 c^2} \right]^{-\frac{1}{2}} \]  

(3.88)

Substituting the above into Equation (3.37) and performing the various cancellations from the normalization with \( Z_\infty \), the Casimir energy becomes

\[ \mathcal{E} = \frac{\hbar c}{2\pi} \int_0^\infty d\kappa \ln \frac{\det \overline{M}}{\det \overline{M}_\infty} \]  

(3.89)

If the RWG basis is chosen as the basis function by which the currents are expanded, then \( \overline{M} \) becomes the traditional EFIE impedance matrix using imaginary frequencies. To find the Casimir force, the gradient of the Casimir energy is taken with respect to the displacement of the \( i \)-th object over which the force is to be found.

\[ \overline{F}_i = -\frac{\hbar c}{2\pi} \int_0^\infty d\kappa \nabla_i \ln \frac{\det \overline{M}}{\det \overline{M}_\infty} = -\frac{\hbar c}{2\pi} \int_0^\infty d\kappa \text{Tr} \left[ \overline{M}^{-1} \cdot \nabla_i \overline{M} \right] \]  

(3.90)

The evaluation of the determinants and traces numerically is troublesome due to the size of the impedance matrix. However, the calculation can be done equivalently using eigenvalue solvers. For the Casimir energy,

\[ \ln \frac{\det \overline{M}}{\det \overline{M}_\infty} = \sum_{i=1}^N \ln \frac{\lambda_i}{\lambda_{\infty,i}} \]  

(3.91)

where \( \lambda \) and \( \lambda_\infty \) are the eigenvalues of \( \overline{M} \) and \( \overline{M}_\infty \) respectively. For the Casimir force,

\[ \text{Tr} \left[ \overline{M}^{-1} \cdot \nabla_i \overline{M} \right] = \sum_{i=1}^N \lambda_i \]  

(3.92)

where \( \lambda \) are the eigenvalues of the generalized eigenvalue problem

\[ \nabla_i \overline{M} \cdot \mathbf{x}_i = \alpha_i \overline{M} \cdot \mathbf{x}_i \]  

(3.93)
The semi-infinite integration over the imaginary wavenumbers can be performed numerically using a Gauss-Laguerre quadrature. This formulation was first done specifically for PEC objects but it has been extended to dielectric objects using an equivalent derivation [9]. The result is that the Casimir energy and force are the same as in Equations (3.89) and (3.90) but the matrix $\mathbf{M}$ is now the PMCHWT impedance matrix.

### 3.4 New Formulation

The RRWJ method provides the framework for easily calculating the Casimir energy and force using pre-existing EFIE MOM codes. However, the framework of the derivation is complicated and rather stringent. The effective action that is used for the path integral is formulated only for the cases of PEC objects where the induced currents are restricted to the surfaces. In addition, the mathematical tool used to equate the path integral over the current coefficients with the determinant of the impedance matrix assumes that the impedance matrix is either positive or negative definite (see Chapter 9 of [70]). To reformulate the impedance matrix for different materials, the derivation must be redone with a new action [9]. In addition, there are some impedance matrices that may be useful but are not positive or negative definite matrices, like the A-EFIE. This section will present a new way of deriving the Casimir energy to address these shortcomings. First, the constraint on the matrix being definite for PEC objects can be eliminated by working backwards from the RRWJ Casimir energy equation. Second, the RRWJ derivation can be avoided altogether and an equivalent result arrived at using a simpler derivation that has the desired flexibility.

These alternative methods come about through the argument principle which allows a contour integral to be related with a summation over the enclosed poles and zeroes. Specifically, the argument principle states that

$$
\frac{1}{2\pi i} \oint \phi(\omega) \frac{d}{d\omega} \ln f(\omega) d\omega = \sum_i \phi(\omega_{0,i}) - \sum_j \phi(\omega_{\infty,j})
$$

(3.94)
where $\omega_{0,i}$ are the zeros and $\omega_{\infty,j}$ are the poles of the function $f(\omega)$ inside the counter-clockwise contour of integration. From the arguments of causality, it is known that the dyadic Green’s function is analytic in the upper-half of the complex frequency plane. The zeroes and poles, if any, lie in the lower-half plane. The expression for the Casimir energy derived by the RRWJ method, given as

$$E = \frac{\hbar c}{2\pi} \int_{0}^{\infty} d\kappa \ln \frac{\det \mathbf{M}(\kappa)}{\det \mathbf{M}_\infty(\kappa)}$$  \hspace{1cm} (3.95)$$

with the elements of the matrix $\mathbf{M}$, being

$$[\mathbf{M}]_{ij} = \int \int f_i(r) \cdot \mathbf{G}_\kappa(r, r') \cdot f_j(r')dr'dr$$ \hspace{1cm} (3.96)$$

are of the same form as the matrix elements of the traditional CEM impedance matrix for the method of moments. Note then that the matrix $\mathbf{M}$ is a projection of the dyadic Green’s function onto a subspace, a matrix representation in other words. The integration path can be unfolded along the imaginary frequency axis to extend from $-\infty$ to $+\infty$ and then closed to enclose the right-hand plane as seen in Figure 3.2. With the normalization of the integrand in Equation (3.95), the integrand approaches zero in the limit as the frequency goes to $\pm \infty$ and $\pm i\infty$. As such, Jordan’s lemma states that the semicircular integration path that closes the imaginary frequency axis evaluates to zero. The closed contour integration still evaluates to the same Casimir energy as the original integral in Equation (3.95). In the argument principle integral, let $\phi(\omega) = \omega$ and

$$f(i\kappa) = \frac{\det \mathbf{M}(\kappa)}{\det \mathbf{M}_\infty(\kappa)}$$ \hspace{1cm} (3.97)$$
Thus, making use of the imaginary frequency where $\omega = ic\kappa$, then

$$ -2\pi i \sum_i \phi(\omega_{0,i}) + 2\pi i \sum_j \phi(\omega_{\infty,j}) = \int_0^\infty \phi(\omega) \frac{d}{d\omega} \ln f(\omega) d\omega = \int_{-\infty}^{\infty} \omega \frac{d}{d\omega} \ln f(\omega) $$

$$ = \int_{-\infty}^{\infty} ic\kappa \frac{d}{icd\kappa} \ln f(ic\kappa) icd\kappa = ic \int_{-\infty}^{\infty} \kappa \frac{d}{d\kappa} \ln f(ic\kappa) d\kappa $$

$$ = ic\kappa \ln f(ic\kappa)|_{-\infty}^{\infty} - ic \int_{-\infty}^{\infty} \ln f(ic\kappa) d\kappa $$

$$ = -ic \int_{-\infty}^{\infty} \ln f(ic\kappa) d\kappa \quad (3.98) $$

In other words,

$$ \int_{-\infty}^{\infty} \ln \left( \frac{\det M(\kappa)}{\det M_\infty(\kappa)} \right) d\kappa = \frac{2\pi}{c} \left[ \sum_i \phi(\omega_{0,i}) + \sum_j \phi(\omega_{\infty,j}) \right] = \frac{2\pi}{c} \left[ \sum_i \omega_i + \sum_j \omega_{\infty,j} \right] \quad (3.99) $$

where $\omega_{0,i}$ are the zeros of $\det M$ and $\omega_{\infty,j}$ are the zeros of $\det M_\infty$. So the Casimir energy found using Equation (3.95) can be related to being equivalent to a summation over the zeros of the determinants of the impedance matrices. The frequencies where the
determinants become zero are also where the impedance matrix becomes singular. As such, one can replace the EFIE impedance matrices with any set of matrices that shares the exact same poles and zeroes.

The above result justifies the use of a larger set of impedance matrices, including ones that are not positive or negative definite, to find the Casimir force but it is still restricted by the fact that we can only do so with matrices that have the same poles and zeroes as the original EFIE impedance matrix. In addition, the derivation still suffers from the unfortunate problem where a new action in the path integral is needed to use materials other than PEC. The solution once again lies in the use of the argument principle.

Note that the Casimir force results from the relative perturbation of the quantum vacuum fields by the objects in question. The objects force the fluctuating vacuum fields to conform to the appropriate boundary conditions and in doing so change the energy density of the vacuum. Equivalently, the presence of the scatterers causes a change in the local density of states in the vacuum field. If one can determine the eigenfrequencies of the field configurations that can satisfy the geometry’s boundary conditions, then the unnormalized energy of the geometry can be found via [43,44]

\[ E_{\text{vac}} = \sum_{\omega} \frac{1}{2} \hbar \omega \quad (3.100) \]

The above is the divergent vacuum energy and as such renormalization of the vacuum energy to avoid divergence is needed. To do so, the modes of the normalizing configuration where, for example, the objects are taken to be infinitely separated, can be subtracted from the original energy density. That is,

\[ \mathcal{E} = E_{\text{vac}} - E_{\text{norm}} = \sum_{i,j} \frac{1}{2} \hbar [\omega_i - \omega_{j,\text{norm}}] \quad (3.101) \]

The above relates back to Equation (3.99) that relates the integral of the RRWJ Casimir energy with the summation over the zeros of the determinants of the impedance matrices. It can be related directly back to the energy defined in Equation (3.100) by choosing
\[ \phi(\omega) = \frac{\hbar}{2} \omega. \] Thus,

\[
\mathcal{E} = \frac{\hbar}{2} \left[ \sum_i \omega_i - \sum_j \omega_{\infty,j} \right] = \frac{1}{2\pi i} \oint \phi(\omega) \frac{d}{d\omega} \ln \frac{f(\omega)}{f_{\text{norm}}(\omega)} d\omega
\]

\[
= \frac{\hbar}{4\pi i} \oint \omega \frac{d}{d\omega} \ln \frac{\det \mathbf{M}(\omega)}{\det \mathbf{M}_\infty(\omega)} d\omega = \frac{\hbar c}{2\pi} \int_0^\infty \ln \frac{\det \mathbf{M}(\kappa)}{\det \mathbf{M}_\infty(\kappa)} d\kappa \quad (3.102)
\]

The use of the argument principle to change the summation over the modes to a contour integration is a well known method and from here it will be necessary to detail a brief literature review of the technique. It was first used for the dielectric case in van Kampen, Nijboer, and Schram’s paper [4]. Van Kampen et al. derived the Casimir force between two dielectric slabs using the argument principle. The choice of the function \( f(\omega) \) was taken as a dispersion relation that describes the allowed modes of the system and equates to zero at an eigenfrequency. Van Kampen et al. then showed that their results were the same as found by Lifshitz [3]. Lifshitz’s process is generally taken as the standard benchmark because his treatment of the problem specifically accounted for lossy media and derived its result using fluctuation dissipation.

There are a few downsides to the van Kampen treatment of the problem. The first was that in the van Kampen et al. paper, retardation effects were ignored. This was later corrected in such papers like Schram’s [46]. The second point is that the dispersion function \( f(\omega) \) can contain branch points in the complex frequency plane. This can be avoided by first assuming that the problem is contained in a finite cavity, finding the energy of the cavity by summing up over all the modes, and then taking the limit of this energy as the cavity becomes infinite [43]. Finally, and this is curiously absent from the paper, is that this interpretation of the energy density requires that the system be lossless.

Going back to the introduction of the Casimir vacuum in B.1, note that the vacuum modes came about by the treatment of the vacuum as being contained within a PEC cavity. The energy density is dictated by the allowed modes of the cavity and taking the size of the cavity to infinity would convert the summation over the modes to a continuous integration over the density of states. However, this interpretation is no longer valid if lossy materials are placed within the cavity as the modes now become complex. The energy of a complex
frequency becomes indeterminate and one expects the summation over the complex modes to yield a complex number.

Despite the inappropriateness of using Equation (3.100) for the Casimir energy in lossy systems, the result derived using the argument principle contour still agrees with more suitable derivations like Lifshitz [41,43,71,72]. A full treatment of the explanation has not been found; however, Sernelius published a paper with an explanation but there has not been much subsequent discussion of the article [72]. Another recent paper by Intravaia and Behunin show that the energy can be taken as the summation over complex frequencies by analyzing the case of a medium coupled to a bath [73]. Barash and Ginzburg in a series of papers found that one could calculate the Casimir energy of a system using an integral like the argument principle [44,71]. The vacuum energy can be described as the energy of a collection of harmonic oscillators [74]. This description is not surprising given our previous treatment of the vacuum but the physical reason is different. A dielectric medium can be treated as a density of harmonic oscillators, specifically by treating the permittivity as a vacuum permittivity added to a distribution of oscillators. The energy of the oscillators that makes up what is again the polarization field can be calculated to find the energy of the vacuum fields using the fluctuation-dissipation theorem. That is, thermal energy couples with the oscillators in the dielectric exciting their fluctuations which the oscillators duly dissipates. The fluctuation-dissipation theorem describes the coupling between the fluctuations of the oscillators and their dissipative properties.

To find the energy of a lossy system, we imbue the oscillators with a loss and find their energy once again. Barash and Ginzburg did this by introducing the idea of coupling with an “auxiliary” oscillator. For example, they imagined encasing the electromagnetic system inside another cavity. The modes of this auxiliary system can be found to be real and taking the summation over the modes gives the resulting energy. For the electromagnetic case, the auxiliary problem is described by the new Faraday’s and Ampere’s laws,

\[
\nabla \times E_a = i \frac{\omega_a}{c} B_a \\
\nabla \times B_a = -i \frac{\omega_a}{c} \epsilon(\omega) E_a
\n\]
Here, $\epsilon$ is the complex permittivity and the frequency $\omega$ is taken as a parameter. The auxiliary system is solved as driven at the frequency $\omega_a$ as indicated by the subscript $a$ on the field vectors. The advantage of the auxiliary system is that the oscillators used to find the energy of the fields are now normal modes as opposed to the eigenmodes of the oscillators in the original lossy system. Thus, the energy can be calculated by summing up the energies of the individual oscillators. Each oscillator represents a single eigenmode of the system where the dispersion relation is $\omega_a^2(\omega) - \omega^2 = 0$. The energy becomes

$$
\mathcal{E} = \frac{i}{2\pi} \sum_a \int_{-\infty}^{\infty} d\omega \phi(\omega, T) \frac{\partial}{\partial \omega} \ln \left[ \omega_a^2(\omega) - \omega^2 \right] = \frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega \phi(\omega, T) \frac{\partial}{\partial \omega} \ln \prod_a \left[ \omega_a^2(\omega) - \omega^2 \right]
$$

$$
= \frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega \phi(\omega, T) \frac{\partial}{\partial \omega} \ln f(\omega)
$$

(3.105)

where $f(\omega)$ is the dispersion relation for the auxiliary system and the energy of the modes, $\phi(\omega, T)$, can include temperature effects. But the auxiliary system is mathematically equivalent to the original problem when $\omega_a = \omega$ when looking at Equations (3.103) and (3.104). So the zeros for the dispersion relation in the right-hand plane are same zeros in a dispersion relation that is designed for the original lossy system. The added bonus of this derivation is that the problems with the branch cuts, like in the van Kampen derivation, no longer exist. It is therefore tempting to try and use Equation (3.105) to find the Casimir energy by enclosing a contour to use the argument principle to generalize the integrand to any equivalent dispersion relation. The problem lies in that the zeros in the left-hand plane should not be included since under these conditions the lossy permittivity behaves like an active media when $\omega_a = -\omega$. In addition, the integration is along the real frequency axis as opposed to the imaginary frequency axis. Still, the Barash-Ginzburg result shows a way of bridging the gap between the Casimir energy derived using fluctuation-dissipation theorem (used to find the energy of the oscillators) with the van Kampen method that does not take into account absorption.

Rosa, Dalvit, and P.W. Milonni published a set of papers that looked at calculating the energy of the electromagnetic fields in various media using various methods, both classical and quantum electrodynamically [74, 75]. Of interest is the fact that the energy of the
fields in a homogeneous lossy dielectric can be shown to be dependent only on the real part of the permittivity as a function of the frequency. That is, assuming that the permittivities satisfy the Kramers-Kronig relation, then the energy of fields in a lossy medium can be calculated using only the knowledge of the real part of the permittivity. The second paper deals with the case of inhomogeneous dielectrics, the case of true interest for the Casimir force. Once again, the authors conclude that the correct energy density of a lossy system can be found without assumption of absorption. This can be done by replacing the complex permittivity and dyadic Green’s functions with real values (specifically by taking the Wick rotation to move to purely imaginary frequencies, more on that below) [75]. In addition, they reaffirm that the energy density can be taken as the integral over the zero-point energies and their density of states. That is, they generalize Equation (3.100) to an infinite volume and define the energy density as

\[ u(r) = \int_{0}^{\infty} \rho(r, \omega) \frac{1}{2} \hbar \omega \]  

(3.106)

where \( \rho(r, \omega) \) represents the local density of states that is dependent upon the Wick rotated permittivity and dyadic Green’s function.

While there has been variance in the detailed explanations of why the argument principle integration gives the correct Casimir energy despite complex modes, authors are generally in agreement in the overall reasoning which we briefly touched above. In essence, the complex permittivity relates the real and imaginary parts using a Hilbert transform called the Kramers-Kronig relation [57]. That is,

\[ \epsilon'(\omega) = \epsilon_0 + \frac{1}{\pi} \int_{-\infty}^{\infty} dx \frac{\epsilon''(x)}{x - \omega} \]  

(3.107)

\[ \epsilon''(\omega) = -\frac{1}{\pi} \int_{-\infty}^{\infty} dx \frac{\epsilon'(x) - \epsilon_0}{x - \omega} \]  

(3.108)

where we have assumed that \( \epsilon(\infty) = \epsilon_0 \). The reasoning behind this is that electromagnetic radiation interacts to lessening degrees at higher frequencies until there is no longer any interaction; this physical condition is used by Casimir to provide a cutoff in his integration of the Casimir energy density [40]. From the above Equations (3.107) and (3.108), one can
see that the imaginary part of the permittivity impresses itself onto the real part via the Hilbert transform. In addition, assuming purely imaginary frequencies, then

\begin{align}
\epsilon'(i\omega) &= \epsilon_0 + \frac{2}{\pi} \int_0^\infty \frac{x\epsilon''(x)}{x^2 + \omega^2} \\
\epsilon''(i\omega) &= 0
\end{align}

(3.109)  
(3.110)

Hence, the permittivity is purely real at imaginary frequencies and includes information about the imaginary part. For this reason, when evaluating the energy density in the integrand of the Casimir energy, the system behaves like it is a lossless system since the complex permittivities become real valued. Thus, despite the fact that the eigenfrequencies are still complex, the contour integral of the argument principle becomes an integration along the imaginary frequency axis that has a real valued integrand.

Two things can be concluded from the previous brief literature review. First, it can be expected that the integration along the imaginary frequency of the energy density for the Casimir energy will always be real valued if the complex permittivities follow Kramers-Kronig. Second, that it has been shown by Rosa et al. in [75] that in an inhomogeneous medium that the energy density can be of the form assumed in Equation (3.100) and that the complex permittivity can be dealt with by the Wick rotation. Thus, through the use of the argument principle, the Casimir energy can be expressed as

\begin{equation}
\mathcal{E} = \frac{\hbar c}{2\pi} \int_0^\infty \ln \frac{f(\kappa)}{f_{\text{norm}}(\kappa)} d\kappa
\end{equation}

(3.111)

where \( f(\kappa) \) is any valid dispersion relation of a lossy system evaluated at \( \omega = i\kappa \).

Physically, the dispersion relation evaluates to zero at the frequencies at which the fluctuating electromagnetic fields can configure themselves to satisfy the relevant boundary conditions. This can be related to the EFIE impedance matrix by noting that the impedance matrix relates the surface currents to the excitation fields via

\begin{equation}
\bar{Z} \cdot \mathbf{J} = \mathbf{V}
\end{equation}

(3.112)
where $\mathbf{Z}$ is the impedance matrix, $\mathbf{J}$ is the vector representation of the currents, and $\mathbf{V}$ is the vector representation of the exciting fields. Since the eigenmodes of interest are those where the fluctuating fields can conform themselves to the boundary conditions, these eigenmodes exist where we can have currents on the surface of the objects without the need for any excitation fields. These natural modes automatically satisfy the geometry’s boundary conditions and thus satisfy the relation

$$\mathbf{Z} \cdot \mathbf{J} = 0 \quad (3.113)$$

Being a singular matrix, it can be concluded that $f(\omega) = \det \mathbf{Z} = 0$ for the frequencies where the vacuum fields can arrange themselves to satisfy the boundary conditions without the need for an external source. Using this new function for the argument principle’s Casimir energy, applying the Wick rotation, and normalizing the energy with the energy of the problem when the objects are infinitely separated, we rederive the RRWJ Casimir energy given in Equation (3.95). Thus, it has been shown how the RRWJ result can be derived using only the argument principle contour integral. The advantage here is that any equivalent impedance matrix can now be used. All that is needed to do is to find a similar matrix relation between the induced sources in a system and the external excitations. This includes matrix problems for arbitrary materials and matrices that are non-symmetric and neither positive or negative definite.

### 3.5 Low Frequency Methods

The EFIE used in the PEC impedance matrix case suffers from the drawback of low frequency breakdown. The problem manifests itself when solving problems involving the impedance matrix at frequencies where the electrical size of the mesh elements becomes very small, around $10^{-8}$ of the wavelength when using the RWG basis [59] at double precision [76, 77]. In Casimir simulations, initial investigations found that the low frequency breakdown would occur between two PEC spheres using dense EFIE matrices.
solved via LU decomposition when the edge lengths transitioned from 1.25 m to 1.00 m at a frequency of 5.0 MHz with spheres of radius of 1 m. While it is true that the scales of interest to researchers are on the order of the micrometer and nanometer, many problems like the case of spheres can be made scale invariant. Thus, we can use meshes of the order of m, \(\mu\)m, or nm and still derive the same general result. Only a broadband, multi-scale algorithm can give the researcher the freedom to choose their scaling and meshing.

Essentially the problem lies in the fact that at low frequencies, the scalar potential contribution to the EFIE becomes overwhelmed by the vector potential contribution due to the finite numerical precision. This causes the EFIE to develop a numerical null-space that creates difficult in numerically inverting the matrix, or equivalently an eigenvalue problem. This presents an obstacle to finding the Casimir energy and force since the calculation involves integrating across the frequency range and solving eigenvalue problems. The case is particularly exacerbated by the fact that the low frequency region of the integrand contributes the most. What follows is a brief investigation into the reason for the low frequency breakdown and solutions to the problem that can be implemented in the calculations.

The EFIE operator when acting on a continuous current is

\[
\mathcal{L}(\mathbf{r}, \mathbf{r}') \mathbf{J}(\mathbf{r}') = i\omega \mu \int_S g(\mathbf{r}, \mathbf{r}') \mathbf{J}(\mathbf{r}') d\mathbf{r}' - \frac{\nabla}{i\omega \epsilon} \int_S g(\mathbf{r}, \mathbf{r}') \nabla' \cdot \mathbf{J}(\mathbf{r}') d\mathbf{r}'
\]

(3.114)

The current can be expanded using the RWG basis and the elements of the impedance matrix for the MOM problem becomes

\[
Z_{m,n} = i\omega \mu \int_{S_m} \Lambda_m(\mathbf{r}) \cdot \int_{S_n} g(\mathbf{r}, \mathbf{r}') \Lambda_n(\mathbf{r}') d\mathbf{r}' d\mathbf{r} \\
+ \frac{1}{i\omega \epsilon} \int_{S_m} \nabla \cdot \Lambda_m(\mathbf{r}) \int_{S_n} g(\mathbf{r}, \mathbf{r}') \nabla' \cdot \Lambda_n(\mathbf{r}') d\mathbf{r}' d\mathbf{r}
\]

(3.115)
The above can be separated into the contributions from the scalar and vector potentials.

\[
L^V_{m,n} = \int_{S_m} \Lambda_m(r) \cdot \int_{S_n} g(r, r') \Lambda_n(r') \, dr \, dr' \tag{3.116}
\]

\[
L^S_{m,n} = \int_{S_m} \nabla \cdot \Lambda_m(r) \int_{S_n} g(r, r') \nabla' \cdot \Lambda_n(r') \, dr \, dr' \tag{3.117}
\]

Dimensional analysis can be carried out to relate the two contributions.

\[
L^V_{m,n} \sim \Delta_m \Delta_n L^S_{m,n} \tag{3.118}
\]

where \(\Delta_i\) is the edge length of the \(i\)-th basis function. As such,

\[
Z_{m,n} \approx \frac{\eta}{ik} \left( 1 - k^2 \Delta_m \Delta_n \right) L^S_{m,n} = \frac{\eta}{ik} \left( 1 - 4\pi^2 \frac{\Delta_m \Delta_n}{\lambda^2} \right) L^S_{m,n} \tag{3.119}
\]

What can be learned from the above is that when the size of the RWG basis function is much smaller than the wavelength, then the contribution from the scalar potential becomes swamped due to the finite numerical precision leaving a near singular impedance matrix. Notice also that if we were to take the case of a meter scaled problem and scale it down uniformly to nanoscale, then the edge lengths would scale by a factor of \(10^{-9}\), while at the same time the bandwidth over which the integration needs to be performed, defined as \(\alpha \kappa Z\) where \(\alpha\) is some constant and \(Z\) the separation of the objects, requires that the frequency scale by \(10^9\). This means that the wavelength scales as \(10^{-9}\) and thus scaling the problem down in size and up in frequency is not going to address the inherent scaling problems that give rise to the low frequency breakdown.

There are a few ways to avoid the low frequency breakdown. One way is with the loop-tree decomposition [14] whereby the loop-tree impedance matrix is the result of a matrix similarity transform of the traditional MOM impedance matrix. This transform, achieved using left and right matrix multiplications, is canceled out in the integrand for the Casimir energy when the determinant is normalized with \(M_\infty\), thus leaving us with a mathematically equivalent expression for the Casimir energy. However, the loop-tree decomposition is not suitable for medium frequency problems like the traditional EFIE. As
such, the user would need to switch between the loop-tree and traditional EFIE during the integration. Fortunately, there is another formulation that is suitable across both the low and mid frequency regions called the Augmented-Electric Field Integral Equation (A-EFIE) [11]. The A-EFIE impedance matrix is not a symmetric definite matrix nor a similarity transform of the EFIE impedance matrix like the loop-tree. Thus, we need to make use of the new derivation that was given in 3.4.

The A-EFIE starts from the normal EFIE formulations whose matrix representation is

\[
\left( ik_0 \eta_0 \nabla + \frac{\eta_0}{ik_0} \mathbf{S} \right) \cdot \mathbf{J} = \mathbf{V}
\]

(3.120)

Assuming that the subspace is spanned by the RWG basis, the problem has \(t\) independent surfaces, \(e\) RWG edges, and \(p\) patches. The vector and scalar matrices above, \(\nabla\) and \(\mathbf{S}\) respectively, and an additional matrix representation of the scalar Green’s function using patch basis, are defined to be

\[
\nabla_{m,n} = \mu_r \int_{S_m} \nabla_m(r) \cdot \int_{S_n} g(r, r') \Lambda_n(r')dS'dS
\]

(3.121)

\[
\mathbf{S}_{m,n} = \epsilon_r^{-1} \int_{S_m} \nabla \cdot \Lambda_m(r) \cdot \int_{S_n} g(r, r') \nabla' \cdot \Lambda_n(r')dS'dS
\]

(3.122)

\[
\mathbf{P}_{m,n} = \epsilon_r^{-1} \int_{S_m} h_m(r) \int_{S_n} g(r, r') h_n(r')dS'dS
\]

(3.123)

where \(g(r, r')\) is the homogeneous Green’s function, \(\Lambda(r)\) is an RWG basis that is not normalized against the edge length and \(h(r)\) is a pulse basis. The patch matrix, \(\mathbf{P}\), can be related to the scalar matrix, \(\mathbf{S}\), using the incidence matrix \(\mathbf{D}\) by

\[
\mathbf{S} = \mathbf{D}^T \cdot \mathbf{P} \cdot \mathbf{D}
\]

(3.124)

By taking into account current continuity,

\[
\overrightarrow{\mathbf{D}} \cdot \mathbf{J} = ik_0 c_0 \mathbf{\rho}
\]

(3.125)
the A-EFIE matrix system can be derived as

\[
\begin{bmatrix}
V & D^T \cdot P \\
D & k_0^2 I
\end{bmatrix} \cdot \begin{bmatrix}
 i k_0 J \\
c_0 \rho
\end{bmatrix} = \begin{bmatrix}
\eta_0^{-1} V \\
0
\end{bmatrix}
\] (3.126)

The given formulations for A-EFIE can be further improved upon, but with the greater flexibility of the derivation, it can easily be seen that the A-EFIE impedance matrix, called \( Z_A \), will have the same function \( f(\omega) = \det Z_A = 0 \) at the eigenfrequencies of the geometry. As such, the A-EFIE impedance matrix can simply be used as a direct replacement for the EFIE impedance matrix. The Casimir energy and force become

\[
E = \frac{\hbar c^2}{2\pi} \int_0^\infty d\kappa \ln \left( \frac{\det Z_A(\kappa)}{\det Z_A,\infty(\kappa)} \right)
\] (3.127)

\[
F = -\frac{\hbar c^2}{2\pi} \int_0^\infty d\kappa \nabla_i \ln \left( \frac{\det Z_A(\kappa)}{\det Z_A,\infty(\kappa)} \right)
\] (3.128)

where \( \nabla_i \) represents the derivatives with respect to a physical displacement of the \( i \)-th object (the one we wish to find the forces acting upon). We can find the integrand to the Casimir energy by solving for the eigenvalues of \( Z_A \) and \( Z_A,\infty \).

\[
\ln \left( \frac{\det Z_A(\kappa)}{\det Z_A,\infty(\kappa)} \right) = \sum_{n=1}^N \ln \left( \frac{\lambda_n}{\lambda_n^\infty} \right)
\] (3.129)

where \( \lambda_n \) and \( \lambda_n^\infty \) are the eigenvalues of \( Z_A \) and \( Z_A,\infty \) respectively. The Casimir force integrand can be expressed as

\[
\nabla_i \ln \det Z_A(\kappa) = \sum_{n=1}^N \alpha_n \hat{r}_i
\] (3.130)

where \( \alpha_n \) are the eigenvalues of the generalized eigenvalue problem

\[
\nabla_i Z_A \cdot \mathbf{x} = \alpha Z_A \cdot \mathbf{x} \hat{r}_i
\] (3.131)

It should be noted that the gradient acting upon \( Z_A \) acts on all elements of the matrix;
requiring that the bottom block matrices become zeros. That is,

\[
\nabla_i \mathbf{Z}_A = \begin{bmatrix} \nabla_i \mathbf{V} & \mathbf{D}^T \cdot \nabla_i \mathbf{P} \\ 0 & 0 \end{bmatrix}
\]

The performance of the normal A-EFIE can be further enhanced by the enforcement of charge neutrality. Keeping with the assumption that there are \( t \) independent surfaces, then the summation of the charges on each surface should equal zero in accordance with charge neutrality. To enforce this explicitly in the matrix problem, one of the unknown charges is dropped from each independent surface and the condition is mathematically added that the dropped unknown charge is equal and opposite to the sum of the surface’s remaining charges. This is done by defining two additional mapping matrices between the original charge vector and the reduced charge vector.

\[
\rho_r = \mathbf{F} \cdot \rho \quad (3.133)
\]
\[
\rho = \mathbf{B} \cdot \rho_r \quad (3.134)
\]

The matrix \( \mathbf{F} \) is of dimensions \((p - t) \times p\) and maps the full charge vector forward to the reduced one by simply removing the rows representing one charge from each independent surface. The matrix \( \mathbf{B} \) is of dimension \( p \times (p - t) \) and projects the reduced charge vector backward to the full one by taking the opposite sign of the sum of the charges on each surface to fill in the missing rows in the full charge vector. The mapping is much easier to implement using matrix-vector product rules as opposed to creating the actual sparse matrices. The new A-EFIE matrix problem becomes

\[
\begin{bmatrix} \nabla \mathbf{V} & \mathbf{D}^T \cdot \mathbf{P} \cdot \mathbf{B} \\ \mathbf{F} \cdot \mathbf{D} & k_0^2 \mathbf{I}_r \end{bmatrix} \cdot \begin{bmatrix} i k_0 \mathbf{J} \\ c_0 \rho_r \end{bmatrix} = \begin{bmatrix} \eta_0^{-1} \mathbf{V} \\ 0 \end{bmatrix}
\]

The above should be the actual implementation used in the Casimir energy and force codes as it was found to have a significant improvement at low frequencies. In the case of two PEC spheres, the condition number of the impedance matrix for the A-EFIE using
Equation (3.126) would surpass the EFIE impedance matrix at around 1 kHz. However, enforcing charge neutrality via Equation (3.135) corrected this problem. For example, at 6.3 Hz, the condition number for the EFIE matrix was $9.85 \cdot 10^8$, the A-EFIE matrix was $1.46 \cdot 10^{14}$, and the charge neutrality A-EFIE was a satisfyingly low $1.41 \cdot 10^3$ in the case of two PEC spheres. Both the RRWJ method and the stress tensor method benefit from using the A-EFIE. Finally, a preconditioner should be used for any iterative matrix solver. One that gives good results was presented in Qian and Chew’s paper and is as follows:

$$
\mathbf{M} = \begin{bmatrix}
\text{diag}(\mathbf{V}) & (\mathbf{F} \cdot \mathbf{D})^T \cdot \left( \mathbf{F} \cdot \text{diag}(\mathbf{P}) \cdot \mathbf{F}^T \right) \\
\mathbf{F} \cdot \mathbf{D} & k_0^2 \mathbf{I}_r
\end{bmatrix} = \begin{bmatrix}
\mathbf{V}_d & \mathbf{D}_r^T \cdot \mathbf{P}_{rd} \\
\mathbf{D}_r & k_0^2 \mathbf{I}_r
\end{bmatrix} \quad (3.136)
$$

$$
\mathbf{M}^{-1} = \begin{bmatrix}
\mathbf{V}_d^{-1} & 0 \\
0 & 0
\end{bmatrix} + \begin{bmatrix}
- \mathbf{V}_d^{-1} \cdot \mathbf{D}_r^T \cdot \mathbf{P}_{rd} \\
\mathbf{I}_r
\end{bmatrix} \cdot \mathbf{\Delta}^{-1} \cdot \begin{bmatrix}
- \mathbf{D}_r \cdot \mathbf{V}_d^{-1} \\
\mathbf{I}
\end{bmatrix} \quad (3.137)
$$

where the sparse Schur complement matrix is

$$
\mathbf{\Delta} = k_0^2 \mathbf{I}_r - \mathbf{D}_r \cdot \mathbf{V}_d^{-1} \cdot \mathbf{D}_r^T \cdot \mathbf{P}_{rd} \quad (3.138)
$$

Both matrix-vector products of $\mathbf{M}$ and $\mathbf{M}^{-1}$ are sparse operations and a fast direct sparse solver can be used to deal with the $\mathbf{\Delta}$ inversion [78]. Using dense matrices with the GMRES iterative solver [79], the inverse problems for the stress tensor method converged on average after 30 iterations but a high tolerance of $10^{-5}$ was needed for the convergence factor for satisfactory results.

**Results**

Using the A-EFIE, a more accurate calculation of the low frequency energy spectrum can be obtained than when using the EFIE. As an example for comparison, results from the RRWJ formulation are used. While in general, double precision can give satisfactory results, Figures 3.3a and 3.3b show that the A-EFIE and EFIE results start to depart from each other as the frequency drops when using single precision for the case of the energy
Figure 3.3: Comparison of the EFIE and A-EFIE results for (a) Casimir energy integrand and (b) Casimir force integrand for two PEC spheres.

Figure 3.4: Comparison of the EFIE and A-EFIE results for (a) Casimir force between 2 PEC spheres and (b) Casimir force integrand for two rounded PEC capsules.
and force integrands between two PEC spheres of radius 1.0 m and separation of 4.0 m. The result from the numeric integral is calculated using the points indicated in the figure and the points with the largest contribution lie in the region of poor performance. The deviation in the calculated integration is a result of the location of the Gaussian points but the error in the EFIE result can be greater than 100% while the A-EFIE result remains in agreement. Figure 3.4b shows a comparison between the single precision A-EFIE and single precision EFIE with double precision results as a benchmark between two rounded PEC capsules of radius 1.0 m, length 6.0 m, and separation of 4.0 m. Finally, Figure 3.4a presents a comparison of the Casimir force for the PEC spheres as a function of the separation.

### 3.6 Large-Scale Methods

In the stress-tensor formulation, we need to solve five inverse matrix problems for each integration point in the surface and frequency integration. Direct methods have a computational complexity of $O(N^3)$ and memory usage of $O(N^2)$ and this presents the primary limiting factor in terms of the size and complexity of the problem. These methods are generally limited to meshes of about 10,000 unknowns for current PCs. Using the RRWJ method, finding the Casimir energy and force, the problem becomes one of solving for eigenvalues. For the Casimir energy we must solve the following problem:

$$\ln \frac{\det \mathbf{M}}{\det \mathbf{M}_\infty} = \sum_{n=1}^{N} \ln \frac{\lambda_n}{\lambda_n^\infty}$$

where $\lambda_n$ and $\lambda_n^\infty$ are the $n$-th eigenvalues of the matrices $\mathbf{M}$ and $\mathbf{M}_\infty$ respectively. For the Casimir force we need to find

$$\nabla_i \ln \det \mathbf{M} = \text{Tr} \left[ \mathbf{M}^{-1} \cdot \nabla_i \mathbf{M} \right] = \sum_{n=1}^{N} \alpha_n \hat{r}_i$$
where $\alpha_n$ is the $n$-th eigenvalue of the generalized eigenvalue problem

$$\nabla_i \mathbf{M} \cdot \mathbf{x}_n = \alpha_n \mathbf{M} \cdot \mathbf{x}_n \hat{r}_i$$  \hfill (3.141)

The computational complexity of these eigenvalue problems, like finding the inverse matrix, is of $O(N^3)$ when they are solved using direct methods and the challenge becomes finding a way to calculate these energies and forces by using a fast matrix-vector product like the fast multipole algorithm (FMA) [12–14]. By using a suitable FMA, the computational complexity and memory size of solving the matrix inversion problem become $O(N \log N)$ as opposed to $O(N^3)$. The FMA used in these calculations is the mixed form-fast multipole algorithm (MF-FMA) [12] that combines the multilevel fast multipole algorithm (MLFMA), used for mid-frequency problems and the low frequency-fast multipole algorithm (LF-FMA) for low frequencies. In this manner a single FMA code can be used across the frequencies of interest for the integrations.

Using the MF-FMA to solve the inverse problem is a suitable direct replacement for the direct methods in the stress-tensor formulation. However, for the RRWJ formulation, the problem exists whereby solving for all the eigenvalues exhaustively, or calculating the inverse product for the force instead, would have a computational complexity of $O(N^2 \log N)$. However, looking at the eigenvalue spectrum of the impedance matrices, one sees that a large part of the spectrum is vanishingly small in comparison to the extremum values. For example, the eigenvalue spectrum can be analyzed for one of the integration points for the force between two PEC spheres. In Figure 3.5a is the eigenvalue spectrum and Figure 3.5b shows the value of the sums between the pairs of the extreme eigenvalues. These figures show that the magnitude of the eigenvalues drops off dramatically after the first dozen or so eigenvalues. In addition, the spectrum is almost symmetric and this symmetry is strengthened by the similarity of the objects. Each eigenvalue of one magnitude has an associated partner of nearly the same magnitude but opposite in sign. Thus, by summing each pair together the magnitude of the sums drops off quickly as well.

This means that a good approximation of the sum of the eigenvalues can be achieved by adding together only a small subset of the eigenvalue pairs. This makes an iterative
Figure 3.5: For the case of the force between two PEC spheres, (a) the eigenvalue spectrum and the (b) values of the sum between opposite pairs of eigenvalues.

eigenvalue solver an attractive solution to the problem of accelerating the RRWJ formulation. Instead of finding all \( N \) eigenvalues, we only need to find a small number, typically on the order of 10-20. Several possible eigenvalue solvers were investigated but the EFIE impedance matrix can make use of a few properties to its advantage. First, the impedance matrix is theoretically symmetric. This is not true in practice because of how the integration between nearby basis functions is handled. For basis functions that are near to each other or overlap, the \(|\mathbf{r} - \mathbf{r}'|\) denominator in the homogeneous Green’s function becomes nearly singular or singular. The singularity is still an integrable singularity and the solution is to calculate the near-field integrations in closed form [80–82]. The closed form integration is often done only for the integration over the basis functions and not for the integration for both basis and testing functions. As such, the off-diagonal near-field elements will be calculated numerically in such a way that they are no longer guaranteed to be symmetric. A simple way of changing this to a symmetric matrix is to sum the impedance matrix with its transpose and take the average.

A second property that can be taken advantage of is that the impedance matrix at imaginary frequencies is negative definite. For the purposes of future calculations, one can take the opposite sign of the impedance matrix to make it positive definite. The sign
change does not affect the resulting energy or force calculations due to the normalization with $\mathbf{M}_\infty$. As such, an iterative eigenvalue solver for symmetric positive definite matrices can be used for the Casimir energy calculation and an iterative generalized eigenvalue solver where the $\mathbf{B}$ matrix is symmetric positive definite can be used for the force calculation. Most iterative solvers make use of finding the inverse of the $\mathbf{B}$ matrix. This can be done using an iterative inverse solver like GMRES [79]. Inverse solvers were rejected due to the fact that it introduces another approximation into the process. When finding eigenvalues, it is only possible to find one or a few eigenvalues at a time. In order to find more, the eigenvalue problem uses the approximated eigenvalues and eigenvectors to deflate the original matrix. For example, the generalized eigenvalue problem is

$$\overline{\mathbf{A}} \cdot \mathbf{x} = \lambda \mathbf{B} \cdot \mathbf{x}$$ (3.142)

Using an iterative method, the first $j$ eigenvalues and eigenvectors are found and to find the next set of $j$ eigenvalues of interest, the previously found values are shifted in the problem’s eigenspace to a region where they are no longer in the space being searched by the eigenvalue solver. This is done by solving the deflated eigenvalue problem,

$$\overline{\mathbf{A}}_D \cdot \mathbf{x} = \lambda \mathbf{B} \cdot \mathbf{x}$$ (3.143)

where $\overline{\mathbf{A}}_D = \overline{\mathbf{A}} + \mathbf{B} \cdot \overline{\mathbf{X}}_j \cdot \Sigma \cdot \overline{\mathbf{X}}_j^* \cdot \mathbf{B}$ with $\overline{\mathbf{X}}_j$ being a matrix whose column vectors are the $j$ eigenvectors of interest and $\Sigma$ is a diagonal matrix with entries of $\lambda_1 + \sigma_1$, $\lambda_2 + \sigma_2$, $\ldots$, $\lambda_j + \sigma_j$. Here, $\sigma$ is the value by which the associated eigenvalue is shifted. If the largest eigenvalues are being found, then $\sigma$ should be chosen to shift the known eigenvalues to a large negative value. In this manner, the largest eigenvalues of the deflated problem become the next block of eigenvalues of interest.

With the matrix deflation, it becomes crucial to accurately estimate the eigenvalues and eigenvectors of the system so that the deflated system’s results do not correspondingly diverge. In addition, each matrix inversion requires its own set of iterations, usually found to be on the order of 10-30 iterations using preconditioners for the Casimir problems. As
such, the inverse solver needs to run with very strict tolerances and the increase in iterations becomes undesirable.

**EFIE Implementation**

With the EFIE, the generalized eigenvalue problem has the advantage of both matrices $\mathbf{A}$ and $\mathbf{B}$ being symmetric and matrix $\mathbf{B}$ can be positive definite. Several iterative eigenvalue solver algorithms have been developed that can address such a problem, but the one that had the most favorable performance was an inverse-free preconditioned Krylov subspace method [83,84]. This Krylov subspace algorithm works by iteratively approximating an eigenvalue and eigenvector by using a Rayleigh-Ritz projection onto a Krylov subspace generated by a shifted matrix.

In more detail, a Krylov subspace is a space that is constructed iteratively using the Krylov sequence $\mathbf{x}_k = \mathbf{A} \cdot \mathbf{x}_{k-1}$ [31]. The Krylov matrix becomes,

$$
K_k = \begin{bmatrix}
\mathbf{x}_0 & \mathbf{A} \cdot \mathbf{x}_0 & \ldots & \mathbf{A}^{k-1} \cdot \mathbf{x}_0
\end{bmatrix}
$$

and the Krylov subspace is

$$
K_k = \text{span}\{\mathbf{x}_0, \mathbf{A} \cdot \mathbf{x}_0, \ldots, \mathbf{A}^{k-1} \cdot \mathbf{x}_0\}
$$

The advantage of using this subspace is that it can be easily built with each new column vector being orthonormalized, usually using a modified Gram-Schmidt method. Given an initial approximation of the eigenpair, the approximation is improved through the Rayleigh-Ritz orthogonal projection, essentially by minimizing the Rayleigh quotient,

$$
\lambda = \frac{\mathbf{x}^T \cdot \mathbf{A} \cdot \mathbf{x}}{\mathbf{x}^T \cdot \mathbf{B} \cdot \mathbf{x}}
$$
The gradient of $\lambda$ with respect to the eigenvector $\mathbf{x}$ is

$$
\nabla_x \lambda = \nabla_x \frac{x^T \cdot \bar{A} \cdot x}{x^T \cdot \bar{B} \cdot x}
= \left( \bar{A}^T \cdot \mathbf{x} + \frac{x^T \cdot \bar{A} \cdot x}{x^T \cdot \bar{B} \cdot x} \right) \frac{1}{x^T \cdot \bar{B} \cdot x} - \frac{x^T \cdot \bar{A} \cdot x \cdot (\bar{B}^T \cdot \mathbf{x} + \bar{B} \cdot \mathbf{x})}{x^T \cdot \bar{B} \cdot x} \cdot \frac{1}{x^T \cdot \bar{B} \cdot x \cdot x^T \cdot \bar{B} \cdot x}
= \left( 2 \bar{A} \cdot \mathbf{x} - \frac{x^T \cdot \bar{A} \cdot x}{x^T \cdot \bar{B} \cdot x} \cdot 2 \bar{B} \cdot \mathbf{x} \right) \frac{1}{x^T \cdot \bar{B} \cdot x}
= 2 \frac{(\bar{A} - \lambda \bar{B})}{x^T \cdot \bar{B} \cdot x} \frac{x}{x^T \cdot \bar{B} \cdot x}
$$

(3.147)

Thus, the correction parameter $r$ is taken to be

$$
r = \frac{(\bar{A} - \lambda \bar{B})}{x^T \cdot \bar{B} \cdot x} \frac{x}{x^T \cdot \bar{B} \cdot x}
$$

(3.148)

which can be used in a steepest descent method to choose a new approximate eigenpair. Using an initial eigenpair $(\lambda_0, \mathbf{x}_0)$, this can be considered as a Rayleigh-Ritz projection method on the subspace $K_1 = \text{span} \{ \mathbf{x}_0, (\bar{A} - \lambda_0 \bar{B}) \cdot \mathbf{x}_0 \}$. Another process for finding the eigenvalue is the inverse iteration. Using the inverse iteration, the new approximation of the eigenvector becomes $\mathbf{x}_1 = (\bar{A} - \lambda_0 \bar{B})^{-1} \cdot \bar{B} \cdot \mathbf{x}_0$. If this inversion is solved using an iterative solver, then the vector $\mathbf{x}_1$ is taken from a Krylov subspace generated by $\bar{A} - \lambda_0 \bar{B}$. This can be seen, for example, from the example of the GMRES algorithm. GMRES solves the problem of $\bar{A} \cdot \mathbf{x} = \mathbf{b}$ by building the Krylov subspace $K_{GMRES}^n = \text{span} \{ \mathbf{r}, \bar{A} \cdot \mathbf{r}, \ldots, \bar{A}^{n-1} \cdot \mathbf{r} \}$. The solution $\mathbf{x}_n$ is chosen from this Krylov subspace in order to minimize the Euclidean norm of the residual $\bar{A} \cdot \mathbf{x}_n - \mathbf{b}$ [85]. Since the new approximation is extracted from the Krylov subspace to solve the linear system, it may not be a good choice of approximating the eigenvector. To address this, Golub et al. chose to expand the search direction using a Krylov subspace.

In other words, the approximate eigenvector is chosen from the larger Krylov subspace of,

$$
K_m = \text{span} \{ \mathbf{x}_k, (\bar{A} - \lambda_k \bar{B}) \cdot \mathbf{x}_k, \ldots, (\bar{A} - \lambda_k \bar{B})^m \cdot \mathbf{x}_k \}
$$

(3.149)

where the user chooses the value $m$ and $k$ indicates the iteration number. The algorithm
starts with the user supplying an initial guess to the eigenpair and the value \( m \). Then in each \( k \)-th iteration, the above Krylov subspace is used as a Rayleigh-Ritz projection method to find the new eigenpair. The matrices \( \mathbf{A}_m \) and \( \mathbf{B}_m \) are created as projections onto this subspace via \( \mathbf{A}_m = \mathbf{Z}_m^T \cdot (\mathbf{A} - \lambda_k \mathbf{B}) \cdot \mathbf{Z}_m \) and \( \mathbf{B}_m = \mathbf{Z}_m^T \cdot \mathbf{B} \cdot \mathbf{Z}_m \). The correction factors for the eigenpairs are found by solving the generalized eigenvalue problem of the projections, that is \( \mathbf{A}_m \cdot \mathbf{v} = \mu \mathbf{B}_m \cdot \mathbf{v} \). The basic algorithm is shown in Algorithm 1 and will find the smallest eigenvalues. It can easily be modified to find the largest eigenvalues using \(-\mathbf{A}\). A \( B \)-orthonormal basis using the Arnoldi method was used to construct \( \mathbf{Z}_m \) as opposed to using a Lanczos method because the \( B \)-orthonormal space incorporates reorthogonalization of the \( \mathbf{Z}_m \) basis. This is done via an Arnoldi method that is similar to a modified Gram-Schmidt process. The construction of the basis space is shown in Algorithm 2 and note that \( \| \mathbf{x} \|_B = (\mathbf{x}^T \cdot \mathbf{B} \cdot \mathbf{x})^{\frac{1}{2}} \) is called the \( B \)-norm.

**Algorithm 1** The basic inverse free Krylov subspace algorithm.

Set initial guess \( \lambda_0 \) and \( \mathbf{x}_0 \) where \( \| \mathbf{x}_0 \| = 0 \).
Set the value \( m \) for the size of the \( m \times m \) matrices.

for \( k = 0 \) to convergence do
  Construct a basis \( \mathbf{Z}_m = [\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_m] \) for \( K_m = \text{span}\{ \mathbf{x}_k, (\mathbf{A} - \lambda_k \mathbf{B}) \mathbf{x}_k, \ldots, (\mathbf{A} - \lambda_k \mathbf{B})^m \mathbf{x}_k \} \);
  Form \( \mathbf{A}_m = \mathbf{Z}_m^T \cdot (\mathbf{A} - \lambda_k \mathbf{B}) \cdot \mathbf{Z}_m \) and \( \mathbf{B}_m = \mathbf{Z}_m^T \cdot \mathbf{B} \cdot \mathbf{Z}_m \);
  Find the smallest eigenpair \((\mu_1, \mathbf{v}_1)\) for \((\mathbf{A}_m, \mathbf{B}_m)\);
  \( \lambda_{k+1} = \lambda_k + \mu_1 \) and \( \mathbf{x}_{k+1} = \mathbf{Z}_m \cdot \mathbf{v}_1 \);
end for

**Algorithm 2** The \( B \)-orthonormal basis, \( \mathbf{Z}_m \), constructed by the Arnoldi method.

\[
\begin{align*}
\mathbf{C}_k &= \mathbf{A} - \lambda_k \mathbf{B}; \\
\mathbf{z}_0 &= \mathbf{x}_k / \| \mathbf{x}_k \|_B; \\
\text{for } i = 0 \text{ to } m-1 \text{ do} & \\
  \mathbf{w}_k &= \mathbf{C}_k \mathbf{z}_i; \\
  \text{for } j = 0 \text{ to } i \text{ do} & \\
    h_{j,i} &= \mathbf{z}_j^T \cdot \mathbf{B} \cdot \mathbf{w}; \\
    \mathbf{w} &= \mathbf{w} - h_{j,i} \mathbf{z}_j; \\
  \text{end for} & \\
  \mathbf{z}_{i+1} &= \mathbf{w} / \| \mathbf{w} \|_B; \\
\text{end for} & 
\end{align*}
\]
**Algorithm 3** The block inverse free Krylov subspace algorithm.

To solve for a block of \( p \) eigenpairs.

Set initial guess \( \vec{X}^{(1)} \) where \( \vec{X}^{(1)*} \cdot \vec{B} \cdot \vec{X}^{(1)} = \vec{I} \).

\( \Theta^{(1)} = \text{diag} \left( \vec{X}^{(1)*} \cdot \vec{A} \cdot \vec{X}^{(1)} \right) \);

for \( k = 1 \) to convergence do

for \( i = 1 \) to \( p \) do

Construct a basis \( \vec{Z}_i \) of \( K_m \left( \vec{A} - \Theta^{(k)} \cdot \vec{B}, \vec{x}^{(k)}_i \right) \);

end for

Orthonormalize \( (\vec{Z}_1 \ldots \vec{Z}_p) \) to obtain \( \vec{Z} \);

Form projection \( \vec{A}_m = \vec{Z}^* \cdot \vec{A} \cdot \vec{Z} \), \( \vec{B}_m = \vec{Z}^* \cdot \vec{B} \cdot \vec{Z} \)

Find the \( p \) smallest eigenpairs \( (\theta_i, \vec{u}_1) \) for \( (\vec{A}_m, \vec{B}_m) \);

\( \Theta^{(k+1)} = \text{diag} (\theta_1, \ldots, \theta_p) \);

\( \vec{X}^{(k+1)} = \vec{Z} \cdot \vec{U}, \quad \vec{U} = (\vec{u}_1 \ldots \vec{u}_p) \);

end for

**Algorithm 4** The preconditioned block \( B \)-orthonormal basis, \( \vec{Z}_m \), constructed by the Arnoldi method.

Set \( \vec{Z}^{(1)*} \cdot \vec{M} \cdot \vec{Z}^{(1)} = \vec{I} \).

for \( j = 1 \) to \( m \) do

\( \vec{W}_j = \vec{P}^{-1} \left( \vec{A} \cdot \vec{Z}_j - \vec{B} \cdot \vec{Z}_j \cdot \Theta \right) \);

for \( i = 1 \) to \( j \) do

\( \vec{H}_{ij} = \vec{Z}_i^* \cdot \vec{M} \cdot \vec{W}_j \);

\( \vec{W}_j = \vec{W}_j - \vec{Z}_i \cdot \vec{H}_{ij} \);

end for

Compute \( \vec{W}_j = \vec{Z}_{j+1} \cdot \vec{H}_{j+1,j} \), the QR factorization of \( \vec{W}_j \) with respect to the \( M \)-inner product.

end for
The algorithm can be further improved by the use of a preconditioner. However, the main problem with the above algorithm is that it only finds one eigenvalue at a time. This is undesirable because it introduces more error into the resulting eigenvalues since each successive eigenvalue works off of a deflated matrix. Each deflation introduces some error into the resulting eigenvalue. In addition, there is a level of repeated overhead in the algorithm that can be streamlined by working on multiple eigenvalues during the same steps. A subsequent paper was published that proposed a block inverse-free preconditioned Krylov subspace method [84]. The resulting algorithm is very similar except that now the diagonal matrix \( \Theta \) contains the estimated eigenvalues and the matrix \( X \) the eigenvectors in its columns. The implemented algorithms are shown in Algorithms 3 and 4. It should be noted that the eigenvalues do not converge on the same iteration. As such, the algorithm was modified so that when an eigenvalue converged, the eigenpair was removed from the \( \Theta \) and \( X \) matrices. This allows the algorithm to save on trying to improve the converged eigenpairs while at the same time keeping the Krylov subspace that was built. Restarting the algorithm whenever an eigenpair is found can be wasteful since the Krylov subspace containing information about the known and unknown algorithms is dropped, forcing work to be done to reconstruct the same information.

The matrix \( M \) in Algorithm 4 was chosen to be the \( B \) matrix and \( P \) is the preconditioner matrix that estimates \( A - \lambda B \) where \( \lambda \) is the estimated value where the block of eigenvalues reside. The preconditioner for the original algorithm was meant to be used with a sparse matrix. This would allow for sparse pseudoinverse techniques like the ILUT or incomplete Cholesky factorization. However, the impedance matrix in EFIE is a dense matrix and the large scale methods only allow for a fast matrix-vector product. There have been preconditioners developed for MLFMA, but it was found that an effective preconditioner could be constructed using an iterative solver. In the Casimir force, the \( A \) matrix is the \( \nabla_i M \) matrix. The gradient of the impedance matrix is going to be a comparatively weak matrix. This is because all of the self interaction entries in the matrix become zero, only the entries representing interactions between the Casimir object and another object are non-zero. In comparison, the \( B \) matrix is the impedance matrix \( M \), which retains the strong diagonal terms due to the self interactions. One can thus estimate
the inverse of the preconditioner as

$$\overline{P}^{-1} = (\overline{A} - \lambda \overline{B})^{-1} \approx -\frac{1}{\lambda} \overline{B}^{-1}$$  \hspace{1cm} (3.150)$$

where $\lambda$ is chosen as the estimated eigenvalue. The matrix product of our $\overline{P}^{-1}$ with $\overline{A} \cdot \overline{Z}_j - \overline{B} \cdot \overline{Z}_j \cdot \Theta$ can be quickly estimated by using GMRES for a few iterations (usually five iterations). It is not necessary to provide the exact inverse but just an estimate for the preconditioner.

The GMRES preconditioner above can be compared with a more traditional preconditioner derived from the near-field matrix used in the FMA. The proposed preconditioner is an $LDL^T$ factorization of the $\overline{A} - \lambda_k \overline{B}$ matrix calculated at the $k$-th step. The reason for this is that the rate of convergence depends upon the eigenvalue spectrum of $\overline{A} - \lambda_k \overline{B}$ and not the eigenvalues of $(\overline{A}, \overline{B})$. Specifically, the convergence at each step is approximately

$$\frac{\lambda_{k+1} - \Lambda_1}{\lambda_k - \Lambda_1} \leq 4 \left( \frac{1 - \sqrt{\psi_0}}{1 + \sqrt{\psi_0}} \right)^{2m} + 4 \left( \frac{1 - \sqrt{\psi_0}}{1 + \sqrt{\psi_0}} \right)^m \left( \frac{||\overline{B}||}{\gamma_2} \right) (\lambda_k - \Lambda_1)^{\frac{1}{2}} + \ldots$$  \hspace{1cm} (3.151)$$

where

$$\psi_0 = \frac{\gamma_2}{\gamma_N}$$  \hspace{1cm} (3.152)$$

where $\Lambda_1$ is the actual eigenvalue being sought and $\gamma_i$ is the $i$-th eigenvalue out of $N$ eigenvalues of $\overline{A} - \lambda_k \overline{B}$. Here, $m$ is the size of the Krylov subspace that is calculated at each outer iteration. Thus, the rate of convergence is at worst linear and depends upon the relative size of the second and last eigenvalues of $\overline{A} - \lambda_k \overline{B}$. A proper preconditioner would then be used to improve the eigenvalue spectrum of $\overline{A} - \lambda_k \overline{B}$. This can be done using a $LDL^T$ factorization of $\overline{A} - \lambda_k \overline{B}$ such that the diagonal matrix $\overline{D}$ is composed of $\pm 1$. This means that the eigenvalues of $\overline{L}_k^{-1} \cdot (\overline{A} - \lambda_k \overline{B}) \cdot \overline{L}_k^T = \overline{D}_k$ are $\gamma_1 = -1$ and $\gamma_2 = \gamma_3 = \cdots = \gamma_N = 1$. Thus, for $m \geq 1$ the convergence becomes quadratic since the first term of Equation (3.151) drops out.
The above scheme requires the calculation of the factorization at each step. However, if $\Lambda_1$ was known, we could use a constant $L_k = L$ from the $LDL^T$ of $\overline{A} - \Lambda_1 \overline{B}$ such that $\overline{D}$ is composed of 0 and 1 and quadratic convergence would be achieved. In practice an incomplete $LDL^T$ factorization of $\overline{A} - \lambda_k \overline{B}$ is used and this can achieve linear convergence but at a rate much better than with the original problem. This was implemented by using the near-field terms of $\overline{A}$ and $\overline{B}$ that are calculated in the FMA tree.

![Comparison of the Convergence Between Different Preconditioners.](image)

Figure 3.6: Various preconditioner results.

The result can be seen in Figure 3.6. Without the use of a preconditioner, the algorithm experiences a slow linear convergence. The ideal preconditioner, that of perfect inversion of $\overline{A} - \lambda_k \overline{B}$, achieves quadratic convergence as expected. However, the performance of the sparse preconditioner constructed from the near-field terms does not achieve convergence. The residue of the resulting iterations fails to improve after a few iterations. An acceptable performance is obtained by using GMRES instead to estimate the inversion of the $\overline{B}$ matrix as the preconditioner. So the preconditioner works as confirmed by the existence of quadratic and improved linear convergence when implementing the preconditioners suggested by Golub. However, our initial attempts at using the near-field terms of $\overline{A}$ and $\overline{B}$ as the basis for an incomplete $LDL^T$ factorization fail. The near-field matrices do not have enough information to work as an adequate preconditioner. Using a GMRES estimate of the inverse of the matrix $\overline{B}$, which contains the far-field terms, performs well.
In the end, results were obtained using the inverse-free preconditioned iterative
eigenvalue solvers. The problem appears when the integration approaches the lower
frequencies, on the order of 1 MHz, whereby the number of inner and outer iterations in
the iterative eigenvalue solver increases greatly. The proofs of the convergence for the
algorithms will not be reproduced here; they can be found in detail in the relevant
papers [83,84], but we will discuss their implications. First, it can be shown that the
iterative eigenvalue method, which generates the estimated eigenvalue $\lambda_k$ and eigenvector
$x_k$ at step $k$, will always converge to an eigenvalue of the matrix pencil $(\mathbf{A}, \mathbf{B})$ and that
$\|(\mathbf{A} - \lambda \mathbf{B}) x_k\| \to 0$. In other words, the algorithm will always converge to an eigenvalue
and the direction of the convergence of the eigenvector is always in the correct direction.
This brings us to our first problem. Using the matrix generated by the direct solvers, the
above conditions can be achieved where the algorithm will converge to a reasonable
arbitrary residue. The speed of this convergence is not an issue here for the moment.
However, when using the FMA for the lower frequencies, convergence is not guaranteed for
a desired residue. This is a bit confusing because one can think of the FMA matrix-vector
product to be $\mathbf{A}'x$ where $\|\mathbf{A}x - \mathbf{A}'x\| = \delta$. That is, the FMA matrix-vector product can
be thought of as a matrix=vector product of an estimate of the actual dense matrix $\mathbf{A}$. As
such, while one may not expect the converged eigenvalue to be the same (indeed, the
converged result is not the same within an increasingly small error), one should expect that
the guarantee of convergence should still hold for the estimated matrix $\mathbf{A}'$.

Investigations suggest that the reason for the lack of convergence using FMA could be
the loss of definiteness of the $\mathbf{B}$ matrix. For these iterative eigenvalue solvers to work, the
$\mathbf{B}$ matrix must be positive definite. Estimating the effectiveness of the FMA matrix by
reconstructing it via matrix-vector products with the columns of the identity matrix, it can
be seen that for low harmonic numbers in the LF-FMA that the $\mathbf{B}$ matrix is indefinite.
Increasing the number of harmonic terms shifts the eigenvalues of $\mathbf{B}$ such that it becomes
weakly then fully positive definite. Unfortunately, we cannot say that a low frequency
solver would help push these negative eigenvalues into positive regions. In addition, we also
cannot say with certainty that these reconstructed matrices represent the desired matrix
within a reasonable amount of error. For example, the sum of the generalized eigenvalues

106
for these reconstructed matrices only approaches the true sum by using at least six harmonic terms for one example problem. Five harmonic terms can be used and the algorithm still obtains convergence to an eigenvalue, but the converged eigenvalues are incorrect. The converged eigenvalues are correct when compared to the effective matrix that is being reproduced via the matrix-vector products. As such, the problem is no longer a convergence problem but a problem where the estimated matrix is too poor of an approximation to provide correct eigenvalues.

The solution at hand is to increase the accuracy of the FMA matrix-vector product at low frequencies. This becomes infeasible since the time required for the matrix-vector product to be calculated for even a modestly sized problem increases to unacceptable lengths. The same problem occurs when using the FMA for the stress tensor problem despite the fact that it is now an inverse problem. To be sure though, the generalized eigenvalue problem is analogous to the inverse problem since one finds the eigenvalues of $B^{-1} \cdot A \cdot x = \lambda x$.

Nevertheless, Figure 3.7 shows the results for the forces between a sphere and plate using the direct matrix vector product in the stress tensor method, FMA in the stress tensor method, and the RRWJ formulation. The RRWJ formulation is the most accurate as the force on the sphere is equal and opposite to the force on the plate to a very high degree of
accuracy. With the stress tensor formulation, the forces agree well enough but only to a few digits of accuracy. This is due to the approximation of the surface integration. Still, the FMA and direct matrix vector product compare very favorably.

Parallel Processing

The algorithms for finding the Casimir force lend themselves easily to parallel processing. In the frequency integration, all Gaussian point integrands are independent of each other. This allows for the calculation of the impedance matrices and solving for the integrand to be done in parallel. In addition, the filling of the impedance matrix can be done in parallel too. For the stress tensor method, the disadvantage of having to integrate across the surfaces can be lessened by evaluating the surface integration points in parallel. Since all of these operations are independent of each other, there are significant gains to be achieved. The parallelization was done using OpenMP and Message Passing Interface (MPI). OpenMP allows for parallelization across shared memory and is used to parallelize the processes on a single node. For example, most modern CPUs are multi-core processing units and many nodes on a supercomputer utilize multiple CPUs that share the same RAM. OpenMP can be used to efficiently parallelize the program across the cores and CPUs that share the same RAM. This is preferable to MPI because MPI runs a copy of itself with each instance, requiring redundant copies of the data to be stored in the same shared memory space. Between nodes, however, MPI is used to parallelize the program. The stress tensor and the RRWJ algorithms have both been parallelized and tested on modest computer clusters using the combined MPI and OpenMP scheme. The RRWJ algorithm suffers from the bottleneck that the eigenvalue solver is not parallelized. As such, it can only be parallelized across the frequency integration while the stress tensor can be parallelized across the frequency and surface integrations. The stress tensor will benefit more from the parallelization and can become competitive with the RRWJ formulation with enough nodes to operate on. Since these problems are embarrassingly parallel, they can experience near linear speedup up to the number of integration points involved.
### 3.7 Large-Scale Method Findings

To find the Casimir energy and force, we need to solve the following equations:

\[
E = \frac{hc}{2\pi} \int_0^\infty d\kappa \ln \frac{\text{det} \, M}{\text{det} \, M_\infty}
\]

\[
F_i = -\nabla_i E = -\frac{hc}{2\pi} \int_0^\infty d\kappa \text{Tr} \left[ M^{-1} \cdot \nabla_i M \right]
\]

The integrands in the above can be found by solving eigenvalue problems.

\[
\ln \frac{\text{det} \, M}{\text{det} \, M_\infty} = \sum_{i=1}^N \ln \frac{\lambda_i}{\lambda_{\infty,i}}
\]

\[
\text{Tr} \left[ M^{-1} \cdot \nabla_i M \right] = \sum_{i=1}^N \lambda'_{i}
\]

where \(\lambda\) and \(\lambda_\infty\) are the eigenvalues of \(M\) and \(M_\infty\) respectively and \(\lambda'\) are the eigenvalues of the generalized eigenvalue problem

\[
\nabla_i M \cdot x_i = \lambda'_i M \cdot x_i
\]

Solving for the energy has not been feasible using large scale methods because of the normalization of the \(M\) matrix by \(M_\infty\). Under these conditions, eigenvalues undergo cancellation leaving only the pairs of eigenvalues that differ to contribute to the overall sum. Looking at the eigenvalue spectrum for a given frequency point, the eigenvalues that contribute to the net sum for the energy are found on the interior of the eigenspectrum. For example, the base 10 logarithm of the eigenspectrum of the \(M\) matrix for two PEC spheres can be seen in 3.8a. From the plot, it can be seen that the magnitude of the eigenvalues falls off rather quickly. However, looking at the plot of the \(\ln \lambda/\lambda_\infty\) in 3.8b, we see that the pairs that contribute to the sum lie in the interior. As such, we cannot find a way to search for these eigenvalues without knowing where we need to target.

On the other hand, the eigenvalue spectrum for the generalized problem for the force is very much like that in Figure 3.8a. As such, we can calculate the Casimir force by solving...
for a small number of eigenvalues that dominate the sum. The problem is that the FMA matrix-vector product (MVP) for the derivative terms requires more terms in the FMA translator to achieve the same accuracy as the MVP for the original impedance matrix. For example, while the original matrix would require 2 harmonics in LF-FMA, the gradient matrix would require around 6-8. While this will not affect the CPU time scaling, it does greatly increase the amount of time devoted to a MVP to the point that the LF-FMA was not feasible for the gradient matrix.

An alternative is to estimate the derivatives by using finite difference.

\[-\nabla_i \ln \frac{\det \overline{M}}{\det \overline{M}_\infty} \approx \frac{1}{\Delta} \ln \frac{\det \overline{M}_-}{\det \overline{M}_+}\] (3.158)

where \(\overline{M}_+\) has the Casimir object displaced by \(\frac{\Delta}{2} \hat{r}\) and \(\overline{M}_-\) has the Casimir object displaced by \(-\frac{\Delta}{2} \hat{r}\). This formulation has the same problem as the original energy formulation, the eigenvalues of interest lie in the interior due to the normalization. Alternatively, we can show that

\[
\frac{\det \overline{M}_-}{\det \overline{M}_+} = \det \left( \overline{M}_- \cdot \overline{M}_+^{-1} \right) = \det \left( \overline{M}_+^{-1} \cdot \overline{M}_- \right)\] (3.159)
Thus, the problem can be recast as finding the eigenvalues of $\mathbf{M}_+^{-1} \cdot \mathbf{M}_-$ and taking the sum of the logarithm of these eigenvalues. In doing so, the eigenvalues are already normalized and the extremum eigenvalues contribute the most. A problem lies with the fact that the eigenvalues will be clustered around 1 since the two matrices are very similar. The resulting eigenvalues will be dominated by the factor of 1 which will give rise to a poor estimation of the summation due to cancellation. This can be rectified by normalizing the eigenvalues by the identity matrix.

$$
\mathbf{M}_- \cdot \mathbf{x} = (\lambda' - \mathbf{I}) \cdot \mathbf{M}_+ \cdot \mathbf{x}
\rightarrow (\mathbf{M}_- - \mathbf{M}_+) \cdot \mathbf{x} = \lambda' \cdot \mathbf{M}_+ \cdot \mathbf{x}
$$

(3.160)

where the integrand is now found as

$$
-\nabla_i \ln \frac{\det \mathbf{M}}{\det \mathbf{M}_\infty} \approx \frac{1}{\Delta} \sum_{i=1}^{N} \lambda'
$$

(3.161)

Now the left-hand side of Equation (3.160) is the finite difference of the gradient. We can then see that this formulation is approximately the same as taking the finite difference of the generalized eigenvalue problem in Equation (3.157). Finally, the actual generalized eigenvalue problem is

$$
(\mathbf{M}_- - \mathbf{M}_+) \cdot \mathbf{x} = \lambda \cdot \mathbf{M} \cdot \mathbf{x}
$$

(3.162)

The eigenvalues here should approximate the very same eigenvalues we found for the exact formulation of the Casimir force. This will require that three matrices are kept in memory. However, note that the $\mathbf{A}$ matrix is composed of $\mathbf{M}_- - \mathbf{M}_+$, which will remove the block diagonal entries of $\mathbf{M}$ that represent the self-interactions (in addition to the lower block matrices of $\mathbf{D}$ and $k_0^2 \mathbf{I}$ in the A-EFIE formulation). As such, we only need to calculate and store the entries that represent interactions between the Casimir and a non-Casimir object. This will mean that we do not need any near-field terms for the FMA matrix for the $\mathbf{A}$ product if we assume that the objects will be sufficiently spaced so that two different
objects do not share the same bounding box. This can reduce the amount of memory and CPU time needed to calculate the $\bar{A}$ MVP.

### Calculating the Eigenvalues

The next step is calculating the eigenvalues. Most iterative generalized eigenvalue solvers require that we solve the problem of $\mathbf{B}^{-1} \cdot \mathbf{x} \rightarrow \mathbf{y}$. Since our $\mathbf{B}$ is $\mathbf{M}$, the impedance matrix, this is an expensive proposition. In addition, solving the inverse problem can only be done as an estimate that may cause the eigenvalues to drift from their true values. This makes the use of inverse free generalized eigenvalue solvers attractive. A few packages were found but when it came to using this for the A-EFIE, only one package was suitable because other algorithms required that the impedance matrix be symmetric positive-definite. The one package that could handle our problem was the Jacobi-Davidson with QZ (JDQZ) [86].

To find the Casimir force, the code iterates over all the integration points for the frequency (typically 5 to 12 integration points due to the smoothness of the integrand). At each frequency point, the matrices are calculated and iteratively solved for the extremum eigenvalues. Because we need multiple eigenvalues and since these eigenvalues are clustered together, it is necessary to use a block iterative solver which solves for a block of $k$ eigenvalues at a time. It has the advantage of using the subspace information for multiple eigenvalues and block solvers are better at resolving the individual eigenvalues within a cluster. However, we do not know how many eigenvalues we need to estimate the sum, nor does the solver always resolve all the eigenvalues in a cluster. For example, if we solve for the four largest and smallest eigenvalues, we ideally would get

$$\lambda_i = \{8.7398609e-5, -7.6385969e-5, 5.172029e-5, 5.16598443e-5, 4.95802663e-5, -4.7685760e-5, -4.7632070e-5, -4.60693075e-5\}.$$ 

The third and fourth eigenvalues are very close and sometimes the eigenvalue solver will only find one of them in a single iteration. Instead of the $5.1659844e-5$ eigenvalue, the solver may find $2.8894328e-5$. This would adversely affect the estimate of the eigenvalue sum since the pairs of extremum eigenvalues are nearly equal in magnitude. Thus, it may require a
second block of eigenvalues to correctly resolve clustered eigenvalues. In addition, for this case, the ideal sum is $2.55978424e - 5$, but the sum of the 8 eigenvalues found above are $2.25859078e - 5$, which gives an error of 11.7\%. If we wish to achieve an error less than 10\%, we would need to find another block of eigenvalues.

Matrix deflation is used to find these multiple blocks. This can be done a few ways, but the deflation method implemented for the $\overline{A}$ matrix is as follows:

$$\overline{A} = \overline{A} - \overline{B} \cdot \overline{X} \cdot \overline{A} \cdot (\overline{B} \cdot \overline{X})^H$$

(3.163)

where $\overline{X}$ are the converged Schur vectors and $\overline{A}$ is a diagonal matrix of the found eigenvalues. This will deflate the found eigenvalues to zero. The converged Schur vectors are the orthonormalized eigenvectors. An iterative modified Gram-Schmidt process is used to orthonormalize the eigenvectors against the $\overline{B}$ norm to make the column vectors of $\overline{X}$ orthogonal and normalized such that $\overline{X} \cdot \overline{B} \cdot \overline{X}^H = \overline{I}$. During the Gram-Schmidt orthogonalization, we also calculate the $\overline{B} \cdot \overline{X}$ product making the process require a one time cost of $k$ MVPs where $k$ is the number of eigenvalues found.

This method of deflation works well but there are some drawbacks. Each successive deflation requires more MVPs to find the eigenvalues than the previous block. The accuracy of the deflated matrix depends on the accuracy of the found eigenvectors. This means that we have to solve for the eigenvalues at a much higher tolerance so that the deflated problem remains accurate. This also means that we can only deflate a few times before the deflated matrix becomes a poor approximation.

This is a problem for JDQZ where it was observed that for an accurate estimate of the eigenvalues we may use a tolerance of $10^{-6}$, but for deflation, we need a tolerance of $10^{-11}$. This greatly increases the number of MVP. In trials using Matlab, it was found that ARPACK [87,88] generally gave a very good estimate of the eigenvectors for a given tolerance and was more robust under deflation than JDQZ. With ARPACK, we chose to solve the standard eigenvalue problem of $\overline{B}^{-1} \cdot \overline{A} \cdot \overline{x} = \lambda \overline{x}$ where we took the $\overline{A}$ and $\overline{B}^{-1}$ as separate MVP using GMRES to find the $\overline{B}^{-1}$ MVP. It was found that the tolerance on the GMRES could be relaxed to $10^{-4}$ but the tolerance on the ARPACK eigenvalue solver still
needed to be very stringent, at most $10^{-12}$. One problem with the ARPACK solver was that it did a poor job of finding the largest (or smallest) eigenvalues exclusively. ARPACK has the tendency to find both the largest and smallest eigenvalues in an iteration despite searching for the largest eigenvalues exclusively. This was solved by simply searching for the $2k$ eigenvalues of largest magnitude. The only disadvantage is that we may not find $k$ pairs as there may be more largest than smallest or vice-versa in a block of $2k$. Still, in general it was found that the inverse free iterative generalized eigenvalue solvers did not perform as well as those that estimated the inverse using an iterative solver.

In summary, the program flow goes like this. We integrate over 5-12 frequency points for the integration. At each frequency point we fill the matrices and find $k$ pairs of eigenvalues. We calculate the sum and check if the sum of the last pair divided by the total sum so far is less than a convergence parameter (currently set to $10^{-2}$). If the sum has converged, we stop; otherwise we calculate the deflation and find the next $k$ pairs. Normally, we only need to find 2 or 3 blocks of eigenvalues. The result is quite accurate. For two PEC spheres using dense matrices of 1404 edges, we find that the estimated force is $4.545841839 e^{-3}$. The ideal force is $4.66042057 e^{-3}$ giving a relative error of 2.46%. This was done using ARPACK with GMRES set to $10^{-4}$ and ARPACK to $10^{-12}$ finding 8 eigenvalues at a time. It may be possible to loosen the tolerances even more if we could reduce the amount of deflation. The problem with loose tolerances is that after one or two deflations, the eigenvalues drift too far for the sum to be convergent. The number of MVPs used for the five frequency points were, in order of increasing frequency, 2098 (2 blocks), 1542 (2 blocks), 1231 (2 blocks), 1049 (2 blocks), and 3173 (5 blocks) for a total of 9093 MVPs. If we loosened our convergence criteria to $10^{-1}$, then we have 4166 MVPs for the same tolerances with a relative error of 4.84%. In the end, we could get away, at least for this specific problem, with doing single blocks of 8 eigenvalues at a time. By loosening the tolerances to $10^{-2}$ for both the GMRES and ARPACK, then the total number of MVPs was 948 for a relative error of 5.65%. It was found that varying the tolerance for the ARPACK from $10^{-2}$ to $10^{-4}$ did not really impact the number of MVPs used or the accuracy of the final solution. The reason for this is probably because ARPACK is converging after a very modest number of iterations of its algorithm, as little as two
iterations. Thus, the Arnoldi space that is being created to find the eigenvalues retains far
more information than we need. JDQZ could not compete with ARPACK because it had
trouble resolving the correct 4 largest and 4 smallest eigenvalues. It sometimes missed one
of the eigenvalues that would have to be accounted for by running deflation.

**Preconditioners**

The final way to improve the number of MVP that we use is by improving the
preconditioner. However, our attempts have generally been unsuccessful. The
preconditioner recommended for A-EFIE is

\[
\mathbf{M} = \begin{bmatrix}
diag(\mathbf{V}) & (\mathbf{F} \cdot \mathbf{D})^T \cdot \left(\mathbf{F} \cdot \text{diag}(\mathbf{P}) \cdot \mathbf{F}^T\right) \\
\mathbf{F} \cdot \mathbf{D} & k_0^2 \mathbf{I}_r
\end{bmatrix} = \begin{bmatrix}
\mathbf{V}_d & \mathbf{D}_r^T \cdot \mathbf{P}_{rd} \\
\mathbf{D}_r & k_0^2 \mathbf{I}_r
\end{bmatrix}
\]  

(3.164)

Efforts were made to improve this by using the tri- and quintuple-diagonal approximations
of \(\mathbf{V}\) and \(\mathbf{P}\) without any improvements. The only improvement that we did manage was
when we corrected the \(\mathbf{D}_r^T \cdot \mathbf{P}_{rd}\) term. Ideally, this block should be \(\mathbf{D}_r^T \cdot \mathbf{P} \cdot \mathbf{B}\). Using the
reduced patch matrix \(\mathbf{P}_{rd}\) correctly gives the diagonal of the \(\mathbf{P} \cdot \mathbf{B}\) product, but misses a
few non-zero rows. However, \(\mathbf{B}\) is easily constructed and being a sparse matrix, it is not
difficult to find the exact \(\mathbf{P}_d \cdot \mathbf{B}\) where \(\mathbf{P}_d\) is the diagonal patch matrix. The sparse matrix
\(\mathbf{V}_d^{-1} \cdot \mathbf{D}_r^T \cdot \mathbf{P} \cdot \mathbf{B}\) is stored in memory and used for the calculation of the preconditioner.
This only achieved a very modest improvement in the number of MVPs.

**Moving to Large Scale**

Large scale computing may be possible, but right now it requires a lot of finesse to achieve
it. It is probable to do it for the two PEC sphere case because we know how many
eigenvalues are needed and thus we can avoid the large costs in deflation from the increase
in the tolerances. If the FMA performs the same as the dense matrix, then we can expect
to get a good estimate of the results using \(O(100)\) MVPs. The original A-EFIE paper
states that the code used around 2 seconds for the preconditioner and 27.1 seconds for the GMRES interactions. Each iteration used one MVP and one preconditioner and this was run on the group’s HPC. According to Intel, the E5365 Xeon 3.0 GHz processor runs at 48 GFLOP. A representative replacement workstation from Dell uses the i7-3770 processor which has 108.8-125 GFLOP. Thus, using a midrange contemporary workstation, we may expect to see an improvement of a factor of two. Using MPI, we can parallelize the program across the frequency calculations. Above, the simulation that found only 8 eigenvalues at each frequency had at most 200 MVP in a frequency. This gives an estimated run time of around 45 minutes.

Despite this feasibility estimate, the overall algorithm is still too costly to implement at the moment. Only with a large amount of fine tuning and a priori knowledge can a problem be made feasible. In the end, it was concluded that large scale efforts along these lines will require additional improvements in computing power, or, more likely, the implementation of a parallel FMA code. The FMA implementation performs poorly because the FMA is designed for far-field calculations while the Casimir force is focused around the local field coupling between objects. The importance of the low frequency results places an emphasis on the accuracy of near field interactions which must be approximated using FMA. In addition, using EFIE or A-EFIE introduces unnecessary normalization and spurious information that has to be removed. This is represented in the cancellation that occurs between the extremum eigenvalue pairs. A more successful algorithm would incorporate the necessary normalizations in the matrix representation itself.

Final Observations

The eigenvalue problem can be recast to show how it is a scattering problem. In doing so, it can be shown where the spurious information is introduced and how we can renormalize the matrices. The new formulation would remove the cancellation in the eigenvalues allowing for more robust reduction in the number of eigenvalues needed and the accuracy with which they need to be found. First, note that the trace of the product of two matrices
is equivalent if we commute the product.

\[ \text{Tr} \left( \mathbf{A} \cdot \mathbf{B} \right) = \text{Tr} \left( \mathbf{B} \cdot \mathbf{A} \right) \quad (3.165) \]

Thus, we can recast the integrand for the force as

\[ \text{Tr} \left( \mathbf{M}^{-1} \cdot \nabla_i \mathbf{M} \right) = \text{Tr} \left( \nabla_i \mathbf{M} \cdot \mathbf{M}^{-1} \right) \quad (3.166) \]

This changes the physical interpretation and the computational complexity of the eigenvalue solver. First, let us assume that our matrix \( \mathbf{M} \) is the A-EFIE impedance matrix:

\[ \mathbf{M} = \mathbf{Z}_A = \begin{bmatrix} \mathbf{V} & \mathbf{D}^T \cdot \mathbf{P} \\ \mathbf{D} & k_0^2 \mathbf{I} \end{bmatrix} \quad (3.167) \]

The original EFIE can be recast as

\[ \mathbf{Z} \cdot \mathbf{J} = \left( \mathbf{V} - \frac{1}{k_0^2} \mathbf{D}^T \cdot \mathbf{P} \cdot \mathbf{D} \right) \cdot \mathbf{J} = \mathbf{b} \quad (3.168) \]

where the elements of \( \mathbf{b} \) are

\[ [\mathbf{b}]_m = -\frac{i}{\omega \mu_0} \int_{\Lambda_m} \Lambda_m(r) \cdot \mathbf{E}^i(r) dr \quad (3.169) \]

where we have assumed freespace. Note that the gradient of the A-EFIE impedance matrix is

\[ \nabla_i \mathbf{M} = \nabla_i \mathbf{Z}_A = \begin{bmatrix} \nabla_i \mathbf{V} & \mathbf{D}^T \cdot \nabla_i \mathbf{P} \\ 0 & 0 \end{bmatrix} \quad (3.170) \]

The zero block matrices in the bottom row of \( \nabla_i \mathbf{Z}_A \) mean that the product of \( \nabla_i \mathbf{Z}_A \cdot \mathbf{Z}^{-1} \) will also have the same zero block matrices. Let us divide up \( \mathbf{Z}_A^{-1} \) into blocks that
correspond to the four block matrices of \( Z_A \), then

\[
\nabla_i Z_A \cdot Z_A^{-1} = \begin{bmatrix}
\nabla_i \nabla \cdot \left[ Z_A^{-1} \right]_{11} + D^T \cdot \nabla_i \Phi \cdot \left[ Z_A^{-1} \right]_{21} & \nabla_i \nabla \cdot \left[ Z_A^{-1} \right]_{12} + D^T \cdot \nabla_i \Phi \cdot \left[ Z_A^{-1} \right]_{22}
\end{bmatrix}
\]

(3.171)

Since we are taking the trace of the above, the trace can be reduced to taking the trace of the upper-left block matrix. Thus,

\[
\text{Tr} \left( \nabla_i Z_A \cdot Z_A^{-1} \right) = \text{Tr} \left( \nabla_i \nabla \cdot \left[ Z_A^{-1} \right]_{11} + D^T \cdot \nabla_i \Phi \cdot \left[ Z_A^{-1} \right]_{21} \right)
\]

(3.172)

If one were to find the trace of the above using an iterative eigenvalue solver, then the dimension of the eigenvalue problem can be reduced from \( e + (p-t) \) to \( e \), where \( e \) is the total number of edges, \( p \) the total number of patches, and \( t \) the total number of removed patches. This can be done equivalently by using the original MVP of \( \nabla_i Z_A \cdot Z_A^{-1} \) with the vector returned by the iterative algorithm padded by zeros. That is, we run the solver for \( e \) unknowns and are asked to find the MVP with the vector \( b \). We still use the MVP of \( \nabla_i Z_A \cdot Z_A^{-1} \) with the vector \( \eta_0 \cdot \left[ \eta_0^{-1} b^T 0^T \right]^T \). This is the same form as the excitation vector that is used to solve for the currents and charges in the A-EFIE. Thus, we can treat the vectors used in the MVP as excitations.

\[
\eta_0 \cdot \nabla_i Z_A \cdot Z_A^{-1} \cdot \begin{bmatrix}
\eta_0^{-1} b \\
0
\end{bmatrix} = \eta_0 \cdot \nabla_i Z_A \cdot \begin{bmatrix}
ik_0 J \\
0 \\
c_0 \rho
\end{bmatrix} = \eta_0 \cdot \begin{bmatrix}
\eta_0^{-1} \nabla_i b^{\text{scat}} \\
0
\end{bmatrix} = \nabla_i b^{\text{scat}}
\]

(3.173)

where \( \nabla_i b^{\text{scat}} \) can be thought of as the change in the scattered electric field on the surface of the objects due to a given incident field under an infinitesimal displacement of the Casimir object. When we solve for the trace, we take the summation of the eigenvalues that satisfy the standard eigenvalue problem

\[
\nabla_i Z_A \cdot Z_A^{-1} \cdot \begin{bmatrix}
x \\
0
\end{bmatrix} = \lambda \begin{bmatrix}
x \\
0
\end{bmatrix}
\]

(3.174)
The eigenvalues can be interpreted as giving the inverse of the change in a given electric field due to the infinitesimal displacement of the Casimir object in response to field itself. That is, the eigenvectors guarantee that the scattered field is the same mode as the incident field. The eigenvalues give a measure of the change in this mode due to the displacement of the Casimir object.

If we reduce the size of the eigenvectors to $e$, then the eigenvalues do not change. Rerunning the same case for two PEC spheres we find that it requires 1069 MVP with the largest number of MVP at a frequency being 240. The relative error was 7.00%. As such, we do not register any appreciable change in the algorithm.

There is an alternative that can help provide normalized eigenvalues. Currently, our eigenvalues come in pairs that are nearly equal but opposite in sign. One way that a matrix can have such a behavior is to have the following form:

$$
\mathbf{M} = \begin{bmatrix}
A & 0 \\
0 & -A + \delta
\end{bmatrix}
$$

(3.175)

In the above, if $\delta = 0$ then we have pairs of eigenvalues of opposite sign and equal magnitude, giving rise to a trace of zero. The perturbation $\delta$ will shift the eigenvalues of one sign by the factor $\delta$. The result is that the trace is now $N\delta$, where $N$ is the number of rows or columns in $\mathbf{M}$. We can still retain this behavior if we have weak off-diagonal block matrices.

$$
\mathbf{M} = \begin{bmatrix}
A & B \\
C & -A + \delta
\end{bmatrix}
$$

(3.176)

If $\hat{B}$ and $\hat{C}$ are weak compared to $\hat{A}$, then the eigenvalues will still be roughly equal and opposite, but the trace remains unchanged. In fact, the trace can be shown to be

$$
\text{Tr} [\mathbf{M}] = \text{Tr} [\hat{A}] + \text{Tr} [-\hat{A} + \delta] = \text{Tr} [\hat{A} - \hat{A} + \delta] = N\delta.
$$

In this manner, by combining the block diagonal elements we remove the normalization of the trace that is pushed onto the eigenvalues.

Take the example of a two object system in EFIE. By taking the trace of the block
diagonals of the resulting Casimir force matrix, it can be shown that

\[ \text{Tr} \left[ \mathbf{M} \right] = \text{Tr} \left[ \mathbf{Z}_{21}^{-1} \cdot \nabla \mathbf{Z}_{12} + \mathbf{Z}_{12}^{-1} \cdot \nabla \mathbf{Z}_{21} \right] = \text{Tr} \left[ \mathbf{Z}_{21}^{-1} \cdot \nabla \mathbf{Z}_{12} - \left( \mathbf{Z}_{21}^{-1} \cdot \nabla \mathbf{Z}_{12} \right)^T \right] \] (3.177)

This shows explicitly the cancellation that must occur in the eigenvalues of the original matrix. It is possible to isolate the block diagonal components and combine them in a MVP that is self-normalized. However, in EFIE and A-EFIE, this requires far more work than it would take to solve the original system. In the end, removing the cancellation in the eigenvalues will require a different formulation than EFIE.
CHAPTER 4

EQUIVALENCE PRINCIPLE ALGORITHM

4.1 Introduction

The equivalence principle algorithm (EPA) is a domain decomposition method (DDM) for integral equations that allows for the simplification of a system by replacing the scatterers with an equivalent surface current [15–17,89,90]. The problem is changed from solving a matrix problem of the currents over the surface of the objects to the problem of solving the currents over the user-defined equivalent surfaces (ES). This is done by relating the radiated fields from the induced currents on the objects via scattering matrices and the interaction between the ES via translation matrices.

Given the scattering problem shown in Figure 4.1a, the total number of unknowns will be equal to \( N_{\text{Total}} = 2N_1 + N_2 + N_4 \). The size of the matrix will be \( N_{\text{Total}}^2 \) and the complexity for finding all the eigenvalues for the Casimir force will be \( N_{\text{Total}}^3 \). Domain decomposition methods, like the EPA, allow for us to subdivide the problem into multiple subproblems, solve these subproblems individually, and then use the resulting information to solve the overall problem. This is done in such a way that the memory and/or computational costs are reduced. For example, we can solve individual problems associated with each object. The total cost of solving the subproblems will be \( 2N_1^3 + N_2^3 + N_4^3 \) which is lower than \( N_{\text{Total}}^3 \). The information from the subproblems will be combined to formulate a severely reduced matrix that can be solved in a trivial amount of time.

The EPA works by enclosing all or part of the objects in one or more fictitious surfaces as shown in Figure 4.1b. The surface of the ES is represented by its own mesh and set of unknowns. Each ES constitutes its own subproblem in the form of a scattering matrix
while the relationship between the ES is related in a translation matrix. The task then becomes solving for the scattering characteristics of the objects enclosed in each ES via the scattering matrices and then solving the overall equations that relate the ES to each other using the translation and scattering matrices. The goal is to have the number of unknowns representing the ES mesh to be less than the number of unknowns representing the mesh of the enclosed scatterers. The final matrix problem that represents the total scattered fields is dependent upon the number of unknowns of the ES, not the enclosed scatterers. In this way, we can achieve a reduced $N_{Total}$ for the final matrix problem.

In other words, the advantage is that the original matrix problem is transformed into an effective matrix problem that moves the unknowns from the surface of the objects to the ES. The ES will generally have a coarser mesh than the mesh for the original object, thus reducing the overall problem size. This is due to the fact that with the ES offset from the surface of the original object, the near field (evanescent field) from the object dies out to some extent before reaching the ES. As such, the fields across the ES are smoother and the equivalent currents to represent these fields can thus be approximated with fewer basis functions. Fine detailed features in the objects can also require mesh densities greater than what would be needed for the ES. Finally, there could be several objects packed within the same ES that would result in a greater total surface area for the objects compared to the
enveloping ES. All of these reasons contribute to the ability for the ES mesh to have significantly fewer unknowns than the enclosed objects.

Another advantage is that the scattering matrix is dependent only upon the objects enclosed by an ES and the geometry of the ES. This allows for the scattering matrix to be reused for identical objects, further reducing the required computation time and memory. By enclosing identical sets of objects with identical ES meshes, the subproblems become the same. Thus, we can solve only one of the identical subproblems and reuse this information. The scattering matrices can also be stored to disk and recalled to construct a variety of scenes without the need of solving for the complex underlying scattering problem. Instead, the new scene can be constructed by calculating the new translation matrices whose number of unknowns are dependent on the ES, not the enclosed objects. Thus, the difficult problem of solving for the scattering matrices of the enclosed objects can be done once. Whenever a scene is constructed based upon the same objects, the scattering data can be recalled and only the reduced matrix data needs to be calculated anew.

The original formulation assumes that the entirety of an object is always enclosed within the ES. However, connected objects can still be split into separate surfaces using the tap basis [17,91]. This is useful in decomposing complex objects that have constituent objects that would benefit from a multi-physics approach. Despite the division of the object, the tap basis connections allow for currents to flow between the ES. Finally, the EPA has the added benefit of being able to be accelerated using FMA [17] and incorporating A-EFIE [92,93].

The EPA thus provides an attractive formulation for the Casimir force. There is the potential for an appreciable reduction in memory and CPU time for problems that involve the repeated use of identical geometries, like the force between multiple spheres, or complex geometry that require high mesh densities. The implementation is greatly simplified because most of the simulations that are modeled are between independent objects. However, connected objects can still be split into separate surfaces using the tap basis [17,91]. This would be useful for more realistic structures like a cantilever switch. In this case, it is desired to find the force on the cantilever, but the cantilever is connected to a stable base in addition to connecting to the substrate when the switch is closed. These
connections allow for currents to flow between the structure of interest and stable supports which one would want to define as separate ES. The EPA has the added benefit of being able to be accelerated using FMA [17] and incorporating A-EFIE [92,93]. Thus, the techniques discussed in previous chapters can still be implemented in EPA.

4.2 EPA Derivation

The EPA is based around the use of the Equivalence Principle which allows us to represent a given electromagnetic field in a volume as a set of electric and magnetic current sources on the surface of the volume. Take the case of the scatterer enclosed in the ES given at the top of Figure 4.2. The scatterer is under the influence of a given incident electromagnetic field. Assuming that \( \hat{n} \) is the outward pointing normal of the ES, we can express the incident electromagnetic field inside or outside the ES using a set of impressed equivalent sources on the ES. In either case, the complementary volume has a null field. This facilitates the domain decomposition by allowing us to deal with the scattering within each ES independently of what lies outside the ES. EPA works by first defining a set of scattering operators that characterize this independent scattering of the objects inside an ES.
ES. Then we derive a set of equations that relate this independent scattering to the desired scattering problem via coupling between the multiple ES. The independent scattering is described by a scattering operator while the coupling is facilitated using a translation operator. The resulting system of equations is the EPA matrix problem that is solved to find the actual scattered fields.

## Scattering and Translation Operators

Given the ES $S$, the field in the region into which the normal vector $\hat{n}$ points toward can be found from the fields along the ES as

$$E_\mathbf{r} = \nabla \times \oint_S d\mathbf{r}' g(\mathbf{r} - \mathbf{r}')\hat{n}' \times E_s(\mathbf{r}') - \frac{1}{i\omega \epsilon} \nabla \times \nabla \times \oint_S d\mathbf{r}' g(\mathbf{r} - \mathbf{r}')\hat{n}' \times H_s(\mathbf{r}')$$

(4.1)

In this case, fields in the region that lies in the opposite direction of $\hat{n}$ are zero. The equivalent currents along the ES will be defined as

$$J_s = \hat{n} \times H_s$$

(4.2)

$$M_s = -\hat{n} \times E_s$$

(4.3)

Thus, the fields become

$$E_\mathbf{r} = -\nabla \times \oint_S d\mathbf{r}' g(\mathbf{r} - \mathbf{r}')M_s(\mathbf{r}') - \frac{1}{i\omega \epsilon} \nabla \times \nabla \times \oint_S d\mathbf{r}' g(\mathbf{r} - \mathbf{r}')J_s(\mathbf{r}')$$

(4.4)

The above can be split into a representation that uses the familiar $\mathcal{L}$ and $\mathcal{K}$ operators.

$$\mathcal{L}^S(\mathbf{r}, \mathbf{r}')J(\mathbf{r}') = -\frac{1}{i\omega \epsilon} \nabla \times \nabla \times \oint_S d\mathbf{r}' g(\mathbf{r} - \mathbf{r}')J_s(\mathbf{r}')$$

(4.5)

$$\mathcal{K}^S(\mathbf{r}, \mathbf{r}')M(\mathbf{r}') = -\nabla \times \oint_S d\mathbf{r}' g(\mathbf{r} - \mathbf{r}')M_s(\mathbf{r}')$$

(4.6)
The resulting fields are represented as

\[
E(r) = \mathcal{L}^S(r, r')J(r') + \mathcal{K}^S(r, r')M(r') \\
H(r) = -\mathcal{K}^S(r, r')J(r') + \frac{1}{\eta^2}\mathcal{L}^S(r, r')M(r')
\] (4.7)

where \( \eta = \sqrt{\frac{\mu}{\epsilon}} \). For the sake of consistency, \( \hat{n} \) will be defined to always be the outward normal. Note that other publications concerning the EPA may choose to have an inward pointing normal resulting in a sign difference but the solved currents remain the same. The above choice for the equivalent surface currents on the ES, given in Equations (4.2) and (4.3), gives rise to the situation where the fields inside the ES are zero and the non-zero fields lie outside the ES, as shown in the right side of Figure 4.2. For the case where the non-zero fields are inside the ES and the null fields are outside, shown in the left side of Figure 4.2, the equivalent currents become

\[
J^\text{out}_s = -\hat{n} \times H_s = -J_s \\
M^\text{out}_s = \hat{n} \times E_s = -M_s
\] (4.9)

The EPA works by finding scattering and translation matrices with the desired outcome that the scattering matrices are smaller than the original matrix describing the enclosed object. The scattering matrix is a matrix of operators that takes the incident currents along an ES and gives the resulting scattered currents along the ES due to the enclosed objects. To start, the incident currents along the ES due to an incident electromagnetic field are defined to be

\[
J^\text{inc}_s = \hat{n} \times H^\text{inc}_s \\
M^\text{inc}_s = -\hat{n} \times E^\text{inc}_s
\] (4.11)
The resulting fields inside the ES, with zero fields outside, are thus

\[ E_{\text{inc}}^{\text{inside}}(r) = -L_s^{\text{s}}(r, r')J_s^{\text{inc}}(r') - K_s^{\text{s}}(r, r')M_s^{\text{inc}}(r') \]  
\[ H_{\text{inc}}^{\text{inside}}(r) = K_s^{\text{s}}(r, r')J_s^{\text{inc}}(r') - \frac{1}{\eta^2} L_s^{\text{s}}(r, r')M_s^{\text{inc}}(r') \]  

(4.13) (4.14)

Knowing the incident fields inside the ES, the scattered currents on the surface of the object can be solved. In theory, the operator that solves for the currents based upon the fields impinged on the surfaces can be any suitable current solver. A MOM solver is particularly suited for how these matrix operators are represented in a matrix representation. Still, it should be theoretically consistent to use other solvers as long as the user can devise a means of converting the input and output of the solver to be consistent with the matrix representation of the translation matrix operators. For the purposes of this chapter, it will be assumed for generality that the enclosed objects are dielectrics and the current solver operator is the PMCHWT operator [51]. Using the above definitions for the \( L \) and \( K \) operators, the PMCHWT operator is the following matrix operator.

\[
Z = \begin{bmatrix}
-L_{\text{Out}}^{s} & -K_{\text{Out}}^{s} \\
K_{\text{Out}}^{s} & -\frac{1}{\eta_{\text{Out}}} L_{\text{Out}}^{s} & -\frac{1}{\eta_{\text{In}}} L_{\text{In}}^{s} 
\end{bmatrix}
\]  
(4.15)

where \( \text{Out} \) and \( \text{In} \) refer to the operators using the material properties of the volume outside and inside the object respectively. Using this operator, the scattered surface currents on the object become

\[
\begin{bmatrix}
J_{\text{scat obj}}^{\text{scat}} \\
M_{\text{scat obj}}^{\text{scat}}
\end{bmatrix} = Z^{-1} \begin{bmatrix}
E_{\text{inc}}^{\text{in}} \\
H_{\text{inc}}^{\text{in}}
\end{bmatrix}
\]  
(4.16)

where the incident fields are found along the surface of the scatterer. If the scatterer is a PEC, then the magnetic currents are set to zero and \( Z \) would be a suitable current solver like the EFIE operator. Knowing the surface currents of the scattered field on the scatterer, the equivalence principle can then give the scattered currents along the ES that gives the equivalent fields outside the ES with null fields inside the ES. The scattered fields
along the ES are

\[ E_{s}^{\text{scat}}(r) = \mathcal{L}^{s}(r, r') J_{\text{obj}}^{\text{scat}}(r') + \mathcal{K}^{s}(r, r') M_{\text{obj}}^{\text{scat}}(r') \]  
(4.17)

\[ H_{s}^{\text{scat}}(r) = -\mathcal{K}^{s}(r, r') J_{\text{obj}}^{\text{scat}}(r') + \frac{1}{\eta^2} \mathcal{L}^{s}(r, r') M_{\text{obj}}^{\text{scat}}(r') \]  
(4.18)

The equivalent scattered currents on the ES become

\[ J_{s}^{\text{scat}} = \hat{n} \times H_{s}^{\text{scat}} \]  
(4.19)

\[ M_{s}^{\text{scat}} = -\hat{n} \times E_{s}^{\text{scat}} \]  
(4.20)

We now arrive at the desired output of the scattering matrix operator. The scattering matrix operator can thus be found by combining the above operations. The scattering matrix that relates the incident currents on the \( i \)-th ES to the scattered currents on the \( i \)-th ES is

\[
\begin{bmatrix}
    J_{s_i}^{\text{scat}} \\
    M_{s_i}^{\text{scat}}
\end{bmatrix} = \begin{bmatrix}
    -\hat{n} \times \mathcal{K}^{s_{ii}} & \frac{1}{\eta^2} \hat{n} \times \mathcal{L}^{s_{ii}} \\
    -\hat{n} \times \mathcal{L}^{s_{ii}} & -\hat{n} \times \mathcal{K}^{s_{ii}}
\end{bmatrix} \cdot Z_{s_{ii}}^{-1} \cdot \begin{bmatrix}
    -\mathcal{L}^{s_{ii}} & -\mathcal{K}^{s_{ii}} \\
    \mathcal{K}^{s_{ii}} & -\frac{1}{\eta^2} \mathcal{L}^{s_{ii}}
\end{bmatrix} \cdot \begin{bmatrix}
    J_{s_i}^{\text{inc}} \\
    M_{s_i}^{\text{inc}}
\end{bmatrix}
\]  
(4.21)

\[
S_{ii} \cdot \begin{bmatrix}
    J_{s_i}^{\text{inc}} \\
    M_{s_i}^{\text{inc}}
\end{bmatrix}
\]  
(4.22)

Note that the scattering matrix operator is split into three matrix operators: the inside-out operator, the current solver operator, and the outside-in operator. The inside-out and outside-in operators act to take a current from the interior of the ES onto the ES and vice-versa. In short, the inside-out and outside-in operators are respectively defined as

\[
T_{ii}^{\text{io}} = \begin{bmatrix}
    -\hat{n} \times \mathcal{K}^{s_{ii}} & \frac{1}{\eta^2} \hat{n} \times \mathcal{L}^{s_{ii}} \\
    -\hat{n} \times \mathcal{L}^{s_{ii}} & -\hat{n} \times \mathcal{K}^{s_{ii}}
\end{bmatrix}
\]  
(4.23)

\[
T_{ii}^{\text{oi}} = \begin{bmatrix}
    -\mathcal{L}^{s_{ii}} & -\mathcal{K}^{s_{ii}} \\
    \mathcal{K}^{s_{ii}} & -\frac{1}{\eta^2} \mathcal{L}^{s_{ii}}
\end{bmatrix}
\]  
(4.24)
For PEC objects, the scattering matrix operator can be reduced to

\[
S_{ii}^{PEC} = \begin{bmatrix} -\hat{n}_i \times K_{si} \\ -\hat{n}_i \times L_{si} \end{bmatrix} \cdot Z_{PEC,ii}^{-1} \cdot \begin{bmatrix} -L_{si} & -K_{si} \end{bmatrix}
\]

(4.25)

The scattered field outside of the ES can be found from these currents. Given two ES with known currents on surface \( j \), the translation matrix operator gives the equivalent currents on the \( i \)-th ES. These translated currents give the field outside the \( i \)-th ES due to the currents on the \( j \)-th ES. (Note: We take the fields to be outside as that is the initial conditions assumed for the incident currents given to the scattering matrix operator.) Given these source currents, the fields on the \( i \)-th ES become

\[
\begin{align*}
E_i(r_i) &= L^S(r_i, r'_j)J_j(r'_j) + K^S(r_i, r'_j)M_j(r'_j) \\
H_i(r_i) &= -K^S(r_i, r'_j)J_i(r'_j) + \frac{1}{\eta^2}L^S(r_i, r'_j)M_j(r'_j)
\end{align*}
\]

(4.26)

(4.27)

where the subscripts \( i \) and \( j \) on the position vectors \( r \) and \( r' \) are to indicate explicitly the associated surfaces that the vectors are taken over. The desired currents are thus

\[
\begin{bmatrix} J_{si} \\ M_{si} \end{bmatrix} = \begin{bmatrix} -\hat{n}_i \times K^s & \frac{1}{\eta^2} \hat{n}_i \times L^s \\ -\hat{n}_i \times L^s & -\hat{n}_i \times K^s \end{bmatrix} \cdot \begin{bmatrix} J_{sj} \\ M_{sj} \end{bmatrix}
\]

\[
= T_{ij} \cdot \begin{bmatrix} J_{sj} \\ M_{sj} \end{bmatrix}
\]

(4.28)

(4.29)

where \( T_{ij} \) is the translation matrix operator.

**Solving for the Scattered Currents**

The scattering matrix operator given in Equation (4.21) is sufficient information for the scattering problem of a single ES. For multiple ES, however, the translation matrix operator is needed to form a system of equations to relate the interaction between the objects contained in each ES. Consider the case where there are two arbitrary
homogeneous objects enclosed in two separate ES. There will be a set of incident and scattered currents on each ES due to the external excitation. The scattered sources on ES 1 are a result of the scattering on ES 1 by the incident sources on ES 1 and the scattered sources on ES 2. That is,

$$\begin{bmatrix}
    J_{s_1}^{scat} \\
    M_{s_1}^{scat}
\end{bmatrix} = \text{Scattering in ES 1 via} \begin{bmatrix}
    J_{s_1}^{inc} \\
    M_{s_1}^{inc}
\end{bmatrix} + \text{Scattering from ES 2 via} \begin{bmatrix}
    J_{s_2}^{scat} \\
    M_{s_2}^{scat}
\end{bmatrix}$$ (4.30)

Using the scattering and translation matrix operators,

$$\begin{bmatrix}
    J_{s_1}^{scat} \\
    M_{s_1}^{scat}
\end{bmatrix} = S_{11} \cdot \begin{bmatrix}
    J_{s_1}^{inc} \\
    M_{s_1}^{inc}
\end{bmatrix} + S_{11} \cdot T_{12} \cdot \begin{bmatrix}
    J_{s_2}^{scat} \\
    M_{s_2}^{scat}
\end{bmatrix}$$ (4.31)

A similar relationship can be found for the scattered sources on ES 2,

$$\begin{bmatrix}
    J_{s_2}^{scat} \\
    M_{s_2}^{scat}
\end{bmatrix} = S_{22} \cdot \begin{bmatrix}
    J_{s_2}^{inc} \\
    M_{s_2}^{inc}
\end{bmatrix} + S_{22} \cdot T_{21} \cdot \begin{bmatrix}
    J_{s_1}^{scat} \\
    M_{s_1}^{scat}
\end{bmatrix}$$ (4.32)

The two equations results in a system of equations represented by the matrix operation

$$\begin{bmatrix}
    \mathcal{I} & -S_{11} \cdot T_{12} \\
    -S_{22} \cdot T_{21} & \mathcal{I}
\end{bmatrix} \cdot \begin{bmatrix}
    J_{s_1}^{scat} \\
    M_{s_1}^{scat}
\end{bmatrix} = \begin{bmatrix}
    S_{11} \cdot \begin{bmatrix}
    J_{s_1}^{inc} \\
    M_{s_1}^{inc}
\end{bmatrix} \\
    S_{22} \cdot \begin{bmatrix}
    J_{s_2}^{inc} \\
    M_{s_2}^{inc}
\end{bmatrix}
\end{bmatrix}$$ (4.33)
where $\mathcal{I}$ is the identity operator. The case for multiple objects can be easily inferred as an extension to the above. For example, three ES create the following matrix problem:

$$
\begin{bmatrix}
\mathcal{I} & -S_{11} \cdot T_{12} & -S_{11} \cdot T_{13} \\
-S_{22} \cdot T_{21} & \mathcal{I} & -S_{22} \cdot T_{23} \\
-S_{33} \cdot T_{31} & -S_{33} \cdot T_{32} & \mathcal{I}
\end{bmatrix}
\begin{bmatrix}
\mathcal{J}_{\text{scat}}_{s1} \\
\mathcal{M}_{\text{scat}}_{s1} \\
\mathcal{J}_{\text{scat}}_{s2} \\
\mathcal{M}_{\text{scat}}_{s2} \\
\mathcal{J}_{\text{scat}}_{s3} \\
\mathcal{M}_{\text{scat}}_{s3}
\end{bmatrix}
= 
\begin{bmatrix}
S_{11} \cdot \mathcal{J}_{\text{inc}}_{s1} \\
S_{22} \cdot \mathcal{J}_{\text{inc}}_{s2} \\
S_{33} \cdot \mathcal{J}_{\text{inc}}_{s3}
\end{bmatrix}
$$

The case can also be changed by removing the EPA around one or more objects. Under these circumstances, the associated right-hand vector becomes the incident electromagnetic field and the scattering matrix operators on the left-hand side disappear with exception for the diagonal term which becomes the appropriate combination of the $\mathcal{L}$ and $\mathcal{K}$ operators that generates the electromagnetic field.

Up to this point, the preceding discussion has been done purely from the operator picture and is not directly translatable to a matrix method. Despite this, we can introduce direct matrix replacements for each of the vectors allowing for the scattering and translation matrices to keep their operator form. However, we will forgo the derivation of the actual matrices as it was subsequently found that the EPA fails at low frequencies and must be replaced.

### 4.3 Incorporation into Casimir Force

It was previously shown that the calculation of the Casimir force can be done using the argument principle by finding an appropriate dispersion relationship for the system. Equation (4.34) shows the relationship between the incident currents on the ES and the resulting scattered currents on the ES. Once again there is a simple matrix representation that relates the currents induced in a system with an external excitation. For the Casimir force, the currents that can exist without an external excitation occur at the
eigenfrequencies of the system. As such, the dispersion relation becomes, for the three ES system,

$$\det \mathbf{M} = 0 \quad (4.35)$$

where

$$\mathbf{M} = \begin{bmatrix}
\mathbf{I}_{11} & -\mathbf{S}_{11} \cdot \mathbf{T}_{12} & -\mathbf{S}_{11} \cdot \mathbf{T}_{13} \\
-\mathbf{S}_{22} \cdot \mathbf{T}_{21} & \mathbf{I}_{22} & -\mathbf{S}_{22} \cdot \mathbf{T}_{23} \\
-\mathbf{S}_{33} \cdot \mathbf{T}_{31} & -\mathbf{S}_{33} \cdot \mathbf{T}_{32} & \mathbf{I}_{33}
\end{bmatrix} \quad (4.36)$$

Note that the normalization condition, when the objects are separated to infinity, would necessitate that the translation matrices become zero (assuming that the object of interest is contained within a single ES). The normalization matrix, $\mathbf{M}_\infty$, simply becomes the identity matrix and thus the Casimir matrix is self-normalizing. The resulting Casimir energy and force then become

$$\mathcal{E} = \frac{\hbar c}{2\pi} \int_0^\infty d\kappa \ln \det \mathbf{M} \quad (4.37)$$

$$\mathbf{F}_i = -\frac{\hbar c}{2\pi} \int_0^\infty d\kappa \nabla_i \ln \det \mathbf{M} \quad (4.38)$$

where $i$ refers to the gradient in the position of the $i$-th object. The advantage of using the self-normalizing matrix is that it may enable the use of an iterative eigenvalue solver for the Casimir energy. This does not affect the Casimir force calculation since the normalization factor goes away due to the gradient.
Comparison to Previous Methods

Further insight into the result can be found from the two ES case. With only two ES, the system of equations when there are no excitations becomes

\[
\begin{bmatrix}
  j_1^{\text{scat}} \\
  m_1^{\text{scat}}
\end{bmatrix}
- \mathbf{S}_{11} \cdot \mathbf{T}_{12} \cdot
\begin{bmatrix}
  j_1^{\text{inc}} \\
  m_1^{\text{inc}}
\end{bmatrix}
= 0
\]

\[
\begin{bmatrix}
  j_2^{\text{scat}} \\
  m_2^{\text{scat}}
\end{bmatrix}
- \mathbf{S}_{22} \cdot \mathbf{T}_{21} \cdot
\begin{bmatrix}
  j_2^{\text{inc}} \\
  m_2^{\text{inc}}
\end{bmatrix}
= 0
\]

Solving for the currents on the second ES, the dispersion relation can be reduced to

\[
(\mathbf{I}_{11} - \mathbf{S}_{11} \cdot \mathbf{T}_{12} \cdot \mathbf{S}_{22} \cdot \mathbf{T}_{21}) \cdot
\begin{bmatrix}
  j_1^{\text{inc}} \\
  m_1^{\text{inc}}
\end{bmatrix}
= 0
\]

The Casimir matrix and the resulting energy then become

\[
\mathbf{M} = \mathbf{I}_{11} - \mathbf{S}_{11} \cdot \mathbf{T}_{12} \cdot \mathbf{S}_{22} \cdot \mathbf{T}_{21}
\]

\[
\mathcal{E} = \frac{\hbar c}{2\pi} \int_0^\infty dk \ln \det (\mathbf{I}_{11} - \mathbf{S}_{11} \cdot \mathbf{T}_{12} \cdot \mathbf{S}_{22} \cdot \mathbf{T}_{21})
\]

The result is the same that was found by Rahi, Emig, Graham, Jaffe, and Kardar (REGJK) [7] for the two object case. In fact, further investigation finds that the scattering matrices on the right-hand side of the original EPA scattering problem can be moved to the left-hand side by left multiplying a block diagonal matrix. The new scattering matrix problem becomes

\[
\begin{bmatrix}
  \mathbf{S}_{11}^{-1} & -\mathbf{T}_{12} & -\mathbf{T}_{13} \\
  -\mathbf{T}_{21} & \mathbf{S}_{22}^{-1} & -\mathbf{T}_{23} \\
  -\mathbf{T}_{31} & -\mathbf{T}_{32} & \mathbf{S}_{33}^{-1}
\end{bmatrix}
\begin{bmatrix}
  j_1^{\text{scat}} \\
  m_1^{\text{scat}} \\
  j_2^{\text{scat}} \\
  m_2^{\text{scat}} \\
  j_3^{\text{scat}} \\
  m_3^{\text{scat}}
\end{bmatrix}
= \begin{bmatrix}
  j_1^{\text{inc}} \\
  m_1^{\text{inc}} \\
  j_2^{\text{inc}} \\
  m_2^{\text{inc}} \\
  j_3^{\text{inc}} \\
  m_3^{\text{inc}}
\end{bmatrix}
\]
Thus, the Casimir matrices are

\[
\mathbf{M} = \begin{bmatrix}
S_{11}^{-1} & -T_{12} & -T_{13} \\
-T_{21} & S_{22}^{-1} & -T_{23} \\
-T_{31} & -T_{32} & S_{33}^{-1}
\end{bmatrix}
\]  

(4.45)

\[
\mathbf{M}_\infty = \text{diag}[S_{11}^{-1}, S_{22}^{-1}, S_{33}^{-1}]
\]  

(4.46)

This is equivalent to the general result found by REGJK although this is not a desirable form to use for numerical stability. The advantage with the EPA formulation is that in REGJK, the translation and scattering matrices represented the translation and scattering of the electromagnetic waves themselves, not the currents. This meant that the bases projected onto the operators to give the matrix representations were limited to plane wave, cylindrical wave, or spherical wave bases. The drawback in their formulation was that a basis may become unsuitable for small separations between the objects. This is due to the scattered fields requiring a large number of basis functions to accurately approximate them. In addition, one basis may become inappropriate at small separations. For example, a cylindrical basis being used with elliptical cylinders becomes unable to be resolved by the basis at very small separations. This is because one of the elliptical cylinders may not fit inside the smallest circular cylinder basis that encloses the other elliptical cylinder.

With the EPA formulation, it is hoped that the basis functions have greater freedom in choice and are better suited for modeling the near fields. One advantage is that the choice of basis function can allow for higher-order basis functions on the ES [17]. In addition, the EPA using the tap basis allows for the Casimir object to be connected to other non-Casimir objects.
4.4 Results

Results have been produced using the original EPA formulation. The case of a single dielectric sphere with a relative permittivity of 4.0 has been examined. The sphere has a radius of 1 m and is impinged upon by a plane wave 5.0 m in wavelength. The results between using PMCHWT to find the RCS and the EPA using a spherical surface of 1.5 m have been compared in Figure 4.3. The case where the sphere and its ES are replicated with a displacement of 1 wavelength is shown in Figure 4.4. In both cases the RCS results agree very well with the results generated by using PMCHWT.

Unfortunately, results for the Casimir force were inconclusive. The mid frequency results matched with the original A-EFIE results, but the low frequency results were erroneous. It is apparent that EPA suffers from the same low frequency breakdown that occurs with EFIE. However, unlike EFIE, using double precision does not alleviate the low frequency breakdown. A true Augmented-Equivalence Principle Algorithm (A-EPA) needs to be used [94,95].

Figure 4.3: A dielectric sphere ($\epsilon_r = 4.0$) of radius 1 m that is enclosed by an ES of radius 1.5 m.
4.5 Augmented-Equivalence Principle Algorithm

While EPA has a number of advantages, it does not serve well for the Casimir problem due to the need for broadband calculations. At low frequencies, the EPA suffers from low frequency breakdown. EFIE was often able to produce adequate results despite the low frequency breakdown by the use of double precision arithmetic, but results from the EPA using double precision were still erroneous despite the use of the high-order field point sampling scheme. The solution is then to use the A-EPA that uses the A-EFIE algorithm to avoid loop-tree decomposition. To implement A-EPA, it is necessary to derive a matrix equation form of the operator picture. Previously, the algorithm was only derived and implemented in a matrix-free algorithm that calculated the translation and scattering operations on the fly which is not suitable for the Casimir force due to the large number of MVP that is needed for the iterative eigenvalue solver. For this derivation we will assume...
that the scatterers are only PEC. To start, the unknowns for the ES are expanded as

\[ X = \begin{bmatrix} J_s \\ \frac{1}{\eta} M_s \\ c \rho^e \\ c \frac{1}{\eta} \rho^m \end{bmatrix} \]  
(4.48)

The \( \mathcal{L} \) operator will be split into two parts associated with the vector and scalar potentials. Thus,

\[
E(r) = \mathcal{L}^s_v(r, r') J(r') + \mathcal{L}^s_s(r, r') \rho^e(r') + \mathcal{K}^s(r, r') M(r')  
(4.49)
\]

\[
H(r) = -\mathcal{K}^s(r, r') J(r') + \frac{1}{\eta^2} \mathcal{L}^e_v(r, r') M(r') + \frac{1}{\eta^2} \mathcal{L}^e_s(r, r') \rho^m(r')  
(4.50)
\]

where

\[
\mathcal{L}^s_v(r, r') J(r') = ik \eta \oint_S dr' g(r - r') J(r')  
(4.51)
\]

\[
\mathcal{L}^s_s(r, r') \rho^e(r') = -c \eta \oint_S dr' \nabla g(r - r') \rho^e(r')  
(4.52)
\]

\[
\mathcal{K}^s(r, r') M(r') = -\oint_S dr' \nabla g(r - r') \times M(r')  
(4.53)
\]

Note that \( \mathcal{K} \) has not changed but has simply been expressed to explicitly show its use of the gradient of the Green’s function. The incident sources along the ES due to an incident electromagnetic field are defined to be

\[
J_s^{inc} = \hat{n} \times H_s^{inc}  
(4.54)
\]

\[
M_s^{inc} = -\hat{n} \times E_s^{inc}  
(4.55)
\]

\[
\rho^{e, inc} = \frac{1}{i \omega} \nabla \cdot J_s^{inc} = \epsilon \hat{n} \cdot E_s^{inc}  
(4.56)
\]

\[
\rho^{m, inc} = \frac{1}{i \omega} \nabla \cdot M_s^{inc} = \mu \hat{n} \cdot H_s^{inc}  
(4.57)
\]
where we have assumed the $e^{-i\omega t}$ time dependence and an outward pointing normal. Thus, the resulting electric field inside the ES is

\[
E_{\text{inside}}^{\text{inc}}(r) = -\mathcal{L}_{s}(r, r')J_{s}^{\text{inc}}(r') - \mathcal{L}_{v}(r, r')\rho^{e,\text{inc}}(r') - \mathcal{K}_{s}(r, r')M_{s}^{\text{inc}}(r')
\] (4.58)

The A-EFIE matrix formulation used will be

\[
\begin{bmatrix}
\nabla & \overline{D}^T \cdot \overline{P} \\
\overline{D} & k^2 \overline{I}
\end{bmatrix}
\begin{bmatrix}
\mathbf{i}k \mathbf{J} \\
\mathbf{c}\rho^e
\end{bmatrix} = \begin{bmatrix}
\frac{1}{\eta} \mathbf{V} \\
0
\end{bmatrix}
\] (4.59)

where we note that the elements of the excitation vector $\mathbf{b}$ is described as

\[
[V]_m = \langle \Lambda_{\text{obj},m}(r), E_{\text{inside}}^{\text{inc}}(r') \rangle
\] (4.60)

The first step in producing a matrix implementation is to construct the outside-in translation matrix for the $i$-th ES. Using the above information, we can find the scattered current and charge coefficients of the enclosed objects via

\[
\begin{bmatrix}
\mathbf{i}k \mathbf{J}_{\text{obj}}^{\text{scat}} \\
\mathbf{e}_{\text{scat}}^e \mathbf{c}_{\text{obj}}\rho
\end{bmatrix} = \overline{Z}_{A,i}^{-1} \cdot \begin{bmatrix}
-\frac{1}{\eta} \nabla_{v,i}^o & -\mathbf{K}_{i}^o & -\frac{1}{\mathbf{c}\eta} \nabla_{s,i}^o & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{bmatrix} \cdot \begin{bmatrix}
\mathbf{j}^{\text{inc}}_i \\
\frac{1}{\eta} \mathbf{m}^{\text{inc}}_i \\
\mathbf{c}\rho^e \mathbf{m}^{\text{inc}}_i \\
\mathbf{e}_{\text{inc}} \mathbf{\rho}^e_i
\end{bmatrix}
\] (4.61)

where the matrices here are placeholders that will be defined based upon the choice of the basis. Previously, we used the RWG basis function as the expansion for the currents and the pulse basis function for the charges. But, the RWG basis function is inadequate for representing the currents on the ES, necessitating the use of the Nyström method. We will use the existing scheme where, for example, the radiated field from an electric current is calculated as

\[
E(r) \approx \sum_{i=1}^{N_P} \sum_{j=1}^{N_I} \left\{ \mathcal{L}_{v}^{s}(r, r')N_j(r', r_j)\mathbf{J}(r_j) + \mathcal{L}_{s}^{s}(r, r')N_j(r', r_j)\rho^{e}(r_j) \right\}
\] (4.62)
where the integrating surface has been divided up into $N_P$ patches. Each of these patches represents the currents using the interpolation function $N_j$ sampled over $N_I$ points. The original code that was adapted for the Casimir force used triangular patches with point sampling.

$$N_j(r', r_j) = w_j \delta(r' - r_j) \quad (4.63)$$

where the weights $w_j$ and positions $r_j$ are the weights and sampling points for the integration over a two-dimensional triangular patch. The currents and charges on the ES will be represented as

$$J(r) = \sum_{m=1}^{N_P} \sum_{n=1}^{N_I} \left\{ j_{n}^{i} \hat{i} + j_{n}^{j} \hat{j} \right\} \quad (4.64)$$

$$\rho_e(r) = \sum_{m=1}^{N_P} \sum_{n=1}^{N_I} \rho_{n}^{e} \quad (4.65)$$

where the unit vectors $\hat{i}$ and $\hat{j}$ are two orthogonal vectors that are tangent to the ES at the given sampling point and $j_{n}^{i}$, $j_{n}^{j}$, and $\rho_{n}^{e}$ are the source coefficients that will be used in Equation (4.61). Thus, the electric field radiated onto the enclosed object from the incident sources on the $m$-th patch will be

$$E_{\text{inside}}^{\text{inc}}(r) = \sum_{n=1}^{N_I} \left\{ -w_j \mathcal{L}_{v}^{s}(r, r_j) \left( j_{n}^{i} \hat{i} + j_{n}^{j} \hat{j} \right) - w_j \mathcal{K}^{s}(r, r_j) \left( m_{n}^{i} \hat{i} + m_{n}^{j} \hat{j} \right) - w_j \mathcal{L}_{s}^{e}(r, r_j) \rho_{n}^{e} \right\} \quad (4.66)$$
The above provides all the framework needed to describe the outside-in translation matrix. The coefficient vectors will be ordered as follows:

\[
\begin{align*}
\mathbf{j}_{m}^{inc} &= \begin{bmatrix}
  j_{s,m,1}^i \\
  j_{s,m,1}^j \\
  j_{s,m,2}^i \\
  \vdots \\
  j_{s,m,N_{sm}}^i \\
\end{bmatrix} \\
\mathbf{m}_{m}^{inc} &= \begin{bmatrix}
  m_{s,m,1}^i \\
  m_{s,m,1}^j \\
  m_{s,m,2}^i \\
  \vdots \\
  m_{s,m,N_{sm}}^i \\
\end{bmatrix} \\
\rho_{m}^{e,inc} &= \begin{bmatrix}
  \rho_{s,m,1}^e \\
  \rho_{s,m,2}^e \\
  \vdots \\
  \rho_{s,m,N_{sm}}^e \\
\end{bmatrix} \\
\rho_{m}^{m,inc} &= \begin{bmatrix}
  \rho_{s,m,1}^m \\
  \rho_{s,m,2}^m \\
  \vdots \\
  \rho_{s,m,N_{sm}}^m \\
\end{bmatrix}
\end{align*}
\]

where \(N_{sm} = N_{P,s,m} \cdot N_{I,s,m}\). The entries of the outside-in matrices become

\[
\begin{align*}
\mathbf{L}_{m}^{oi} &= \begin{bmatrix}
  \mathbf{L}_{s}^{s}(r,r_{n}) \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \rangle \\
  \mathbf{L}_{s}^{s}(r,r_{n}) \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \rangle \\
  \vdots \\
  \mathbf{L}_{s}^{s}(r,r_{n}) \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \rangle \\
\end{bmatrix} \\
\mathbf{K}_{m}^{oi} &= \begin{bmatrix}
  \mathbf{K}_{s}^{s}(r,r_{n}) \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \rangle \\
  \mathbf{K}_{s}^{s}(r,r_{n}) \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \rangle \\
  \vdots \\
  \mathbf{K}_{s}^{s}(r,r_{n}) \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \rangle \\
\end{bmatrix} \\
\mathbf{L}_{s}^{oi} &= \begin{bmatrix}
  \langle \Lambda_{obj,m}(r), \mathbf{w}_{n} \mathbf{L}_{s}^{s}(r,r_{n}) \rangle \\
\end{bmatrix}
\end{align*}
\]

With the above outside-in translation matrix, the A-EFIE current solver will find the coefficients of the electric currents expanded over the RWG basis and the electric charge expanded over the pulse basis. The scattered electromagnetic field on the \(i\)-th ES due to the scattered electric current and electric charge on the enclosed objects is given as

\[
\begin{align*}
\mathbf{E}_{s}(r) &= \sum_{i=1}^{N_{e}} \mathbf{L}_{s}^{s}(r,r_{n}) j_{scat}^{i} \Lambda_{obj,i}(r_{n}) + \sum_{i=1}^{N_{p}} \mathbf{L}_{s}^{s}(r,r_{n}) \rho_{e,scat}^{i} \iota_{obj,i}(r_{n}) \\
\mathbf{H}_{s}(r) &= \sum_{i=1}^{N_{e}} -\mathbf{K}_{s}^{s}(r,r_{n}) j_{scat}^{i} \Lambda_{obj,i}(r_{n})
\end{align*}
\]

where there are \(N_{e}\) edges and \(N_{p}\) patches on the enclosed objects. The equivalent sources on the ES are the same as before and by using the Nyström method with the delta
function gives the following inside-out translation matrix:

\[
T^{i,o} = \begin{bmatrix}
-\frac{1}{ik} K^{i,o}_{\hat{n} \times} & 0 \\
-\frac{1}{ik\eta} \tilde{L}^{i,o}_{v,\hat{n} \times} & -\frac{1}{c\eta} \tilde{L}^{i,o}_{s,\hat{n} \times} \\
\frac{1}{ik\eta} \tilde{L}^{i,o}_{v,\hat{n} \cdot} & \frac{1}{c\eta} \tilde{L}^{i,o}_{s,\hat{n} \cdot} \\
-\frac{1}{ikc} K^{i,o}_{\hat{n} \cdot} & 0 \\
\end{bmatrix}
\] (4.73)

where the elements of the matrices are given by

\[
\begin{align*}
\left[ L^{i,o}_{v,\hat{n} \times} \right]_{m,n} &= \left[ \hat{v}_m \cdot \hat{n} \times \langle \mathcal{L}_v(r_m, r') \rangle, \Lambda_{obj,n}(r') \rangle \right] \\
\left[ L^{i,o}_{s,\hat{n} \times} \right]_{m,n} &= \left[ \hat{v}_m \cdot \hat{n} \times \langle \mathcal{L}_s(r_m, r') \rangle, \Lambda_{obj,n}(r') \rangle \right] \\
\left[ K^{i,o}_{\hat{n} \times} \right]_{m,n} &= \left[ \hat{v}_m \cdot \hat{n} \times \langle \mathcal{K}_s(r_m, r') \rangle, \Lambda_{obj,n}(r') \rangle \right] \\
\left[ L^{i,o}_{v,\hat{n} \cdot} \right]_{m,n} &= \hat{n} \cdot \langle \mathcal{L}_v(r_m, r') \rangle, \Lambda_{obj,n}(r') \rangle \\
\left[ L^{i,o}_{s,\hat{n} \cdot} \right]_{m,n} &= \hat{n} \cdot \langle \mathcal{L}_s(r_m, r') \rangle, \Lambda_{obj,n}(r') \rangle \\
\left[ K^{i,o}_{\hat{n} \cdot} \right]_{m,n} &= \hat{n} \cdot \langle \mathcal{K}_s(r_m, r') \rangle, \Lambda_{obj,n}(r') \rangle 
\end{align*}
\] (4.74-4.79)

where \( r_m \) are the locations of the sampling points in the ES patches. Thus, the full scattering matrix becomes

\[
S_{ii} = \begin{bmatrix}
-K^{i,o}_{\hat{n} \times,ii} & 0 \\
-\frac{1}{\eta} \tilde{L}^{i,o}_{v,\hat{n} \times,ii} & -\frac{1}{c\eta} \tilde{L}^{i,o}_{s,\hat{n} \times,ii} \\
\frac{1}{\eta} \tilde{L}^{i,o}_{v,\hat{n} \cdot,ii} & \frac{1}{c\eta} \tilde{L}^{i,o}_{s,\hat{n} \cdot,ii} \\
-K^{i,o}_{\hat{n} \cdot,ii} & 0 \\
\end{bmatrix}
\begin{bmatrix}
\frac{1}{ik} & 0 \\
0 & I \\
\end{bmatrix}
\begin{bmatrix}
\mathbb{Z}^{-1} A_i \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
-\frac{1}{\eta} \tilde{L}^{oi}_{v,ii} & -K^{oi}_{ii} & -\frac{1}{c\eta} \tilde{L}^{oi}_{s,ii} & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\] (4.80)
For the translation operator, start with the expressions for the electromagnetic field on one ES from another.

\[
E(r) = \mathcal{L}_v^s(r, r')J(r') + \mathcal{L}_s^s(r, r')\rho^s(r') + \mathcal{K}^s(r, r')M(r') \tag{4.81}
\]

\[
H(r) = -\mathcal{K}^s(r, r')J(r') + \frac{1}{\eta^2}\mathcal{L}_v^s(r, r')M(r') + \frac{1}{\eta^2}\mathcal{L}_s^s(r, r')\rho^m(r') \tag{4.82}
\]

Using the same relationships for the scattered fields given in Equations (4.54) through (4.57), the translation operator matrix, incorporating the scaling of the sources, becomes

\[
\mathbf{T}_{ij} = \begin{bmatrix}
-K_{\hat{n} \times, ij} & \frac{1}{\eta} \mathbf{I}_{v, \hat{n} \times, ij} & \mathbf{0} & \frac{1}{\eta} \mathbf{I}_{s, \hat{n} \times, ij} \\
\frac{1}{\eta} \mathbf{I}_{v, \hat{n} \times, ij} & -K_{\hat{n} \times, ij} & \mathbf{0} & \frac{1}{\eta} \mathbf{I}_{s, \hat{n} \times, ij} \\
\mathbf{0} & \frac{1}{\eta} \mathbf{I}_{v, \hat{n} \times, ij} & K_{\hat{n}, ij} & \mathbf{0} \\
-K_{\hat{n}, ij} & \frac{1}{\eta} \mathbf{I}_{v, \hat{n} \times, ij} & \mathbf{0} & \frac{1}{\eta} \mathbf{I}_{s, \hat{n}, ij}
\end{bmatrix} \tag{4.83}
\]

where the matrix elements are given by

\[
[\mathbf{I}_{v, \hat{n} \times}]_{m,n} = \begin{bmatrix}
w_n\hat{i}_m \cdot \hat{n} \times \mathcal{L}_v^s(r_m, r_n)\hat{i}_{n} \\
w_n\hat{j}_m \cdot \hat{n} \times \mathcal{L}_v^s(r_m, r_n)\hat{j}_n
\end{bmatrix} \tag{4.84}
\]

\[
[\mathbf{I}_{s, \hat{n} \times}]_{m,n} = \begin{bmatrix}
w_n\hat{i}_m \cdot \hat{n} \times \mathcal{L}_s^s(r_m, r_n)\hat{i}_{n} \\
w_n\hat{j}_m \cdot \hat{n} \times \mathcal{L}_s^s(r_m, r_n)\hat{j}_n
\end{bmatrix} \tag{4.85}
\]

\[
[\mathbf{K}_{\hat{n} \times}]_{m,n} = \begin{bmatrix}
w_n\hat{i}_m \cdot \hat{n} \times \mathcal{K}^s(r_m, r_n)\hat{i}_{n} \\
w_n\hat{j}_m \cdot \hat{n} \times \mathcal{K}^s(r_m, r_n)\hat{j}_n
\end{bmatrix} \tag{4.86}
\]

\[
[\mathbf{I}_{v, \hat{n}}]_{m,n} = \begin{bmatrix}
w_n\hat{i} \cdot \mathcal{L}_v^s(r_m, r_n)\hat{i}_{n} \\
w_n\hat{j} \cdot \mathcal{L}_v^s(r_m, r_n)\hat{j}_n
\end{bmatrix} \tag{4.87}
\]

\[
[\mathbf{K}_{\hat{n}}]_{m,n} = \begin{bmatrix}
w_n\hat{i} \cdot \mathcal{K}^s(r_m, r_n)\hat{i}_{n} \\
w_n\hat{j} \cdot \mathcal{K}^s(r_m, r_n)\hat{j}_n
\end{bmatrix} \tag{4.88}
\]

\[
[\mathbf{I}_{s, \hat{n}}]_{m,n} = \begin{bmatrix}
w_n\hat{i} \cdot \mathcal{L}_s^s(r_m, r_n)\hat{i}_{n}
\end{bmatrix} \tag{4.89}
\]
4.6 Casimir Results Using A-EPA

Using the above implementation of the A-EPA using point sampling, excellent results were achieved for both the Casimir energy and force. The A-EPA was used to produce the impedance matrix representing the integrand for the Casimir energy. As stated above, the A-EPA matrix is normalized and thus we can use an iterative eigenvalue solver, in this case ARPACK, to find the largest and smallest eigenvalues to perform the pairwise additions. It was found that a very modest number of eigenvalues were generally needed, around 6-16. It should be noted that the self-normalization properties disappear when the tap basis is used. The normalization can be performed by using an iterative solver like GMRES to perform the $\mathbf{M}_\infty^{-1}$ matrix-vector product. Using the tap basis, the number of unknowns represented by the taps should be low, allowing for the iterative solver to have a low cost. The Casimir force performs better because it was found that the eigenvalues represent the eigenvalue pairs that would be found for the EFIE or A-EFIE case. Thus, the A-EPA Casimir force does not suffer from cancellation over the sum of its eigenvalues allowing for fewer eigenvalues to reach convergence.

To find the gradient of the impedance matrix for the Casimir force, a simple finite difference was taken whereby the Casimir object was displaced within the ES. This means that, for example in the case of a three ES system where one ES contains only the Casimir object, the eigenvalue problem can be cast as

$$\nabla_i \mathbf{M} = \begin{bmatrix} 0 & -\nabla_i \mathbf{S}_{11} \cdot \mathbf{T}_{12} & -\nabla_i \mathbf{S}_{11} \cdot \mathbf{T}_{13} \\ 0 & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

(4.90)

The above makes the matrix vector product a very efficient implementation. Since we can solve either the standard eigenvalue problem of $\mathbf{M}_\infty^{-1} \cdot \nabla_i \mathbf{M}$ or $\nabla_i \mathbf{M} \cdot \mathbf{M}_\infty^{-1}$, the eigenvalue problem can be reduced so that only the eigenvector entries associated with the sources along the $i$-th ES must be non-zero. The standard eigenvalue problem for the Casimir force
using A-EPA is thus

\[
\begin{bmatrix}
0 & -\nabla I_{11} \cdot T_{12} & -\nabla I_{11} \cdot T_{13} \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
I_{11} & -S_{11} \cdot T_{12} & -S_{11} \cdot T_{13} \\
-S_{22} \cdot T_{21} & I_{22} & -S_{22} \cdot T_{23} \\
-S_{33} \cdot T_{31} & -S_{33} \cdot T_{32} & I_{33}
\end{bmatrix}^{-1}
\begin{bmatrix}
x_1 \\
0 \\
0
\end{bmatrix}
= \lambda
\begin{bmatrix}
x_1 \\
0 \\
0
\end{bmatrix}
\tag{4.91}
\]

where \(x_1\) is the full eigenvector. The inverse can be accurately calculated using an iterative solver due to the fact that the A-EPA matrix is so well-conditioned. Thus, despite the inherent approximations in the use of finite difference for the gradient, an iterative solver for the inverse matrix-vector product, and the use of an iterative eigenvalue solver, the final calculated force compares very favorably with direct methods while at the same time showing a reduction in CPU time and memory. It should also be noted that the scaling of the sources is important to the stability of the iterative eigenvalue solver. Without the scaling, the charges and currents will differ by many orders of magnitude. This introduces irrecoverable numerical problems in the iterative eigenvalue solver.

As a proof of concept, the original case of two PEC spheres used in the A-EFIE calculations were reproduced. It was found, through trial and error, that the lowest number of unknowns to represent the A-EPA sources were achieved when the ES were placed 0.25 m from the surface of the sphere of radius 1 m, the patches had edge lengths of \(\frac{1}{3}\) of the length of the ES cube, and each patch had six sampling points. In general, the distances are independent of wavelength since many of the calculations are taken at such low frequencies that the dimensions involved are sub-wavelength. Instead, it was found that the overall number of unknowns for a given cube size was more important. While the use of more sampling points or a finer mesh would allow the ES to be closer to the object, the increase in the number of unknowns was too much. Thus, with a separation between the ES and the object of 0.25 m, the minimum separation that can be simulated would be 0.5 m. For smaller simulations, the ES were kept the same size, placed side by side, and
the PEC spheres were moved in the interior of the ES. It was found that results remained accurate for the same number of unknowns down to a separation of 0.1 m.

The PEC spheres were identical spheres generated with average edge lengths of 0.25 m. This resulted in 2,370 total unknowns. The results were generated from 4.0 m to 0.6 m of separation using the same $S$ matrix for both of the PEC spheres. The results took 5 minutes and 16 seconds for the Casimir energy and 9 minutes and 22 seconds for the Casimir force using a quad core Intel Core i7 (Q740) laptop. Almost the entirety of the code took advantage of OpenMP in the matrix filling and in the matrix products using the Intel Math Kernel Library. The Casimir energy required 69 MB and the force 114 MB of memory to hold the matrices with 3,456 unknowns (the Casimir force solved for an eigenvector of size 1,728 unknowns). Next, the results from 0.5 m to 0.1 m were calculated using unique scattering matrices. The energy results took 7 minutes and 38 seconds using 92 MB and the force took 14 minutes and 19 seconds using 114 MB. This is to be compared with the A-EFIE implementation that took 4 hours and 50 minutes for the energy and 2 hours and 5 minutes for the force using 120 MB. The reason why the A-EFIE is so much more time-consuming is that the direct eigenvalue solver takes much longer to solve for the eigenvalues than the iterative solver. In addition, the eigenvalue solver was not parallelized like the matrix-vector products. Although the time spent calculating the matrix-vector products in the eigenvalue solver is so short that its parallelization has a negligible impact.

The accuracy of the results is excellent, with errors less than 10% for all the points greater than 0.1 m and often with errors on the order of 1%. The results can be seen in Figures 4.5a and 4.5b. Overall, it took 12 minutes and 54 seconds for the energy and 23 minutes and 41 seconds for the force. The number of unknowns for the A-EPA was higher than for the A-EFIE but that is to be expected for such a simple object. The advantage of the A-EPA is the ability to reduce the representation of complex contained objects. One example would be investigating the case of the force between two corrugated plates. The corrugation would increase the number of unknowns for the volume of the object. In addition, previous investigations into corrugated plates made use of 2D approximations. They did not take into account edge effects nor were they able to allow for unaligned corrugations.
Figure 4.5: Comparison of the EFIE and A-EFIE results for (a) Casimir energy and (b) Casimir force between 2 PEC spheres.

We took the example of a corrugated plate, shown in Figure 4.6, that was originally designed as a quarter-wavelength corrugation at 200 MHz. In general, the object is generated with a mesh that can be scaled according to the desired length scale. So the plate is 5.8 units square, 1.5 tall, a fin width of 0.1 units with the depth of the corrugation being 0.75 units and 0.375 units wide where units can be m or nm for example. Using an edge length of 0.3 units, or $\lambda/5$ when using meters as the units of measurement, the mesh has 5,316 patches and 8,823 edges. This gives us 14,129 unknowns for one plate and 28,278 unknowns for two plates using the A-EFIE formulation resulting in a matrix of 5.96 GB and 11.92 GB of memory needed for the force calculation.

Using the optimum ES conditions found for the PEC sphere, the ES was located 0.25 units away from the plate, had an edge length of around 1.5 units, and used seven sampling points per patch. The resulting ES mesh comprised 96 patches resulting in an A-EPA matrix of 8,064 unknowns, an eigenvector of size 4,032 unknowns, and a memory requirement of 620 MB for the Casimir force. This represents a reduction of 71% in the number of unknowns and 95% in memory. With such a large matrix needed to be inverted, there were concerns that the LU decomposition in the inner scattering solver would not be accurate. The accuracy of the LUD was confirmed by the use of iterative refinement of the
solution. After solving for the inverse matrix product, the solution is fed into a GMRES routine that uses the original matrix and right-hand side to iteratively improve the LUD solution to a small tolerance, taken as $10^{-10}$. The RCS results between the unrefined and refined solutions were identical, confirming that the LUD was still accurate with such a large matrix.

The plates were simulated over a distance between 0.55 and 4.0 units. Normally we would take five Gaussian points for the frequency but we would like to make use of the ability to store and reuse the scattering matrices. If we choose 12 frequency points for the Gaussian quadrature for the typical bandwidth of $kZ = 7.0$ at $Z = 0.5$ units, then 5 of the 12 integration points will still lie within the bandwidth specified by a 5-point integration at $Z = 4.0$ units. Thus, we can generate the scattering matrices for the 12 integration points and reuse them for any separation between 0.5 and 4.0 units. We could improve our efficiency by also noting that the two corrugated plates are identical with the second plate undergoing a rotation and translation. So we could generate two scattering matrices (one of them being the gradient matrix) and generate the appropriate translation matrix to represent the second plate. This was not done with these results, but it remains an option to be implemented in the future.

To generate all the scattering matrices for the 12 integration points took 650 minutes and 4.36 GB of total memory for the scattering and translation matrices. Then, each
frequency point would calculate the translation matrices and 16 eigenvalues to estimate the sum. This took an average 3.7 minutes per force calculation. The results at a micrometer scale are shown in Figures 4.7a and 4.7b for the cases of the corrugations being aligned and anti-aligned although using proper scaling the problem is scale invariant. Results were also obtained at the m and nm scales that were identical. This confirms that the algorithm is truly multi-scale. There is decent agreement with the PFA results which validates the results and shows that the A-EPA can handle objects large enough to approximate infinite plates. While a rectangular corrugation of this type can thus be approximated using the PFA, other corrugations like a sinusoidal may not, but by extension from these results should be well simulated by the A-EPA. In addition, the previous two-dimensional methods used to calculate the force between corrugated plates would not be able to handle the anti-aligned case making this a completely novel simulation.

Figure 4.7: The A-EPA results for the (a) aligned and (b) anti-aligned corrugated plates compared to the PFA results for μm scale dimensions.
The work presented in the previous chapters represents only a brief proof of concept for the use of advanced computational electromagnetics in calculating the Casimir force. By using the argument principle, any number of appropriate matrix methods can now be used in calculating the Casimir force. This will open up a wide array of possible geometries for simulation. With the techniques presented in here, the A-EPA could be further developed.

The main drawback to the Nyström method is that the higher ordered basis functions make it difficult to change the density of the ES mesh without a large change in the associated number of unknowns. This was dealt with by the use of a simple point collocation sampling. However, a simple linear basis function, like the rooftop basis, may be a suitable substitute. In addition, the limitations in using the RWG basis function could be addressed by using the Tangential-EPA (T-EPA) formulation [96, 97].

Currently, the equivalent currents are associated to the fields along the ES via the following matrix operator problem:

\[
\begin{bmatrix}
I & 0 \\
0 & I
\end{bmatrix} \cdot \begin{bmatrix}
J \\
M
\end{bmatrix} = \begin{bmatrix}
\hat{n} \times H \\
-\hat{n} \times E
\end{bmatrix}
\]

(5.1)

This relationship is a desirable matrix problem because it has a trivial solution. However, the RWG basis is a poor choice for testing \(\hat{n} \times\) terms [98]. Instead, a more numerically accurate result would be obtained with tangentially projected operators. Assuming that \(\hat{n}\) is always the outward pointing normal of the ES, then it is our desire that given a field distribution, we wish to find the equivalent currents on the ES that give rise to the same field outside the ES with a null field inside the ES. The relationship between the currents along the ES and the fields on the ES can be found and is similar to that already described
with the main difference being that because the fields are evaluated on the ES, then a factor of $\frac{1}{2}$ arises due to the principle value integral [51]. The resulting relationship between the fields and currents is

$$
\begin{bmatrix}
\mathcal{L}^{es_{ii}} & \mathcal{K}^{es_{ii}} \\
-\mathcal{K}^{es_{ii}} & \frac{1}{\eta^2} \mathcal{L}^{es_{ii}}
\end{bmatrix}
\begin{bmatrix}
J \\
M
\end{bmatrix}
= \frac{1}{2}
\begin{bmatrix}
E \\
H
\end{bmatrix}
$$

(5.2)

where $es_{ii}$ means that the operator takes its source and observation points to be both along the $i$-th ES. Middle frequency tests of the T-EPA show that it has comparable accuracy with the Nyström method that was originally developed. The T-EPA will have the advantage of more flexibility in adjusting the density of the ES meshes in addition to making it easier to account for the singularity terms in the Green’s function and its gradient. For the Casimir force, an augmented version of the T-EPA will need to be developed but the augmentation should follow the same steps described in this work.

Another area ripe for development is the case of inhomogeneous objects as currently, only homogeneous objects have been simulated using MOM. Volume integral equation methods can easily be substituted or the use of impedance boundary conditions can be used to represent coatings on nanoparticles. This would investigate how thin surface coatings can affect the force between small objects. For example, surface plasmon modes could give rise to interesting effects.

Finally, a large number of MEMS devices are constructed on a semiconductor substrate using lithographic techniques. The substrate, or other layers, could be modelled using the dyadic Green’s function for layered media. The argument principle method once again shows how we can trivially demonstrate that the DGLM can be used in a matrix method. A simple starting point would be to use the closed form case of an infinite PEC sheet. One could easily calculate the attraction of a PEC sphere with the infinite sheet as an initial proof of concept. The interpolation technique described in this thesis would provide an excellent speedup for these problems. When calculating the Casimir force, one could precalculate the tabulation tables for the problem at the frequencies of interest and over the possible object separations of interest in a manner similar to what was done in the A-EPA
simulations. Since the information could be used repeatedly, it should greatly reduce the computation time needed. Dyadic Green’s functions for spherically layered geometries may be useful for the case of a spherical nanoparticle with various coatings. However, one would first wish to judge whether or not the modes would penetrate deeply into the object to warrant using a true layered model over a thin film coating approximation.
CHAPTER 6

CONCLUSION

The preceeding chapters on Casimir force calculations outlined different formulations and the brief work that has been done to extend their functionality. The three main goals were to extend the functionality for arbitrary objects and materials, extend the codes to large scale problems, and provide a new formulation to support these extensions. The new formulation has been derived and a proof of concept has been done with incorporating A-EFIE as a low frequency method. At the same time, while achieving a broadband FMA algorithm has failed, we have provided the groundwork for a future FMA implementation. A method for implementing iterative solvers has been devised and proven while the numerical cost of a MF-FMA implementation has been estimated. True large scale results may be achieved in the future through the use of a parallelized FMA code and more advanced CPUs. Finally, the class of objects that can be simulated has been extended through the use of DDM with the A-EPA.

The use of the EPA presented difficult technical problems as was demonstrated in the implementation of the original RWG based EPA. Using PMCHWT, we were able to find the Casimir force between homogeneous dielectric objects. However, the limitations of the EPA with low frequency breakdown prevented its use in solving for the Casimir force. It was necessary to augment the EPA to handle low frequencies and include a Nyström method to handle the high frequencies and ES of arbitrary shape. In addition, the original high order Nyström sampling has the drawback that the ES mesh is very low density but the number of unknowns per patch is high. This makes it difficult to efficiently tune the number of unknowns in the A-EPA matrix since a minimal increase in the patch density causes a very large increase in the number of unknowns. This was addressed by omitting the high order interpolation function and using Dirac deltas. In the end, we demonstrated
accurate results down to small separations for length scales from meter to nanometer. We also simulated complex, novel geometries that have not been done before nor would be realizable using the original A-EFIE dense methods.

In summary, the primary contribution of this work has been to integrate boundary element methods into the argument principle technique. Previous uses of the argument principle have shown that the use of scattering methods can simulate a wide variety of geometries and materials. However, each of these methods required a tailored formulation for a specific class of problems. That is, a dispersion relation is derived based upon the given geometry and materials and this dispersion relation is expressed using a closed form equation or matrix representation. But there was no cohesive method of deriving the dispersion relation for all objects and materials. However, the class of BEM used in CEM can simulate practically any geometry and material. This allows for a single algorithm to be used across a wide array of problems without having to specifically reformulate it for new geometries. The avoidance of the path integral makes the argument principle approach far more accessible to researchers and the CEM techniques are more accessible as well due to the fact that they are rooted in classical electromagnetics. All of this helps to extend the class of problems that now can be simulated while greatly lowering the knowledge barrier into the subject.
A.1 Quasi-Static Subtraction

It should be noted that the handling of the quasi-static subtraction requires some additional discussion. The first problem that arises is that in the equations for the quasi-static subtraction term, shown in Equations (2.69) through (2.74), there exists an unavoidable singularity. Only when we are calculating when the source and observation points coincide on a boundary, i.e. \( r_s = Z = 0 \), will this singularity arise. The singularity exists in the first-order and second-order derivatives, namely Equations (2.70) and (2.71). These singularities are contained in the quasi-static subtraction terms and do not exist in the integrals used to fill the tabulation grids used in the interpolations. As such, the performance of the interpolation is not affected by the existence of these singularities. However, the user needs to consider how to handle the quasi-static subtraction at these singularities. The singularity arising from the second-order derivative of the quasi-static subtraction term, Equation (2.71), can be ignored. This is because the closed form of the quasi-static term is equivalent to the direct contribution from the source. As such, the correction of this quasi-static subtraction term can be handled by the user outside of the DGLM interpolation code.

The singularity from the first-order derivative of the quasi-static subtraction term, Equation (2.70), cannot be ignored. However, the singularity is still integrable; it is a natural log. The singularity also does not arise during the integrations used in a standard MOM code. Using the RWG basis, the triangular mesh must only have edges along the boundaries between layers for penetrating scatterers. The numeric quadrature rules used
will generally only choose points on the interior of the surface of integration. As such, since
the singularity would only lie on the boundaries of the surface, the singularity is not called
by the numeric integration. This would only happen if the entire patch was in the plane of
the boundary. The first-order derivative singularity only arises in $g_{z1}$ and $g_{z2}$. The inner
products associated with these terms, shown in Equation (2.17), take the dot product of
the test or basis currents with $\hat{z}$. If we are calculating the surface integral between a patch
and itself that is in the plane of a boundary, then this dot product is zero. Thus, the
singularity is not an issue in most implementations and if necessary, it is an integrable
singularity that the user can handle.

In addition, the evaluation of the integrals in Equations (2.69) through (2.74) can be
done in an efficient manner. The integral of $\frac{e^{ikmR-1}}{R}$ in Equations (2.69) and (2.70) can be
tabulated and interpolated using a one-dimensional interpolation. This requires a minimal
amount of memory and CPU time and the interpolation can be done using the existing
interpolation algorithms. The exponential integral used in Equations (2.72) and (2.73) can
be computed using one of the many fast algorithms that are available [32]. Taken on the
whole, when we consider the case where the source and observation points can lie in the
same layer, we will populate the tabulation grids with the values of $g_1$ and $g_2$, where $g_2$ has
been treated with quasi-static subtraction. In addition to constructing these tabulation
grids, we will also create a tabulation grid for the integral of $\frac{e^{ikmR-1}}{R}$. When filling the
impedance matrix, we will interpolate the values of $g_1$ and $g_2$ and add back in the
quasi-static subtraction term for $g_2$ when necessary. In doing so, we only need to calculate
interpolations and exponential integrals. Since the exponential integral can be computed
quickly using the appropriate algorithm, the calculation time of the DGLM will be fairly
constant and independent of the background inhomogeneity. The calculation time will be
mainly dependent upon the number of points used for the interpolation.
A.2 Interpolation

When using the one-dimensional interpolations, the user will need to specify which 
$z$-planes the tabulation grids will need to be calculated for and which $z$-plane is required 
for each interpolation. This is done by passing in an array containing the values of the 
unique $z$ positions that the scatterer lies in. For example, in the simulation of the two 
circular patches, we confined the scatterers to the planes $z = 0$ cm and $z = 0.07874$ cm. 
Thus, we passed on the values 0 and 0.0007874 to the interpolation code and then specified 
which plane the source and observation points were in when we called for an interpolation. 
The latter was done by referencing the index of the array that was passed in. The use of a 
one-dimensional interpolation will also require the calculation and tabulation of the 
derivatives of $g^{TE,TM}$. Normally we would estimate the derivatives with respect to $z$ and $z'$ 
using interpolation but since we do not need to interpolate along the $z$ or $z'$ dimensions, we 
need to calculate, tabulate, and then interpolate the derivatives. This will require an 
additional tabulation grids.

The rest of the error plots for the results for $g^{TM}$ and $g^{TE}$ and their derivatives can be 
found in Figures A.1a through A.8. Figures A.9a through A.14b show the error results in 
the same manner as previously presented for the Green’s functions used to fill the 
impedance matrix.
Figure A.1: Comparison of the (a) mean error and the (b) root mean squared error for $\frac{\partial g^{TM}}{\partial z}$ for varying step sizes and number of points used for interpolation.

Figure A.2: Comparison of the (a) maximum error for $\frac{\partial g^{TM}}{\partial z}$ and the (b) mean error for $\frac{\partial^2 g^{TM}}{\partial z^2}$ for varying step sizes and number of points used for interpolation.
Figure A.3: Comparison of the (a) root mean squared error and the (b) maximum error for $\partial^2 g^{TM}/\partial z^2$ for varying step sizes and number of points used for interpolation.

Figure A.4: Comparison of the (a) mean error and the (b) root mean squared error for $g^{TE}$ for varying step sizes and number of points used for interpolation.
Figure A.5: Comparison of the (a) maximum error for \( g^{TE} \) and the (b) mean error for \( \frac{\partial g^{TE}}{\partial z} \) for varying step sizes and number of points used for interpolation.

Figure A.6: Comparison of the (a) root mean squared error and the (b) maximum error for \( \frac{\partial g^{TE}}{\partial z} \) for varying step sizes and number of points used for interpolation.
Figure A.7: Comparison of the (a) mean error and the (b) root mean squared error for $\frac{\partial^2 g^{TE}}{\partial z^2}$ for varying step sizes and number of points used for interpolation.

Figure A.8: Comparison of the maximum error for $\frac{\partial^2 g^{TE}}{\partial z^2}$ for varying step sizes and number of points used for interpolation.
Figure A.9: Comparison of the (a) mean error and the (b) root mean squared error for $g_{ss}$ for varying step sizes and number of points used for interpolation.

Figure A.10: Comparison of the (a) maximum error for $g_{ss}$ and the (b) mean error for $g_{zz}$ for varying step sizes and number of points used for interpolation.
Figure A.11: Comparison of the (a) root mean squared error and the (b) maximum error for $g_{zz}$ for varying step sizes and number of points used for interpolation.

Figure A.12: Comparison of the (a) mean error and the (b) root mean squared error for $g_{z1}$ and $g_{z2}$ for varying step sizes and number of points used for interpolation.
Figure A.13: Comparison of the (a) maximum error for $g_{z1}$ and $g_{z2}$ and the (b) mean error for $g_{\phi}$ for varying step sizes and number of points used for interpolation.

Figure A.14: Comparison of the (a) root mean squared error and the (b) maximum error for $g_{\phi}$ for varying step sizes and number of points used for interpolation.
APPENDIX B

CASIMIR FORCE BACKGROUND DISCUSSIONS

The original derivation of the Casimir force was done by analyzing the change in the electromagnetic zero point energy [40]. This derivation begins by analyzing the quantum electrodynamics (QED) description of an electromagnetic field. It can be shown that the electromagnetic field in QED can be described as a harmonic oscillator and as a consequence, must have a non-zero ground state energy [43].

B.1 Quantum Vacuum

The quantum electrodynamic picture is derived starting from Maxwell’s equations for a field in a source-free region [43,63].

\[ \nabla \cdot \mathbf{E} = 4\pi \rho \]  
\[ \nabla \cdot \mathbf{B} = 0 \]  
\[ \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \]  
\[ \nabla \times \mathbf{B} = \frac{1}{c} \left( \frac{\partial \mathbf{E}}{\partial t} + 4\pi \mathbf{J} \right) \]

Using the appropriate vector and scalar potentials, we will define the magnetic flux density as

\[ \mathbf{B} = \nabla \times \mathbf{A} \]
In order to fully determine the vector potential $\mathbf{A}$, we will choose the Coulomb gauge whereby

$$\nabla \cdot \mathbf{A} = 0$$  \hspace{1cm} (B.6)

Substituting the vector potential form of the magnetic flux density into Equation (B.3) tells us that the expression $\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}$ must have zero curl. This implies that the expression must be equal to the gradient of some potential $\phi$. As such,

$$\mathbf{E} = -\nabla \phi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}$$ \hspace{1cm} (B.7)

In the absence of any sources, $\phi = \mathbf{J} = 0$. Noting the

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = -\nabla^2 \mathbf{A},$$

we can use Equation (B.4) to find

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0$$ \hspace{1cm} (B.8)

We will assume monochromatic solutions and, using separation of variables, define the vector potential to be

$$\mathbf{A}(\mathbf{r},t) = \alpha(t)\mathbf{A}_0(\mathbf{r}) + \alpha^*(t)\mathbf{A}_0^*(\mathbf{r}) = \alpha(0)e^{-i\omega t}\mathbf{A}_0(\mathbf{r}) + \alpha^*(0)e^{i\omega t}\mathbf{A}_0^*(\mathbf{r})$$ \hspace{1cm} (B.9)

Thus, the vector $\mathbf{A}_0(\mathbf{r})$ satisfies the Helmholtz equation,

$$\nabla^2 \mathbf{A}_0(\mathbf{r}) + k^2 \mathbf{A}_0(\mathbf{r}) = 0$$ \hspace{1cm} (B.10)

where the wave number $k = \omega/c$ and $\ddot{\alpha}(t) = -\omega^2 \alpha(t)$. The resulting electric and magnetic fields are given by

$$\mathbf{E}(\mathbf{r},t) = -\frac{1}{c} \left[ \dot{\alpha}(t)\mathbf{A}_0(\mathbf{r}) + \dot{\alpha}^*(t)\mathbf{A}_0^*(\mathbf{r}) \right]$$ \hspace{1cm} (B.11)

$$\mathbf{B}(\mathbf{r},t) = \alpha(t) \nabla \times \mathbf{A}_0(\mathbf{r}) + \alpha^*(t) \nabla \times \mathbf{A}_0^*(\mathbf{r})$$ \hspace{1cm} (B.12)
To derive the Hamiltonian, we will find the total electromagnetic energy in terms of our vector potential. However, we first must note that

\[ \int_V dr \left[ \nabla \times A_0(r) \right]^2 = k^2 \int_V dr A_0(r)^2 \]  

(B.13)

The derivation of Equation (B.13) can be found in Appendix B.4 and can be similarly applied in the case of \( \left[ \nabla \times A_0^*(r) \right]^2 \) and \( |\nabla \times A_0(r)|^2 \). The resulting Hamiltonian for electromagnetic fields is

\[
H_F = \frac{1}{8\pi} \int_V dr \left( E^2 + B^2 \right) = \frac{1}{8\pi c^2} \dot{\alpha}(t)^2 \int_V dr A_0(r)^2 + \frac{1}{8\pi c^2} \dot{\alpha}^*(t)^2 \int_V dr A_0^*(r)^2 \\
+ \frac{1}{4\pi c^2} |\dot{\alpha}(t)|^2 \int_V dr |A_0(r)|^2 + \frac{1}{8\pi} \alpha(t)^2 \int_V dr |\nabla \times A_0(r)|^2 \\
+ \frac{1}{8\pi} \alpha^*(t)^2 \int_V dr |\nabla \times A_0^*(r)|^2 + \frac{1}{4\pi} |\alpha(t)|^2 \int_V dr |\nabla \times A_0(r)|^2
\]  

(B.14)

Noting Equation (B.13), the fact that \( \dot{\alpha}(t)^2 = -\omega^2 \alpha(t)^2 \) and assuming that the mode function \( A_0(r) \) is normalized such that

\[ \int_V dr |A_0(r)|^2 = 1 \]  

(B.15)

we can reduce Equation (B.14) to

\[ H_F = \frac{k^2}{2\pi} |\alpha(t)|^2 \]  

(B.16)

We will define the classical real quantities

\[
q(t) = \frac{i}{c^2 \sqrt{4\pi}} \left[ \alpha(t) + \alpha^*(t) \right] \\
p(t) = \frac{k}{c \sqrt{4\pi}} \left[ \alpha(t) - \alpha^*(t) \right]
\]  

(B.17, B.18)

which allow us to rewrite Equation (B.16) as

\[ H_F = \frac{1}{2} \left( p^2 + \omega^2 q^2 \right) \]  

(B.19)
The above satisfies Hamilton’s equations of motions where it is required that

\[
\dot{q} = \frac{\partial H}{\partial p} = \{q, H\} \quad (B.20)
\]

\[
\dot{p} = -\frac{\partial H}{\partial q} = \{p, H\} \quad (B.21)
\]

This gives us an avenue to convert the classical Hamiltonian to a quantum one by quantizing the Poisson brackets into the canonical commutation relations. The classical and quantum Hamiltonian for a harmonic oscillator is expressed as

\[
H_{\text{osc}} = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2 \quad (B.22)
\]

where \( p \) and \( q \) are operators in the quantum Hamiltonian. The quantum operators of the harmonic oscillator share the same form as the classical Hamilton equations.

\[
\dot{q} = \frac{1}{i\hbar} \left[ q, \hat{H} \right] = \frac{\hat{p}}{m} \quad (B.23)
\]

\[
\dot{p} = \frac{1}{i\hbar} \left[ p, \hat{H} \right] = -m\omega^2 \hat{q} \quad (B.24)
\]

It should be noted here that we are working in the Heisenberg picture as opposed to the Schrödinger picture. In the Heisenberg picture, the operators carry a time dependence and the eigenstates are time independent. In the Schrödinger picture, the operators are time independent and the eigenstates evolve in time. Analysis of the harmonic oscillator allows us to rewrite the \( \hat{p} \) and \( \hat{q} \) operators in the form of lowering and raising operators [43] [99] which will be defined as

\[
\hat{a} = \frac{1}{\sqrt{2m\hbar\omega}} (-i\hat{p} + m\omega\hat{q}) \quad (B.25)
\]

\[
\hat{a}^\dagger = \frac{1}{\sqrt{2m\hbar\omega}} (i\hat{p} + m\omega\hat{q}) \quad (B.26)
\]

Here, \( \hat{a} \) and \( \hat{a}^\dagger \) are the lowering and raising operators and when acting on an eigenstate will change the eigenstate to the lower or higher adjacent eigenstate respectively. In addition,
the resulting energy levels are described as

\[ E_n = \left( n + \frac{1}{2} \right) \hbar \omega, \quad n = 0, 1, 2, \ldots \]  

(B.27)

It is obvious that the classical Hamiltonian for the electromagnetic field is reminiscent of the classical and quantum Hamiltonian for the harmonic oscillator. One can easily show that the Hamilton equations are similar too as \( \dot{\hat{q}} = \hat{p} \) and \( \dot{\hat{p}} = -\omega^2 \hat{q} \). The electromagnetic Hamiltonian is thus equivalent to a harmonic oscillator with a mass \( m = 1 \) and we redefine the classical variables to be

\[
\begin{align*}
\hat{\alpha}(t) &= \sqrt{\frac{2\pi \hbar c^2}{\omega}} \hat{a}(t) \\
\hat{\alpha}^*(t) &= \sqrt{\frac{2\pi \hbar c^2}{\omega}} \hat{a}^\dagger(t)
\end{align*}
\]

(B.28)

(B.29)

This replaces the classical vector potential with the operator

\[
A(r, t) = \sqrt{\frac{2\pi \hbar c^2}{\omega}} [\hat{a}(t)A_0(r) + \hat{a}^\dagger(t)A_0^*(r)]
\]

(B.30)

The resulting operators for the electric and magnetic fields then become

\[
\begin{align*}
\hat{E}(r, t) &= i\sqrt{2\pi \hbar \omega} [\hat{a}(t)A_0(r) + \hat{a}^\dagger(t)A_0^*(r)] \\
\hat{B}(r, t) &= \sqrt{\frac{2\pi \hbar c^2}{\omega}} [\hat{a}(t)\nabla \times A_0(r) + \hat{a}^\dagger(t)\nabla \times A_0^*(r)]
\end{align*}
\]

(B.31)

(B.32)

The Hamiltonian can be written into the same form as the harmonic oscillator.

\[
\hat{H}_F = \hbar \omega \left( \hat{a}^\dagger \hat{a} + \frac{1}{2} \right)
\]

(B.33)

Thus, the resulting energy levels are the same as in Equation (B.27). Now, the integer \( n \) denotes the number of photons in the field mode described by the state \( |n\rangle \). It is important to note that the vacuum state, \( |0\rangle \), is a ground state that contains no photons but still has an energy of \( \frac{1}{2} \hbar \omega \). Due to the fact that \( \hat{a} |n\rangle = \sqrt{n} |n - 1\rangle \) and the eigenstates are

168
orthogonal, $\langle n | \hat{a} | n \rangle$. This means that the expectation values of the fields will be zero.

$$\langle E(r, t) \rangle = \langle B(r, t) \rangle = 0 \quad \text{(B.34)}$$

Thus, the electric and magnetic field vectors will fluctuate with a mean value of zero in the state $|n\rangle$ despite the fact that the field has a constant non-zero energy.

### B.2 Origins of the Casimir Force

What we have shown above in our treatment of the quantum vacuum is that photon number states are not eigenstates of the electric and magnetic field operators. So while in the zero photon state, the vacuum state, the expectation value of the fields is zero. However, the fields still have a non-zero variance. These field fluctuations are still subject to their macroscopic classical behavior, confining them to satisfying the boundary conditions one derives from Maxwell’s equations. The reasoning from this can be seen from the quantum electrodynamic treatment of polarizable molecules in the quantum vacuum [43]. The Casimir-Polder force describes the intermolecular force between molecules due to quantum electrodynamics [2] and can be derived using the quantum vacuum. The fluctuating fields of the quantum vacuum induce a polarization in the molecules present. The induced fluctuating dipole moments in the molecules interact with each other with the aggregate effect being that the energy due to these dipole moments is dependent upon the relative position of the molecules. This spatial dependence in the energy of the system gives rise to an intermolecular force. The induction of these fluctuating dipole moments has a direct parallel with the classical electrodynamics of linear materials. That is, the electric displacement field is defined as

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \quad \text{(B.35)}$$

where $\mathbf{P}$ is the polarization density and reflects the density of the dipole moments induced by the applied $\mathbf{E}$ field. Likewise, the induced fluctuating dipole moments from the vacuum
fields can be thought of as another polarization density where the vacuum fields are our applied $E$ field. As such, if we take a bulk volume of polarizable molecules, we can encapsulate the effects of the induced dipole moments macroscopically in the displacement field $D$. Hence, the energy contained in the macroscopic electromagnetic fields is the energy of the vacuum fields and their induced moments in the materials. As we can see in Equation (B.33), the vacuum energy is non-zero and is divergent for increasing frequency. As such, the total energy of our system will be divergent and requires normalization to a finite quantity. This can be done by simply removing the energy of the vacuum field itself by normalizing by the energy of the system when the objects are at infinite separation. To demonstrate this more clearly, we can take the example of two parallel infinite PEC plates.

**B.3 Parallel Plates**

With the defining of the energy of the electromagnetic fields in Equation (B.27), we can now calculate the energy of the vacuum state and the changes that occur when objects are placed in the vacuum. This can be used to find the Casimir force acting on the objects, as in the case of two parallel perfectly conducting plates [40,43]. We will assume that there are two plates located in the $x - y$ plane at the locations $z = 0$ and $z = d$ with side lengths $L$. The appropriate field modes here will be confined by the rectangular parallelepiped of sides $L_x = L_y = L$ and $L_z$. The boundary conditions for the electric field require that the tangential components vanish on the walls of the conducting plates. This will allow us to define the vector potential, $\mathbf{A}(\mathbf{r}) = A_x(\mathbf{r})\hat{x} + A_y(\mathbf{r})\hat{y} + A_z(\mathbf{r})\hat{z}$, as

$$A_x(\mathbf{r}) = a_x \cos(k_x x) \sin(k_y y) \sin(k_z z) \quad (B.36)$$

$$A_y(\mathbf{r}) = a_x \sin(k_x x) \cos(k_y y) \sin(k_z z) \quad (B.37)$$

$$A_z(\mathbf{r}) = a_x \sin(k_x x) \sin(k_y y) \cos(k_z z) \quad (B.38)$$
with the wave numbers

\[ k_x = \frac{\ell \pi}{L}, \quad k_y = \frac{m \pi}{L}, \quad k_z = \frac{n \pi}{L} \]  \hspace{1cm} (B.39)

Neglecting the details of the derivation and the requirements of normalization, the zero-point energy of the system is dependent only on the number of modes and their frequencies. These frequencies are derived from the wave numbers of each mode.

\[ \omega_{\ell mn} = k_{\ell mn} c = \pi c \sqrt{\frac{\ell^2}{L^2} + \frac{m^2}{L^2} + \frac{n^2}{L_z^2}} \]  \hspace{1cm} (B.40)

The zero-point energy of the fields inside the cavity will be

\[ \sum_{\ell,m,n}' \frac{1}{2} \frac{\hbar \omega_{\ell mn}}{\omega_{\ell mn}} = \sum_{\ell mn}' \frac{\pi \hbar c}{\sqrt{\frac{\ell^2}{L^2} + \frac{m^2}{L^2} + \frac{n^2}{L_z^2}}} \]  \hspace{1cm} (B.41)

The factor of 2 is due to the fact that each mode has two standing waves and the prime indicates that the factor of 1/2 is needed when one of the mode integers is zero. In our given situation, \( L_z = d \) and we can assume that \( L \) is very large in comparison with \( d \). If \( L \) is large enough, we can regard \( k_x \) and \( k_y \) as continuous variables and replace the summations over them with integrals. Thus,

\[ E(d) = \sum_{\ell,m,n}' \frac{1}{2} \frac{\hbar \omega_{\ell mn}}{\omega_{\ell mn}} = \frac{L^2}{\pi^2} \frac{\hbar c}{\sqrt{\frac{\ell^2}{L^2} + \frac{m^2}{L^2} + \frac{n^2}{L_z^2}}} \int_0^\infty dk_x \int_0^\infty dk_y \int_0^\infty dk_z \frac{k_x^2 + k_y^2 + \frac{n^2 \pi^2}{d^2}}{\sqrt{k_x^2 + k_y^2 + k_z^2}} \]  \hspace{1cm} (B.42)

The original vacuum state can be described as being when the two plates are at infinite separation, in which case we can also replace the summation over \( k_z \) with a continuous integration.

\[ E(\infty) = \frac{L^2}{\pi^2} \frac{\hbar c}{\pi} \frac{d}{\pi} \int_0^\infty dk_x \int_0^\infty dk_y \int_0^\infty dk_z \frac{k_x^2 + k_y^2 + k_z^2}{\sqrt{k_x^2 + k_y^2 + k_z^2}} \]  \hspace{1cm} (B.43)
The potential energy of the entire system is thus the difference between the energy when the plates are separated by a distance \( d \) and by a distance that is infinitely large.

\[
U(d) = \frac{L^2 \hbar c}{\pi^2} \left[ \sum_n \int_0^\infty dk_x \int_0^\infty dk_y \sqrt{k_x^2 + k_y^2 + \frac{n^2 \pi^2}{d^2}} 
\right.
\vphantom{\int_0^\infty} \left. - \frac{d}{\pi} \int_0^\infty dk_x \int_0^\infty dk_y \int_0^\infty dk_z \sqrt{k_x^2 + k_y^2 + k_z^2} \right]
\] (B.44)

The evaluation of the integrals can be better facilitated by the use of polar coordinates \( u \) and \( \theta \) in the \( k_x - k_y \) plane. In the polar coordinates \( u \) and \( \theta \), Equation (B.44) becomes

\[
U(d) = \frac{L^2 \hbar c}{\pi^2} \left( \frac{\pi}{2} \right) \left[ \sum_n \int_0^\infty duu \sqrt{u^2 + n^2 \pi^2} \right.
\vphantom{\int_0^\infty} \left. - \int_0^\infty \frac{d}{\pi} \int_0^\infty duu \sqrt{u^2 + k_z^2} \right]
\] (B.45)

In addition, in order to achieve a finite result, we will introduce a cutoff function \( f(k) = f(\sqrt{u^2 + k_z^2}) \). The cutoff function will be defined such that \( f(k) = 1 \) for \( k << k_m \) and \( f(k) = 0 \) for \( k >> k_m \). Casimir notes that this cutoff will not adversely affect the solution since very short wavelengths, like X-rays, will pass through the conducting plate with minimum interference. Thus, the very short wavelength contributions to the zero-point energy are very small. Milonni suggests thinking of the limit \( k_m \) being on the order of \( 1/a_0 \) where \( a_0 \) is the Bohr radius. The assumption due to the cutoff is that the Casimir effect is primarily a low-frequency, nonrelativistic effect. Using the change of variables where \( x = u^2 d^2 / \pi^2 \) and \( \kappa = k_z d / \pi \), we arrive at

\[
U(d) = \frac{L^2 \hbar c}{4 \pi^2} \left( \frac{\pi^3}{d^3} \right) \left[ \sum_n \int_0^\infty dx \sqrt{x + n^2} f \left( \frac{\pi}{d} \sqrt{x + n^2} \right) 
\right.
\vphantom{\int_0^\infty} \left. - \int_0^\infty d\kappa \int_0^\infty dx \sqrt{x + \kappa^2} f \left( \frac{\pi}{d} \sqrt{x + \kappa^2} \right) \right]
\] (B.46)

The above is now an equation of the form

\[
U(d) = \frac{\pi^2 \hbar c}{4 d^3} L^2 \left[ \frac{1}{2} F(0) + \sum_n F(n) - \int_0^\infty d\kappa F(\kappa) \right]
\] (B.47)
Using the Euler-MacLaurin summation formula \([30]\), the above can be reduced to

\[
U(d) = \frac{\pi^2 \hbar c}{4d^3 L^2} \left[ -\frac{1}{12} F'(0) + \frac{1}{720} F'''(0) \ldots \right] \tag{B.48}
\]

It can be shown that \(F'(0) = 0\), \(F'''(0) = -4\), and all the higher derivatives \(F^{(n)}(0)\) vanish assuming all the derivatives of the cutoff function vanish at \(\kappa = 0\). Thus, the closed form equation for the Casimir energy becomes

\[
U(d) = -\frac{\pi^2 \hbar c}{720d^3} L^2 \tag{B.49}
\]

The force is the negative gradient of the energy and is

\[
F(d) = -\frac{\pi^2 \hbar c}{240d^4} L^2 \tag{B.50}
\]

where the direction of the force makes the force attractive between the two plates.

### B.4 Derivation of Equation (B.13)

The following derivation is taken nearly verbatim from Milonni [43]. The general identity

\[
\nabla \cdot (F + G) = G \cdot \nabla \times F - F \cdot \nabla \times G,
\]

together with the Coulomb gauge condition \(\nabla \cdot A_0 = 0\), implies

\[
(\nabla \times A_0)^2 = \nabla \times (A_0 \times \nabla \times A_0) + A_0 \cdot \nabla \times (\nabla \times A_0)
\]

\[
= \nabla \times (A_0 \times \nabla \times A_0) + A_0 \cdot [\nabla (\nabla \cdot A_0) - \nabla^2 A_0]
\]

\[
= \nabla \times (A_0 \times \nabla \times A_0) - A_0 \cdot \nabla^2 A_0
\]

\[
= \nabla \times (A_0 \times \nabla \times A_0) + k^2 A_0^2 \tag{B.51}
\]
when we use the fact that $A_0$ satisfies the Helmholtz equation shown in Equation (B.10). Then the divergence theorem implies

$$\int_V dr [\nabla \times A_0]^2 = \oint dS \hat{n} \cdot A_0 \times (\nabla \times A_0) + k^2 \int_V dr A_0(r)^2$$

$$= k^2 \int_V dr A_0(r)^2 \quad (B.52)$$

since the surface integral vanishes as a consequence of the assumed periodic boundary condition on $A_0(r)$. 
APPENDIX C

NORMALIZATION IN THE STRESS TENSOR METHOD

The stress tensor method derived by Xiong et al. performs renormalization by finding the
excited currents on the Casimir object when the other objects are removed to infinity.
These excited currents are then subtracted from the currents found for the original system.
That is, assuming that the Casimir object is contained in the upper left block \((\mathbf{Z}_{11})\) of the
impedance matrix, then

\[
\mathbf{J}' = \mathbf{J} - \mathbf{J}^0 = \begin{bmatrix} \mathbf{J}_1 - \mathbf{J}_1^0 \\ \mathbf{J}_2 \end{bmatrix}
\]  

where

\[
\mathbf{ZJ} = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix}
\]  

\[
\mathbf{Z}^0\mathbf{J}^0 = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{J}_1^0 \\ \mathbf{J}_2^0 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{0} \end{bmatrix}
\]

We can see that \(\mathbf{J}_2^0\) is trivially zero and that the second matrix problem is equivalent to the
problem of \(\mathbf{Z}_{11}\mathbf{J}_1^0 = \mathbf{V}_1\).

The current problem with this method is that the cancellation occurs in the solved
currents. The advantage of performing this cancellation is to help remove some of the
strong images and singularities that the point sources induce on the Casimir object. But
this is not done until after we have already had to solve for currents under these
undesirable excitations. That is, the singularities and strong images are created by the
excitations in the vector \(\mathbf{V}_1\) shown in Equation (C.2). If we were to normalize the Casimir
energy before solving for the currents $J'$, then we would expect that the resulting matrix equation would have the following excitation vector,

$$V' = V - V_\infty = \begin{bmatrix} 0 \\ V'_2 \end{bmatrix} \quad (C.4)$$

where $\bar{Z}J' = V'$. Physically, what we expect is that the renormalization of the Casimir energy would remove all the currents directly excited on the Casimir object by the point source excitations. These currents are represented by $J^0_1$. Instead, the remaining excitations would arise from the reflections from the non-Casimir objects that cause perturbation currents on the surface of the Casimir object from the scattering of the point source excitations off of the Casimir object. In short, after normalization we would expect that the effective excitation would be sourced from the non-Casimir objects.

This can be done using renormalization. Note that $\bar{Z}J = V$, $\bar{Z}^0J^0 = V^0$ and $J' = J - J^0$ as described above in Equations (C.1) through (C.3). Then,

$$V' = \bar{Z}J' = V - \bar{Z}J^0 = \begin{bmatrix} 0 \\ V_2 - \bar{Z}21J^0_1 \end{bmatrix} \quad (C.5)$$

Since we hypothesized that $V' = V - V_\infty$, we find that

$$V_\infty = \begin{bmatrix} V_1 \\ \bar{Z}21J^0_1 \end{bmatrix} = \bar{Z} \begin{bmatrix} J^0_1 \\ 0 \end{bmatrix} \quad (C.6)$$

Thus we can see that the new excitation to find the final currents matches our physical predictions. The effective excitation is one that is only sourced on the non-Casimir objects. This allows us to solve the problem in a different order than before. Before, we first solved $J = \bar{Z}^{-1}V$ and then $J^0 = \bar{Z}^{0-1}V^0$ and then found $J'$ via $J - J^0$. However, this meant that we included the singularities and strong images in both of our scattering problems. Now we can solve the problem by first solving $J^0 = \bar{Z}^{0-1}V^0$ and then $J' = \bar{Z}^{-1}[V - V_\infty]$. This means that we can truly remove these singularities and strong images when we solve one of the scattering problems. In addition, since $V_\infty = \bar{Z}J^0$, we find that the excitation for the
second scattering problem helps to reduce the impact of the numerical errors from the singularities and images that are present in calculating $J^0$. This is because we have pushed the fields from $J^0$ farther away onto our non-Casimir objects. This may help by removing some of the evanescent parts of the scattered field from $J^0$ and thus lessen the impact of numerical errors in calculating $J^0$.

By rearranging the renormalization of the Casimir energy, it is hoped that we can help reduce the numerical instability that has been present in the algorithm thus far. While we still have to deal with the hypersingularities and strong images when we solve for the $J^0$ currents, it is hoped that by moving the fields of these currents onto the non-Casimir objects we help soften the impact of their numerical error when we calculate the normalized currents $J'$. 

The set of simulation results seen in Figures C.1 is a proof of concept and shows a comparison of the results using the old and new right-hand side method. The plot shows the relative error in the Casimir force compared to using the old method with an iterative solver tolerance of $10^{-6}$. The new right-hand side's results using a tolerance of $10^{-4}$ are closer to the results using a stricter tolerance than the old method. Thus, the new method allows for more accurate results and shorter computation times.
REFERENCES


[38] H. Falkenhagen, *Physik. Z.*, vol. 21, p. 87, 1922.


