Simulating Surface Charge Effects in Carbon Nanotube Templated Ionic Crystal Growth

Todd Gingrich
University College
University of Oxford

A thesis submitted for the degree of
Master of Science by Research

Trinity 2010

To George, Steve, Hannah, Harry, Jerry, and Nate for getting me started.

Statement of Originality

The work presented in this thesis is my own. Information and figures which are derived from other sources are referenced in the text.

Signature: ________________________ Date: _______________

Acknowledgements

First and foremost I would like to acknowledge my supervisor, Dr Mark Wilson. Mark introduced me to interesting problems and let me attack them with a remarkable degree of freedom, never complaining when I decided to sample a few other things on the side, be they physics lectures or a day at Wimbledon. He was consistently approachable and available, a supervisory rarity that I valued greatly. I am also indebted to Prof David Logan for welcoming me into the department, helping me to fashion a rather uncommon academic arrangement that suited my needs, and for introducing me to the beauty of (elementary) many-body theory.

Special thanks goes to Rhodes House and the staff that made me feel at home, especially John, Bob, Colin, Sheila, and Mary. Thanks of course to the Wardens and the Trust for the financial support as well. My gratitude must also be extended to the gardeners, who provided me with a most excellent location for reading, lunch, and bocce.

My time in Oxford would not have been so pleasant were it not for some great friends I made along the way. Ben Eidelson and Clara Blättler deserve special recognition for tolerating my antics in our Cowley Road flat. Abhilash Mishra was so kind as to talk shop with me from time to time, a privilege I truly enjoyed. I'm wary of an explicit list for fear of omissions, but a number of good friends joined me in picking strawberries, traveling, catching crayfish, and playing euchre. They should know who they are.

I owe my Caltech professors for setting me up with the basics. I cannot imagine a four year education that would have better prepared me to mix a little math, physics, chemistry, and computers. Thanks also to the Caltech friends - Leo, Matt, Tom, Ashok, Alexandra, Ila, Elette, among others - who responded to random emails seeking clarity in math, physics, programming, and life.

Finally, I must acknowledge Mom, Dad, Rebecca, Mark, Kari, and the rest of my family, who have patiently accepted the fact that my interest in science often sends me far away. While I occasionally give them grief, my parents' support has been consistent (and not just because they wanted an excuse to visit Europe, though I'm sure that didn't hurt). Their love means a lot.

Abstract

A model is adopted to simulate molecular dynamics near metallic materials. Gaussian charges are induced to maintain the equipotential constraint of the metal. A regularization of the required Gaussian charge Ewald summation is performed using convergence factor methods. A particularly simple expression for the electrostatic energy is obtained, developed for a simple test case, and finally used to construct a classical model for metallic rigid carbon nanotubes (CNTs). Parameterization of the model is discussed and compared to both continuum electrostatic theory and to density functional representations of the metal surface. The model enables the study of the direct filling of CNTs by molten salts. A binary molten salt, modeled using a Born-Mayer potential, is observed to form inorganic nanotubes (INTs) within the CNTs, and these structures are seen to be insensitive to the treatment of the CNT as a metal or insulator. The INT formation is however sensitive to the addition of a voltage bias to the metallic CNT. By considering INT growth mechanisms driven by an external bias, it is shown that the distribution of computationally observed INT morphologies is path-dependent due to kinetic barriers. As such, the distribution of structures is out of equilibrium and therefore cannot be understood through the free energies.

Contents

Appendices

List of Figures

List of Tables

Introduction

Much attention has been given to the broad field of nanotechnology. The interest stems from the fact that ordered structures on the nanoscale often have structural and electronic properties that are quite distinct from bulk materials. From an engineering perspective these properties could be useful, yet our ability to utilize nanostructures hinges upon our ability to construct such devices. Given the small scale, it is likely that formation of future nanodevices will rely heavily on the ability to bias self-assembly. One common proposal is that self-assembly could be controlled by providing a suitable template. In this thesis we present computational experiments aimed at better understanding how much control can reasonably be obtained for the particular templated growth problem of molten salt crystalization inside a carbon nanotube (CNT). Theoretical work predicts that the internal salt could crystallize into a number of distinct inorganic nanotubular structures, providing an ideal model system on which to explore morphological control of self-assembled structures.

For this templated growth problem there are a handful of macroscopically controllable parameters that could conceivably influence the crystalization. The size, shape, and electronic structure of the template can be controlled by selecting a particular CNT morphology, the temperature and pressure may be varied, or a potential bias can be applied to the tube. The effects of tuning the temperature, pressure, and CNT radius have been previously explored within computational models. [5–7, 66–73] We introduce a classical model of a metallic carbon nanotube, which allows us to also investigate the impact of the CNT electronic structure in both floating and electric potential-biased tubes. The potential-biased growth exhibits a radically different mechanism than previously studied charge-neutral simulations. The mechanistic differences are shown to impact the probability of forming the different inorganic nanotubes (INTs), indicating that the crystalization cannot be understood merely through equilibrium statistical mechanics. The final distribution of structures is not in equilibrium because kinetic barriers trap the system in one of many local minima. The macroscopic controls are not sufficient to strongly bias the system toward one particular minimum, a feature which is likely to be general to inorganic systems whose interactions lack specificity.

The thesis is structured as follows. We begin by reviewing in Chapter 1 the prior work on INT formation within CNTs and on classical descriptions of metals. To simulate a conducting electrode we must adapt the Ewald summation procedure in a manner similar to Reed *et al.* [40], so in Chapters 2 and 3 we turn our attention to the method of Ewald summation using the approaches of Bertaut [4] and of de Leeuw, Perram, and Smith [15], respectively. The energy expressions derived in Chapters 2 and 3 are used to develop the numerical classical dynamics schemes of Chapter 4, with particular attention given to the numerical methods necessary for handling dynamics of the induced charge density that builds up on a conducting CNT. These methods are compared with continuum classical electrostatics in Chapter 5, where the influence of the metal model's single parameter is studied. Finally, in Chapter 6 we present and discuss the results of the numerical experiments, ultimately concluding that the chaotic classical dynamics results in a distribution of INT morphologies. The macroscopic control parameters can influence this distribution by altering the mechanism of filling, but the distribution remains broadly spread over several distinct INTs, indicating that the controls are too crude to faithfully grow a single INT.

It is worth noting that the intermediate chapters are densely packed with mathematics, and a reader may well resent the excessive detail. It seems customary for authors to avoid this level of detail and present rather cursory derivations devoid of the grungy intermediate algebra. This approach has the advantage of presenting a compact set of equations without the distractions of the messy calculations, which often simplify into much more appealing forms. Unfortunately the result is that many calculations come across as mystical, and perhaps worst of all it is not clear how to simply extend the results to a slightly more general problem. Chapter 3 presents a natural (and actually a fairly trivial) extension of Ewald electrostatics to Gaussian charge distributions, but the simplicity of the extension is only made obvious when the full derivation of the point charge problem is made laboriously clear. For this reason, I felt strongly that it was appropriate to include intermediate calculations almost to a gruesome level. Especially tedious peripheral calculations have been relegated to appendices, but the main text still includes significantly more algebraic manipulations than is standard. Readers who find this distracting are advised to simply skip to the final line of a derivation to see the compact form.

Chapter 1

Background

Since they were discovered, carbon nanotubes (CNTs) have excited scientists with their potential to display unique physiochemical properties. As is generally the case, the structural morphology of the CNTs determines the physical properties, so to make sense of the diverse properties it is natural to study the equally diverse tubular geometries. Much work has been devoted to the direct utilization of the CNT's structural and electronic properties, but the tubes can also serve as templates for the selective growth of new materials. One can imagine using CNTs to bias other inorganic materials into forming pseudo-one-dimensional structures grown up the middle of the tube. Indeed, it has been observed experimentally through high resolution transmission electron microscopy (HRTEM) studies that metal halide molten salts can permeate the nanotubes and crystallize inside the pore. [34, 52–55] Altering the morphology of the CNT influences the internal inorganic structure, but the effect of the template is complex. Wilson showed that changing the radius of the CNT impacts the energetics of inorganic crystallite growth in a straightforward manner, which we will shortly describe further. [70, 72] The CNT morphology affects not only the tube radius, but also the electronic properties of the tube. Roughly one-third of the possible CNT structures are metallic, and the metallicity of the tube could also be expected to impact the growth of ionic structures. To probe these effects we must incorporate a classical metallicity into Wilson's model, something we do by using the method of Siepmann and Sprik [49] as implemented by Reed *et al.* [40] Before extending these methods, it is useful to review the pertinent prior work. First, a cursory introduction to CNT morphology and electronic structure is provided, followed by discussion of the experimental and theoretical studies of CNT filling. We then shift gears and review simulation of metal surfaces to understand how metallic tubes can be simulated. The Siepmann and Sprik method we will use introduces numerical complications related to the computation of the long-range Coulomb energies, the review of which spills over into the following chapters.

1.1 Carbon Nanotube Morphology

The carbons in CNTs are three-coordinate, so if the tube were sliced down its length and unrolled, the carbons would lie on a hexagonal graphene grid. Conversely, to understand the stable tubular structures, one need only enumerate the ways a hexagonal lattice can be periodically rolled to give a tube. We denote the (non-orthogonal) principal lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, and note that the vector wrapping around the cylinder perpendicular to its length must be a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$ with integer coefficients. This circumferential vector is known as the chiral vector, $\mathbf{C}_h$, and its restriction to integer superpositions of $\mathbf{a}_1$ and $\mathbf{a}_2$ ensures that the tip and tail of $\mathbf{C}_h$ both rest on carbon sites, which will be mapped onto each other upon folding. In fact all pairs (n, m) with $\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2$ give rise to a stable tube when the endpoints of $\mathbf{C}_h$ are folded around to connect. The natural classification of the CNT morphologies is then provided by specifying the components n and m of $\mathbf{C}_h$. Some properties of the tube can be computed directly from this classification. For example, the diameter is given by

$$d = \frac{|\mathbf{C}_h|}{\pi} = \frac{a\sqrt{3\left(n^2 + m^2 + nm\right)}}{\pi}, \tag{1.1}$$

where a is the C-C bond length. It is conventional to also define a chiral angle, θ, which gives the angle at which the nearest non-bonded neighbor spirals around the tube. Like the diameter, θ is a simple function of n and m,

$$\theta = \cos^{-1}\left(\frac{2n + m}{2\sqrt{n^2 + m^2 + nm}}\right) \tag{1.2}$$

A much more complete discussion of the morphology as well as the group theoretic consequences is provided in Saito, Dresselhaus, and Dresselhaus. [46] We reproduce their illustration of the geometry in Figure 1.1 to make the definitions of $\mathbf{C}_h$ and θ more clear.

1.2 Carbon Nanotube Electronic Structure

Just as the CNT structure can be derived from that of graphene, the electronic structure can also be related. A simple tight binding analysis of graphene yields

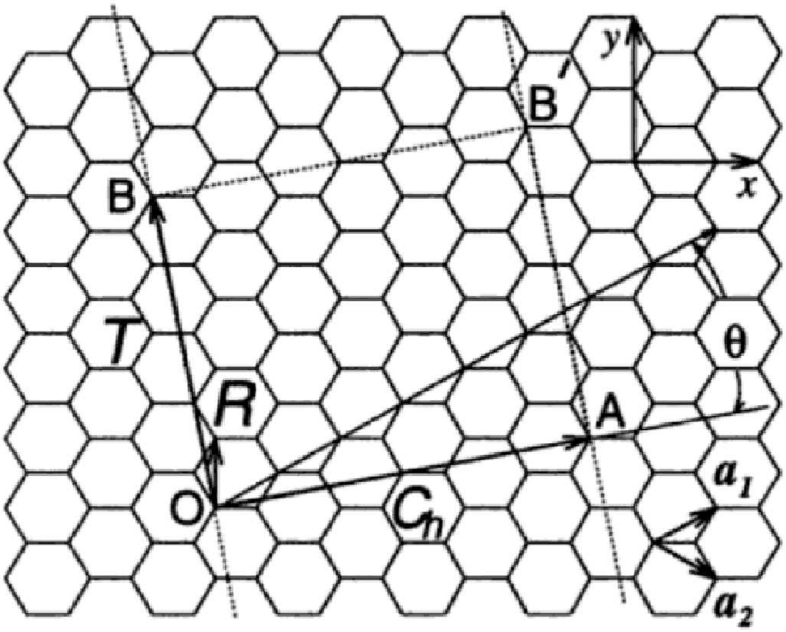


Figure 1.1: The structure of a carbon nanotube can be thought of as arising from a rolled hexagonal lattice. For example connecting points O to A and B to B' yields a (4,2) nanotube with translation vector **T**, chiral vector $\mathbf{C}_h$, and chiral angle θ. This figure is reproduced from Reference [46]

the well known result that graphene has a zero band gap.[1] This can be seen by Figure 1.2, which shows the band energy of the π and π^* bands as a function of the two dimensional wavevector. This wavevector is two dimensional since the graphene lattice has two infinitely replicated directions. In contrast, nanotubes contain only one infinite direction (**T** in Figure 1.1). By associating the ends of $\mathbf{C}_h$, only values of the wavevector, **k**, which satisfy the periodicity in the direction of $\mathbf{C}_h$ are allowed. The zone-folding method for approximating CNT electronic structure then identifies the nanotube's 1D bands as projections of the 2D bands in Figure 1.2 onto the discrete set of lines through reciprocal space which satisfy the $\mathbf{C}_h$ periodicity. Under this approximation one can show that the 1D subbands will have zero band gap if the restricted **k** vectors run through the graphene conical intersections. It can be shown that this is the case if and only if $n - m$ is evenly divisible by three. [16, 36, 46] Assuming the CNT electronic structure can be understood by graphene's tight binding solution, one-third of the CNT morphologies are then expected to be metallic while the remainder are semiconducting. Tubes with small radii do not

---

[1] It is technically a "zero band gap semiconductor" or a "semimetal" rather than a metal, but this is not important for our purposes.

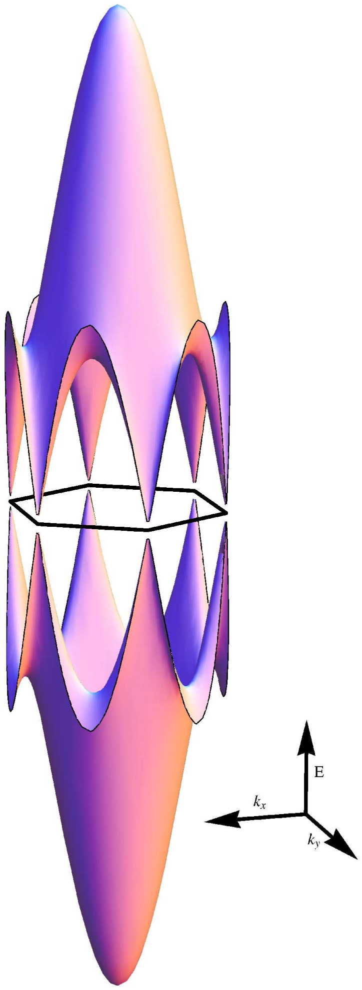


Figure 1.2: The first Brillouin zone of the π and π^* bands of a tight binding model of graphene. Note the conical intersections at the vertices of the Brillouin zone at which the band gap vanishes. When 1D sub-bands cut through these vertices the resultant tube is (to first approximation) a metal.

perfectly satisfy this rule because the high curvature mixes the sp^2 and p_z orbitals. [38] Further refinements can be made by extending the graphene tight binding model beyond nearest neighbor [42], but these improvements to the zone-folding method are irrelevant to this work. The critical aspect is that CNTs are either conducting or semiconducting as a function of morphology.

1.3 CNT Filling

High resolution transmission electron microscopy (HRTEM) has enabled a series of experimental investigations into the structure of a variety of CNT-enclosed metal cation salts: NiO, Bi_2O_3, V_2O_5, MoO_3 [1–3, 11, 61], UCl_4 [50], AgCl/AgBr [53, 55], KI [34, 54], BaI_2 [51], CoI_2 [37], Sb_2O_3 [24], metal trihalides [21–23], PbI_2 [19], HgTe [10], $ZrCl_4$ [8], and PbO [29]. Ionic salts are of particular interest because the enclosed structures can be formed simply by allowing a molten salt to diffuse into the CNT [34, 53–55] since these materials have surface tensions which are small enough that the CNT is not crushed. [66] The salts have been observed to form crystals which can roughly be described as rock salt [34, 54] or wurtzite structures [53, 55], but some bond distances are notably distorted from the ordinary bulk structures. The distortions of the rock salt KI structures were reproduced and rationalized by molecular dynamics simulation and energy minimization calculations [67, 73]. Further simulations have shown that the rock salt and wurtzite structures can both be understood as members of a large class of pseudo-one-dimensional structures, which cannot all be directly related to bulk crystal structures. [66, 68–70] This more general classification, based on the folding of square and hexagonal lattices, predicts the existence of stable inorganic nanotubes (INTs), which have yet to be experimentally observed. [68–72]

1.3.1 Stable Structures

The rock salt KI structures were originally described as $2 \times 2 \times \infty$ and $3 \times 3 \times \infty$ crystallites. It was observed that the most stable enclosed structure would depend strongly on the radius of the enclosing CNT. Using a particularly simple Lennard-Jones description of the CNT potential and Born-Mayer electrostatics, Wilson showed that the energy of stable crystallites indeed depends strongly on the CNT radius. [66] A phase diagram of the most stable structures as a function of CNT radius, reproduced here as Figure 1.3, shows this radius-dependence, but it also demonstrates that the particular CNT morphology (i. e. $(n, 0)$ versus (n, n) CNTs) is not influential between tubes of comparable radii. Rather than considering the KI crystals as a fragment of a bulk crystal, Wilson proposed that it could be understood as the folding of a square net, much like the folding of hexagonal nets gives rise to nanotubes (see Section 1.1). [69] This construction unifies the observation of both rock salt and wurtzite fragments, as the latter can be considered to result from folding a hexagonal net in exactly the same way as CNTs. The only difference is that in the ionic structures the

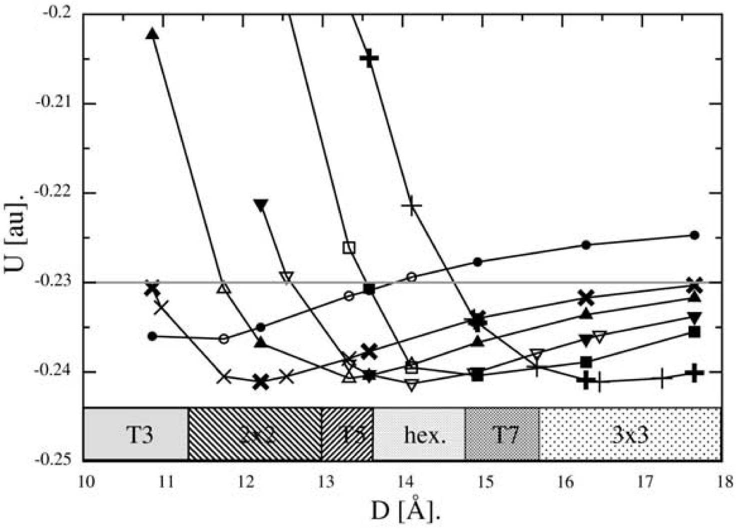


Figure 1.3: Calculated phase diagram for KI enclosed in carbon nanotubes as a function of the CNT radius. Unsurprisingly as the CNT radius increases larger crystallites are stabilized. Perhaps more surprising is the diversity of structures, which are more compactly classified in Wilson's later work. Thicker symbols are CNTs of the form (n, n), while the other symbols are $(n, 0)$. Note that these CNT morphological differences do not significantly influence the energy calculations. This figure is reproduced from Reference [66]

bipartite nature of the lattice is significant since anions and cations are interspersed. The net-folding construction is especially useful for classifying the ionic structures with a compact nomenclature. As with CNTs themselves the chiral vector uniquely defines the structure when combined with a specification of whether the underlying lattice is square or hexagonal. In addition to the convenient unified nomenclature, the construction suggests that the experimentally observed structures may be part of a much larger class of stable tubular structures of the form $(n, m)_X$, where $X = sq$ or *hex* denotes the underlying lattice.

While such inorganic nanotubes have not been experimentally observed as inclusions of CNT, both ionic and covalent INTs have been constructed by other methods from a variety of materials including MoS_2, WS_2 [60], $NiCl_2$ [27], BN, and BN/C composites [32, 58]. The classical ionic potential model that predict the KI deformations is general and transferable, such that it can also explore the stability of INTs. The potential model parameters can be adapted to predict either six-coordinate or four-coordinate bulk structures, an adaptation that physical corresponds to changing ionic radii. [68, 70] These parameters give rise to structures which are built upon four-coordinate square nets and three-coordinate hexagonal nets respectively. Wilson showed that when the potential model's parameters are tuned to values that would give rise to bulk wurtzite, INTs spontaneously form inside CNTs, provided the CNT radius is sufficiently large. [68, 70, 71] A phase diagram of these INTs can be constructed in the same way it was for KI. For reference, the phase diagram is reproduced from Reference [70] as Figure 1.4. Phase diagrams are computed based on both energy minimization calculations and molecular dynamics simulations. These two diagrams differ slightly, most significantly in that the molecular dynamics simulations sometimes yield multiple different INTs grown within the same CNT. Simple analytic models based upon the stresses of folding a hexagonal net and the Lennard-Jones interaction with a cylindrical tube can be used to identify several low-lying INT structures [70], but a distribution of these structures is observed by the molecular dynamics, indicating that entropic and/or kinetic effects are significant. By comparing the density of vibrational states in $(n, 0)$ and (n, n) tubes, Wilson suggests that entropic differences between the INTs is negligible and that the relative lack of $(n, 0)$

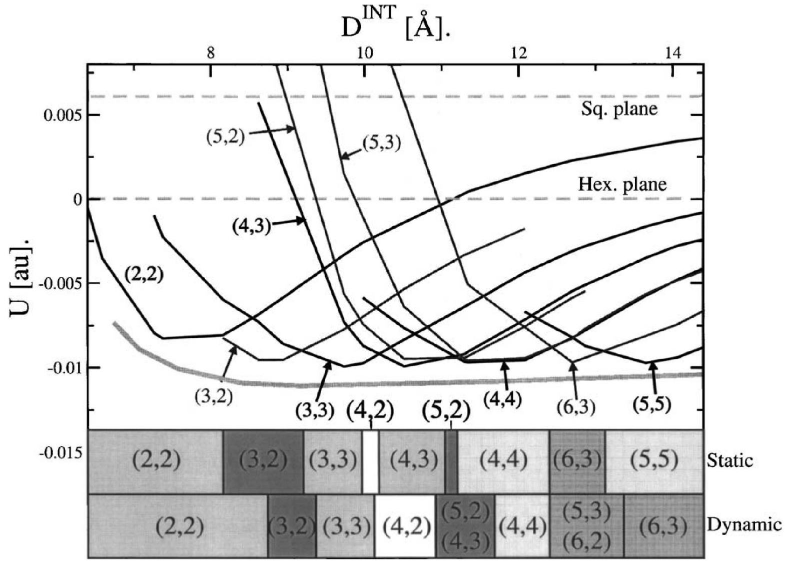


Figure 1.4: Calculated phase diagram for a generic MX ionic salt enclosed in carbon nanotubes as a function of the CNT radius. Curves show the energy of particular INTs inserted into a CNT of the given radius as computed by energy minimization techniques. The "static" phase diagram across the bottom gives the minimum energy INT for that CNT radius. The "dynamic" phase diagram is the structure(s) that resulted from molecular dynamics simulation. This figure is reproduced from Reference [70]

type tubes is largely a kinetic effect. [72] Introduction of a voltage bias allows for a distinct mechanism of INT formation, providing a method to better explore this kinetic effect, something we will discuss in Chapter 6.

1.3.2 Dynamics of INT Formation

In addition to providing information about energy minima, the computational model provides dynamics of the filling events. As will be described in Chapter 6, typical molecular dynamics simulations allow a capped, empty CNT to equilibrate while immersed in a molten salt. The caps are removed and molecular dynamics is carried out in either the NVT or NPT ensemble. Because the liquid is charged and therefore highly structured, a significant liquid reorganization is required before charge-neutral chains begin to diffuse into the tube. This reorganization tends to take on the order of 30 ps, but the reorganization time is a function of CNT size and is widely variant since the liquid reorganization requires rare fluctuations that lead to the development of a loop which threads into the CNT. Once the appropriate initial fluctuation occurs, the salt rapidly wets and fills the CNT. [70] Characteristic filling profiles of the number of enclosed ions versus time are reproduced in Figure 1.5 from Reference [72].

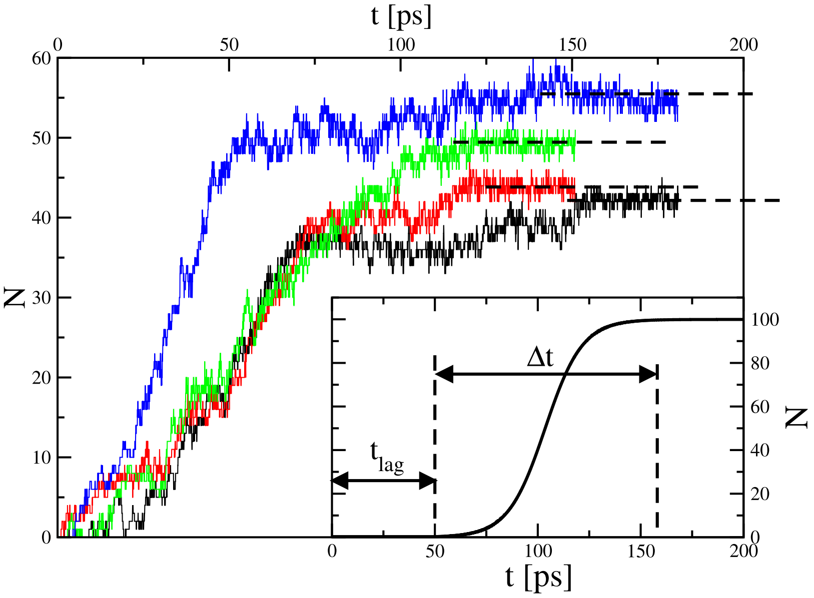


Figure 1.5: Typical filing profiles showing the number of ions inside the CNT as a function of time. The characteristic time lag, t_{lag} reflects a liquid reorganization time, which increases as the CNT radius decreases. This figure is reproduced from Reference [72]

Wilson has postulated that the particular way in which the initial charge-neutral loop threads its way into the CNT may account for the relative lack of $(n, 0)$ tubes observed by computer simulation. [72] He has also observed that subsequent diffusion of ions into the tube proceeds rapidly as ions are effectively passed down an already formed chain. [71] As mentioned, the relative importance of these mechanistic details can be better understood by comparing to crystal growth within a voltage-biased nanotube, which does not proceed by the diffusion of charge-neutral loops.

1.4 Simulation of Metallic Surfaces

Wilson's results that we have discussed pertain to molecular dynamics of an ionic salt in a rigid Lennard-Jones tube. Bishop and Wilson have also extended these studies to a flexible model of CNTs [5–7], but notably absent from all of the prior simulation work is the CNT electronic structure's influence on INT formation. Metallic CNTs support induced surface charge densities and the forces exerted by this surface charge must be included in any simple classical model of a metal tube. If the nanotube could be treated as a continuous surface, an image charge (or more generally Green's function) approach could be applied to compensate for the induced charge. [30] Indeed, the work of Cui can provide an analytic description for the behavior of ions near a cylindrical metal surface.[2] [13] Unfortunately the analytic Green's function approach is only valid far from an equipotential cylindrical surface since the atomic detail of the metal surface cannot be ignored for distances less than 7.5 Å. [25] At this range, the atomic corrugations in the equipotential surface cause the image charge approach to break down. [31] This break down can be understood by the fact that classical continuum electrostatics assumes a smoothness of the metal surface on all scales, but a molecular surface deviates from this idealized surface geometry.

Rather than approach the induced charge problem from a fully analytic approach, we will use classical methods that take advantage of computational power to explicitly describe the surface charge density. By directly treating the surface charge density

---

[2] It is worth noting that the standard image charge problems (Section 5.2.1) suggest that the method is more useful than it generally is. When the conducting surface is planar or spherical the surface charge density induced by point charges can be replaced by a collection of image charges on the other side of the metal surface which are themselves point charges. The power of the method arises from the fact that calculating the force due to the surface charge would require *integrating* over that surface charge density, but the image charge approach allows one to merely *sum* over image point charges. All other geometries of metal surfaces require that the image charges be a more complicated charge distribution, not just a single point charge. Integrating over this (often complicated) image charge distribution is not necessarily any easier than just dealing with the surface charge distribution in the first place.

as a grid of point charges, it was shown that the essential physics of a point charge interacting with an Al surface could be captured down to 2.5 Å as compared with density function theory (DFT) calculations. [18] In an approach first suggested by Siepmann and Sprik, the surface charge can instead be expanded onto a basis of Gaussian charges centered on each metal site, which should better represent the delocalized nature of metallic charge. [49] The magnitudes of these Gaussian charges are adjusted to minimize the Coulomb energy subject to an equipotential constraint which is imposed site-wise. The constraint that the electric potential be equal only at the metal sites, not along a continuous surface, is purposely weaker than the traditional analytic description of an equipotential metal surface. It is the "weakness" of this constraint that allows the described method to work since the Gaussian charges provide a natural way of imposing a corrugated equipotential surface. The model has the added benefit of applying regardless of the metal geometry, such that it can be extended to the flexible nanotubes of Bishop and Wilson. This extension is not carried out in this thesis, but it is a natural future step.

To compute the surface charge density and the forces on the ions, the Coulomb energy must be derived. Because the Coulomb potential is particularly long-ranged, the interaction cannot be simply truncated. A standard technique for bypassing this problem is to use the method of Ewald summation. [17, 20] The Ewald summation expression for the Coulomb energy of a mixed system consisting of both the Gaussians (metal sites) and point charges (ions) introduces small complications that are not present in the normal point charge calculations. Reed *et al.* provide the general framework for the Gaussian extension of Ewald summation, but we correct an error to yield a physically obvious result. [40] The next three chapters are devoted to the analytic details of implementing the Siepmann and Sprik model. First we review a standard presentation of Ewald summation, noting how it could be adapted to describe the energy between Gaussian charges. Next we present a more formal computation that places the Ewald procedure on firm mathematical ground while providing a computationally efficient form of the energy. Finally we demonstrate in Chapter 4 how this energy is used to solve for the induced charge density. So, without further ado, let's review classic Ewald summation.

Chapter 2

Review of Ewald Summation using the Bertaut Method

2.1 Motivation

Calculating the electrostatic interactions of a finite number of charges is a trivial matter. The total interaction can be decomposed into interactions between pairs of charges, which is described by the ordinary r^{-1} Coulomb potential. Due to computational limitations, however, typical molecular simulations cannot treat a system with sufficiently many independent atoms to appropriately approximate the bulk. In order to avoid undesirable finite system boundary effects, simulations of the condensed phase routinely make use of periodic boundary conditions to mimic an extended system. [20] While periodic boundary conditions succeed in eliminating the surface of a simulation cell, they introduce an infinite number of periodic images for each charge, thereby complicating the electrostatics. Specifically, the electric potential at site i for a periodic array of point charges is given by

$$\phi(\mathbf{r}_i) = {\sum_{j,\mathbf{n}}}' \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|}, \tag{2.1}$$

where the charge q_j is located at $\mathbf{r}_j$. We assume a cubic unit cell with length L and primitive unit cell vectors $\mathbf{n}$. The prime means the $j = i, \mathbf{n} = 0$ term is not summed over. In other words a point charge doesn't contribute to its own potential, but the periodic images of that point charge do contribute ($j = i$, $\mathbf{n} \neq 0$ terms). It seems that Equation (2.1) should be sufficient to determine the Coulomb energy for a periodically replicated unit cell, but in practice we must truncate the periodic array and add a tail correction that approximates the average behavior of the truncated terms. This scheme is simple for short-range potentials, such as the Lennard-Jones potential [20, pg 35], but in the case of the Coulomb potential the tail correction would be of the form

$$\phi_{tail} \propto \frac{N\rho}{2} \int_{r_c}^{\infty} dr \left(\frac{1}{r}\right) 4\pi r^2,$$

where r_c is the cut-off radius, N is the number of particles in the simulation cell, and ρ is the average particle density. This tail correction integral clearly diverges since the $1/r$ potential is too long-ranged.

The difficulty lies in the fact that we are trying to study a charge distribution which extends out to infinity. If we truncate the system at any finite size, we neglect an infinite number of point charges outside the cut-off radius, each contributing a $1/r$ potential that falls off too slowly to merely be neglected. Rather than applying a cut-off in real space, a Fourier space cut-off could naturally include the infinite spacial extent, offering some hope of convergence. The Fourier sum, however, converges slowly and it is actually a clever mixture of real space and Fourier space terms that allows one to express the potential as the sum of two rapidly convergent series. This decomposition into a real space and reciprocal space sum was first shown by Ewald in 1921. [17] Perhaps the clearest physically motivated derivation follows the approach of Bertaut [4] which has been frequently reproduced. [15, 20] We reproduce this approach yet again with slightly more detail than is customary in order to make the derivation as straightforward as possible. This method will be extended to Gaussian charges in Section 2.8, but that result will only be useful if the Gaussians are sufficiently broad. To treat the general Gaussian charge problem we will use the more formal approach of de Leeuw *et al.* in Chapter 3. [15]

2.2 Basic Approach

In real space, Poisson's equation is simply

$$\nabla^2 \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}),$$

with ρ and ϕ being the charge density and electric potential respectively. Note that the linearity of this equation allows us to break the problem up into superimposed charge distributions.

$$\nabla^2 (\phi_1 + \phi_2) = \nabla^2\phi_1 + \nabla^2\phi_2 = -4\pi\rho_1 - 4\pi\rho_2 = -4\pi(\rho_1 + \rho_2)$$

If one can find the potential associated with both ρ_1 and ρ_2 individually then the potential of the superimposed charge distribution is automatically provided by the linearity of the Poisson equation. The Bertaut method makes use of this decomposition by considering the infinitely replicated charge distribution, $\rho(\mathbf{r})$, as a superposition of mathematically convenient terms. Specifically, our periodic array of point charges can be re-expressed as

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_{j,\mathbf{n}} q_j \delta(\mathbf{r}-\mathbf{r}_j) \\ &= \left[\sum_{j,\mathbf{n}} q_j \left(\delta\left(\mathbf{r}-\mathbf{r}_j\right) - \left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}-\mathbf{r}_j|^2}\right)\right] + \sum_{j,\mathbf{n}} q_j \left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}-\mathbf{r}_j|^2},\end{aligned} \tag{2.2}$$

where $\delta(x)$ is the Dirac delta function and α is just a parameter whose importance will be seen later. This superposition of charge densities is depicted pictorially in Figure 2.1. The δ-function on site i does not contribute to its own potential, so the charge seen by the site is a superposition of δ-functions on all other sites. The charge distribution giving rise to this potential seen by site i is then

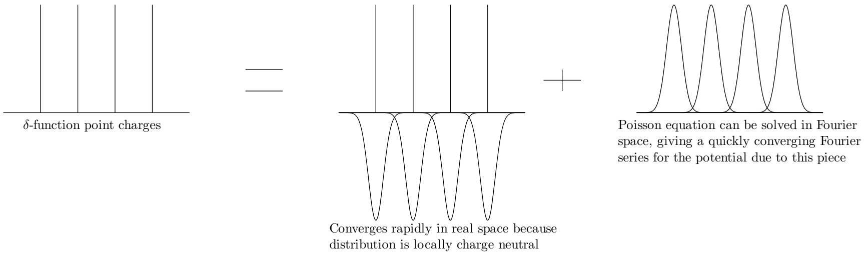


Figure 2.1: The δ-function charge distribution is split into a superposition of a short-range charge neutral term and a periodic term amenable to Fourier analysis

$$\begin{aligned}\rho(\mathbf{r}) &= {\sum_{j,\mathbf{n}}}' q_j \delta(\mathbf{r}-\mathbf{r}_j) \\ &= \left[{\sum_{j,\mathbf{n}}}' q_j \left(\delta\left(\mathbf{r}-\mathbf{r}_j\right) - \left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}-\mathbf{r}_j|^2}\right)\right] + \sum_{j,\mathbf{n}} q_j \left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}-\mathbf{r}_j|^2} \\ &\quad - q_i \left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}-\mathbf{r}_i|^2} \\ &\equiv \rho_{screened} + \rho_{fourier} + \rho_{self},\end{aligned} \tag{2.3}$$

where the prime denotes that site i is removed from the sum. This is shown pictorially in Figure 2.2.

The sum of Gaussians includes site i to preserve the periodicity since this term will

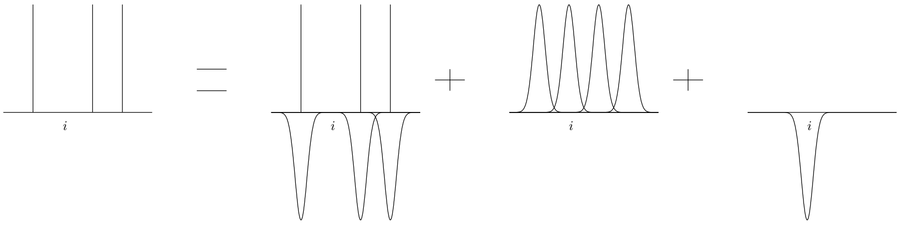


Figure 2.2: No δ-function is on site i since the charge does not contribute to its own potential. The charge-neutral term also excludes the charge on site i, but the charge is included in the Gaussian term since the function must remain periodic for Fourier analysis. This requires a correction term of an extra Gaussian.

be handled by a Fourier series. This requires an extra negative Gaussian on site i, which gives rise to what is known as the self-interaction. The potential on site i can be similarly decomposed into

$$\phi_\delta = \phi_{screened} + \phi_{fourier} + \phi_{self}, \tag{2.4}$$

where each term is the solution to the Poisson equation for the corresponding charge density term from Equation (2.3).

2.3 Energy

From the potential one can directly compute energies. Before going into the mathematical details of finding $\phi_{screened}$, $\phi_{fourier}$, and ϕ_{self}, we derive standard expressions for the energy given the potential. Of course in the case of two point charges, the energy is given by $U = q\phi$, where ϕ is the potential due to the other charge. Equivalently, and more generally, we can sum over both particles by writing $U = \frac{1}{2}\sum q_i\phi(\mathbf{r}_i)$, with the factor of one half needed to prevent double counting. Since the potential is just the charge over the distance, this can be written as

$$U = \frac{1}{2}\sum_{i\neq j}\frac{q_i q_j}{r_{ij}}$$

Moving from point charges to the continuum we naturally have

$$\begin{aligned} U &= \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' \ \frac{\rho(\mathbf{r}')\rho(\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|} \\ &= \frac{1}{2}\int d\mathbf{r} \ \rho(\mathbf{r})\left(\int d\mathbf{r}' \ \frac{\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\right) \\ &= \frac{1}{2}\int d\mathbf{r} \ \rho(\mathbf{r})\phi(\mathbf{r}) \end{aligned} \tag{2.5}$$

For the case of point charges with $\rho(\mathbf{r}) = \sum_i q_i\delta(\mathbf{r}-\mathbf{r}_i)$, the integral simply reduces to our familiar expression

$$U = \frac{1}{2}\sum_i q_i\phi_\delta(\mathbf{r}_i), \tag{2.6}$$

where $\phi_\delta(\mathbf{r}_i)$, the potential due to the point charges, can be exactly re-expressed as the superposition of contributions due to different mathematically convenient charge distributions as shown in Equation (2.4). Hence,

$$U = \frac{1}{2}\sum_i q_i\phi_{fourier}(\mathbf{r}_i) + \frac{1}{2}\sum_i q_i\phi_{self}(\mathbf{r}_i) + \frac{1}{2}\sum_i q_i\phi_{screened}(\mathbf{r}_i) \tag{2.7}$$

The total energy is obtained when the sums in Equation (2.7) extend over all periodic charges, but the infinite system must have infinite energy. We really seek the energy per unit cell, which is just found by restricting the sums to the unit cell. Note also that in moving from Equation (2.5) to Equation (2.6) we simply treated the point charges as δ-functions, but when we need the potential we split the charges up into a superposition of terms via Equation (2.3). The important point here is that the $\rho(\mathbf{r})$ and $\rho(\mathbf{r}')$ of Equation (2.5) are treated differently. One term is treated with the straightforward δ-functions while the other utilizes the non-trivial superposition of Equation (2.3).

2.4 Poisson Equation in Fourier Space

We already know how to solve the Poisson equation for point charges (δ-functions).

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} q_i\delta(\mathbf{r}-\mathbf{r}_i) \Rightarrow \phi(\mathbf{r}) = \sum_{i=1}^{n}\frac{q_i}{4\pi\,|\mathbf{r}-\mathbf{r}_i|}$$

When we have charge distributions of different functional forms, a Fourier representation of the Poisson equation is useful. To find this we start by finding the Fourier series of $\phi(\mathbf{r})$ and $\rho(\mathbf{r})$. The Poisson summation formula yields

$$\phi(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}}\tilde{\phi}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}$$

where $\mathbf{k} = \left(\frac{2\pi}{L}\right)\mathbf{l}$, $\mathbf{l}$ are the lattice vectors in Fourier space, and the Fourier components, $\tilde{\phi}(\mathbf{k})$, are then given by $\tilde{\phi}(\mathbf{k}) = \int_V d\mathbf{r}\phi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}$. One can similarly find the Fourier series for $\rho(\mathbf{r})$,

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}}\tilde{\rho}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}},$$

with $\tilde{\rho}(\mathbf{k}) = \int_V d\mathbf{r}\rho(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}$. Hence, starting with the Poisson equation in real space we have,

$$\begin{aligned}
\nabla^2\phi(\mathbf{r}) &= -4\pi\rho(\mathbf{r}) \\
\Rightarrow \nabla^2\left(\frac{1}{V}\sum_{\mathbf{k}}\tilde{\phi}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\right) &= -\frac{4\pi}{V}\sum_{\mathbf{k}}\tilde{\rho}(\mathbf{k})e^{i\mathbf{r}\cdot\mathbf{k}} \\
\Rightarrow \nabla\left(\frac{1}{V}\sum_{\mathbf{k}}\tilde{\phi}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}i\mathbf{k}\right) &= -\frac{4\pi}{V}\sum_{\mathbf{k}}\tilde{\rho}(\mathbf{k})e^{i\mathbf{r}\cdot\mathbf{k}} \\
\Rightarrow -\sum_{\mathbf{k}}\frac{\mathbf{k}\cdot\mathbf{k}}{V}\tilde{\phi}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}} &= -\frac{4\pi}{V}\sum_{\mathbf{k}}\tilde{\rho}(\mathbf{k})e^{i\mathbf{r}\cdot\mathbf{k}} \\
\Rightarrow -\frac{1}{V}\sum_{\mathbf{k}}k^2\tilde{\phi}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}} &= -\frac{4\pi}{V}\sum_{\mathbf{k}}\tilde{\rho}(\mathbf{k})e^{i\mathbf{r}\cdot\mathbf{k}}
\end{aligned}$$

Since this must hold for arbitrary $\rho(\mathbf{r})$, and thus for arbitrary $\rho(\tilde{\mathbf{k}})$,

$$k^2\tilde{\phi}(\mathbf{k}) = 4\pi\tilde{\rho}(\mathbf{k}). \tag{2.8}$$

Note that in Fourier space, differential operators do not appear in the Poisson equation. In fact, it is trivial to deduce the k-space potential given the k-space charge distribution. This provides a route to deducing the real-space potential of a periodic charge distribution. One can expand the charge distribution as a Fourier series, deduce the k-space potential, then transform the potential back into real space via an inverse Fourier transform.

2.5 Periodic Superposition of Gaussians

Suppose the real space charge distribution consists of a single point charge at the origin, $\rho_1(\mathbf{r}) = q\delta(\mathbf{r})$. In Fourier space,

$$\tilde{\rho}_1(\mathbf{k}) = \int_V d\mathbf{r}\ \rho_1(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}} = q\int_V d\mathbf{r}\ \delta(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}} = q$$

From Equation (2.8) we then have the k-space form of the potential due to a real-space point charge.

$$\tilde{\phi}(\mathbf{k}) = \frac{4\pi q}{k^2} = q\tilde{g}(\mathbf{k}), \tag{2.9}$$

where $\tilde{g}(k) = \frac{4\pi}{k^2}$ is the Green's function since it is the Fourier space solution to the operator equation $\nabla^2\phi(\mathbf{r}) = -4\pi\delta(\mathbf{r})$. Now consider a collection of point charges, $\rho_2(\mathbf{r}) = \sum_i q_i\delta(\mathbf{r}-\mathbf{r}_i)$. The k-space distribution is given by

$$\tilde{\rho}_2(\mathbf{k}) = \int_V d\mathbf{r} \sum_{i=1}^{n} q_i\delta(\mathbf{r}-\mathbf{r}_i)e^{-i\mathbf{k}\cdot\mathbf{r}} = \sum_{i=1}^{n} q_i e^{-i\mathbf{k}\cdot\mathbf{r}_i} \tag{2.10}$$

Plugging this k-space charge distribution into Equation (2.8) gives the k-space potential due to a superposition of point charges.

$$\tilde{\phi}_2(\mathbf{k}) = \frac{4\pi}{k^2}\sum_{i=1}^{n} q_i e^{-i\mathbf{k}\cdot\mathbf{r}_i} = \tilde{g}(\mathbf{k})\tilde{\rho}_2(\mathbf{k})$$

Finally, we want to find the potential due to a periodic superposition of Gaussians,

$$\rho_3(\mathbf{r}) = \sum_{j=1}^{n}\sum_{\mathbf{n}} q_i \left(\frac{\alpha^2}{\pi}\right)^{3/2} \exp\left(-\alpha^2\left|\mathbf{r}-(\mathbf{r}_j+\mathbf{n}L)\right|^2\right).$$

This could be tackled explicitly, but it is simpler to utilize the convolution theorem, which states

$$f_1(x) = f_2(x) \star f_3(x) \equiv \int dx' f_2(x')f_3(x-x') \qquad \Rightarrow \qquad \tilde{f}_1(k) = \tilde{f}_2(k)\tilde{f}_3(k)$$

In words, convolutions in real space are products in Fourier space. We can recognize ρ_3 as a sum of δ-functions convolved with Gaussians.

$$\rho_3(\mathbf{r}) = \sum_i q_i\gamma(\mathbf{r}-\mathbf{r}_i) = \int d\mathbf{r}'\ \gamma(\mathbf{r}')\sum_i q_i\delta(\mathbf{r}-\mathbf{r}_i-\mathbf{r}') = \gamma(\mathbf{r}) \star \rho_2(\mathbf{r}),$$

where $\gamma(\mathbf{r})$ is a Gaussian centered at zero along with its periodic images,

$$\gamma(\mathbf{r}) = \sum_{\mathbf{n}}\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}+L\mathbf{n}|^2}.$$

Because $\rho_3(\mathbf{r}) = \gamma(\mathbf{r}) \star \rho_2(\mathbf{r})$, the convolution theorem implies $\tilde{\rho}_3(\mathbf{k}) = \tilde{\gamma}(\mathbf{k})\tilde{\rho}_2(\mathbf{k})$. Application of Equation (2.8) gives

$$\tilde{\phi}_3(\mathbf{k}) = \tilde{g}(\mathbf{k})\tilde{\rho}_3(\mathbf{k}) = \tilde{g}(\mathbf{k})\tilde{\gamma}(\mathbf{k})\tilde{\rho}_2(\mathbf{k}) \tag{2.11}$$

We already have expressions for $\tilde{g}(\mathbf{k})$ and $\tilde{\rho}_2(\mathbf{k})$, but it remains to find $\tilde{\gamma}(\mathbf{k})$.

$$\begin{aligned}\tilde{\gamma}(\mathbf{k}) &= \sum_{\mathbf{n}}\int_V d\mathbf{r}\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}+L\mathbf{n}|^2}e^{-i\mathbf{r}\cdot\mathbf{k}} = \int_{\text{all space}} d\mathbf{r}\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}|^2}e^{-i\mathbf{r}\cdot\mathbf{k}} \\ &= \int_{\text{all space}} d\mathbf{r}\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\left(\alpha\mathbf{r}-\frac{i\mathbf{k}}{2\alpha}\right)^2}e^{-k^2/4\alpha^2} = e^{-k^2/4\alpha^2} \end{aligned} \tag{2.12}$$

Combining equations (2.10), (2.11), and (2.12) yields

$$\tilde{\phi}_1(\mathbf{k}) = \frac{4\pi}{k^2} \sum_{j=1}^{n} q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \exp\left(-\frac{k^2}{4\alpha^2}\right) \tag{2.13}$$

Now we Fourier transform this back into position space to find $\phi_3(\mathbf{r})$.

$$\begin{aligned}
\phi_3(\mathbf{r}) &= \frac{1}{V} \sum_{\mathbf{k}\neq 0} \phi_3(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \\
&= \frac{1}{V} \sum_{\mathbf{k}\neq 0} \frac{4\pi}{k^2} \sum_{j=1}^{n} q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} e^{i\mathbf{k}\cdot\mathbf{r}} \exp\left(-\frac{k^2}{4\alpha^2}\right) \\
&= \frac{1}{V} \sum_{\mathbf{k}\neq 0} \frac{4\pi}{k^2} \sum_{j=1}^{n} q_j e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}_j)} \exp\left(-\frac{k^2}{4\alpha^2}\right) \qquad (2.14)
\end{aligned}$$

This is the contribution to the electrostatic potential from the periodic Gaussian charge distribution. It is expressed as a Fourier series in k, and due to the exponential term it converges quickly provided α is sufficiently small. This potential allows one to compute the Coulomb energy of the Fourier term via Equation (2.7), in which ϕ_3 was referred to as $\phi_{fourier}$.

$$U_{fourier} = \frac{1}{2} \sum_{\mathbf{k}\neq 0} \sum_{i,j=1}^{n} \frac{4\pi q_i q_j}{V k^2} e^{i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)} \exp\left(-\frac{k^2}{4\alpha^2}\right)$$

It is customary to take this one step further by recognizing the "structure factor" from scattering theory, $S(\mathbf{k}) = \sum_i q_i e^{i\mathbf{k}\cdot\mathbf{r}_i}$.

$$U_{fourier} = \frac{1}{2} \sum_{\mathbf{k}\neq 0} \frac{4\pi}{k^2} |S(\mathbf{k})|^2 \exp\left(-\frac{k^2}{4\alpha^2}\right) \tag{2.15}$$

This final step is not mere cosmetics. The double sum over n objects has been replaced by the square modulus of a single sum, reducing the computational scaling from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$.

2.6 Correcting for the Self-interaction

To find the self-interaction part of Equation (2.7), we must compute the potential at $\mathbf{r}_i$ due to the Gaussian charge distribution, $q_i \left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2|\mathbf{r}-\mathbf{r}_i|^2}$. We will place $\mathbf{r}_i$ at the origin without loss of generality. It may seem that the potential due to this charge distribution has already been found in Equation (2.14), but we are now dealing with an isolated Gaussian. The Fourier series expression is only helpful in the case of periodically replicated Gaussians. To deal with the isolated Gaussian, we solve the Poisson equation in position space. Using spherical symmetry,

$$\nabla^2\phi = -\frac{1}{r}\frac{\partial^2(r\phi(r))}{\partial r^2}$$

$$\Rightarrow \frac{1}{r}\frac{\partial^2(r\phi(r))}{\partial r^2} = -4\pi\rho(r) = -4\pi q_i\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2 r^2}$$

$$\begin{aligned}\Rightarrow \frac{\partial(r\phi(r))}{\partial r} &= -\int_{\infty}^{r} dr\ 4\pi r q_i\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2 r^2} = -\int_{\infty}^{r} d(r^2)2\pi q_i\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2 r^2}\\ &= \frac{2\pi}{\alpha^2}q_i\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2 r^2}\end{aligned}$$

Integrating each side again gives

$$\begin{aligned}r\phi(r) &= \int_0^r dr\ \frac{2\pi}{\alpha^2}q_i\left(\frac{\alpha^2}{\pi}\right)^{3/2} e^{-\alpha^2 r^2} = \int_0^r dr\ \frac{2\alpha}{\sqrt{\pi}}q_i e^{-\alpha^2 r^2}\\ &= \int_0^{\alpha r} du\ \frac{2}{\sqrt{\pi}}q_i e^{-u^2} = q_i \mathrm{erf}(\alpha r)\end{aligned}$$

where $\mathrm{erf}(x) \equiv \int_0^r dx\ \frac{2}{\sqrt{\pi}}e^{-x^2}$. So

$$\phi_{Gaussian}(\mathbf{r}) = \frac{q_i \mathrm{erf}(\alpha\mathbf{r})}{\mathbf{r}} \tag{2.16}$$

We evaluate this potential at the origin via a limiting procedure.

$$\phi_{Gaussian}(r=0) = \lim_{r\to 0}\frac{q_i \mathrm{erf}(\alpha r)}{r} = \lim_{r\to 0}\frac{q_i\int_0^{\alpha r} dx\ \frac{2}{\sqrt{\pi}}e^{-x^2}}{r} = \lim_{r\to 0}\frac{q_i\frac{2}{\sqrt{\pi}}e^{-r^2}\alpha}{1} = \frac{2q_i\alpha}{\sqrt{\pi}}$$

the second to last step being a combination of L'Hopitâl's Rule and the fundamental theorem of calculus. We are now equipped to evaluate the extra self-interaction term from Equation (2.7). Hence,

$$U_{self} = \frac{1}{2}\sum_{i=1}^{n} q_i\phi_{self}(\mathbf{r}_i) = -\frac{1}{2}\sum_{i=1}^{n} q_i\left(\frac{2q_i\alpha}{\sqrt{\pi}}\right) = -\left(\sum_{i=1}^{n} q_i^2\right)\frac{\alpha}{\sqrt{\pi}}, \tag{2.17}$$

where the minus sign arises from the fact that the point charge and the Gaussian charge have opposite signs as is clear from Figure 2.2.

2.7 Screened Term

We already know the potential due to a point charge: $\phi(r) = \frac{q_i}{r}$ and now from Equation (2.16) we know the potential due to a Gaussian charge. So the potential due to a superposition is just

$$\phi_{screened}(r) = \frac{q_i}{r} - \frac{q_i}{r}\mathrm{erf}(\alpha r) = \frac{q_i}{r}\mathrm{erfc}(\alpha r)$$

where $\mathrm{erfc}(x) \equiv 1 - \mathrm{erf}(x)$. Hence the energy for this charge-neutral term is

$$U_{screened} = \frac{1}{2}\sum_{i \neq j} \frac{q_i q_j \mathrm{erfc}(\alpha r_{ij})}{r_{ij}}, \tag{2.18}$$

where the sum over $i \neq j$ in principle means the sum over the periodic images as well as the simulation cell. In practice a real-space cutoff distance is imposed since the screened potential will converge rapidly.

Combining equations (2.15), (2.17), and (2.18) gives the total energy for our original system of point charges:

$$U = \frac{1}{2V}\sum_{\mathbf{k} \neq 0} \frac{4\pi}{k^2} \left|S(\mathbf{k})\right|^2 e^{-k^2/4\alpha^2} - \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{n} q_i^2 + \frac{1}{2}\sum_{i \neq j} \frac{q_i q_j \mathrm{erfc}(\alpha r_{ij})}{r_{ij}} \tag{2.19}$$

2.8 Ewald Summation with Gaussian Charges

The progression from Equation (2.5) to (2.6) hinged upon the system consisting only of point charges. Suppose instead that charges at positions $\mathbf{R}_i$ are Gaussian with $\rho(\mathbf{r}) = \sum_i Q_i \left(\frac{\eta^2}{\pi}\right)^{3/2} e^{-\eta^2|\mathbf{r}-\mathbf{R}_i|^2}$. Then Equation (2.5) becomes

$$U = \frac{1}{2}\sum_{i=1}^{N} \int d\mathbf{r}\ Q_i \left(\frac{\eta^2}{\pi}\right)^{3/2} e^{-\eta^2|\mathbf{r}-\mathbf{R}_i|^2}\phi(\mathbf{r}), \tag{2.20}$$

where $\phi(\mathbf{r})$ describes the potential due to the periodic system of Gaussian charges. We already have computed this potential as a series over k in Equation (2.14), so

$$\begin{aligned}
U &= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\sum_{i,j}^{N}\frac{4\pi Q_iQ_j e^{-k^2/4\eta^2}}{k^2}\left(\frac{\eta^2}{\pi}\right)^{3/2}\int d\mathbf{r}\; e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_j)}e^{-\eta^2|\mathbf{r}-\mathbf{R}_i|^2} \\
&= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\sum_{i,j}^{N}\frac{4\pi Q_iQ_j e^{-k^2/4\eta^2}}{k^2}\left(\frac{\eta^2}{\pi}\right)^{3/2}\int d\mathbf{r}\; e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_i)}e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)}e^{-\eta^2|\mathbf{r}-\mathbf{R}_i|^2} \\
&= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\sum_{i,j}^{N}\frac{4\pi Q_iQ_j e^{-k^2/4\eta^2}}{k^2}\left(\frac{\eta^2}{\pi}\right)^{3/2} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)}\int d\mathbf{r} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_i)}e^{-\eta^2|\mathbf{r}-\mathbf{R}_i|^2} \\
&= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\sum_{i,j}^{N}\frac{4\pi Q_iQ_j e^{-k^2/4\eta^2}}{k^2}\left(\frac{\eta^2}{\pi}\right)^{3/2} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)}\int d\mathbf{r}\; e^{-\left(\eta(\mathbf{r}-\mathbf{R}_j)-\frac{i\mathbf{k}}{2\eta}\right)^2}e^{-\frac{k^2}{4\eta^2}} \\
&= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\sum_{i,j}^{N}\frac{4\pi Q_iQ_j e^{-k^2/4\eta^2}}{k^2}e^{-\frac{k^2}{4\eta^2}}\left(\frac{\eta^2}{\pi}\right)^{3/2} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)}\int d\mathbf{r}\; e^{-\eta^2\mathbf{r}^2} \\
&= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}|S(\mathbf{k})|^2 e^{-k^2/2\eta^2}, \qquad (2.21)
\end{aligned}$$

As before $S(\mathbf{k}) = \sum_i Q_i \exp(i\mathbf{k}\cdot\mathbf{R}_i)$ is the structure factor. This is very similar to the Fourier series term in the point charge Ewald sum but now the parameter η is thought of as a real spread in the charge distribution (well the reciprocal of the spread) whereas α before was an arbitrary parameter for mathematical convenience. One would choose α such that it is sufficiently small for the Fourier term to converge quickly but sufficiently large for the screened term to also converge. In contrast, setting η directly controls the way we choose to represent the real charge distribution. It determines the physical extent of the Gaussians that represent charge on the metal sites.

If $\eta^2/2 > \alpha^2$ then the sum of Equation (2.21) will converge slower than the reciprocal space sum of Equation (2.19). Assuming these two sums share the same k-space cut-off, the truncation error of the Gaussian system will exceed the error of the associated point charge system. This problem, which is particularly important at short-range, can be resolved by shifting some of the Gaussian system's energy calculation into real space. Unfortunately, the method for shifting this energy calculation into real space is not clear in this formulation of the problem. As seen in Chapter 3, the Ewald sum calculation of de Leeuw *et al.* makes this alteration transparent. So while the energy of Equation (2.21) is technically correct, it is not computationally desirable. The calculation has been provided here for future comparison, not for actual use.

One could also use the Fourier representation of the Gaussian potential to derive the energy of interaction between a point charge and the Gaussian system. This energy would also be expressed entirely in Fourier space and would suffer from the same shortcomings as Equation (2.21). Rather than derive another useless expression, we proceed directly to the de Leeuw method.

Chapter 3

Formal Ewald Summation

While Equation (2.19) is indeed the commonly used electrostatic energy of the Ewald sum, the derivation and justification of the previous chapter is sloppy. The Bertaut approach completely ignores whether or not the original sum of Equation (2.1) is even convergent. The sum, as it turns out, is only conditionally convergent [15], and as formally shown in Chapter 3 of Rudin's *Principles of Mathematical Analysis*, an infinite conditionally convergent series is ill-defined. [44] For example, consider the infinite alternating harmonic series,

$$S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots$$

Regrouping the terms of this series could yield

$$S_1 = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \dots = \sum_{k=1}^{\infty} \frac{1}{2k\,(2k-1)},$$

but an alternative regrouping results in

$$S_2 = \left(1 - \frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{6} - \frac{1}{8}\right) + \dots = \sum_{k=1}^{\infty} \frac{1}{4k\,(2k-1)} = \frac{S_1}{2}.$$

This simple example conveys the diabolical nature of conditionally convergent series. Unlike their convergent counterparts, the order of summation matters. In fact the Riemann series theorem states that by merely altering the order of summation, the series can be made to sum to any arbitrary value. [44] This is deeply troubling for a physical sum, which clearly must converge to a well-defined value. We therefore need to select which ordering we mean when we write down a sum. A normal gut reaction to S_2 is that it is the wrong ordering, but one must formalize what is so wrong about it. The problem is that a finite truncation of S_2 contains twice as many negative as positive terms, a clear difference from a finite truncation of S as originally written.

For a sum with physical meaning we regularize the sum by choosing the ordering of the infinite sum whose finite truncations correspond to physically realistic truncations of the system.

To specify the order of summation of our periodically replicated simulation cell we must recall what the sum given by Equation (2.1) physically represents. It arises from an infinitely replicated crystal, but an infinitely replicated crystal is only an approximation for a very large crystal. Hence we expect the physical sum must be truncated at some real space cut-off. Only charges closer than the cut-off radius are part of the physical crystal. To order the terms of the sum appropriately, one must sum over spherical shells of increasing radii. That way a finite truncation of the sum looks like a large spherical crystal.

Ordering of the sum is only one of two major problems presented by the conditional convergence of the sum. Standard algebraic tricks (differentiating under an integral, switching the order of sums and integrals, etc.) are valid if and only if a series is uniformly convergent. Since these methods are desirable for certain calculations, one must introduce a convergence factor, making the series formally uniform. One can then obtain well-defined limits of the uniform series as the convergence factor vanishes. The method is illustrated below closely following the work of de Leeuw *et al.* [15]

3.1 Convergence Factors

Consider a conditionally convergent series, $\sum_{n=1}^{\infty} c_n$. Rather than study this series directly, we study an extension of the series as a function of the parameter s.

$$L(s) = \sum_{n=1}^{\infty} c_n f_n(s), \tag{3.1}$$

where $\{f_n(s)\}$ is a monotonically decreasing sequence of continuous functions on the domain $s \geq 0$ with range $[0,1]$ and $f_n(0) = 1 \ \forall \ n$. In other words, we have the restrictions

(i) $f_n(s)$ continuous for $s \geq 0$

(ii) $f_n(0) = 1 \ \forall \ n$

(iii) $f_{n+1}(s) \leq f_n(s) \ \forall \ s \geq 0, \ \forall \ n$ (monotonically decreasing sequence for any given s)

(iv) $0 \leq f_n(s) \leq 1 \; \forall \; n, \; \forall \; s \geq 0$

As de Leeuw *et al.* show, $L(s)$ is uniformly convergent on $s \geq 0$. This seems paradoxical since setting $s = 0$ returns the original conditionally convergent series. The catch is that constraint (iii) on f selects an ordering of the terms of the series. It is not possible to reorder terms while still satisfying (iii) and this constraint carries over smoothly in the $s \to 0$ limit. Hence

$$\sum_{n=1}^{\infty} c_n = \lim_{s \to 0} L(s), \tag{3.2}$$

where the ordering of c_n on the left hand side is such that $f_{n+1}(s) \leq f_n(s)$.

3.2 Point Charge Coulomb Energy

The conditionally convergent Coulomb energy per cell of an infinite point charge system is given by

$$U = \frac{1}{2} \sum_{\mathbf{n}}{}' \left(\sum_{i,j} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \right),$$

where $\mathbf{n}$ sums over the lattice vectors in multiples of the box length (i.e. $\mathbf{n} = (ka, lb, mc)$ with $a, b, c \in \mathbb{Z}$). Unlike before, the unit cell is not assumed to be cubic, so $\mathbf{n}$ is defined slightly differently than it was in Chapter 2. The prime on the sum denotes that the $\mathbf{n} = 0, i = j$ term is excluded as that term corresponds to a point charge interacting with itself. We enforce a spherical order of summation by defining

$$U = \lim_{s \to 0} U(s) \text{ with } U(s) = \frac{1}{2} \sum_{\mathbf{n}}{}' \left(\sum_{ij} \frac{q_i q_j e^{-s|\mathbf{n}|^2}}{|\mathbf{r}_{ij} + \mathbf{n}|} \right), \tag{3.3}$$

which orders the nearest copies first such that $e^{-s|\mathbf{n}|^2} \geq e^{-s|\mathbf{m}|^2}$ for $|\mathbf{n}| \leq |\mathbf{m}|$ in accordance with (iii). As $U(s)$ is uniformly convergent on $s \geq 0$, the order of summation can now be interchanged. Hence,

$$U(s) = \frac{1}{2} \sum_{ij} q_i q_j \sum_{\mathbf{n}}{}' \frac{e^{-s|\mathbf{n}|^2}}{|\mathbf{r}_{ij} + \mathbf{n}|}$$

The $|\mathbf{r}_{ij} + \mathbf{n}|^{-1}$ term must be cleverly split into short-range and long-range contributions. This is achieved by representing $|\mathbf{r}_{ij} + \mathbf{n}|^{-1}$ as an integral over a dummy variable via the identity

$$\frac{1}{r} = \frac{1}{\sqrt{\pi}} \int_0^{\infty} dt \; t^{-1/2} \exp\left(-r^2 t\right). \tag{3.4}$$

Proof. Apply the $u-$substitution $u^2 = r^2 t$. Hence $2u\ du = r^2 dt$. So

$$\frac{1}{\sqrt{\pi}}\int_0^\infty dt\, t^{-1/2}\exp\left(-r^2 t\right) = \frac{1}{\sqrt{\pi}}\int_0^\infty du\, \frac{2re^{-u^2}u}{ur^2} = \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{du}{r}e^{-u^2} = \frac{2}{\sqrt{\pi}}\frac{\sqrt{\pi}}{2}\frac{1}{r} = r^{-1}.$$

□

Application of the identity gives

$$|\mathbf{r}_{ij}+\mathbf{n}|^{-1} = \frac{1}{\sqrt{\pi}}\int_0^\infty dt\ t^{-1/2}\exp\left(-t\,|\mathbf{r}_{ij}+\mathbf{n}|^2\right)$$

and

$$U(s) = \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' \frac{1}{\sqrt{\pi}}\int_0^\infty dt\ t^{-1/2}\exp\left(-s\,|\mathbf{n}|^2 - t\,|\mathbf{r}_{ij}+\mathbf{n}|^2\right). \tag{3.5}$$

The integral over dummy variable t may be split up into two parts, one with t large and one with t small corresponding respectively to short-range and long-range contributions of r^{-1}. For clarity in referring to these split up integrals, we define $V^{[a,b]}\left(s,\mathbf{n},\mathbf{r}_{ij}\right)$ as follows.

$$V^{[a,b]}\left(s,\mathbf{n},\mathbf{r}_{ij}\right) = \frac{1}{\sqrt{\pi}}\int_a^b dt\ t^{-1/2}\exp\left(-s\,|\mathbf{n}|^2 - t\,|\mathbf{r}_{ij}+\mathbf{n}|^2\right) \tag{3.6}$$

Hence $U(s)$ is decomposed into four different terms.

$$\begin{aligned}
U(s) &= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' V^{[0,\infty)}\left(s,\mathbf{n},\mathbf{r}_{ij}\right) \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' V^{[\alpha^2,\infty)}\left(s,\mathbf{n},\mathbf{r}_{ij}\right) + \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' V^{[0,\alpha^2]}\left(s,\mathbf{n},\mathbf{r}_{ij}\right) \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' V^{[\alpha^2,\infty)}\left(s,\mathbf{n},\mathbf{r}_{ij}\right) + \frac{1}{2}\sum_{i,j} q_i q_j \sum_{\mathbf{n}\neq 0} V^{[0,\alpha^2]}\left(s,\mathbf{n},\mathbf{r}_{ij}\right) \\
&\qquad + \frac{1}{2}\sum_{i,j} q_i q_j V^{[0,\alpha^2]}\left(s,0,\mathbf{r}_{ij}\right) - \frac{1}{2}\sum_i q_i^2 V^{[0,\alpha^2]}\left(s,0,0\right) \\
&\equiv U_1(s) + U_2(s) + U_3(s) - U_4(s)
\end{aligned} \tag{3.7}$$

3.2.1 $U_1(s)$

A straightforward u-substitution with $t = u^2 \left|\mathbf{r}_{ij} + \mathbf{n}\right|^{-2}$ allows $U_1(s)$ to be directly computed.

$$\begin{aligned}
U_1(s) &= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' V^{[\alpha^2,\infty)}\left(s, \mathbf{n}, \mathbf{r}_{ij}\right) \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' \frac{1}{\sqrt{\pi}} \int_{\alpha^2}^{\infty} dt\; t^{-1/2} \exp\left(-s\left|\mathbf{n}\right|^2 - t\left|\mathbf{r}_{ij} + \mathbf{n}\right|^2\right) \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' \frac{e^{-s|\mathbf{n}|^2}}{\sqrt{\pi}} \int_{\alpha^2}^{\infty} dt\; \frac{\exp\left(-t\left|\mathbf{r}_{ij} + \mathbf{n}\right|^2\right)}{\sqrt{t}} \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' \frac{e^{-s|\mathbf{n}|^2}}{\sqrt{\pi}} \int_{\alpha|\mathbf{r}_{ij}+\mathbf{n}|}^{\infty} du\; \frac{\exp\left(-u^2\right) 2u\left|\mathbf{r}_{ij} + \mathbf{n}\right|}{\left|\mathbf{r}_{ij} + \mathbf{n}\right|^2 u} \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' \frac{2e^{-s|\mathbf{n}|^2}}{\left|\mathbf{r}_{ij} + \mathbf{n}\right|\sqrt{\pi}} \int_{\alpha|\mathbf{r}_{ij}+\mathbf{n}|}^{\infty} du\; e^{-u^2} \\
&= \frac{1}{2}\sum_{i,j} q_i q_j {\sum_{\mathbf{n}}}' \frac{e^{-s|\mathbf{n}|^2} \operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij} + \mathbf{n}\right|\right)}{\left|\mathbf{r}_{ij} + \mathbf{n}\right|},
\end{aligned} \tag{3.8}$$

The s-dependence of $U_1(s)$ is particularly well-behaved, so the $s \to 0$ limit will be easy to take without concern about divergences (stay tuned). The sum is traditionally treated with a real-space cut-off under the minimum image convention. We will write

$$\lim_{s\to 0} U_1(s) = \frac{1}{2}\sum_{i\neq j} q_i q_j \frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|}, \tag{3.9}$$

where it is understood that $|\mathbf{r}_{ij}|$ is the distance to the minimum image and only terms with $|\mathbf{r}_{ij}| < r_c$ are included in the sum.

3.2.2 $U_2(s)$

$U_2(s)$ is a long-range energy due to $\mathbf{k} \neq 0$ terms. As initially defined,

$$U_2(s) = \frac{1}{2\sqrt{\pi}}\sum_{i,j} q_i q_j \sum_{\mathbf{n}\neq 0} \int_0^{\alpha^2} dt\; t^{-1/2} \exp\left(-s\left|\mathbf{n}\right|^2 - t\left|\mathbf{r}_{ij} + \mathbf{n}\right|^2\right) \tag{3.10}$$

For large t we were able to neglect distant $\mathbf{n}$ terms since sufficiently large t implies $\exp\left(-t\left|\mathbf{r}_{ij} + \mathbf{n}\right|^2\right) << \exp\left(-t\left|\mathbf{r}_{ij}\right|^2\right)\ \forall\ \mathbf{n} \neq 0$. For small t, $\exp\left(-t\left|\mathbf{r}_{ij} + \mathbf{n}\right|^2\right)$ does not decay rapidly, so many $\mathbf{n}$ terms contribute to the sum. The sum over $\mathbf{n}$ converges much faster if the (negative) coefficient of $\mathbf{n}$ is large, but this coefficient is t, which is small over $[0, \alpha^2]$. The Jacobi imaginary transformation converts an infinite sum over an exponential in t to one over an exponential in t^{-1}, which will allow much more rapid convergence of the sum than would be provided from a more straightforward attack. First, we state and prove the Jacobi imaginary transformation.

Claim (Jacobi's Imaginary Transformation):

$$\sum_{k=-\infty}^{\infty} \exp\left(-\left(x+ka\right)^2 t\right) = \frac{1}{a}\left(\frac{\pi}{t}\right)^{1/2} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{\pi^2 k^2}{a^2 t} + \frac{i2\pi kx}{a}\right). \tag{3.11}$$

Proof. Let the left hand side be $f(x) = \sum_{k=-\infty}^{\infty} \exp\left(-\left(x+ka\right)^2 t\right)$. Because the sum is infinite, it follows that $f(x+a) = f(x)$. Since f is periodic in x with period a, it can be expanded in a Fourier series,

$$\sum_{k=-\infty}^{\infty} \exp\left(-\left(x+ka\right)^2 t\right) = \sum_{k=-\infty}^{\infty} c_k \exp\left(\frac{2\pi ikx}{a}\right)$$

Projecting out the l^{th} term of the sum gives

$$\begin{aligned}
&\int_0^a dx \exp\left(-\frac{2\pi ilx}{a}\right) \sum_{k=-\infty}^{\infty} \exp\left(-\left(x+ka\right)^2 t\right) = \int_0^a dx \; c_l \\
\Rightarrow &\sum_{k=-\infty}^{\infty} \int_0^a dx \exp\left(-\left(x+ka\right)^2 t - \frac{2\pi ilx}{a}\right) = ac_l \\
\Rightarrow &\sum_{k=\infty}^{\infty} \int_{ka}^{(k+1)a} dx \exp\left(-x^2 t - \frac{2\pi ilx}{a}\right) = ac_l \\
\Rightarrow &\sum_{k=-\infty}^{\infty} \int_{ka}^{(k+1)a} dx \exp\left(-\left(x\sqrt{t} + \frac{\pi il}{a\sqrt{t}}\right)^2\right) \exp\left(-\frac{\pi^2 l^2}{a^2 t}\right) = ac_l \\
\Rightarrow &\int_{-\infty}^{\infty} dx \exp\left(-\left(x\sqrt{t} + \frac{\pi il}{a\sqrt{t}}\right)^2\right) \exp\left(-\frac{\pi^2 l^2}{a^2 t}\right) = ac_l \\
\Rightarrow &\frac{\sqrt{\pi}}{\sqrt{t}} \exp\left(-\frac{\pi^2 l^2}{a^2 t}\right) = ac_l
\end{aligned}$$

Plugging this expression for c_l into the Fourier Series yields

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} \exp\left(-\left(x+ka\right)^2 t\right) &= \sum_{k=-\infty}^{\infty} c_k \exp\left(\frac{2\pi ikx}{a}\right) \\
&= \frac{1}{a}\left(\frac{\pi}{t}\right)^{1/2} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{\pi^2 k^2}{a^2 t} + \frac{i2\pi kx}{a}\right).
\end{aligned}$$

□

Before this result can be applied to Equation (3.10), we perform some cosmetic algebra on the exponential of the integrand. Observe that

$$\begin{aligned}
-s\left|\mathbf{n}\right|^2 - t\left|\mathbf{r}_{ij}+\mathbf{n}\right|^2 &= -t\left|\mathbf{r}_{ij}\right|^2 - (s+t)\left|\mathbf{n}\right|^2 - 2t\mathbf{n}\cdot\mathbf{r}_{ij} \\
&= -\frac{t(t+s)}{t+s}\left|\mathbf{r}_{ij}\right|^2 - 2t\mathbf{n}\cdot\mathbf{r}_{ij} - (s+t)\left|\mathbf{n}\right|^2 \\
&= \left(-\frac{t^2}{t+s} - \frac{st}{t+s}\right)\left|\mathbf{r}_{ij}\right|^2 - 2t\mathbf{n}\cdot\mathbf{r}_{ij} - (s+t)\left|\mathbf{n}\right|^2 \\
&= -(s+t)\left|\mathbf{n}+\frac{t\mathbf{r}_{ij}}{t+s}\right|^2 - \frac{st\left|\mathbf{r}_{ij}\right|^2}{t+s} \qquad (3.12)
\end{aligned}$$

so

$$U_2(s) = \frac{1}{2\sqrt{\pi}}\sum_{i,j} q_i q_j \sum_{\mathbf{n}\neq 0}\int_0^{\alpha^2} dt\; t^{-1/2}\exp\left(-\frac{st\left|\mathbf{r}_{ij}\right|^2}{s+t}\right)\exp\left(-(s+t)\left|\mathbf{n}+\frac{t\mathbf{r}_{ij}}{t+s}\right|^2\right) \qquad (3.13)$$

Application of the Jacobi imaginary transformation to the second exponential of the integrand yields

$$\begin{aligned}
U_2(s) =& \frac{1}{2abc\sqrt{\pi}}\sum_{i,j} q_i q_j \sum_{(k,l,m)\neq(0,0,0)}\int_0^{\alpha^2} dt\; t^{-1/2}\exp\left(-\frac{st\left|\mathbf{r}_{ij}\right|^2}{s+t}\right)\left(\frac{\pi}{s+t}\right)^{3/2} \\
&\times \exp\left[-\frac{\pi^2}{s+t}\left(\frac{k^2}{a^2}+\frac{l^2}{b^2}+\frac{m^2}{c^2}\right)\right]\exp\left[\frac{i2\pi t}{t+s}\left(\frac{x_{ij}k}{a}+\frac{y_{ij}l}{b}+\frac{z_{ij}m}{c}\right)\right],
\end{aligned}$$

Recall that $\mathbf{n} = (ka, lb, mc)$ in this chapter. Defining the reciprocal lattice vectors, $\mathbf{k} = 2\pi\left(\frac{k}{a},\frac{l}{b},\frac{m}{c}\right)$ allows this to be rewritten as

$$\begin{aligned}
U_2(s) = \frac{1}{2abc\sqrt{\pi}}\sum_{i,j} q_i q_j \sum_{\mathbf{k}\neq 0}\int_0^{\alpha^2} dt\; t^{-1/2}\left(\frac{\pi}{s+t}\right)^{3/2}\exp\left(-\frac{st\left|\mathbf{r}_{ij}\right|^2}{s+t}\right) \\
\times \exp\left(-\frac{\left|\mathbf{k}\right|^2}{4(s+t)}+\frac{it\mathbf{k}\cdot\mathbf{r}_{ij}}{t+s}\right) \qquad (3.14)
\end{aligned}$$

We must now manipulate this expression into an expansion asymptotic in s such that the $s\to 0$ limit can be found. Let

$$v^2 = \frac{t}{s(t+s)}\left|\frac{\mathbf{k}}{2}+is\mathbf{r}_{ij}\right|^2 = \frac{t}{s(t+s)}\left(\frac{\left|\mathbf{k}\right|^2}{4}+\frac{i\mathbf{k}\cdot\mathbf{r}_{ij}s}{2}-s^2\left|\mathbf{r}_{ij}\right|^2\right)$$

$$\Rightarrow 2v\,dv = \left|\frac{\mathbf{k}}{2}+is\mathbf{r}_{ij}\right|^2\left(\frac{s(t+s)-st}{s^2(t+s)^2}\right)dt = \frac{2t^{1/2}(t+s)^{3/2}}{s^{1/2}\left|\frac{\mathbf{k}}{2}+is\mathbf{r}_{ij}\right|}dv$$

Observe then that

$$\begin{aligned}\exp\left(v^2\right) &= \exp\left(-\frac{st\left|\mathbf{r}_{ij}\right|^2}{t+s} + \frac{i\mathbf{k}\cdot\mathbf{r}_{ij}t}{t+s} + \frac{t\left|\mathbf{k}\right|^2}{4s(t+s)}\right) \\ &= \exp\left(\frac{\left|\mathbf{k}\right|^2}{4s}\right)\exp\left(-\frac{st\left|\mathbf{r}_{ij}\right|^2}{t+s} + \frac{i\mathbf{k}\cdot\mathbf{r}_{ij}t}{t+s} - \frac{\left|\mathbf{k}\right|^2}{4(t+s)}\right)\end{aligned}$$

Inserting this relationship into Equation (3.14) gives

$$\begin{aligned}U_2(s) &= \frac{\pi}{2abc}\sum_{i,j} q_i q_j \sum_{\mathbf{k}\neq 0}\int_0^{\alpha^2} dt\ t^{-1/2}\left(\frac{\pi}{s+t}\right)^{3/2}\exp\left(-\frac{\left|\mathbf{k}\right|^2}{4s}\right)\exp\left(v^2\right) \\ &= \frac{\pi}{2abc}\sum_{i,j} q_i q_j \sum_{\mathbf{k}\neq 0}\int_0^{\gamma}\frac{2dv}{s^{1/2}\left|\frac{\mathbf{k}}{2}+is\mathbf{r}_{ij}\right|}\exp\left(-\frac{\left|\mathbf{k}\right|^2}{4s}\right)\exp\left(v^2\right) \\ &= \frac{\pi}{abc}\sum_{i,j} q_i q_j \sum_{\mathbf{k}\neq 0}\frac{\exp\left(-\frac{\left|\mathbf{k}\right|^2}{4s}\right)}{s^{1/2}\left|\frac{\mathbf{k}}{2}+is\mathbf{r}_{ij}\right|}\int_0^{\gamma} dv\ \exp\left(v^2\right), \qquad (3.15)\end{aligned}$$

where $\gamma = \left(\alpha\left|\frac{\mathbf{k}}{2}+is\mathbf{r}_{ij}\right|\right)/\sqrt{s\left(\alpha^2+s\right)}$. An asymptotic series for the above integral is given in equation (3.18) of de Leeuw *et al.* [15]

$$\int_0^{\omega/\sqrt{s}} dv\ \exp\left(v^2\right) \sim \frac{s^{1/2}}{2\omega}e^{\omega^2/s}\left[1+\frac{s}{2\omega^2}+\mathcal{O}\left(s^2\right)\right],$$

where $\omega = \sqrt{s}\gamma$ relates de Leeuw's parameter ω to my parameter γ. Inserting this asymptotic expansion into Equation (3.15) and expanding the expression out to neglect terms first order or higher in s, one finds

$$
\begin{aligned}
U_2(s) &\sim \frac{\pi}{abc} \sum_{i,j,\mathbf{k}\neq 0} \frac{q_i q_j \exp\left(-\frac{|\mathbf{k}|^2}{4s}\right)}{\left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right| 2\omega} \exp\left(\frac{\omega^2}{s}\right) \left[1 + \frac{s}{2\omega^2} + \mathcal{O}\left(s^2\right)\right] \\
&\sim \frac{\pi}{2abc} \sum_{i,j,\mathbf{k}\neq 0} \frac{q_i q_j}{\left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right| \omega} \exp\left(\frac{\alpha^2 \left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right|^2}{\left(\alpha^2 + s\right) s} - \frac{|\mathbf{k}|^2}{4s}\right) \left[1 + \frac{s}{2\omega^2} + \mathcal{O}\left(s^2\right)\right] \\
&\sim \frac{\pi}{2abc} \sum_{i,j,\mathbf{k}\neq 0} \frac{q_i q_j}{\left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right| \omega} \exp\left(\frac{\alpha^2 |\mathbf{k}|^2}{4(\alpha^2 + s)s} + \frac{\alpha^2 i\mathbf{k}\cdot\mathbf{r}_{ij}}{\alpha^2 + s} - \frac{\alpha^2 s \left|\mathbf{r}_{ij}\right|^2}{\alpha^2 + s} - \frac{|\mathbf{k}|^2}{4s}\right) \\
&\qquad \times \left[1 + \frac{s}{2\omega^2} + \mathcal{O}\left(s^2\right)\right] \\
&\sim \frac{\pi}{2abc} \sum_{i,j,\mathbf{k}\neq 0} \frac{q_i q_j}{\left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right| \omega} \exp\left(\frac{|\mathbf{k}|^2}{4s}\left(\frac{\alpha^2}{\alpha^2 + s} - 1\right) + i\mathbf{k}\cdot\mathbf{r}_{ij} + \mathcal{O}(s)\right) \\
&\qquad \times \left[1 + \frac{s}{2\omega^2} + \mathcal{O}\left(s^2\right)\right] \\
&\sim \frac{\pi}{2abc} \sum_{i,j,\mathbf{k}\neq 0} \frac{q_i q_j}{\left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right| \omega} \exp\left(\frac{|\mathbf{k}|^2}{4s}\left(-\frac{s}{\alpha^2 + s}\right) + i\mathbf{k}\cdot\mathbf{r}_{ij} + \mathcal{O}(s)\right) \\
&\qquad \times \left[1 + \frac{s}{2\omega^2} + \mathcal{O}\left(s^2\right)\right] \\
&\sim \frac{\pi}{2abc} \sum_{i,j,\mathbf{k}\neq 0} \frac{q_i q_j}{\left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right| \omega} \exp\left(\frac{-|\mathbf{k}|^2}{4\left(\alpha^2 + s\right)} + i\mathbf{k}\cdot\mathbf{r}_{ij} + \mathcal{O}(s)\right) \left[1 + \frac{s}{2\omega^2} + \mathcal{O}\left(s^2\right)\right]
\end{aligned}
\tag{3.16}
$$

Noting that

$$\lim_{s\to 0} \omega = \lim_{s\to 0} \frac{\alpha \left|\frac{\mathbf{k}}{2} + is\mathbf{r}_{ij}\right|}{\sqrt{\alpha^2 + s}} = \frac{|\mathbf{k}|}{2},$$

we can easily take the limit as $s \to 0$ of Equation (3.16).

$$\lim_{s\to 0} U_2(s) = \frac{\pi}{2abc} \sum_{i,j} q_i q_j \sum_{\mathbf{k}\neq 0} \frac{4}{|\mathbf{k}|^2} \exp\left(-\frac{|\mathbf{k}|^2}{4\alpha^2} + i\mathbf{k}\cdot\mathbf{r}_{ij}\right) \tag{3.17}$$

Recognizing $\sum_i q_i e^{i\mathbf{k}\cdot\mathbf{r}}$ as the structure factor, $S(\mathbf{k})$, and abc as the volume, V, we have

$$\lim_{s\to 0} U_2(s) = \frac{1}{2V} \sum_{\mathbf{k}\neq 0} \frac{4\pi}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \left|S\left(\mathbf{k}\right)\right|^2. \tag{3.18}$$

3.2.3 $U_3(s)$

To compute the $\mathbf{n} = 0$ contribution, we start by treating $U_3(s)$ the same way as $U_2(s)$. That is to say the Jacobi imaginary transformation is applied then $\mathbf{n} = 0 \Rightarrow \mathbf{k} = 0$ is inserted into Equation (3.14) yielding

$$U_3(s) = \frac{\pi}{2abc} \sum_{i,j} q_i q_j \int_0^{\alpha^2} dt\ t^{-1/2} (t+s)^{-3/2} \exp\left(- \frac{st\,|\mathbf{r}_{ij}|^2}{s+t} \right) \tag{3.19}$$

Let $u = \frac{t}{s+t} \Rightarrow du = \frac{s}{(s+t)^2} dt$. So

$$\begin{aligned}
U_3(s) &= \frac{\pi}{2abc} \sum_{i,j} q_i q_j \int_0^{\alpha^2/(s+\alpha^2)} du\ \frac{(s+t)^2}{s} t^{-1/2} (t+s)^{-3/2} \exp\left(- \frac{st\,|\mathbf{r}_{ij}|^2}{s+t} \right) \\
&= \frac{\pi}{2abc} \sum_{i,j} q_i q_j \int_0^{\alpha^2/(s+\alpha^2)} du\ \frac{(s+t)^{1/2}}{s\ t^{1/2}} \exp\left(-su\,|\mathbf{r}_{ij}|^2 \right) \\
&= \frac{\pi}{2abcs} \sum_{i,j} q_i q_j \int_0^{\alpha^2/(s+\alpha^2)} du\ u^{-1/2} \exp\left(-su\,|\mathbf{r}_{ij}|^2 \right) \\
&= \frac{\pi}{2abcs} \sum_{i,j} q_i q_j \int_0^{\alpha^2/(s+\alpha^2)} \frac{du}{\sqrt{u}} \left[1 - su\,|\mathbf{r}_{ij}|^2 + \frac{s^2 u^2\,|\mathbf{r}_{ij}|^4}{2} - \ldots \right] \\
&= \frac{\pi}{2abcs} \sum_{i,j} q_i q_j \int_0^{\alpha^2/(s+\alpha^2)} du\ \left[\frac{1}{\sqrt{u}} - s\sqrt{u}\,|\mathbf{r}_{ij}|^2 + \frac{s^2 u^{3/2}\,|\mathbf{r}_{ij}|^4}{2} - \ldots \right] \\
&= \frac{\pi}{2abcs} \left[2u^{1/2} - \frac{2}{3} s u^{3/2}\,|\mathbf{r}_{ij}|^2 + \mathcal{O}\left(s^2\right) \right] \Bigg|_0^{\alpha^2/(s+\alpha^2)} \\
&= \frac{1}{2abc} \sum_{i,j} q_i q_j \left[\frac{2\pi\alpha}{s\sqrt{s+\alpha^2}} - \frac{2\pi}{3}\,|\mathbf{r}_{ij}|^2 \frac{\alpha^3}{(s+\alpha^2)^{3/2}} + \mathcal{O}(s) \right]
\end{aligned} \tag{3.20}$$

This expression reveals the fundamental divergence of the $s \to 0$ limit, which is glossed over by handwavy derivations of Ewald summation. However, in a neutral system, $\sum_i q_i = 0$, so for finite s this first term will exactly vanish. The implication is that the sum can only be regularized if the unit cell is neutral. This was long believed to be the case, but an alternate view of the $\mathbf{k} = 0$ term has revealed that non-neutral systems can indeed by handled in a nearly identical manner. [28] Hence

$$U_3(s) = -\frac{\pi}{3V} \sum_{i,j} q_i q_j \left[|\mathbf{r}_{ij}|^2 \frac{\alpha^3}{(s+\alpha^2)^{3/2}} + \mathcal{O}(s) \right]$$

and

$$\lim_{s\to 0} U_3(s) = -\frac{\pi}{3V} \sum_{i,j} q_i q_j\,|\mathbf{r}_{ij}|^2 . \tag{3.21}$$

This last term, the *surface term*, is often related to the dipole moment of the primitive cell [56]. Actually, the term should be associated with the itinerant dipole moment [9] as made clear by the recent work of Herce *et al.* [28] The contribution of this surface term to most thermodynamic quantities can be shown to vanish if the infinitely replicated system is embedded in an $\epsilon = \infty$ dielectric. [35] This is the so called *tin foil* boundary condition, which yields the simple Bertaut expression. The tin foil boundary condition is commonly employed, though Smith has warned that a virial term of the asymptotic calculation does not vanish in the $\epsilon \to \infty$ limit. [57]

3.2.4 $U_4(s)$

The energy $U_4(s)$ corresponds to the self interaction energy with $\mathbf{r}_{ij} = 0$, $\mathbf{n} = 0$. As defined,

$$\begin{aligned} U_4(s) &= \frac{1}{2}\sum_i q_i^2 V^{[0,\alpha^2]}(s,0,0) \\ &= \frac{1}{2\sqrt{\pi}}\sum_i q_i^2 \int_0^{\alpha^2} dt\ t^{-1/2} \\ &= \sum_i q_i^2 \frac{t^{1/2}}{\sqrt{\pi}}\bigg|_0^{\alpha^2} = \frac{\alpha}{\sqrt{\pi}}\sum_i q_i^2 \end{aligned} \tag{3.22}$$

since the exponential of $V^{[0,\alpha^2]}(s,\mathbf{n},\mathbf{r}_{ij})$ vanishes when $\mathbf{n} = \mathbf{r}_{ij} = 0$.

3.2.5 Bringing It All Together

Recombining the energy terms and taking the $s \to 0$ limit gives

$$\begin{aligned} \lim_{s\to 0} U(s) &= \lim_{s\to 0} U_1(s) + \lim_{s\to 0} U_2(s) + \lim_{s\to 0} U_3(s) - \lim_{s\to 0} U_4(s) \\ &= \frac{1}{2}\sum_{i\neq j} q_i q_j \frac{\operatorname{erfc}(\alpha\,|\mathbf{r}_{ij}|)}{|\mathbf{r}_{ij}|} + \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi e^{-k^2/4\alpha^2}}{k^2}\,|S(\mathbf{k})|^2 \\ &\quad - \frac{\pi}{3V}\sum_{i,j} q_i q_j\,|\mathbf{r}_{ij}|^2 - \frac{\alpha}{\sqrt{\pi}}\sum_i q_i^2 \end{aligned} \tag{3.23}$$

This is the traditional Ewald sum energy we found previously with the exception of the additional surface term, which can be neglected for the tin foil boundary condition. This derivation places the Ewald summation procedure on rigorous, firm ground, while also forming the basis for extensions beyond point charges.

3.3 Gaussian Charge Coulomb Energy

The de Leeuw *et al.* derivation of Ewald summation is particularly amenable to computing the energies between Gaussian charges. This was shown by Reed, Lanning, and Madden in a slab geometry. [40] It can be extended to periodic replication in all three dimensions in the same way. In doing so we expose an error in the energy given by Reed *et al.* The correction to their slab calculation is discussed at length in Appendix B.

The pairwise energy between Gaussians is

$$U_{pair} = \frac{\eta^6}{2\pi^3} \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \frac{Q_i \exp\left[-\eta^2 \left|\mathbf{r}' - \mathbf{r}_i\right|^2\right] Q_j \exp\left[-\eta^2 \left|\mathbf{r}'' - \mathbf{r}_j\right|^2\right]}{\left|\mathbf{r}'' - \mathbf{r}' + \mathbf{r}_{ij}\right|}, \tag{3.24}$$

where η^6/π^3 is the product of the two Gaussian normalization factors and the factor of $1/2$ prevents double counting. As in the case of the point charges, summing an infinite periodic system is conditionally convergent. The same $\exp\left(-s\left|\mathbf{n}\right|^2\right)$ convergence factor can be applied to define $U(s)$, which is then uniform on $s \geq 0$.

$$U(s) = \frac{\eta^6}{2\pi^3} \sum_{i,j,\mathbf{n}} \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \frac{Q_i Q_j \exp\left(-\eta^2\left(\left|\mathbf{r}' - \mathbf{r}_i\right|^2 + \left|\mathbf{r}'' - \mathbf{r}_j\right|^2\right)\right) \exp\left(-s\left|\mathbf{n}\right|^2\right)}{\left|\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}' + \mathbf{n}\right|}. \tag{3.25}$$

Notably there is no prime on the sums as the charge density centered on site i interacts with itself. As in the point charge problem, we begin by applying Equation (3.4) in order to write

$$\begin{aligned} \left|\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}' + \mathbf{n}\right|^{-1} &= \frac{1}{\sqrt{\pi}} \int_0^\infty dt\ t^{-1/2} \exp\left(-t\left|\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}' + \mathbf{n}\right|^2\right) \\ &= \frac{1}{\sqrt{\pi}} \int_0^\infty dt\ t^{-1/2} \exp\left[-t\left(x_{ij} + x'' - x' + ka\right)^2\right] \\ &\quad \times \exp\left[-t\left(y_{ij} + y'' - y' + lb\right)^2\right] \exp\left[-t\left(z_{ij} + z'' - z' + mc\right)^2\right] \end{aligned} \tag{3.26}$$

Additionally we make use of the following identity to write the Gaussian densities as integrals over dummy variables. Claim:

$$\exp\left(-\left|\mathbf{r}\right|^2 \eta^2\right) = (2\pi)^{-3} \frac{\pi^{3/2}}{\eta^3} \int_{-\infty}^{\infty} \exp\left(-\frac{\left|\mathbf{v}\right|^2}{4\eta^2} + i\mathbf{v}\cdot\mathbf{r}\right) d\mathbf{v}. \tag{3.27}$$

Proof. We start with the right hand side and proceed by completing the square.

$$
\begin{aligned}
&\frac{\pi^{3/2}}{(2\pi\eta)^3}\int_{-\infty}^{\infty} d\mathbf{v}\ \exp\left(-\frac{|\mathbf{v}|^2}{4\eta^2}+i\mathbf{v}\cdot\mathbf{r}\right)\\
&\quad=\frac{\pi^{3/2}}{(2\pi\eta)^3}\int_{-\infty}^{\infty} d\mathbf{v}\ \exp\left[-\left(\frac{\mathbf{v}}{2\eta}-i\mathbf{r}\eta\right)^2-|\mathbf{r}|^2\eta^2\right]\\
&\quad=\frac{\pi^{3/2}}{(2\pi\eta)^3}\exp\left(-|\mathbf{r}|^2\eta^2\right)\int_{-\infty}^{\infty} d\mathbf{v}'\ \exp\left(-\frac{|\mathbf{v}'|^2}{2\eta}\right)\\
&\quad=\frac{\pi^{3/2}}{(2\pi\eta)^3}\exp\left(-|\mathbf{r}|^2\eta^2\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dv''_x dv''_y dv''_z\,(2\eta)^3\exp\left(-|\mathbf{v}''|^2\right)\\
&\quad=\frac{\pi^{3/2}}{(2\pi\eta)^3}\exp\left(-|\mathbf{r}|^2\eta^2\right)(2\eta)^3\left(\sqrt{\pi}\right)^3=\exp\left(-|\mathbf{r}|^2\eta^2\right)
\end{aligned}
$$

□

So applying equations (3.26) and (3.27) to (3.25),

$$
\begin{aligned}
U(s)=\frac{\eta^6}{\pi 2\sqrt{\pi}}\frac{\pi^3}{(2\pi)^6\eta^6}&\int_0^\infty dt\int_{\mathbb{R}^3} d\mathbf{r}'\int_{\mathbb{R}^3} d\mathbf{r}''\int_{\mathbb{R}^3} d\mathbf{v}\int_{\mathbb{R}^3} d\mathbf{w}\ t^{-1/2}\sum_{i,j}\sum_{k,l,m=-\infty}^{\infty} Q_iQ_j\\
&\times\exp\left(-s|\mathbf{n}|^2\right)\exp\left[-t\left(x_{ij}+x''-x'+ka\right)^2-t\left(y_{ij}+y''-y'+lb\right)^2\right]\\
&\times\exp\left[-t\left(z_{ij}+z''-z'+mc\right)^2-\frac{|\mathbf{v}|^2}{4\eta^2}+i\mathbf{v}\cdot\mathbf{r}'-\frac{|\mathbf{w}|^2}{4\eta^2}+i\mathbf{w}\cdot\mathbf{r}''\right] \qquad (3.28)
\end{aligned}
$$

Here is where things get tricky. If we were to naively set $s \to 0$ here we could follow the route of Reed *et al.* We first follow this path and rederive a result very similar to that of Reed *et al.* In Appendix A the convergence factor calculation is handled in gory detail. Ultimately the naive approach is valid because the problem can be reduced to integrals which have already been carefully evaluated in the point charge problem.

3.4 Naive $s = 0$ Approach

If s were nonzero, then the $\exp\left(-s|\mathbf{n}|^2\right)$ term of Equation (3.28) would prevent the straightforward application of the Jacobi imaginary transformation since the periodic terms would be of the form

$$\sum_{k=-\infty}^{\infty}\exp\left[-t\left(x+ka\right)^2-sk^2\right].$$

With $s = 0$, these infinite sums can easily be directly transformed via Equation (3.11).

$$
\begin{aligned}
U =&\frac{1}{2^7\pi^5 abc}\int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w} \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}''\ t^{-2} \sum_{i,j} \sum_{k,l,m=-\infty}^{\infty} Q_i Q_j \exp\left[-\frac{|\mathbf{v}|^2 + |\mathbf{w}|^2}{4\eta^2}\right] \\
&\times \exp\left[-\frac{\pi^2 k^2}{a^2 t} + \frac{i2\pi k x_{ij}}{a}\right] \exp\left[-\frac{\pi^2 l^2}{b^2 t} + \frac{i2\pi l y_{ij}}{b}\right] \exp\left[-\frac{\pi^2 m^2}{c^2 t} + \frac{i2\pi m z_{ij}}{c}\right] \\
&\times \exp\left[i\left(\frac{2\pi k}{a} + w_x\right) x''\right] \exp\left[i\left(\frac{2\pi l}{b} + w_y\right) y''\right] \exp\left[i\left(\frac{2\pi m}{c} + w_z\right) z''\right] \\
&\times \exp\left[-i\left(\frac{2\pi k}{a} - v_x\right) x'\right] \exp\left[-i\left(\frac{2\pi l}{b} - v_y\right) y'\right] \exp\left[-i\left(\frac{2\pi m}{c} - v_z\right) z'\right]
\end{aligned}
\tag{3.29}
$$

The integrals over $\mathbf{r}'$ and $\mathbf{r}''$ reduce to three dimensional δ-functions due to the completeness of the Fourier basis.
Claim:

$$
(2\pi)^{-1} \int dx\ \exp\left[i\left(k - k'\right) x\right] = \delta\left(k - k'\right) \tag{3.30}
$$

Proof. Let $F\left\{f(k)\right\}$ denote the Fourier transform.

$$
\begin{aligned}
F\left\{\delta\left(k - k'\right)\right\} &= \int_{-\infty}^{\infty} dk\ e^{ikx}\delta(k - k') = e^{ik'x} \\
\Rightarrow & F^{-1}\left\{e^{ik'x}\right\} = \delta\left(k - k'\right) \\
\Rightarrow & \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\ e^{ik'x} e^{-ikx} = \delta\left(k - k'\right) \\
\Rightarrow & \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\ e^{i(k'-k)x} = \delta\left(k - k'\right)
\end{aligned}
$$

□

Converting from $\mathbf{n}$ to $\mathbf{k}$ and utilizing the δ-functions allows Equation (3.29) to be written compactly as

$$
\begin{aligned}
U &= \frac{\pi}{2abc}\int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w}\ t^{-2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{v}|^2 + |\mathbf{w}|^2}{4\eta^2}\right] \\
&\qquad \times \exp\left[-\frac{|\mathbf{k}|^2}{4t} + i\mathbf{k}\cdot\mathbf{r}_{ij}\right] \delta\left(\mathbf{v} - \mathbf{k}\right) \delta\left(\mathbf{w} + \mathbf{k}\right) \\
&= \frac{\pi}{2abc}\int_0^\infty dt\ t^{-2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{2\eta^2}\right] \exp\left[-\frac{|\mathbf{k}|^2}{4t} + i\mathbf{k}\cdot\mathbf{r}_{ij}\right] \\
&= \frac{\pi}{2abc}\int_0^\infty dt\ t^{-2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4}\left(\frac{1}{t} + \frac{2}{\eta^2}\right) + i\mathbf{k}\cdot\mathbf{r}_{ij}\right]
\end{aligned}
\tag{3.31}
$$

Now consider the substitution

$$t' = \frac{t\eta^2}{\eta^2 + 2t} \qquad \Rightarrow \qquad dt' = \frac{\eta^2(\eta^2 + 2t) - 2t\eta^2}{(\eta^2 + 2t)^2} dt = \frac{\eta^4}{(\eta^2 + 2t)^2} dt = \left(\frac{t'}{t}\right)^2 dt$$

So $t'^{-2} dt' = t^{-2} dt$. The bounds of the new integral over t' are given by

$$\lim_{t \to 0} t' = \lim_{t \to 0} \left(\frac{t\eta^2}{\eta^2 + 2t} \right) = 0$$

and

$$\lim_{t \to \infty} t' = \lim_{t \to \infty} \left(\frac{t\eta^2}{\eta^2 + 2t} \right) = \frac{\eta^2}{2}$$

This final limit is mistakenly set to infinity in the Reed, Lanning, and Madden paper. The correct substitution yields

$$U = \frac{\pi}{2abc} \int_0^{\eta^2/2} dt' \; t'^{-2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4t'} + i\mathbf{k} \cdot \mathbf{r}_{ij} \right] \tag{3.32}$$

This integral is of the same form as a naive $s = 0$ substitution into the quantity $U_2(s) + U_3(s)$ from equations (3.14) and (3.19) if the α^2 from before is replaced by $\eta^2/2$. I describe the substitution as naive because the convergence factor will be slightly different and it is not immediately obvious that the results will carry over. Appendix A performs the detailed convergence factor calculation, confirming that the $s \to 0$ limit of equations (3.17) and (3.22) will remain valid here. On a handwavy level, the naive treatment is sufficient because the ordering of the Gaussians is identical to the ordering of the point charge system.

$$U = \frac{1}{2V} \sum_{\mathbf{k} \neq 0} \frac{4\pi}{k^2} \exp\left(-\frac{k^2}{2\eta^2} \right) |S(\mathbf{k})|^2 - \frac{\pi}{3V} \sum_{i,j} Q_i Q_j |\mathbf{r}_{ij}|^2 \tag{3.33}$$

In a system with tin foil boundary conditions, this second term vanishes and the energy is seen to be exactly the same as that found previously in Equation (2.21). Unfortunately, if $\frac{\eta^2}{2} > \alpha^2$ then the k-space sum will converge less rapidly than the point charge system's k-space sum. In practice it is likely that the same k-space cutoff would be used for point charge and Gaussian subsystems, so the energy in its current form would suffer from significant truncation error at short range. To rectify this issue, some of the energy calculation can be shifted into real space by splitting up the integral of Equation (3.32).

$$\begin{aligned} U = {} & \frac{\pi}{2abc} \int_0^{\alpha^2} dt' \; t'^{-2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4t'} + i\mathbf{k} \cdot \mathbf{r}_{ij} \right] \\ & + \frac{\pi}{2abc} \int_{\alpha^2}^{\eta^2/2} dt' \; t'^{-2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4t'} + i\mathbf{k} \cdot \mathbf{r}_{ij} \right] \end{aligned} \tag{3.34}$$

The first of these integrals becomes the k-space sum, but the second integral we compute in real space. Inverting the Jacobi imaginary transformation on this second integral gives

$$\begin{aligned}
I_2 &= \frac{\pi}{2abc}\int_{\alpha^2}^{\eta^2/2} dt'\ t'^{-2}\sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4t'} + i\mathbf{k}\cdot\mathbf{r}_{ij}\right] \\
&= \frac{1}{2\sqrt{\pi}}\int_{\alpha^2}^{\eta^2/2} dt'\ t'^{-1/2}\sum_{i,j} Q_i Q_j \sum_{\mathbf{n}} \exp\left(-t'\left|\mathbf{r}_{ij}+\mathbf{n}\right|^2\right) \\
&= \frac{1}{2\sqrt{\pi}}\int_{\alpha^2}^{\infty} dt'\ t'^{-1/2}\sum_{i,j} Q_i Q_j {\sum_{\mathbf{n}}}' \exp\left(-t'\left|\mathbf{r}_{ij}+\mathbf{n}\right|^2\right) + \frac{1}{2\sqrt{\pi}}\int_{\alpha^2}^{\eta^2/2} dt'\ t'^{-1/2}\sum_{i} Q_i^2 \\
&\quad - \frac{1}{2\sqrt{\pi}}\int_{\eta^2/2}^{\infty} dt'\ t'^{-1/2}\sum_{i,j} Q_i Q_j {\sum_{\mathbf{n}}}' \exp\left(-t'\left|\mathbf{r}_{ij}+\mathbf{n}\right|^2\right) \qquad (3.35)
\end{aligned}$$

where the $\mathbf{n}=0$, $i=j$ term has been treated separately. The first and third integrals are simply the naive $s=0$ value of $U_1(s)$ as given in Equation (3.8) except with α^2 replaced by $\eta^2/2$ for the third integral. The second integral can easily be computed explicitly. Hence (3.34) can be rewritten as

$$\begin{aligned}
U &= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi e^{-k^2/4\alpha^2}}{k^2}\left|S\left(\mathbf{k}\right)\right|^2 + \frac{1}{2}\sum_{i\neq j} Q_i Q_j \frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right) - \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}}\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} \\
&\quad + \left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}}\right)\sum_i Q_i^2 \\
&= U_{point\ charge} - \frac{1}{2}\sum_{i\neq j} Q_i Q_j \frac{\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}}\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} + \frac{\eta}{\sqrt{2\pi}}\sum_i Q_i^2. \qquad (3.36)
\end{aligned}$$

3.5 Combined Point Charge and Gaussian Charge System

The energy of interaction between the point charge and Gaussian charge subsystems can be computed in an entirely analogous manner. Considering only the $s=0$ calculation for simplicity,

$$U = \frac{\pi}{abc}\int_0^{\infty} dt\ t^{-2}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{\mathbf{k}} q_i Q_j \exp\left[-\frac{|\mathbf{k}|^2}{4}\left(\frac{1}{t}+\frac{1}{\eta^2}\right) + i\mathbf{k}\cdot\mathbf{r}_{ij}\right], \qquad (3.37)$$

which is similar to Equation (3.31). The factor of 1/2 is dropped since the sums are over different sets and therefore do not double count. The term in the exponential is also altered slightly since there is only a single Gaussian involved in the interaction. The natural substitution for the integral is $t' = t\eta^2/\left(t+\eta^2\right)$, so by analogy to (3.36) the energy (with tin foil boundary conditions) is

$$U = \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi e^{-k^2/4\alpha^2}}{k^2}\left(S_1^*\left(\mathbf{k}\right)S_2\left(\mathbf{k}\right)+S_2^*\left(\mathbf{k}\right)S_1\left(\mathbf{k}\right)\right) \\ + \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)-\operatorname{erfc}\left(\eta\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|}, \tag{3.38}$$

with $S_1(\mathbf{k})$ and $S_2(\mathbf{k})$ being the structure factors for the point charge and Gaussian charge subsystems respectively. The factor of 1/2 has re-entered through the symmetrization of the product of the structure factors. The total Coulomb energy of the combined system is then

$$U = \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}\left|S_1\left(\mathbf{k}\right)+S_2\left(\mathbf{k}\right)\right|^2\exp\left(-\frac{k^2}{4\alpha^2}\right)+\left(\frac{\eta}{\sqrt{2\pi}}-\frac{\alpha}{\sqrt{\pi}}\right)\sum_{i=1}^{N}Q_i^2-\frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{n}q_i^2 \\ +\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}q_iq_j\frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|}+\sum_{i=1}^{n}\sum_{j=1}^{N}q_iQ_j\frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)-\operatorname{erfc}\left(\eta\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} \\ +\frac{1}{2}\sum_{i=1}^{N}\sideset{}{'}\sum_{j=1}^{N}Q_iQ_j\frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)-\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}}\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|}. \tag{3.39}$$

Chapter 4

Energy Minimization and Classical Dynamics

4.1 The Basic Model

Armed with the potential energy of a mixed system of Gaussian charges and point charges, Equation (3.39), we can now address methods for simulating ionic dynamics near the model metallic surface. The metal is composed of a discrete collection of metal sites at positions $\{\mathbf{R}_i\}$ carrying a Gaussian charge with integrated charge $\{Q_i\}$. The ions, with positions $\{\mathbf{r}_i\}$ and magnitudes $\{q_i\}$, are treated as rigid point charges. As the Gaussians represent a metallic surface, that surface must be an electrostatic equipotential, but the notion of an atomic surface is significantly different from the surfaces of classical electrostatics. In classical electrostatics one talks of surfaces which are macroscopic and can therefore be well represented by geometric idealizations (i.e. a plane, a cylinder, a sphere). On the atomic scale it is no longer reasonable to believe that the equipotential surface will adopt such a simple geometry. Rather, it should be corrugated, reflecting the influence of the nuclei. Indeed the corrugations are precisely the reason a Gaussian charge was introduced to represent the charge distribution on a metal site. Each site will be constrained to have a fixed Gaussian spread, but the sites are collectively metallic because they are allowed to redistribute the charge amongst themselves until the sites (not the geometric idealized surface) are equipotential. The Gaussian spread is the single parameter entering the model, and unlike the screening parameter of Ewald summation, this spread is physically significant. It provides a measure of how tightly the nuclei attract the electron distribution. The choice of an appropriate parameter will be discussed at length in Chapter 5.

The Coulomb potential at site i is $\partial U_c/\partial Q_i$, so the equipotential constraint is equivalent to

$$\frac{\partial U_c}{\partial Q_i} = \frac{\partial U_c}{\partial Q_j} \;\; \forall \; i, j, \tag{4.1}$$

where U_c is the Coulomb energy given in Equation (3.39). Reed *et al.* consider a system which is coupled to external electrodes, capable of sourcing or sinking charge to maintain a constant potential. [40] This scenario simplifies Equation (4.1) since each site must be held to the externally applied potential, V_0. So the constant potential constraint is

$$\frac{\partial U_c}{\partial Q_i} = V_0 \;\; \forall \; i \tag{4.2}$$

Put another way, the charges rearrange themselves to minimize

$$U_T = U_c - V_0 \sum_i Q_i \tag{4.3}$$

The second term represents the interaction between the charges on the metal and the external system holding the metal at the potential V_0. Equation (4.2) is satisfied when U_T is minimized with respect to the Q_i's since the energy minimization implies $\partial U_T / \partial Q_i = 0$. An isolated system can be treated similarly, but as it has no external forcing term ($V_0 = 0$). Additionally, the isolated system must maintain a constant total charge. Hence the constant charge problem requires a minimization of U_c subject to the constraint $\sum_i Q_i = Q_{const}$.

The *electrostatic* treatment of the model metal is then reduced to this minimization problem. Given point charges $\{q_i, \mathbf{r}_i\}$ and metal sites at $\{\mathbf{R}_i\}$, the induced charges on the metal sites can be found by minimizing Equation (4.3), or in the case of a floating electrode, by minimizing Equation (4.3) subject to a linear constraint that the total charge remain constant.

Electrodynamics could present a more challenging problem, but the Born-Oppenheimer approximation applies. Metallic charges should rearrange effectively instantaneously in response to the relatively slow atomic motion. As is standard in molecular dynamics simulations, the equations of motion must be discretized and numerically integrated.[1] [20] At every time step of the atomic motions, the charge dynamics is assumed to have already achieved a stationary distribution given by the electrostatic minimization. Coulombic forces on the atoms are provided by the gradient of U_T (Appendix C) using the energy minimizing value of the Q_i's.

---

[1] The details of a numerical integration of Newton's laws has been the subject of countless books, papers, and thesis chapters. For the purposes of this thesis, it is understood that these techniques are ubiquitous and they are therefore not described in any sort of detail.

4.2 The Quadratic Form

As shown, the metallic nature of the surface requires nothing more than a minimization of U_T to find the induced charges. While U_c appears daunting in Equation (3.39), it is greatly simplified by observing that the energy is a quadratic form in the Q_i's. Writing the Q_i's as the vector Q, U_T can be written as [2]

$$U_T(Q) = \frac{1}{2} Q^T A Q - b^T Q + c, \tag{4.4}$$

where

$$\begin{aligned}
A_{ij} &= \frac{1}{V} \sum_{\mathbf{k} \neq 0} \frac{4\pi}{k^2} e^{i\mathbf{k}\cdot\mathbf{R}_{ij} - k^2/4\alpha^2} + \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{R}_{ij}\right|\right) - \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}} \left|\mathbf{R}_{ij}\right|\right)}{\left|\mathbf{R}_{ij}\right|} + 2\delta_{ij}\left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}}\right) \\
&= \frac{8\pi}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{\cos\left(\mathbf{k} \cdot \mathbf{R}_{ij}\right)}{k^2} e^{-k^2/4\alpha^2} + \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{R}_{ij}\right|\right) - \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}} \left|\mathbf{R}_{ij}\right|\right)}{\left|\mathbf{R}_{ij}\right|} \\
&\qquad + 2\delta_{ij}\left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}}\right) \qquad (4.5) \\
b_i &= -\frac{1}{2V} \sum_{\mathbf{k} \neq 0} \frac{4\pi}{k^2} \left[e^{i\mathbf{k}\cdot\mathbf{R}_i} S_1(\mathbf{k})^* + e^{-i\mathbf{k}\cdot\mathbf{R}_i} S_1(\mathbf{k})\right] e^{-k^2/4\alpha^2} - \\
&\qquad \sum_j^n q_j \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{R}_i - \mathbf{r}_j\right|\right) - \operatorname{erfc}\left(\eta \left|\mathbf{R}_i - \mathbf{r}_j\right|\right)}{\left|\mathbf{R}_i - \mathbf{r}_j\right|} + V_0 \\
&= -\frac{4\pi}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{2 \operatorname{Re}\left[e^{i\mathbf{k}\cdot\mathbf{R}_i} S_1(\mathbf{k})^*\right]}{k^2} e^{-k^2/4\alpha^2} \\
&\qquad - \sum_j^n q_j \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{R}_i - \mathbf{r}_j\right|\right) - \operatorname{erfc}\left(\eta \left|\mathbf{R}_i - \mathbf{r}_j\right|\right)}{\left|\mathbf{R}_i - \mathbf{r}_j\right|} + V_0 \\
&= -\frac{8\pi}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{e^{-k^2/4\alpha^2}}{k^2} \left(\cos\left(\mathbf{k} \cdot \mathbf{R_i}\right) \operatorname{Re}\left[S_1(\mathbf{k})\right] + \sin\left(\mathbf{k} \cdot \mathbf{R_i}\right) \operatorname{Im}\left[S_1(\mathbf{k})\right]\right) - \\
&\qquad \sum_j^n q_j \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{R}_i - \mathbf{r}_j\right|\right) - \operatorname{erfc}\left(\eta \left|\mathbf{R}_i - \mathbf{r}_j\right|\right)}{\left|\mathbf{R}_i - \mathbf{r}_j\right|} + V_0 \qquad (4.6) \\
c &= \frac{4\pi}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{e^{-k^2/4\alpha^2}}{k^2} \left|S_1(\mathbf{k})\right|^2 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n q_i q_j \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} - \frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^n q_i^2 \qquad (4.7)
\end{aligned}$$

Provided A is positive definite, a single minimum exists, so we assume this to be the case for the remainder of this chapter. This assumption will be verified in the next

---

[2] So it is clear that Q is not in $\mathbb{R}^3$, it is not printed in boldface like the positions and wavevectors are. Q and b are vectors with length N, A is an $N \times N$ and c is a scalar.

chapter. Let us now proceed to the unconstrained minimization of U_T to solve the constant potential problem. That solution will then be tweaked slightly to perform the constrained minimization of the constant charge simulation.

4.3 Constant Potential Minimization

Several viable options exist for performing the minimization of U_T: matrix inversion, steepest descent minimization, a Car-Parrinello approach, and conjugate gradient minimization. For completeness, we survey all four methods, but conjugate gradient minimization should generally be the weapon of choice.

4.3.1 Matrix Inversion

With A positive definite, the unconstrained minimum of a quadratic form is easily shown to be at $Q = A^{-1}b$. [48] The simulations described in this thesis are performed with a modest number of metal sites (typically < 1000). Inversion of a 1000×1000 real, symmetric matrix is exceptionally fast, so on first glance it appears that $Q = A^{-1}b$ can be rapidly and trivially computed by explicit matrix inversion. The catch is that each matrix element of A requires a fairly expensive k-space sum. Since these sums are limiting, calculating A scales as $\mathcal{O}(N^2)$ in the number of metal sites. The gradient of Equation (4.3) is given by

$$\frac{\partial U_T}{\partial Q_i} = AQ - b, \tag{4.8}$$

As shown in Appendix C, this gradient can be calculated with $\mathcal{O}(N)$ k-space sums by computing the structure factor then squaring its modulus. Amazingly, this means that given some Q, the product AQ can be computed much faster than all of the matrix elements of A could be found. We will show that this property usually make iterative minimization schemes more appealing than the explicit matrix calculation.

However, if the metal site positions are a constant of the simulation (rigid electrode), A will also be a constant of the simulation. In such a situation, one could perform a single $\mathcal{O}(N^2)$ calculation at the beginning of the simulation and store the inverted matrix A^{-1}. Thereafter each exact minimization would require only the $\mathcal{O}(N)$ calculation of b at every time step. Furthermore, many model electrodes will have a high degree of symmetry, leading many of the $\mathbf{r}_{ij}$ vectors, and therefore many of A's k-space sums to be identical. This means sufficient symmetry would allow A to be cleverly computed cheaper than $\mathcal{O}(N^2)$. Even the best iterative methods will require roughly ten of the $\mathcal{O}(N)$ AQ products per time step, so the matrix treatment is faster when a rigid electrode will be simulated for more than $N/10$ dynamical time steps.

A subtle caveat is that A depends on the cell box through the k-space sums. Even if the $\mathbf{R}$'s are fixed, a constant pressure simulation would cause A to vary across time steps. These effects will likely be negligible, but one should investigate more closely before blindly applying the matrix inversion scheme to a constant pressure simulation.

4.3.2 Steepest Descent

As alluded to previously, iterative minimization methods can be advantageous because they make use of the squared structure factor to compute ∇U_T with $\mathcal{O}(N)$ k-space sums. The simplest such scheme is the method of steepest descent. Starting with some seed value of Q, we move "downhill" along the steepest possible path, $-\nabla U_T$. The optimal step in this direction is chosen by minimizing the energy along this search direction, a so-called line search. This procedure is repeated until the norm of the gradient drops below some (very small) threshold. If nothing were known about the functional form of the energy then the line search could be performed by a minimum bracketing algorithm such as Brent's Method. [39]

Because U_T is known to be quadratic, the line search can be performed much more quickly without ever re-evaluating the energy along the search direction. The idea can be motivated simply. The critical observation is that given a quadratic form, the energy projected onto any search direction will be a parabola. Let us denote this projected energy by

$$f(x) = \frac{1}{2}ax^2 + bx + c.$$

We are given an initial seed point, x_0 and the directional derivative at that point, $f'(x_0)$. Since the parabola's slope is a linear equation, knowledge of x_0, $f'(x_0)$, and the curvature, a, is sufficient to determine how far from the minimum x_0 lies. Specifically,

$$x_{min} = x_0 - \frac{f'(x_0)}{a}.$$

There is a small jump to compute a (the one dimensional curvature along the search direction) from A (the curvature matrix). Readers interested in those details are directed to Shewchuk's excellent paper. [48] Even with these optimal step sizes, steepest descent convergence can be slow. The difficulty arises from the fact that a move along the direction of steepest descent can result in oscillations within a steep trough. This phenomena, depicted in Figure 4.1, is the reason steepest descent minimization will be avoided in practice.

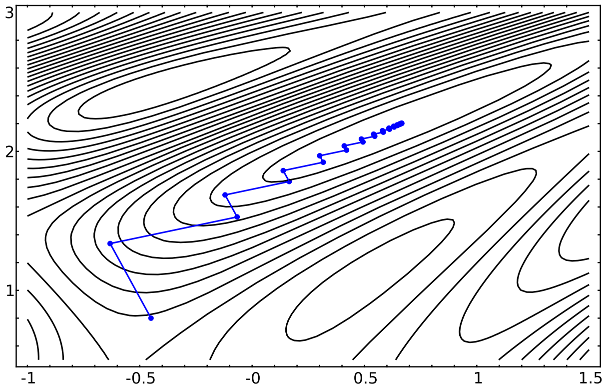


Figure 4.1: Steepest descent trajectory on 2d surface. Note the oscillations along the trough.

4.3.3 Car-Parrinello Method

Steepest descent minimization is disfavored because it converges so slowly. Updating Q along its direction of steepest descent can be resuscitated by using a Car-Parrinello (CP) approach. To simplify our dynamics, we previously invoked the Born-Oppenheimer approximation. Now we actually want to take a step backwards and reconsider the charge dynamics to be on equal footing with the atomic dynamics. This is not to say the Born-Oppenheimer approximation is invalid. Rather we want to generate some fictitious (and physically incorrect) charge dynamics that at every time step lies "close enough" to the true U_T-minimizing value of Q. Though the value of Q at every time step contains error, it is possible for these errors to be effectively unbiased such that the statistics of the $\mathbf{r}$ trajectory is unaffected. We perform this magic by extending the Lagrangian to include a fictitious kinetic energy term, $\frac{1}{2}\sum_i M_i \dot{Q}_i^2$. M_i is referred to as a fictitious mass of the charge magnitudes, but this is nothing more than an aphysical parameter to control rates of convergence. The Lagrange multiplier methods of Appendix D give rise to the equations of motion

$$\begin{cases} m_i \ddot{\mathbf{r}}_i = -\frac{\partial U_c}{\partial \mathbf{r}_i} - \sum_m \lambda_m(t) \frac{\partial g_m}{\partial \mathbf{r}_i} \\ M_i \ddot{Q}_i = -\frac{\partial U_c}{\partial Q_i} + V_0 - \sum_m \lambda_m(t) \frac{\partial g_m}{\partial Q_i} \\ g_m = \frac{\partial U}{\partial Q_m} - V_0 = 0 \end{cases} \tag{4.9}$$

where the constant potential constraint is imposed by the constraining forces λ_m and the $\mathbf{R}$'s and q's have been assumed stationary. Now suppose the numerical integration of these differential equations is performed with a very small time step. If the system begins with the constant potential constraints satisfied then after a very small time step the constraints must remain very nearly satisfied. This means that for a sufficiently small time step, the λ's are very small relative to the $\partial U/\partial \mathbf{r}_i$ and $\partial U/\partial Q_i$ terms. We therefore choose to neglect the constraint forces entirely by setting $\lambda = 0$.

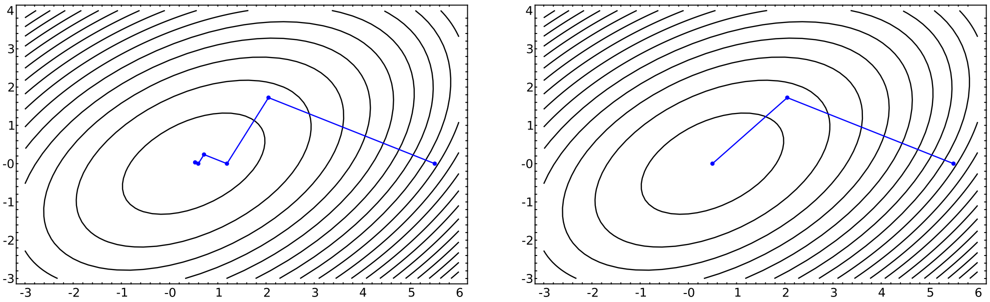
$$\begin{cases} m_i \ddot{\mathbf{r}}_i = -\frac{\partial U_c}{\partial \mathbf{r}_i} \\ M_i \ddot{Q}_i = -\frac{\partial U_c}{\partial Q_i} + V_0 \end{cases} \tag{4.10}$$

In one time step the error is negligible, but the error propagates and one would expect that after many such time steps the constraint will no longer be satisfied. After all, we have just removed the very forces that impose the constraint. The saving grace is that when $\lambda = 0$ in equations (4.9), a force is applied to Q along the direction of steepest descent. If the charges start in an already minimized position (i.e. start on the adiabatic surface) and the time steps are sufficiently small, then several steepest descent steps can be taken before the ions move appreciably. Because ions will move slowly relative to the Q_i dynamics, the ions will effectively experience a force that is averaged out over the past several Q_i configurations. While the constant potential constraint is not rigorously upheld, the zig-zagging nature of steepest descent steps means the system often overshoots the minimum in opposing directions. The consequence is that though the equipotential constraint is not rigorously enforced, $\left\langle \frac{\partial U_c}{\partial Q_i} - V_0 \right\rangle = 0$, where $\langle \cdots \rangle$ denotes a temporal averaging over several time steps during which the ions do not move appreciably. It is for this reason that the statistics of the $\mathbf{r}$ trajectories are not biased. The success of the CP method hinges upon using a sufficiently short time step and an effective mass, M_i, that allows the charges to react quickly to changes in the ion positions. Ensuring that there is no systematic drift away form the constrained surface is a crucial and delicate endeavor. These concerns are addressed in detail in a review by Remler and Madden. [43]

The method can be improved still further by thermostating the Q system to a low temperature with a Nosé-Hoover bath. The thermostat strongly forces the charges into their energetic minimum to ensure adiabaticity. Critically, this can be done independently from the thermostating of the ionic degrees of freedom such that the charges are held at low temperature even when the atoms are held at a high temperature. The thermostating techniques are discussed at length in Frenkel and Smit. [20] For still more detail, the work of Tuckerman *et al.* provides a mathematically satisfying explanation of non-Hamiltonian dynamics as a means of sampling the NVT and NPT ensembles. [62, 63]

4.3.4 Conjugate Gradient Method

We saw that directly computing A allows for the exact minimization by matrix inversion. We also observed that iterative schemes exist and are capable of converging to a minimum more cheaply than A's matrix elements can be calculated. Unfortunately we have no guarantee that the steepest descent trajectory will converge more quickly than an explicit matrix inversion could have done the job. The power and beauty of the conjugate gradient (CG) approach is that it is an iterative and exhaustive search. Like steepest descent minimization, CG can be terminated when it falls within a threshold of the minimum, but unlike steepest descent, CG exactly finds the minimum in at most N steps.[3] Each CG step requires an $\mathcal{O}(N)$ gradient calculation, so convergence is guaranteed in $\mathcal{O}(N^2)$. This is formally the same cost as the matrix approach, but CG is actually much cheaper as it can be stopped once it is "close enough." In fact, it is typical that CG only requires a handful of steps before sufficient convergence is achieved. This is a general observation for CG, but it is especially true for our application because Q for each time step is seeded with Q_{min} from the previous time step



(a) Steepest Descent

(b) Conjugate Gradient

Figure 4.2: Comparison of steepest descent and conjugate gradient minimizations in two dimensions. The difference between rates of convergence becomes even more pronounced as the dimensionality increases.

The method is outlined here, but a few of the details are not fully proven. These details are simple to show, and they are all included in Shewchuk's complete, accessible explanation of the CG method. [48]

---

[3] The minimum is rigorously found only for exact arithmetic. In practice round-off error can prevent the precise minimum from being found. This round-off problem is generally a concern only for especially high dimensional systems.

Assume we seed Q with the value $Q_{(0)}$, but the true minimum lies at Q_{min}.[4] Since $Q \in \mathbb{R}^N$ it must be possible to move from $Q_{(0)}$ to Q_{min} in a series of N steps along any arbitrary basis vectors of the space. We will treat the basis vectors as a sequence of N search directions, and we need a way to compute the optimal step size in each of these directions. In general these basis vectors are hopelessly coupled together. We cannot optimize the step size along one basis vector when we don't already know the optimal steps along all of the other directions. We state without proof that the problem is decoupled when a set of A-orthogonal basis vectors, $\left\{d_{(i)}\right\}$, are used as the search directions.[5] By A-orthogonal it is meant that

$$d_{(i)}^T A d_{(j)} = 0 \quad \text{if } i \neq j \tag{4.11}$$

In Equations (31) and (32) of his paper, Shewchuk shows that optimal steps take the form

$$Q_{(i+1)} = Q_{(i)} + \alpha_{(i)} d_{(i)} \tag{4.12}$$

$$\alpha_{(i)} = \frac{d_{(i)}^T A e_{(i)}}{d_{(i)}^T A d_{(i)}} \tag{4.13}$$

where $e_{(i)} = Q_{(i)} - Q_{min}$ is the error in the i^{th} step. This value of $e_{(i)}$ is unknown.[6] However, $Ae_{(i)}$ can be known since this is the gradient at point $Q_{(i)}$, which we can compute from Appendix C.

$$Ae_{(i)} = AQ_{(i)} - AQ_{min} = \left(\nabla U\left(Q_{(i)}\right) + b\right) - \left(\nabla U\left(Q_{min}\right) + b\right) = \nabla U\left(Q_{(i)}\right) \tag{4.14}$$

The A-orthogonal basis is used precisely for this reason. It allows the optimal step sizes to be related to the gradient evaluated at the current point. Given the A-orthogonal set of search directions, N steps of the form of Equation (4.12) can be taken to complete the minimization. This is true because the A-orthogonality means a step along one search direction never "undoes" the optimality of the previous step. The final challenge is to generate such a set of search directions.

In the same way that a Gram-Schmidt procedure generates a set of orthogonal basis vectors from an arbitrary basis, a conjugate Gram-Schmidt procedure can generate a set of A-orthogonal basis vectors, $\left\{d_{(i)}\right\}$ from an arbitrary basis, $\left\{u_{(i)}\right\}$.

$$d_{(i)} = u_{(i)} - \sum_{j=1}^{i-1} \frac{u_{(i)}^T A d_{(j)}}{d_{(j)}^T A d_{(j)}} d_{(j)}, \tag{4.15}$$

---

[4] The value of the vector Q will be updated iteratively and we denote the i^{th} iteration $Q_{(i)}$. This should not be confused with Q_i, which is the charge on the i^{th} metal site.

[5] This is shown in (34) and (35) of Ref. [48].

[6] If it were known, the problem would be solved since $Q_{min} = Q_{(i)} - e_{(i)}$.

where the second term just subtracts out components in the direction of the previous $d_{(j)}$'s. While this procedure works, it is expensive when performed on an arbitrary basis. Each new vector must remain A-orthogonal (conjugate) to all of the preceding vectors, so the conjugate Gram-Schmidt process requires all previously generated vectors to be remembered. The CG approach applies conjugate Gram-Schmidt to a very special basis, the basis formed by the successive gradient vectors, $\{r_1, r_2, \ldots, r_N\}$.[7] Because the basis is formed by the sequence of gradients, there is a recurrence relation.

$$r_{(i+1)} = Ae_{(i+1)} = A\left(e_{(i)} + \alpha_{(i)}d_{(i)}\right) = Ae_{(i)} + \alpha_{(i)}Ad_{(i)} = r_{(i)} + \alpha_{(i)}Ad_{(i)} \qquad (4.16)$$

Now recall that our A-orthogonal directions are in a sense decoupled. Optimizing a given direction does not undo the prior optimizations. This property implies that the $(i+1)^{\text{th}}$ gradient, $r_{(i+1)}$, is orthogonal to the prior search space.[8] As $r_{(i)}$ lies in this prior search space, $r^T_{(i+1)}r_{(i)} = 0$. But the recurrence relation of Equation (4.16) allows this to be rewritten as

$$r^T_{(i+1)}\left(r_{(i-1)} + \alpha_{(i-1)}Ad_{(i-1)}\right) = 0 \qquad (4.17)$$

For the same reason that $r^T_{(i+1)}r_{(i)} = 0$, $r^T_{(i+1)}r_{(i-1)} = 0$, giving

$$\alpha_{(i-1)}r^T_{(r+1)}Ad_{(i-1)} = 0 \qquad (4.18)$$

Hence $r_{(i+1)}$ is automatically A-conjugate to $d_{(i-1)}$. The recurrence relation could have just as easily been used a second time in Equation (4.17).

$$r^T_{(i+1)}\left(r_{(i-2)} + \alpha_{(i-1)}Ad_{(i-1)} + \alpha_{(i-2)}Ad_{(i-2)}\right) = 0$$

$$\Rightarrow \alpha_{(i-1)}r^T_{(i+1)}Ad_{(i-1)} + \alpha_{(i-2)}Ad_{(i-2)} = 0$$

$$\Rightarrow \alpha_{(i-2)}Ad_{(i-2)} = 0,$$

where (4.18) was used for the last step. This procedure can be repeated to demonstrate that $r_{(i+1)}$ is A-orthogonal to the entire set $\left\{d_{(1)}, \ldots, d_{(i-1)}\right\}$. Symbolically,

$$r^T_{(i+1)}Ad_{(j)} = 0 \;\; \forall\, j \leq i-1 \qquad (4.19)$$

---

[7] These gradient vectors are commonly known as the residuals. As shown by (4.14), the gradient is equivalent to $Ae_{(i)}$, which is the residual for the closely associated problem $Ax = b$. That is to say, $r_{(i)} = Ae_{(i)} = Ax_{(i)} - b$.

[8] If the gradient contained any components which were not orthogonal it would imply that the prior search space had not been optimized

This set of relations greatly simplifies the conjugate Gram-Schmidt procedure. Returning to Equation (4.15) and inserting the set $\{r_{(i)}\}$ for the arbitrary basis $\{u_{(i)}\}$ gives

$$d_{(i)} = r_{(i)} - \sum_{j=1}^{i-1} \frac{r_{(i)}^T A d_{(j)}}{d_{(j)}^T A d_{(j)}} d_{(j)} \tag{4.20}$$

Equation (4.19) demonstrates that the numerator within the sum vanishes unless $j = i - 1$, leaving

$$d_{(i)} = r_{(i)} - \frac{r_{(i)}^T A d_{(i-1)}}{d_{(i-1)}^T A d_{(i-1)}} d_{(i-1)} \tag{4.21}$$

A couple of clever algebraic tricks allows this to be further reduced.[9]

$$d_{(i)} = r_{(i)} - \frac{r_{(i)}^T r_{(i)}}{r_{(i+1)}^T r_{(i+1)}} d_{(i-1)} \tag{4.22}$$

The important point is that the new search direction, $d_{(i)}$ is given by a linear combination of the gradient and the previous search direction. No other prior directions are required, so the generation of the A-orthogonal search directions is fast and easy. This completes the algorithm. Each new search direction is generated from Equation (4.22) and a step is taken along that direction as determined by Equation (4.13). Pseudocode for the entire procedure is given in Appendix E

4.4 Constant Charge Constraint

It was shown that an isolated constant charge simulation corresponds to the minimization of U_c (equivalently U_T with V_0 set to zero) subject to the constraint $\sum_i Q_i = Q_{const}$. We adapt each of the four minimization methods to this constrained problem. As before, the conjugate gradient approach is generally preferred.

4.4.1 Matrix Inversion with Constant Charge

The constant charge restriction just adds a single linear constraint to the quadratic form. Explicitly, we write the constant charge constraint as $v^T Q = Q_{const}$, where $v^T = (1, 1, \ldots, 1)$ and Q_{const} is the value of the constant charge.[10] Appendix D reviews how Lagrange multipliers allow one to convert the constrained extremization

---

[9] Left multiplication of Equation (4.16) by $r_{(i+1)}^T$ and shifting of the index shows $r_{(i)}^T A d_{(i-1)} = \frac{1}{\alpha_{(i-1)}} \left(r_{(i)}^T r_{(i)} \right)$. Inserting $\alpha_{(i-1)}$ from Equation (4.13) and making use of another orthogonality relation completes the transformation.

[10] The notation here is less than ideal because $Q \in \mathbb{R}^N$ and $Q_{const} \in \mathbb{R}$

into an equivalent unconstrained problem. Application of Equation (D.1) reveals that our constrained minimum is the unconstrained minimum of

$$U(Q,\lambda) = \frac{1}{2}Q^T A Q - b^T Q + c + \lambda\left(v^T Q - Q_{const}\right) \tag{4.23}$$

The critical point is given by $Q = A^{-1}(b + \lambda v)$. To solve for λ left multiply this relation by v^T, recalling that $v^T Q = Q_{const}$ (our constant charge constraint). Hence $Q_{const} = v^T A^{-1} b + \lambda v^T A^{-1} v$. Rearrangement yields

$$\lambda = \frac{Q_{const} - v^T A^{-1} b}{v^T A^{-1} v}.$$

So, the Q which minimizes U_T subject to the constant charge constraint is given by

$$Q = A^{-1}\left(b + \frac{Q_{const} - v^T A^{-1} b}{v^T A^{-1} v} v\right) \tag{4.24}$$

4.4.2 Constant Charge Steepest Descent

As with the steepest descent algorithm, the constrained steepest descent iteratively updates $Q^{(i)}$ by adding a step of size α in the search direction d.[11] This d is $-\nabla U_T$ in the unconstrained problem, but in the constrained problem the search direction must be adapted to satisfy the constraint. To determine the proper adaptation, we consider the equivalent unconstrained problem of Equation (4.23). In the extended (Q,λ) space a steepest descent algorithm would move along the search direction

$$d_i = -\frac{\partial U(Q,\lambda)}{\partial Q_i} = -\frac{\partial U_c}{\partial Q_i} - \lambda v. \tag{4.25}$$

To solve for the search direction, we must eliminate λ by making use of the constant charge constraint. Suppose the minimization is initialized at some $Q^{(0)}$, the value of Q at the next minimization step is

$$Q^{(1)} = Q^{(0)} + \alpha d.$$

From the constant charge constraint, $\sum_i Q_i^{(0)} = \sum_i Q_i^{(1)}$, requiring that $\alpha \sum_{i=1}^N d_i = 0$. Inserting (4.25) yields

$$d_i = -\frac{\partial U_c}{\partial Q_i} + \frac{1}{N}\sum_{i=1}^{N} \frac{\partial U_c}{\partial Q_i} \tag{4.26}$$

A line search can be performed as in the unconstrained problem to solve for the optimal step size α.

---

[11] As one final notational point, this $Q^{(i)}$ is the same as what was called $Q_{(i)}$ before. The switch in notation is to accommodate the expression $Q_i^{(0)}$ here whereas before we wanted to accommodate $Q_{(i)}^T$

4.4.3 Constant Charge Car-Parrinello

The Lagrange multiplier $\mu(t)$ can be introduced in order to enforce the constant charge constraint, $h = \sum_i Q_i = Q_{const}$. The effective Lagrangian gains a $\mu(t)h$ term, so the new equations of motion are

$$\begin{cases} m_i \ddot{\mathbf{r}}_i = -\frac{\partial U_c}{\partial \mathbf{r}_i} - \mu(t)\frac{\partial h}{\partial \mathbf{r}_i} \\ M_i \ddot{Q}_i = -\frac{\partial U_c}{\partial Q_i} - \mu(t)\frac{\partial h}{\partial Q_i} \\ h = \sum_i Q_i = Q_{const} \end{cases} \tag{4.27}$$

Since h does not depend on spacial coordinates, $\frac{\partial h}{\partial \mathbf{r}_i} = 0$. Additionally $\frac{\partial h}{\partial Q_i} = 1$.

$$\begin{cases} m_i \ddot{\mathbf{r}}_i = -\frac{\partial U_c}{\partial \mathbf{r}_i} \\ M_i \ddot{Q}_i = -\frac{\partial U_c}{\partial Q_i} - \mu(t) \\ h = \sum_i Q_i = Q_{const} \end{cases} \tag{4.28}$$

These equations of motion still contain $\mu(t)$, the force on a charge required to maintain the constant charge constraint. Critically the $\mu(t)$ that is added to the force on each charge magnitude Q_i is independent of i. For a discretized numerical integration of the equations of motion, the constraint forces can be solved fairly easily. Letting $-\frac{\partial U_c}{\partial Q_i} = F_{Q_i}$,

$$\ddot{h} = \sum_i \ddot{Q}_i = \sum_i \left(\frac{F_{Q_i}}{M_i} - \frac{\mu(t)}{M_i} \right)$$

Since $h = Q_{const}$, $\ddot{h} = 0$ and

$$\mu(t) = \frac{\sum_i \left(\frac{F_{Q_i}}{M_i} \right)}{\sum_i \left(\frac{1}{M_i} \right)}, \tag{4.29}$$

where M_i is the ficitious mass of the i^{th} induced charge. If one reasonably sets all fictitious masses equal, this expression can be further simplified to

$$\mu(t) = \frac{\sum_i F_{Q_i}}{N}. \tag{4.30}$$

Hence the CP constant charge equations of motion are

$$\begin{cases} m_i \ddot{\mathbf{r}}_i = -\frac{\partial U_c}{\partial \mathbf{r}_i} \\ M_i \ddot{Q}_i = -\frac{\partial U_c}{\partial Q_i} + \frac{1}{N}\sum_j \left(\frac{\partial U_c}{\partial Q_j} \right) \end{cases} \tag{4.31}$$

4.4.4 Constant Charge with Conjugate Gradient Minimization

The constant charge steepest descent minimization did not take steps in the direction of the gradient. The presence of the constraint meant that the search direction was actually the gradient projected onto the constrained surface. This prevents the algorithm from exploring the part of the space that corresponds to a changing total charge. The same projection of the gradient onto the feasible surface is needed for conjugate gradient minimization with a linear constraint. The required algorithm is proven by Shariff. [47] We adapt his results to the constant charge constraint.

Observe that the charge vector can be expressed as

$$\begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_N \end{pmatrix} = \frac{1}{N} \begin{pmatrix} N-1 & -1 & \dots & -1 \\ -1 & N-1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & N-1 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_N \end{pmatrix} + \frac{1}{N} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_N \end{pmatrix} \tag{4.32}$$

The matrices are not full rank. The second matrix projects out the fully symmetric component corresponding to changes in the total charge, so the first matrix must be the one that projects onto the feasible search space. In Shariff's notation, this is H. In the unconstrained CG minimization the gradients, $\{r_{(i)}\}$ are conjugate Gram-Schmidt-ed to generate the search directions. In the constrained problem we use the projection of these gradients onto the feasible search space, $\{Hr_{(i)}\}$. This leads to a slightly different calculation of the optimal step size $\alpha_{(i)}$, and a slightly different linear combination of $Hr_{(i)}$ and $d_{(i-1)}$ is needed to compute the next search direction $d_{(i)}$. These differences are shown in Shariff's paper and can be found in the pseudocode of Appendix E.

Chapter 5

Selection of the Parameter η

The metal model outlined in Section 4.1 introduces a single physical parameter, η, to control the size of the Gaussian metal sites. The other parameter, the Ewald screening parameter α, adjusts the partition between real and reciprocal space, influencing the speed of convergence of these two sums. Provided both sums are converged, the energy is independent of α, so the parameter is numerical and not physical. In contrast, η is not a mere numerical tool. Its selection represents a physical statement about how tightly the nuclei bind charge density. The metallic equipotential constraint allows Gaussians to exchange charge with each other, but the model forces the local charge distribution to be Gaussian around all sites. In essence, the model is a rudimentary classical density functional theory with a highly restrictive basis of a single Gaussian per metal site. Because the basis set is so small (relative to normal DFT basis sets), the accuracy of the model will depend significantly on the choice of an appropriate η.

In this chapter we address the selection of η in three parts. First, we explore how the choice of η affects the convergence properties of the numerical minimization. It was observed in Section 4.2 that the minimization of the quadratic form requires the matrix A from Equation (4.4) to be positive definite, and we discuss how this requirement influences the choice of η. Next we discuss the impact of η on the agreement with continuum electrostatic descriptions of metals. By studying the textbook example of a conducting sphere, we show that the continuum results will be given in the macroscopic limit irrespective of the choice of η, thereby challenging the method of Siepmann and Sprik's original parameterization of the model. [49] Instead, we parameterize the model based on microscopic considerations. Indeed, since the choice of η affects only short-range forces it is only natural that the fitting of η depend upon the local charge distribution on the metal.

5.1 Positive Definiteness of A

In the absence of a clear method for analytically determining the conditions on η that make A (as given by Equation (4.5)) positive definite, we numerically compute A for a square grid of metal sites. The 100 site grid was constructed with 0.5 au lattice spacing and the calculations were performed with $\alpha = 0.25$ au^{-1} in a cubic periodic box of length 50 au. The k-space sums were truncated at $k_{max} = 8$ such that only terms with $k^2 \leq 64$ were retained. The value of the matrix A is independent of metal charge magnitudes and of the entire ionic system. It is just a function of the metal geometry that expresses the coupling between metal sites. These couplings in turn determine the stability of the minimization.

Positive definite matrices have strictly positive eigenvalues, so it is sufficient to examine the behavior of A's minimum eigenvalue. Existence of both positive and negative eigenvalues implies that the space has a saddle point, so minimization techniques will be unable to identify a global minimum. If, however, the matrix is positive definite, the techniques outlined in Chapter 4 will all apply. The minimum eigenvalue of the 10×10 grid's A matrix is plotted in Figure 5.1. Several critical observations can be made. First, the matrix is not positive definite when η drops below a critical value of approximately $\sqrt{2}\alpha$. The significance of this value was discussed after Equation (3.33). When η is less than this critical value, the metal site Gaussians are so large that it no longer makes sense to shift any of the calculation to real space. Instead, an alternative form given by Equation (3.33) should be used. Secondly, the existence of small eigenvalues means there is a degree of freedom which converges very slowly to the minimum. Figure 5.1 makes it clear that increasing the value of η will result in more rapid convergence of the worst case scenario. When the minimum eigenvalue is especially small (but still positive), numerical error could prevent the minimization, so for numerical purposes, it is generally best that η be "sufficiently large," where unfortunately sufficiently large depends on the geometry of the metal sites.

It is, of course, desirable to have an analytic way of thinking about the nature of Figure 5.1. Using very weak bounds, one can understand that there must exist some critical value of η below which the minimum eigenvalue is negative and above which it is guaranteed to be positive. To show this, we make use of the fact that a necessary (but not sufficient!) condition for A to be positive definite is that $A_{ii} > 0$. Looking back at Equation (4.5), it is clear that the diagonal elements of A are independent of

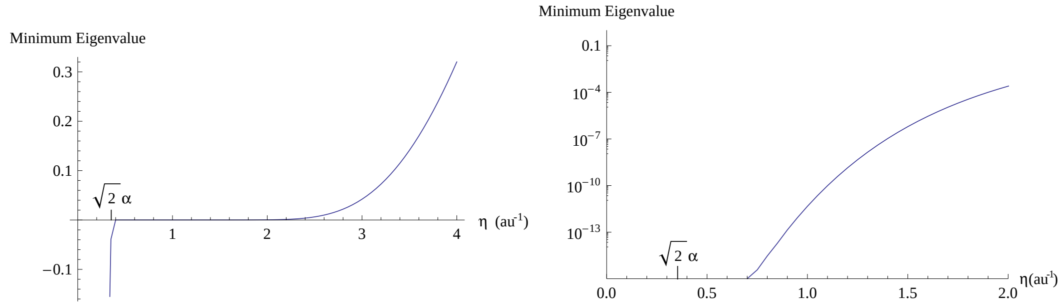


Figure 5.1: Minimum eigenvalue of the matrix A, whose matrix elements are given by Equation (4.5) for a 10×10 grid of metal sites as described in the text. When all eigenvalues are positive, the global minimum of the quadratic form is given by $Q = A^{-1}b$. The log plot (right) emphasizes that as η approaches $\sqrt{2}\alpha$ from above, the minimum eigenvalue remains positive but can be very small, indicating that complete convergence could be very slow.

the metal site positions and are given by

$$A_{ii} \approx 2\left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}}\right) + \frac{8\pi}{V} \sum_{k_1>0} \sum_{k_2,k_3} \frac{e^{-k^2/4\alpha^2}}{k^2} \tag{5.1}$$

The k-space sum can be approximated by converting the sum into an integral in the exact same way that the three dimensional free space density of states is computed.[1] Converting the sum to an integral introduces error, the most significant of which comes from the region near $|k| = 0$ and above $|k| = k_{max}$. Because the k-space sum is truncated and excludes the $k = 0$ term, we integrate from $|k| = (3/(4\pi))^{1/3}$ to $|k| = k_{max}$ to give

$$\begin{aligned} A_{ii} &= 2\left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}}\right) + \frac{2\alpha}{\sqrt{\pi}}\left[\operatorname{erfc}\left(\frac{3^{1/3}\pi^{2/3}}{4^{1/3}\alpha L}\right) - \operatorname{erfc}\left(\frac{\pi k_{max}}{\alpha L}\right)\right] \\ &= \frac{2\eta}{\sqrt{2\pi}} - \frac{2\alpha}{\sqrt{\pi}}\left[\operatorname{erf}\left(\frac{3^{1/3}\pi^{2/3}}{4^{1/3}\alpha L}\right) + \operatorname{erfc}\left(\frac{\pi k_{max}}{\alpha L}\right)\right] \end{aligned} \tag{5.2}$$

So as an exceptionally weak bound, $A_{ii} < 0$ and therefore A is guaranteed to *not* be positive definite when

$$\eta \le \sqrt{2}\alpha\left[\operatorname{erf}\left(\frac{3^{1/3}\pi^{2/3}}{4^{1/3}\alpha L}\right) + \operatorname{erfc}\left(\frac{\pi k_{max}}{\alpha L}\right)\right]. \tag{5.3}$$

Inserting the parameters that generate Figure 5.1 into this bound guarantees at least one negative eigenvalue for $\eta \le 0.0633$. Clearly this is a weak bound, but the point

---

[1] See any good solid state physics textbook

is that there must be a value of η below which the minimum eigenvalue is negative, so the behavior of Figure 5.1 is not particularly concerning.

It is also worth noting that the diagonal elements can be made arbitrarily large while not increasing the off-diagonal matrix elements by simply increasing η. A is necessarily positive definite if the diagonal exceeds the sum of the absolute value of the other matrix elements in the row. This tells us that A will necessarily be positive definite when η exceeds some threshold. These bounds shed some light on the general structure of Figure 5.1, but they are not particularly useful in evaluating whether or not a particular value of η will be sufficiently large to provide a stable minimum. Relying on the numerical results for the planar metal and also for a spherical metal discussed later, the location of the critical η at which the minimum eigenvalue equals zero is quite robust with respect to metal geometry. Typical simulations will use α sufficiently small that η comfortably exceeds $2\sqrt{\alpha}$, so we will be content assuming A to be positive definite.

5.2 Comparison with Continuum Electrostatics

For the metal model to be useful, it must converge to continuum electrostatic theory at long range. The convergence arises naturally from the construction of the discrete metal surface since inter-site interactions are built upon the Coulomb r^{-1} potential. Continuum electrostatics can be derived from this potential and some vector calculus in the limit that the discretization of the metal surface becomes infinite and smooth. Our model does not take this limit, but at long distances it should rapidly converge, acting as though it were a continuous metallic surface. To confirm that the macroscopic theory is properly returned, we focus on the textbook example of a conducting sphere.

5.2.1 The Image Charge Problem

The characteristic of a metal surface is that it maintains a constant electric potential. This is because the metal allows instantaneous redistribution of charge. Any nonzero gradient of the potential would force charge to redistribute until a stationary charge distribution is maintained. When the electronic motion has stopped, the charge distribution is stabilized to be whatever distribution that forces the metal surface to be equipotential. When a conductor is present in an electrostatics problem, interactions with the conductor's stationary surface charge density must be added to the ordinary interactions between the fixed charges of the problem. For a small class of problems there exists a particularly elegant shortcut to determining the effects of these induced surface charges - the method of image charges. We utilize this approach to obtain simple analytic results to which our metal model can be compared.

To understand the method of images we briefly review elementary electrostatic theory. The electric potential is merely given by the solution to Poisson's equation (in cgs units),

$$\nabla^2 \phi = 4\pi\rho.$$

Specification of the fixed charges, ρ, exactly determines ϕ as the solution of the partial differential equation provided the boundary conditions are specified. Electrostatics in infinite free space is straightforward because of the trivial boundary conditions: ϕ is a constant at infinity. Introduction of a metal surface alters the boundary conditions, complicating the solution of Poisson's equation. In general this can present computational difficulties, but potential theory offers a shortcut by guaranteeing that the solution of Poisson's equation is unique. Any charge distribution that satisfies the boundary conditions will give rise to the correct potential, even if the charge distribution was obtained by nonphysical trickery.

As the simplest possible example, consider a point charge at position $x = d$ near a planar metal surface ($x = 0$). Rather than explicitly solving for the surface charge distribution on the metal, we note that the plane $x = 0$ is necessarily equipotential if we introduce an equal and opposite charge at $x = -d$. This extra charge is

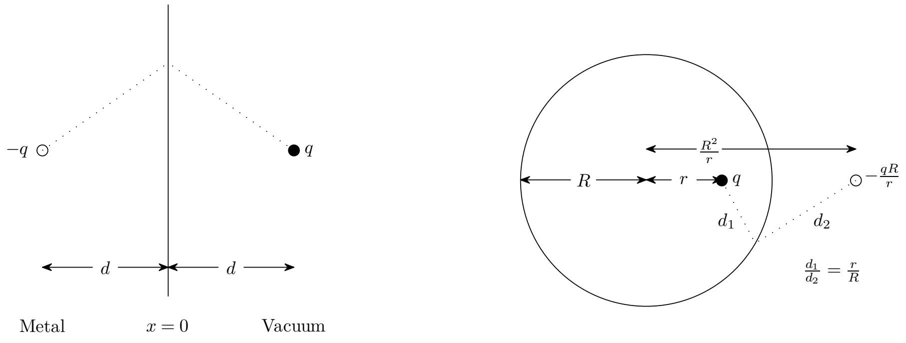


Figure 5.2: Diagram of image charge problems. The $+q$ charge induces surface charge at the metal interface. This surface charge collectively exerts a force exactly equal to the force that would be exerted by the image charge that is shown. This is guaranteed by a uniqueness theorem since symmetry requires that the potential vanishes exactly for every point on the $x = 0$ surface (left) or on the sphere (right).

referred to as the image charge, and by uniqueness, the $x \geq 0$ electric potential of the point charge/image charge system is identical to that given by the original point charge/metal problem. The point charge and image charge are particularly easy to work with since the attractive force is given by the r^{-2} force law. The energy is similarly found from the point charge/image charge system, though there is an extra factor of 1/2 that must be included since the electric fields are not supported in the metal half of the space ($x < 0$). [59]

The image charge method is particularly serendipitous in the case of a planar metal because the symmetry allows the surface charge effects to be described by the inclusion of a single ficitious point charge. This is actually a special case of a point charge in a conductive sphere (sphere radius tending to ∞), which can also be treated with the addition of a single image charge. Consider a point charge with magnitude q positioned at radius r inside a conducting sphere of radius R. It can be shown that the potential everywhere on the sphere will exactly vanish if a point charge with magnitude $-qR/r$ is placed a distance R^2/r from the center of the sphere along the same ray that extends from the sphere's center to the original point charge. [30] Given this observation, the radial force acting on the inner point charge due to the conducting sphere can easily be computed as the force between the point charge and the image charge.

$$F_{\hat{r}} = -\frac{q^2 R r}{\left(R^2 - r^2\right)^2} \tag{5.4}$$

Remembering to include the extra factor of 1/2, the energy can be found in equally effortless fashion.

$$U = -\frac{q^2 R}{2\left(R^2 - r^2\right)} \tag{5.5}$$

These analytic results provide the ideal comparison for the model metal surface.

5.2.2 Calculations on a Conducting Sphere

A grounded conducting sphere of radius 7.94 Å was modeled with 2000 Gaussian charges centered along a golden section spiral so as to achieve approximately equal spacing between sites. [45] A single point charge with magnitude e was placed inside the sphere, and the conjugate gradient method was used to minimize the interaction energy given by Equation (3.39).[2] Periodic images were positioned far away from

---

[2] In the continuum problem the total induced charge is $-q$, so constrained conjugate gradient minimization was also performed to restrict the induced charge to be $-q$. As expected, no appreciable difference was observed between the constrained and unconstrained minimization since the constant charge calculation was constrained to the "correct" value as predicted by continuum theory.

each other by using a large cubic box with 42.33 Å side lengths, and the Ewald screening parameter was set to $\alpha = 0.132$ Å^{-1}. The interaction energy and radial force experienced by the point charge were computed for several values of η over the range of point charge positions. Figure 5.3 shows the induced charge distribution for $\eta = 4.7$ Å^{-1} with the point charge placed 3.97 Å from the center of the sphere. As made clear from Figure 5.4, the energies and forces rapidly converge to the analytic solution at long range with little sensitivity to η. At short range, the choice of η exerts considerable influence because it constrains the induced charge density to be bound more or less tightly to the atomic sites.

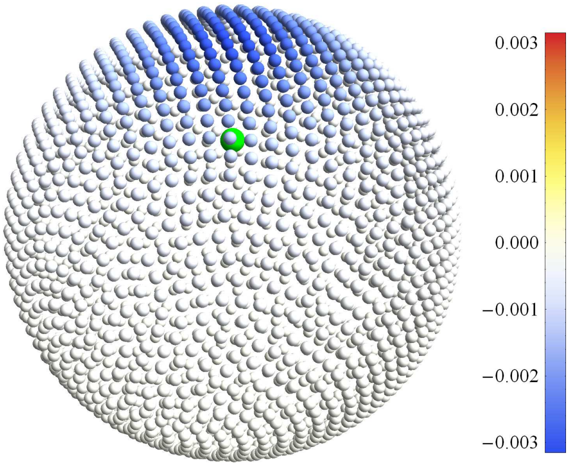


Figure 5.3: Induced charge distribution on a 7.94 Å radius sphere composed of 2000 Gaussian sites ($\eta = 4.7$ Å^{-1}) with a unit point charge (shown as a larger circle) positioned 3.97 Å from the center of the sphere. The color coded charge distribution is given in units of the point charge. The Gaussian sites are shown as circles with radius η^{-1}.

An effective sphere radius can be defined by fitting the long-range portion ($r < R/3$) of Figure 5.4 to the continuum energy expression, Equation (5.5). Figure 5.5 summarizes the η-dependence of this effective radius, which equals the physical radius when $\eta \approx 8$ Å^{-1}. Equality of the physical and the effective long-range image planes implies that the model conductor will behave precisely like a spherical conductor at long range. This equivalence motivated Siepmann and Sprik's initial parameterisation, which has followed through to subsequent works. [40, 41, 64, 65] However, we note that our model conductor is not intended to be a smooth sphere on the atomic scale. Indeed, the corrugated equipotential is considered a desirable aspect of the model, and so the metal's capacitance can be expected to deviate slightly from the

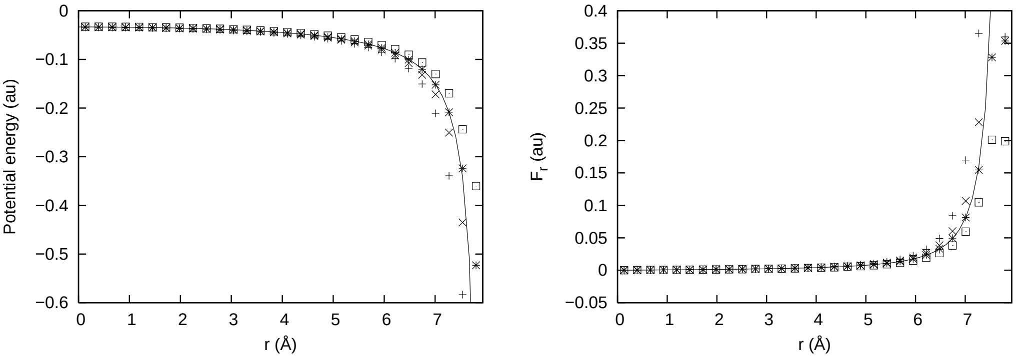


Figure 5.4: Potential energy of (left), and radial force acting upon (right), a point charge a distance r from the center of a 7.94 Å radius conductive sphere. The calculation is performed with the equipotential constraint enforced over 2000 sites of the sphere. The analytic image sphere result is given by the solid line. Computational results are shown for $\eta = 1.89$ Å^{-1} (+), 3.78 Å^{-1} (×), 5.67 Å^{-1} ($\ast$), and 7.56 Å^{-1} ($\boxdot$).

capacitance of a perfect sphere. As a result, it is not necessary for η to be tuned to a value that forces perfect agreement with the results of a perfect continuum sphere. What is essential is that the metal appears to behave like a metal at distances for which the discrete nature of the surface becomes insignificant. It is clear from Fig. 5.4 that the force on a long range point charge is insensitive to η. Fig. 5.5 expresses this insensitivity in a different form. The differences between effective and geometric image sphere radii are so small that they are insignificant to a long-range point charge. A notable exception occurs as η is increased towards the infinite limit, at which point the Gaussians become point charges with infinite self-energies. As η increases, the effective sphere radius also increases without bound. Therefore one must pass to longer and longer range before forces will asymptotically approach the continuum limit, with $\eta = \infty$ failing to reproduce long-range electrostatics at any finite range. For this reason, the numerical analysis was restricted to modest values of η, but this restriction can be rationalised on physical grounds. A metal's electronic density is delocalized on the atomic scale, which should be interpreted as the intersite distance, so large η can be disregarded as inappropriate representations of delocalized charge.

We conclude that a point charge far from our model electrode experiences a force which essentially acts like the ideal smooth conductor at long range for a wide range of sufficiently delocalized Gaussian spread. However, Fig. 5.4 makes it clear that short-range forces depend strongly on η, suggesting that the parameterisation should be determined by the short-range behavior. In describing the charge density as a

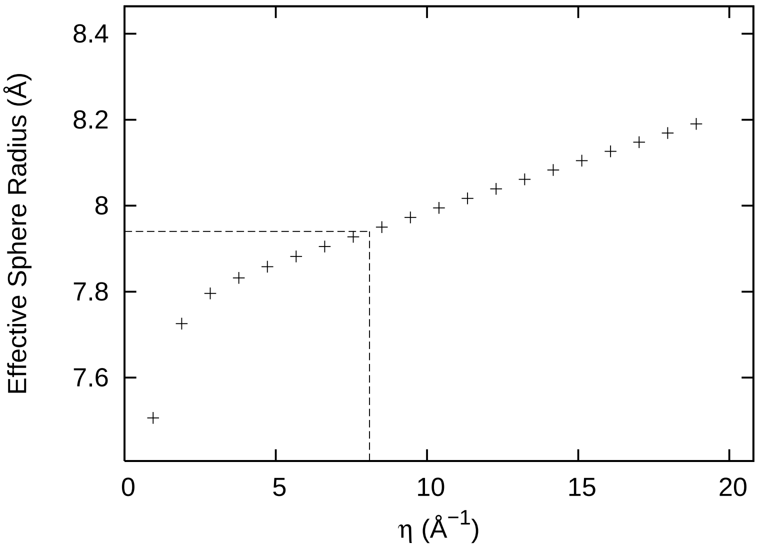


Figure 5.5: An effective image sphere radius, calculated by fitting the asymptotic ($r < R/3$) behavior observed in Figure 5.4 to the continuum image charge energy given by Equation (5.5). The effective sphere radius differs from the physical radius of 7.94 Å except when $\eta \approx 8$ Å^{-1} (as indicated by the dashed lines). Agreement at $\eta \approx 8$ Å^{-1} does not imply that this is the optimal value of the parameter η (see text).

superposition of Gaussians on each site, it is implicitly assumed that a Gaussian is a reasonable approximation of the local electronic structure and that the dominant excitations correspond to redistribution of delocalized charge along a conductor. If the model is to be used to generate short-range interactions with the metal surface, we must fit η to the metal's electronic structure as we attempt to do in Section 5.3.

5.2.3 Calculations on a Potential Biased Sphere

It is trivial to add a potential bias to the computational model via Equation (4.6). The potential bias allows the total charge to fluctuate, so the unit cell is not generally neutral. Herce *et al.* justify the exclusion of the $\mathbf{k} = 0$ term from the reciprocal space sum even in the case of non-neutral cells [28], so the net charge does not affect the calculation of forces and energies.[3] By charging up the sphere described in the last section, we confirm that the metal model also captures the correct continuum behavior of biased electrodes.

The key observation is that the potential bias adds or removes charge from a conducting sphere, but that this extra charge redistributes itself evenly of the sphere

---

[3] It does, however, affect the virial so constant pressure simulations need to compensate for the pressure due to non-neutrality.

surface. The surface charge distribution can be considered to just be the superposition of the unbiased ($V = 0$) distribution plus an extra evenly dispersed voltage-induced charge distribution. Because this "extra" charge is uniform over the hollow sphere, Gauss' Law can be used to show that it exerts no force on the enclosed point charge. Hence, even for the charged sphere, classical continuum theory predicts

$$F_{\hat{r}} = -\frac{q^2 R r}{\left(R^2 - r^2\right)^2}. \tag{5.6}$$

The energy, however, is altered by the potential bias. An extra energy contribution of $Q_{ind}V$ must be included, where Q_{ind} is the induced charge. The additional energy can be re-expressed in terms of the capacitance.

$$U = -\frac{q^2 R}{2\left(R^2 - r^2\right)} + \frac{1}{2} C V^2. \tag{5.7}$$

For an isolated sphere the capacitance is given by the ideal capacitance of a hollow sphere, $C_s = R$ in atomic units. The calcuations present in Figures 5.6 and 5.7 were performed with periodic boundary conditions, so the pertinent capacitance is that of an infinitely replicated set of spheres. This capacitance was not solved analytically, so the curves were fitted in Figures 5.6 and 5.7 with $C = 2R$. This analytical model is compared to the radial force and energies for +2V and +5V bias in the same way as before.[4] It was coincidental that the capacitance was roughly an integer multiple of the hollow sphere capacitance, as demonstrated by altering the box length of the cell. Figure 5.8 demonstrates that for an isolated sphere (entirely real space calculation) Equation (5.7) may be used with $C = R$, but capacitance would not be so simple to analytically compute with periodic boundary conditions. Figure 5.8 effectively provides the numerical calculation of these capacticances, so an analytic expression is not necessary to proceed.

The implication from the preceding calculations is that the potential bias does not affect the forces experience by a point charge. This is only the case when the point charge is located inside of a conducting sphere. A charge outside the sphere would experience a force acting as if the extra charge is concentrated at the center of the sphere, so the potential bias can alter the forces in a meaningful way. Similarly, a potential bias affects the forces experienced by a point charge inside of a non-spherical conductor. The force cancellation that causes a point charge to not feel the excess charge on the sphere is unique to the spherical geometry. For this reason, adding a potential bias to a nanotube is expected to alter the forces on ions inside the tube.

---

[4] These voltages are cited in SI units, but the equations (like all other equations in this thesis) are given assuming atomic units. The voltage conversion from SI to au requires multiplication by $4\pi\epsilon_0 a_0/e \approx 0.03674$.

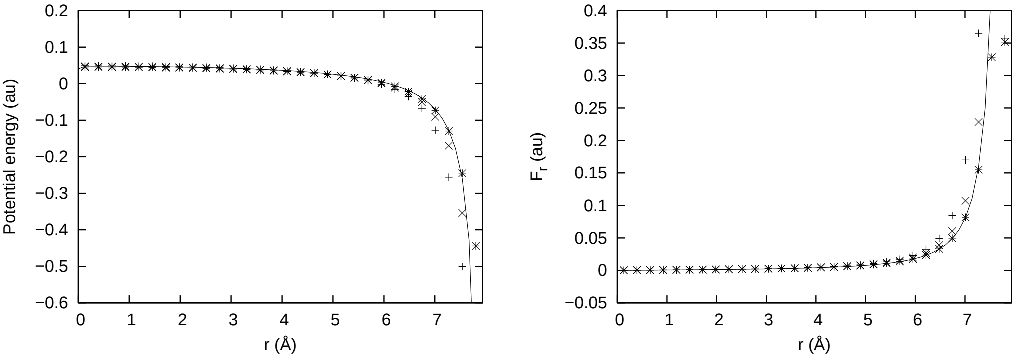


Figure 5.6: Potential energy of (left), and radial force acting upon (right), a point charge a distance r from the center of a 7.94 Å radius conductive sphere held to +2V. The analytic image sphere plus capacitance result is given by the solid line. Computational results are shown for $\eta = 1.89$ Å^{-1} (+), 3.78 Å^{-1} (×), and 5.67 Å^{-1} ($\ast$).

5.3 Fitting η

5.3.1 Back of the Envelope Approximation

The atomic structure of single walled nanotubes (SWNT) is characterized by the folding of a graphene sheet about a particular chiral vector. This construction is also used to compute SWNT electronic structure within the tight binding model. The tube's electronic structure can be directly related to the structure of the π graphene bands. [36] As such, the conductive band of the nanotube arises largely from the $2p_z$ carbon orbitals. Curvature-induced mixing of the sp^2 hybridized orbitals is expected to perturb the π bands slightly, but a crude approximation of the SWNT's electron distribution perpendicular to the tube's surface can be provided by assuming the conductive electrons will reside in hydrogen-like $2p$ orbitals. A rough value of the parameter η can be found by solving for the Gaussian distribution that shares the same expected electron radius as a carbon $2p$ orbital.

Using the hydrogen-like carbon ($Z = 6$) atomic wavefunction,

$$\frac{\langle 2p\,|r|\,2p\rangle}{\langle 2p|2p\rangle} = \frac{5}{6}\text{ au.} \tag{5.8}$$

The expected radius for the Gaussian distribution is

$$\langle r\rangle = \frac{2}{\eta\sqrt{\pi}}. \tag{5.9}$$

The two distributions share a value of $\langle r\rangle$ when $\eta = 1.35$ au^{-1} ($\eta = 2.55$ Å^{-1}). The radial distribution function of this Gaussian and the $2p$ orbital are shown in Figure

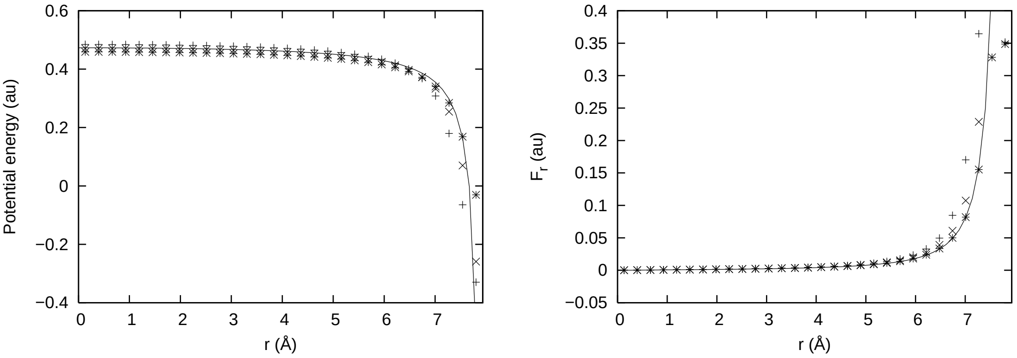


Figure 5.7: Potential energy of (left), and radial force acting upon (right), a point charge a distance r from the center of a 7.94 Å radius conductive sphere held to +5V. The analytic image sphere plus capacitance result is given by the solid line. Computational results are shown for $\eta = 1.89$ Å^{-1} (+), 3.78 Å^{-1} (×), and 5.67 Å^{-1} (∗).

5.9 to demonstrate that the Gaussian is a good approximation for the atomic orbital. One should not expect this value to be the optimal parameter, but it provides a reasonable estimate that can be expected to be well within an order of magnitude of the "correct" value of η.

5.3.2 Comparison with Full DFT Calculations

A more accurate value of η can be obtained by parameterizing to agree with an all-electron density functional theory calculation of a conducting nanotube. Meunier *et al.* performed such a calculation on a (9,0) capped CNT with a 6-31g* basis set to parameterize a continuum electrostatic model. [33] Their results, reproduced in Figure 5.11a, give the electrostatic potential due to a point charge and the induced charge on the nanotube. The exact details of the calculation are not explicitly stated in the paper by Meunier *et al.*, but from Figure 5.11a's "free space" curve of a bare point charge potential it was deduced that the point charge was positioned 10.45 Å from the center of the nanotube axis. It is furthermore assumed that Meunier's voltage perpendicular to the tube lies on the plane containing the point charge.

To compare directly with the DFT calculations, the Gaussian charge model was used to compute the induced charge density and electrostatic potential for the same nanotube/point charge system. The calculation was performed without periodic replicas, making the summing of the pairwise Coulomb potentials trivial. Induced charge distributions were computed by the matrix inversion schemes described in Sections

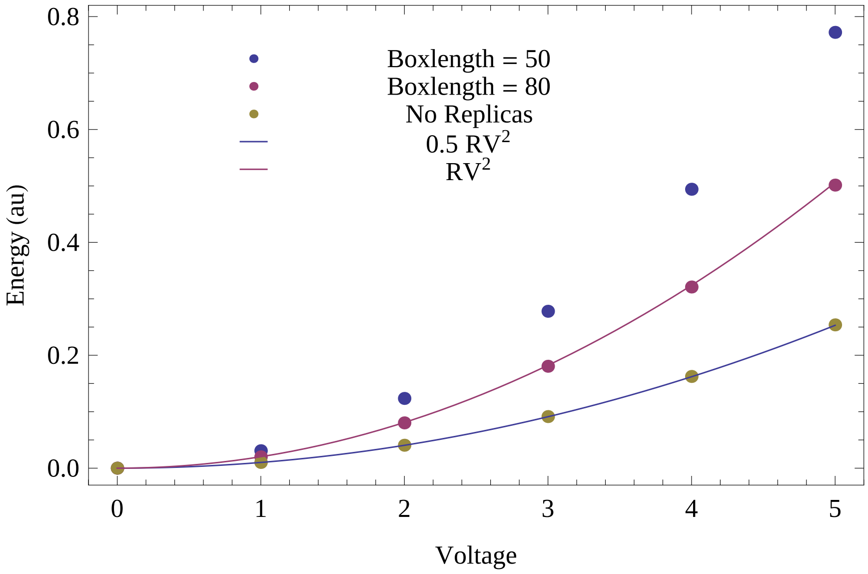


Figure 5.8: Energy of capacitive charging of conductive sphere under potential bias. With no periodic replicas the energy is $CV^2/2$ as expected with $C = R$ for the ideal sphere. In periodically replicated systems, the effective conductor is actually the set of all periodic images separated from each other by some box length. The initial value of the box length (80 au) coincidentally falls near RV^2, but by changing the box length it is clear that the capacitance is dependent on system size. This is consistent with what one would expects since the shrinking box forces the metallic sphere and its replica closer together.

4.3.1 and 4.4.1 for a grounded and conducting nanotube respectively. The results are plotted with variable η in Figure 5.11b.

The comparison in Figure 5.11 demonstrates that the Gaussian induced charge model captures the general behavior of the DFT calculation. The floating tube screens the region inside the tube from the point charge potential, while the grounded tube also screens the region on the other side of the tube.[5] The value of $\eta = 2.55$ Å^{-1} suggested in Section 5.3.1 results in weak electric fields that point in the wrong direction on the inside of the tube. Increasing η by roughly a factor of two rectifies this problem. The larger value of η can be rationalized on physical grounds. Since covalent bonds link the metal sites, they lie closer to each other than to the point charge. Altering the size of the Gaussian on these metal sites most significantly impacts the coupling between nearby metal sites, which affects the degree to which metal sites induce charge in their neighbors. Hence it is the spread of the charge distribution

---

[5] Note that the CT curves from Meunier *et al.* are results from a continuum theory. There is no DFT calculation for a grounded tube.

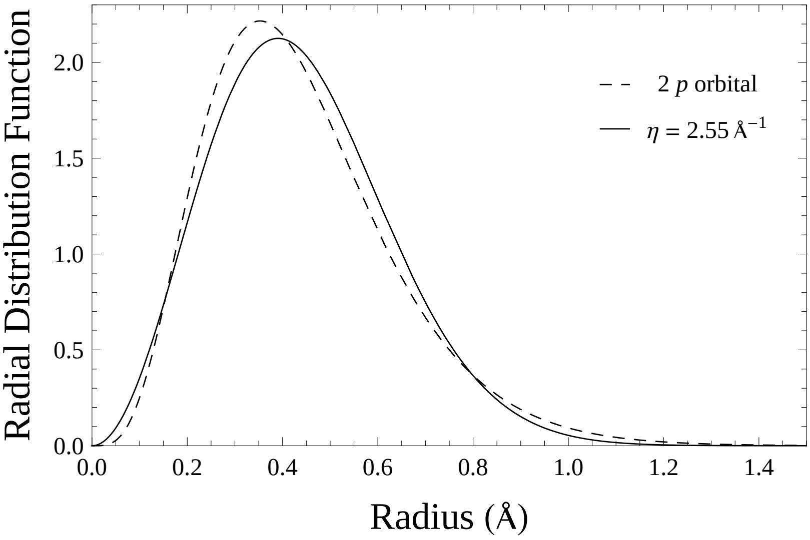


Figure 5.9: Radial distribution functions for a $2p$ hydrogen-like carbon orbital and for a Gaussian distribution with $\eta = 2.55$ Å^{-1}. The distributions share the same value of $\langle r \rangle$, providing a rough estimate of the parameter η.

in the plane parallel to the metal surface, which most strongly influences the overlap between neighboring metal sites. A $2p_z$ orbital extends outward perpendicular to this surface, so the expected spread parallel to the surface is smaller than the spread computed in Section 5.3.1, corresponding to a larger value of η.

Even with this improved value of η, the exact behavior at the surface of the tube does not match the DFT calculation. This is neither surprising nor disturbing. In application, a Lennard-Jones potential will prevent ions from venturing too close to the nanotube wall. For the Gaussian charge model to be useful, it need only reproduce the slope of the electrostatic potential away from the walls, something that is achieved by $\eta \approx 5$ Å^{-1}. Future work aims to perform more DFT calculations to accurately parameterize η, but ultimately η is just a parameter. The work that follows is aimed at exploring a class of dynamics in a chaotic system, not to mimic the exact parameters for a particular material. For this reason, choosing $\eta = 5$ Å^{-1} will be sufficient to proceed.

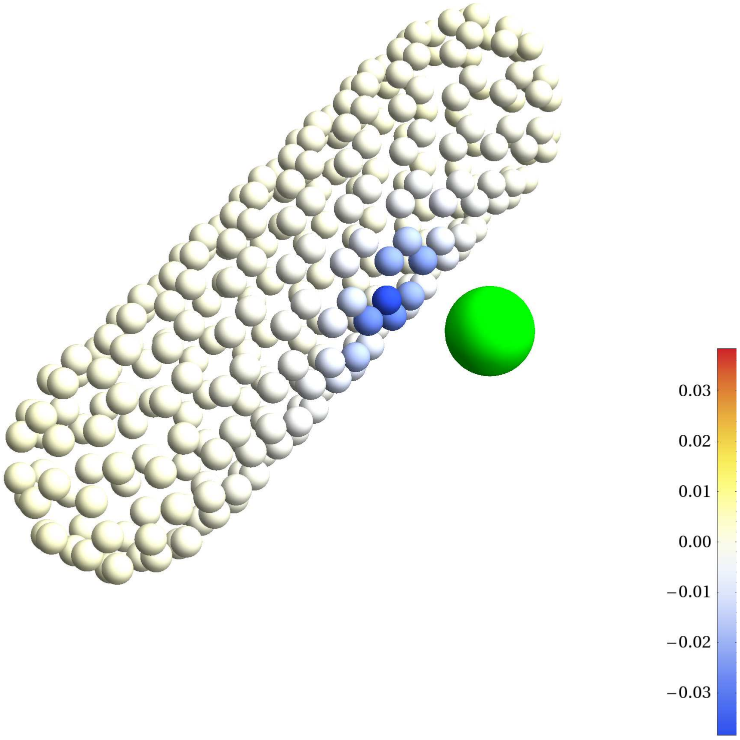
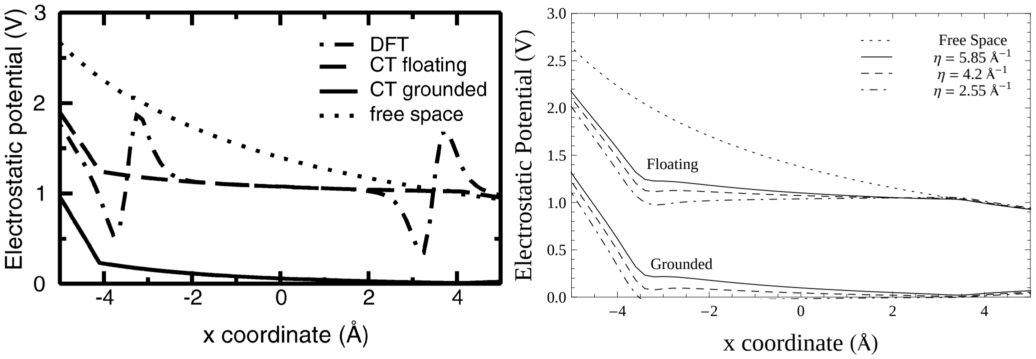


Figure 5.10: Induced Gaussian charges on a capped (9,0) charge-neutral carbon nanotube due to a point charge. The shown distribution is computed with $\eta = 2.55$ Å^{-1} and the Gaussian sites are plotted with a radius η^{-1}. Charge magnitudes are given in units of the point charge.



(a) Full DFT Calculation (b) Gaussian Induced Charge

Figure 5.11: Comparison between a full DFT calculation (left) and the classical calculation with Gaussian surface charges (right) of induced charge in a capped (9,0) nanotube due to a point charge positioned 10.45 Å off axis. The DFT figure, reproduced from Ref [33], shows the same general screening as our classical model for both floating and grounded tubes.

Chapter 6

Inorganic Nanotube Formation

Having developed the numerical methods for treating a CNT as a classical metal, we are now in a position to explore how the metallic nature influences the structure and dynamics of a molten salt as it fills the tube. A charge-neutral floating CNT as well as a potential-biased tube can be considered, giving rise to two distinct mechanisms of filling. Floating tubes fill much like the uncharged rigid tubes that have been previously studied. [70, 71] Application of a +10 V bias introduces excess charge into the simulation cell, which allows ions to enter the CNT without requiring charge-neutrality. Not only are the dynamics of filling different under these conditions, the stable INT structures are also distinct. As we will see, they can be rationalized as multiwalled concentric tubes formed of separate anion INTs and cation INTs. By later removing the potential bias, the system can be relaxed into a charge-neutral INT like those that form directly within floating tubes. This provides an alternative physically realizable[1] mechanism for INT formation, allowing an investigation into the relevance of the filling dynamics on the ultimate INT structures.

6.1 Methods

6.1.1 Potential Model

Like earlier work [6], the molten salt is modeled with a Born-Mayer potential of the form

$$U(r_{ij}) = B_{ij} e^{-a_{ij} r_{ij}} + \frac{Q_i Q_j}{r_{ij}} - \frac{C_6^{ij}}{r_{ij}^6},$$

with the parameters given in Table 6.1. Q_i and Q_j are the formal charges of the ions, $+1$ and -1 for M and X species respectively. The parameters B_{ij} and a_{ij} modulate

---

[1] By physically realizable it is meant that this type of manipulation is within the capabilities of the set of crude macroscopic control parameters discussed in the Introduction.

Born-Mayer Parameters	M-X	M-M	X-X
B_{ij}	8.68	1.1456	61.66
a_{ij}	1.55	1.55	1.55
C_6	2.09	0.0763	115.987

Table 6.1: Born-Mayer parameters for the rigid ion model used in filling simulations. The parameters, given in atomic units, are chosen to favor a four-coordinate bulk structure, which is expected to favor three-coordinate hexagonal two-dimensional nets.

the atomic repulsions while C_6^{ij} terms tune the dispersion interaction. The salt does not correspond to a particular metal halide, rather the values are chose to stabilize a four-coordinate bulk structure. [70, 74] In this way the resultant structures can be expected to be representative of some classes of MX salts while not necessarily being in quantitative agreement with any one salt. While initial study of these INTs used a polarizable ion model [70, 71], more recent work has shown that INT formation is also characteristic of the simpler rigid ion model employed here. [6, 7]

Interactions between the ions and the tube are described by a simple Lennard-Jones potential,

$$U(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right],$$

with parameters given in Table 6.2. As described elsewhere [6], these Lennard-Jones

Lennard-Jones Parameters	C-M	C-X
ϵ_{ij}	57.9	79.4
σ_{ij}	3.4025	3.735

Table 6.2: Parameters for the Lennard-Jones potential used to model interaction between the ions and carbon atoms of the CNT. The parameters, given in degrees Kelvin and Angströms, are taken from interactions between graphite and Noble gasses isoelectronic to K^+ and I^- for C-M and C-X respectively.

parameters are taken from interactions between graphite and Noble gasses isoelectronic to K^+ and I^-. Induced metal surfaces for the floating CNTs are treated as outlined in Chapter 4 using the constrained conjugate gradient minimization. The value of η was set to 1 au^{-1} for initial simulations on a (19,0) CNT. The (11,11) CNT in Section 6.6 adopts the value of η given in Section 5.3.2.

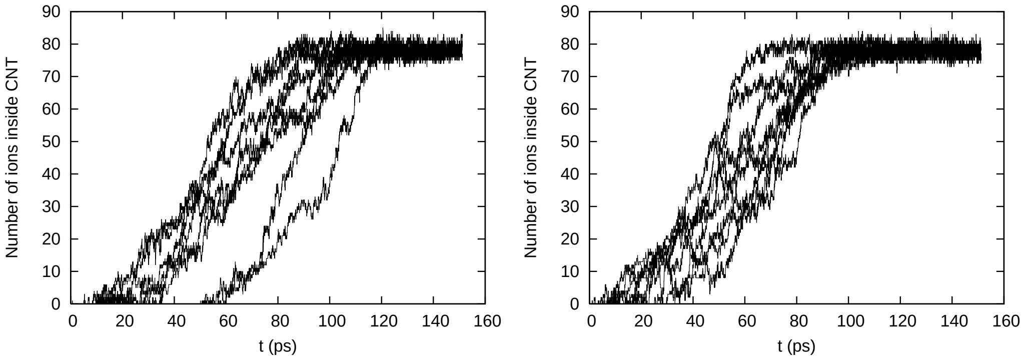
6.1.2 Equilibration

Filling simulations are initiated from an equilibrated liquid state. The equilibration procedure is straightforward. A cubic simulation cell of an equal number of M^+ and X^- ions was generated (864 ions of each type in a box with side lengths of 35.7 Å). Molecular dynamics in the NPT ensemble was carried out using 25 au time steps with the pressure barostated to zero and temperature thermostated to 800 K (relaxation times of 6.0 ps for both the thermostat and barostat). After at least 30 ps of simulation, ions were deleted in a charge-neutral manner out of a cylindrical section of the simulation cell to make room for an inserted CNT. The ends of the CNT were blocked off by a partial graphite layer and at least 30 ps of NPT dynamics were carried out to equilibrate the salt plus CNT system. Distinct equilibrated states were generated by cutting out the cylindrical region at different times in the dynamics of the full molten salt system. Even small changes in the ion-deletion time resulted in different sets of deleted ions, ensuring that the final equilibrated states are statistically independent.

6.2 Filling of a (19,0) CNT

For ease of comparison with prior results [70, 71], we begin by discussing the filling of a (19,0) CNT treated with and without induced charges. It should be noted that a (19,0) CNT is actually semiconducting, but these studies aim to understand the impact of adding metallicity to the simple nonmetallic model. This can be probed even for a tube which is not actually metallic, though the subsequent work naturally focuses on a CNT morphology which is indeed metallic. The equilibration procedure was carried out with both metallic (floating) and nonmetallic representations of the (19,0) CNT using eight distinct salt configurations. Filling was initiated by abruptly removing the graphite covers, allowing ions to enter the tube. The Coulomb liquid is strongly ordered, so filling requires the collective motion of many ions. The typical motif of filling is depicted in Figure 6.1a, showing a charge-neutral loop that threads its way into the tube. Formation of this loop requires a particular class of fluctuation in the liquid structure, and there is often a lag in the filling time due to the rareness of this fluctuation. Once the filling is initiated it proceeds rapidly. These observations agree with a prior report [72], and are well summarized by monitoring the number of ions in the tube as a function of simulation time, the so-called filling profile. Comparison of the filling profiles between the metallic and nonmetallic models of (19,0)

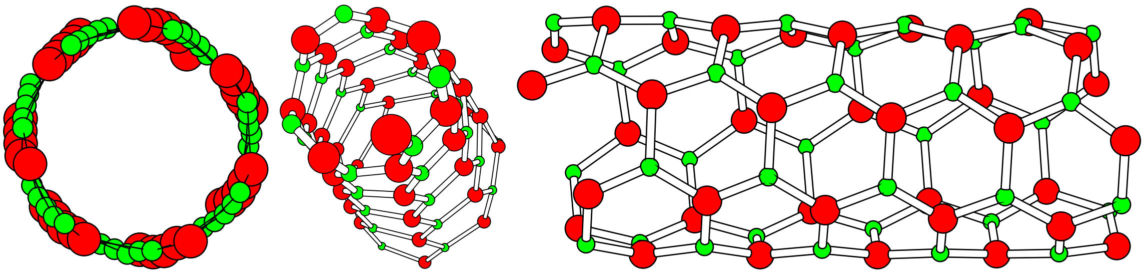
CNTs demonstrates that the inclusion of metallicity does not radically alter the filling behavior. Two of the eight metallic filing simulations have anomalously long lag times compared to the others, but based on a more complete analysis of lag times in the nonmetallic tubes [72], occasional long lag times are to be expected merely due to the rare liquid fluctuations required for filling to initiate. Many more simulations would need to be run to determine if differences between the filling profiles is actually statistically significant, but this is not a path I chose to pursue further.



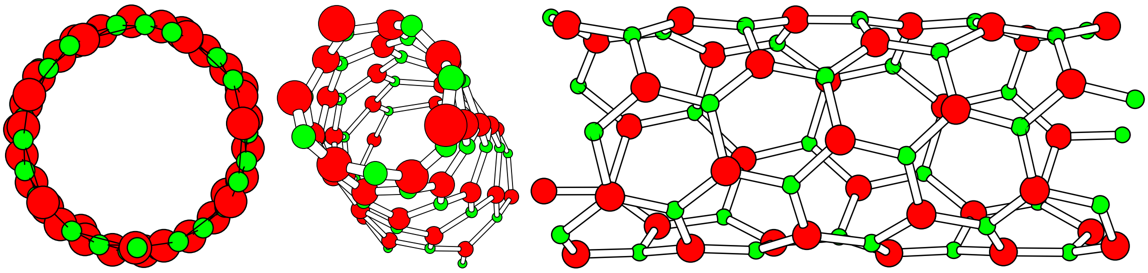
(a) Metallic (b) Nonmetallic

Figure 6.1: Comparison of the filling profiles of metallic and nonmetallic (19,0) CNTs. The rate of filling appears virtually the same. Since filling depends on initial fluctuations of the ionic liquid, slow initial filling is expected and has been regularly observed. [72] These filling simulations are not sufficient to provide a statistically significant comparison of the average lag time and filling rates.

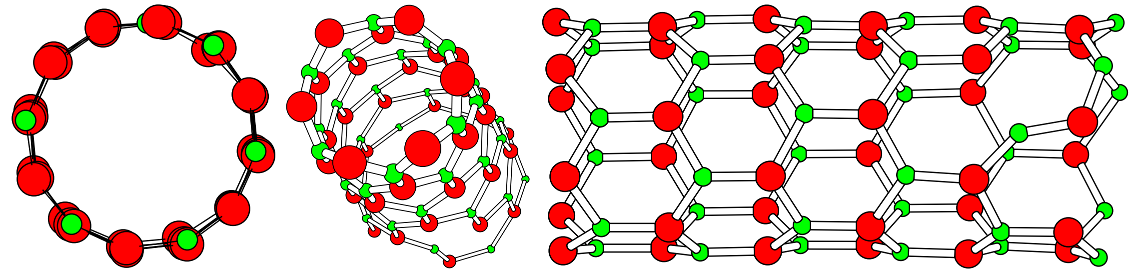
As is clear from the filling profile, the tubes became full after around 100 ps. The final structures adopt INT conformations that oscillate around some local minimum of the energy. To identify this minimum the positions of the ions inside the CNT were averaged for the last several picoseconds of dynamics. Since the structures are stable, the results are very insensitive to the averaging time. Common average structures are depicted in Figure 6.2. In agreement with earlier work on nonmetallic tubes [72], several stable tubes were observed, in this case both (4,3) and (5,2) INTs formed within the metallic and nonmetallic tubes. As summarized in Table 6.3, the proportion of (4,3) and (5,2) tubes was not influenced by the metallicity of the tube, though there are far too few simulations to make any statistically significant statement about filling of (19,0) CNTs. We will return to these considerations within (11,11) CNTs, where a larger sample size was simulated. First, let us consider more carefully why the metallicity effects are not more significant.



(a) (4,3) INT



(b) (5,2) INT



(c) (6,0) INT

Figure 6.2: Inorganic nanotube morphologies which have been observed in computer simulation to form within (19,0) and (11,11) CNTs. Basic geometric properties of these tubes are provided in Table 6.3.

6.3 Why Is Metallicity a Minor Perturbation for a Floating Tube?

The charge induced on a nanotube due to a single point charge was shown in Figure 5.10. As intuition suggests, the permanent charge induces an opposite charge on the nanotube, resulting in an attractive force. In an ionic system containing many positive and negative point charges, the induced charge can be expected to be oscillatory in sign, with positive induced charge closest to the anions. Since induced charge is complementary to the ionic charge, the ions are attracted to the nanotube just as a single ion is attracted in Figure 5.10. Why then does this attractive force seem to have little effect on the floating metallic tube? The answer lies in the small magnitude of the induced charges. As shown in Figure 6.4 and 6.7a, the magnitude of induced

INT	INT Properties		Filling Frequencies				
			(19,0)		(11,11)		
	Radius (Å)	Chiral Angle	N	F	N	F	V
(4,3)	3.50	25.3	7	7	73	64	23
(5,2)	3.60	16.1	1	1	18	25	25
(6,0)	3.46	0.0	0	0	7	9	36
(6,1)	3.78	7.6	0	0	2	0	12
(4,4)	3.99	30.0	0	0	0	2	4

Table 6.3: Physical properties and frequency of observation of INTs forming inside a (19,0) and an (11,11) CNT. The N, F, and V columns give the frequency of observation in direct filling simulations in which the CNT electronic structure was ignored (nonmetallic), treated as a floating metal tube, and treated as a voltage biased metal respectively. From each liquid state metallic and nonmetallic simulations were initiated, but the trajectories quickly diverged such that comparisons between the metal and nonmetal simulations must be based on the statistics of a large number of simulations. INT radii are computed by the usual tube-folding procedure [Equation (1.1)] assuming a M-X distance of 2.09 Å, the average bond distance from the simulations. These calculations neglect the curvature of the INT. More accurate radii can be computed from the asymptotic expressions of Cox and Hill's polyhedral construction [12], but the differences between cylindrical folding and polyhedral construction was within the sample deviation of the computationally observed radii. The final column sums to 99 rather than 100 because the final trajectory had not yet converged to an INT at the time of writing.

charges is less than 0.04 au for the most charged of the metal sites during both the initial filling stages as well as after the CNT is full. The Lennard-Jones repulsions keep the ions about 2.2 Å from the CNT surface, which Figure 6.3 shows is roughly equal to the average M-X bond distance. For any given ion the induced charges are about the same distance away as the neighboring ions, which have nearly 100 times greater charge magnitudes. It is no surprise then that given the small induced charge magnitudes, the INT structures are not strongly affected by metallicity of the CNT. The forces due to CNT induced charge may serve to renormalize the Lennard-Jones parameters slightly, but the basic structures will be the same.

It remains to explain why the induced charge magnitudes are so small. As the CNT lattice is bipartite, one could imagine the model admitting a surface charge distribution with large charge magnitudes and anti-ferromagnetic long-range order (AFLRO). Such a distribution would be induced by a crystalline surface due to its alternating permanent charges. Because the induced charge would be complementary to this crystalline surface, the crystal would be attracted, thereby inducing still greater

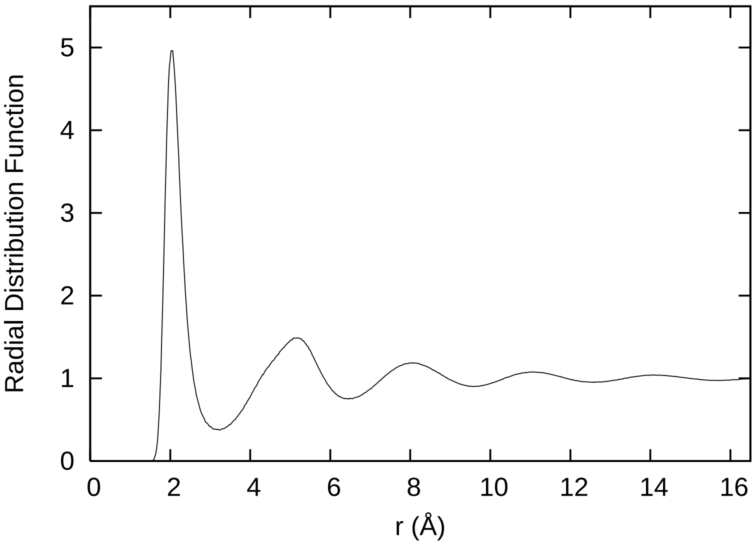


Figure 6.3: Radial distribution function for the distance between M and X species. The distribution function is computed from a box of MX liquid with a filled (19,0) CNT inside the box. Note that the average nearest-neighbor separation is about 2.09 Å. The peak at twice this distance is skewed because it includes correlations between ions inside and outside the tube. The fact that the skewing is minimal demonstrates that the distance between ions and the Lennard-Jones centers roughly equals the M-X separation.

charge magnitudes while preserving the AFLRO and the CNT charge neutrality. Given this argument one might reasonably expect the final charge distribution to be of this AFLRO form, yet Figure 6.4 clearly shows this is not the case. To make sense of the simulation results one must recall that the metal sites of the model have an on-site self-energy, which physically stems from the repulsion of electron density confined to the Gaussian site. To obtain an order of magnitude estimate for the effect of this on-site term we consider a minimal model for induced charge, that of a single metal site 4 au away from a single point charge. We compute the energies in real space for simplicity, but this can be justified as the molten salt will screen long-range contributions. Hence the dominant energy terms for any metal site is the real-space interaction with itself, its metal neighbors, and the nearest ions. As we seek an order of magnitude estimate, it suffices to initially neglect metal neighbors and consider a single ion with charge -1.

$$U(Q) = -\frac{Q\mathrm{erf}\,(\eta r)}{r} + \frac{Q^2\eta}{\sqrt{2\pi}},$$

where η is as always the inverse of the Gaussian spread. The Lennard-Jones potential maintains a metal-ion separation of about 4 au, so we fix r to be 4 au. For

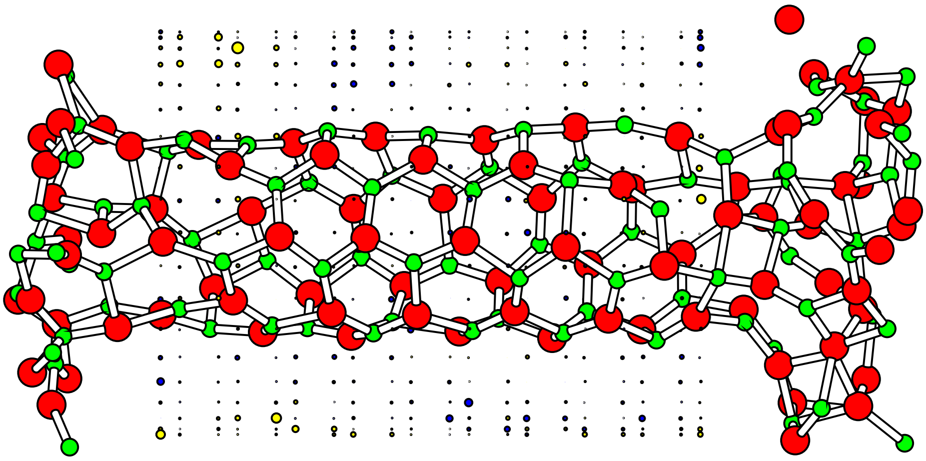


Figure 6.4: Molecular dynamics snapshot of an INT inside the floating (19,0) CNT. Ions outside the CNT have been cropped out for clarity, but they also influence the induced charge. Anions are shown in red, cations in green, positive induced charge in blue, and negative induced charge in yellow. The induced charge magnitudes are indicated by the radius of the plotted ball. This is just a plotting tool as all of the metal sites share the same Gaussian spread. For reference the largest yellow ball corresponds to an induced charge of -0.0376 au.

our values of η, $\mathrm{erfc}(4\eta) \approx 1$, so the energy minimizing value of Q is easily found to be $Q \approx \sqrt{2\pi}/(8\eta)$. For $\eta = 1$ au^{-1}, the value used for (19,0) CNT simulations, the induced charge in this reduced model is $Q \approx 0.31$ au. While this is a full order of magnitude greater than any of the observed induced charges, the result was obtained by neglecting neighboring metal sites, which are closer to each other than to the ions. The single ion of our reduced model would also induce charges in the neighboring metal sites, which would reduce the value of Q on any one site, at least within the regime of the metal-ion distance we have considered. This simplified model demonstrates that the small magnitudes of induced charges can be thought of as a consequence of the on-site energy, but this on-site energy is linearly related to the parameter η. Does this mean a poorly chosen value of η will scale the total induced charge? The answer is no. The magnitudes of the Gaussians of charge can be expected to scale, but this scales both the positive and negative induced charges. As detailed in Chapter 5, this is not to say that η does not matter, but its influence is not as catastrophic as it may first appear.

The on-site energy arises from the microscopic features of the model (the fact that charge density is locally described by Gaussians). The microscopic aspects of the model, however, are the most tenuous, so it is conceivable that the observations are sensitive to the model. Luckily there exists another physical reason to expect

AFLRO to be broken up, namely the incommensurability of lattices. The C-C bond lengths are incommensurate with the M-X lengths. As such, the metal sublattices cannot perfectly complement the salt crystal sublattices. Even if the crystal were commensurate with the carbon lattice, the carbon sites also interact with salt ions outside the tube. These ions are disordered, but even if they were not, the cylindrical geometry actually precludes both internal and external structures from being simultaneously commensurate with the metal lattice.

6.4 Are the Dynamics Overdamped?

Given that metallicity is not a dominant factor for floating CNTs, we temporarily revert to the nonmetallic CNT model to try to better understand why several distinct INTs are formed. Wilson's analytic model approximates the energy of the different nanotubes as a function of the difference between CNT and INT radii. [70] The discrete set of INT morphologies restrict the radii to a finite number of possible radii, and Wilson's expression allows one to estimate the relative energies of these structures. If the systems were in thermal equilibrium, one would expect the distribution of structures to satisfy the Boltzmann distribution, heavily favoring the observation of the lowest energy INT structures. While it has been observed that only relatively low energy INTs are formed, the frequency of observations of each morphology does not obey the Boltzmann distribution. As an extreme example Wilson noted that the analytic model predicts the formation of (7,0) and (5,3) INTs formed within (12,12) CNTs to be isoenergetic, yet the (5,3) INT is observed to form far more frequently.[2] Estimating the free energies by approximating the entropic contribution of INT vibrational modes does not account for the observed distribution, as the vibrational entropies of the different INTs are nearly equal. [72]

Given these observations one must conclude that the system is not in equilibrium or that an entropic contribution is not being included in the free energy estimates. The former is very plausible given the modest simulation times and large barriers to interconversion between the INTs. The metastable structures that result are merely local minima in which the dynamics are trapped for all reasonable simulation time scales. Which minimum is discovered depends on the course of the dynamics, not on equilibrium properties. Within this view of INT formation, it is desirable to identify a growth mechanism, which would illustrate the characteristic atomic motions that

---

[2] Wilson's simulations actually examined filling of (12,12), (21,0), (13,1), (14,10), (15,9), and (18,5) CNTs, all of which have radii between 16.30 and 16.47 Å. For simplicity I just mention the (12,12) result, but the observations are even more general.

result in the approach of a given minimum. Generally this mechanism can be thought of as a curve through configuration space connecting the structures of the initial and final states in the most likely manner - in other words it is the minimum free energy path (MFEP). Hamiltonian dynamics, however, exists in phase space, not position space, so a minimum energy path lies in the full phase space. That is to say that a mechanism is not fully provided by the sequential positions of the degrees of freedom, it also includes the sequential momenta. While the MFEP is formally embedded in the full phase space, it is often the case that the momentum contributions average out. The positions can carry the relevant information such that the dynamics starting at $(\vec{x}, \vec{p}_1)$ is essentially identical to the dynamics starting at $(\vec{x}, \vec{p}_2)$. When this is the case we say the system is overdamped.

Overdamped dynamics are easier to deal with because one can then discuss an intuitive picture of a mechanism as the "movie" of the atomic coordinates. Using this approach Wilson suggested that the relative lack of $(n, 0)$ INTs could be explained by the initial filling mechanism of the simulation. [72] To further probe this hypothesis it is essential to determine whether or not the dynamics is overdamped. More explicitly, we must determine whether or not information of the atomic coordinates (but not momenta) after some initial filling dynamics is sufficient to determine the final structure. The procedure is quite simple. We select two trajectories from the previously described filling of a (19,0) CNT - one trajectory which formed a (4,3) INT and one which formed a (5,2) tube. Every 30.2 ps of that trajectory we copy the atomic positions and spawn 20 replicas, which are given random momenta consistent with the Maxwell-Boltzmann distribution. We then carry out the dynamics of these replica until they form INTs. A schematic of the procedure for the (4,3) trajectory is given in Figure 6.5. The full results, summarized in Table 6.4, demonstrate that the early dynamics does not definitively bias the trajectory to a single local minimum. Even when the tube is nearly full, introducing a dynamical time step with randomized velocities is enough of a kick to push the system into a different basin of attraction. The final tubes are in fact stable and can be well-defined by atomic positions alone, as evidenced by the fact that replicas which split off at late times (after the INT was already formed in the parent trajectory) always formed the same INT as the parent. In contrast, the momenta clearly matter for earlier dynamics, so it is dangerous to think of the system in terms of an overdamped mechanism. Put another way, the initial filling mechanism correlates poorly with the INT that eventually forms.

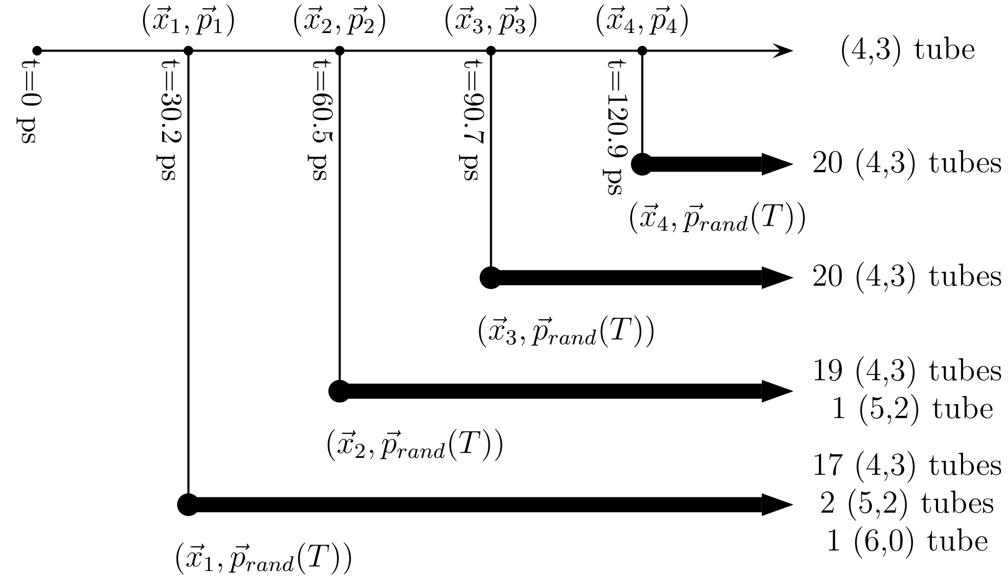


Figure 6.5: Schematic diagram of the series of filling experiments summarized in Table 6.4a. The line across the top represents the original trajectory, which formed a (4,3) tube. The thick arrows represent twenty replicas, each starting with the positions of the original trajectory but with randomized velocities drawn from the Maxwell-Boltzmann distribution. The same procedure was carried out on a trajectory which formed a (5,2) tube, the results of which appear in Table 6.4b.

6.5 Potential Biased Growth in a (19,0) Tube

It is clear that there exist several competing metastable states, so it is unlikely that a CNT can serve as an effective template that selectively grows a single INT morphology. The filling simulations on floating tubes have shown that adding metallicity to the model does not significantly alter the competition between metastable INTs. There remains one more control parameter that can be macroscopically adjusted in a metallic tube. If the CNT were attached to a lead, a potential bias could be added, effectively dumping extra electrons onto the CNT (or removing them). The addition of this bias is natural within the mathematical framework of the metal model, though the neutrality of the simulation cell is broken as discussed in Section 5.2.3. Carrying out the dynamics in the NVT ensemble with a +10 V bias on the tube results in rapid filling as the excess positive charge on the tube attracts X^- ions so strongly that they enter the tube without regard to charge-neutrality. These anions wet the inside of the tube, thereby attracting the cations to form a concentric tube. A filling profile in Figure 6.6 demonstrates the rapidity of this filling relative to unbiased-CNT filling of

Time of Random Perturbation (ps)	Ions in CNT at Perturbation Time	Final Structure (4,3)	(5,2)	(6,0)
30.2	16	17	2	1
60.5	64	19	1	0
90.7	77	20	0	0
120.9	77	20	0	0

(a) Perturbations on a trajectory which formed a (4,3) INT

Time of Random Perturbation (ps)	Ions in CNT at Perturbation Time	Final Structure (4,3)	(5,2)	(6,0)
30.2	20	14	6	0
60.5	75	8	11	1
90.7	79	0	20	0
120.9	78	0	20	0

(b) Perturbations on a trajectory which formed a (5,2) INT

Table 6.4: To probe whether or not the system is overdamped, velocities were randomized at four equally spaced points in time along two different trajectories. In the absence of the randomized velocities the two trajectories led to a (4,3) INT and a (5,2) INT. The tables indicate the time along the original trajectory at which the velocities were perturbed, the number of ions that were inside the CNT at this time, and the number of subsequent trajectories out of twenty which converged to (4,3), (5,2), and (6,0) structures. The significance of the velocities on the final structure indicates that the dynamics, even when the CNT is nearly full, cannot be thought of as overdamped.

Figure 6.1.

A snapshot of the filling mechanism (Figure 6.7b) reveals that the initial filling does not proceed via a charge-neutral loop. Wilson's proposed mechanism for explaining the lack of (6,0) INTs is based upon these charge-neutral loops, so voltage-biased filling provides a route to test the hypothesis that kinetic effects account for the distribution of INTs. By filling the tube under voltage bias then removing that bias to allow the structure to relax into an INT, a distinct physically-motivated mechanism can be probed to see if that mechanism is significant in the final INT structures. In the next sections this is done with the filling of an (11,11) CNT.

6.6 Filling of an (11,11) CNT

To generate more statistically significant results the aforementioned experiments were repeated with an (11,11) CNT, which is metallic and has nearly the same radius as

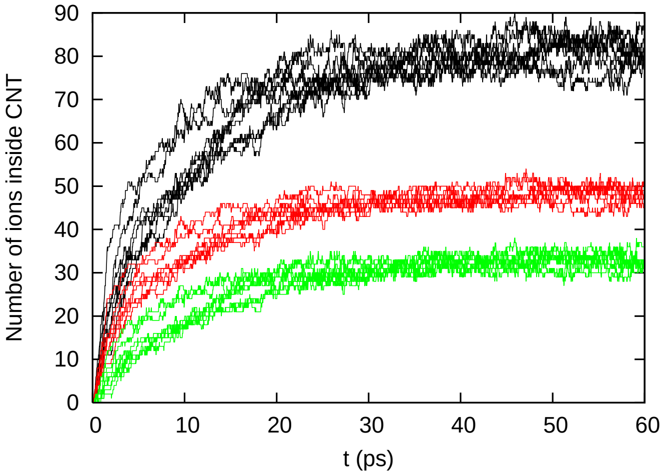
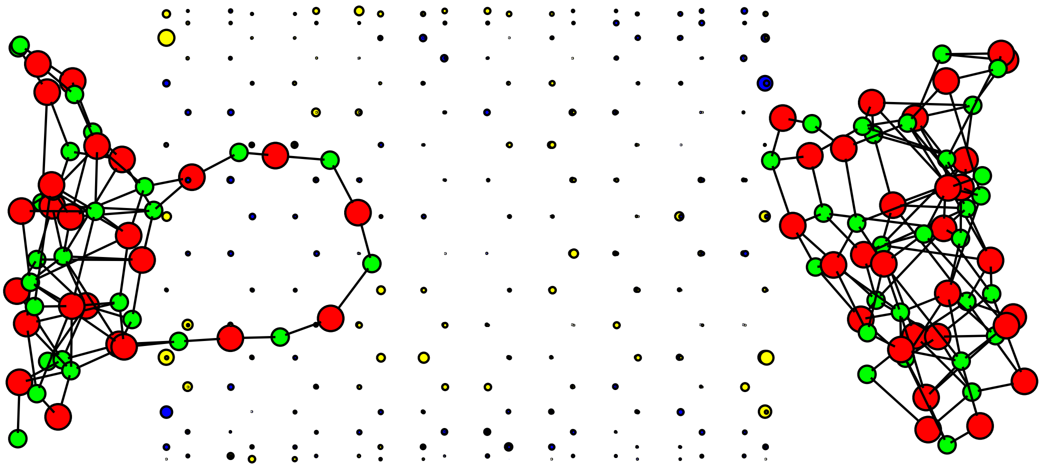


Figure 6.6: Filling profile of a +10 V biased (19,0) CNT. Note that the initial filling is far more rapid that the filling of an unbiased tube since the initial filling mechanism is no longer dependent on collective motions. The addition of charge on the CNT breaks the charge-neutrality of the filling ions. This is seen by comparing the filling profiles of the X^- ions (red) and the filing profile of the M^+ ions (green). The black curve is the profile of the total number of ions.

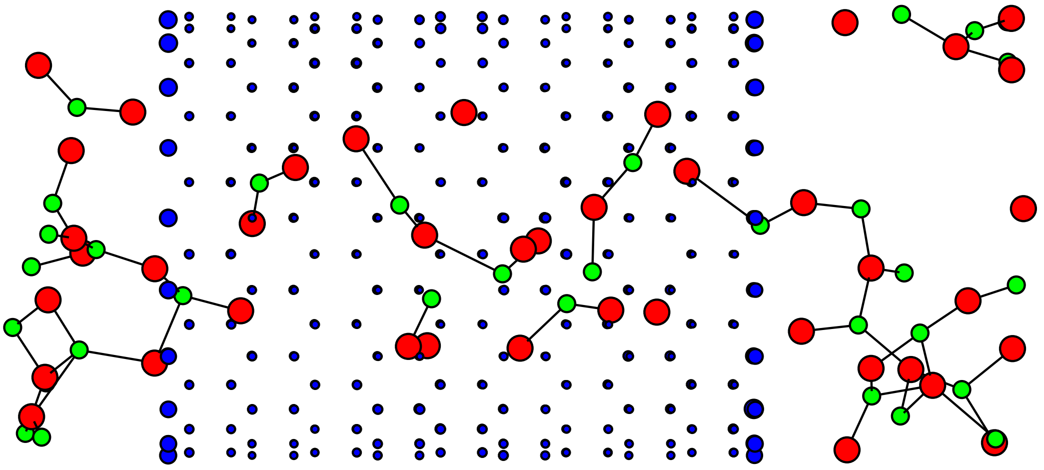
the (19,0) tube. All filling simulations were performed in the NVT ensemble for consistency. Metallic Gaussian charges with $\eta = 5.0$ Å^{-1} were used in accordance with the results of Section 5.3.2. One hundred independent liquid configurations were generated and equilibrated as described in Section 6.1.2. Configurations had roughly 700 ions of each species in a box with side lengths of about 35.0 Å. To ensure rapid filling, extra ions were added for a total of 800 of each species. This was performed by increasing the box size then using NPT dynamics to compress it to a final box length of 33.0 Å. Because the virials for the charged systems have not been calculated, the NPT dynamics occurred without a voltage bias. After the simulation cell was compressed, three copies of the system were equilibrated for 15 ps: one with a nonmetallic tube, one with a floating tube, and one with +10 V applied to a metallic tube. Following the equilibration the graphite layers covering the tube ends were removed to initiate filling as described for the (19,0) CNT. In the case of the nonmetallic and floating CNTs, 120.9 ps of dynamics was simulated to yield stable CNTs. In rare cases[3] the initial CNT filling was especially slow such that 120.9 ps of dynamics was not sufficient to yield a stable structure. In these cases the simulations were continued until a stable INT was obtained. Dynamics on the voltage-biased tubes were also carried out for 120.9 ps to yield structures that will be discussed further in the next

---

[3] Four such cases for the nonmetallic tube



(a) Floating CNT



(b) +10 Volt CNT

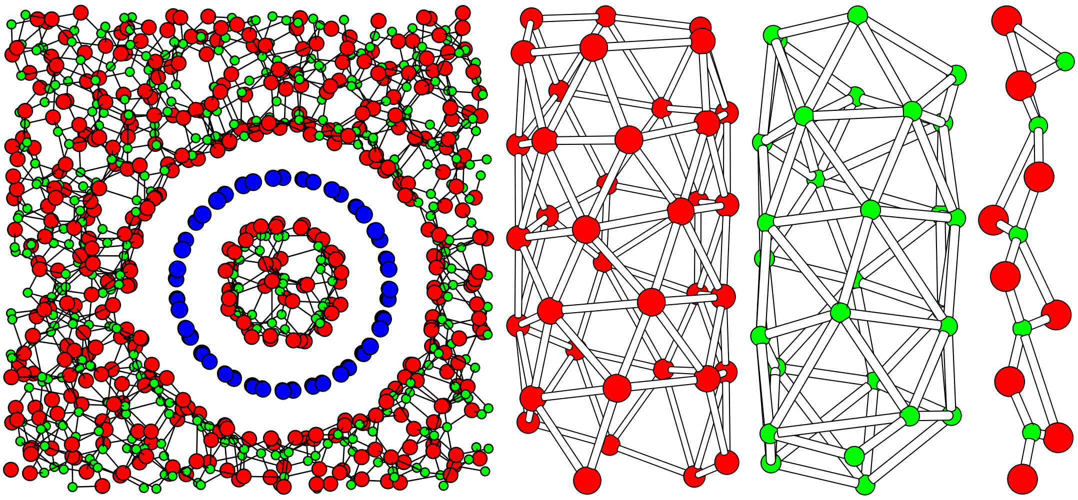
Figure 6.7: Comparison of the filling mechanisms of floating and +10 V biased (19,0) CNTs. The floating tube fills with a charge neutral loop, which is topologically constrained and could potentially direct the crystallite away from particular INT morphologies [72]. The voltage biased filling is less constrained as it is based upon independent fragments. The relative importance of initial filling mechanisms can be probed by comparing the structures arising from these two types of fillings. Green and red circles represent M and X respectively. Blue (yellow) circles denote positive (negative) charge, with the magnitude indicated by the circle's radius. The induced charges are scaled different between the two figures for ease of viewing. The largest yellow circle in (a) corresponds to a -0.0359 au Gaussian charge while the largest blue circle in (b) is a 0.2643 au Gaussian charge. The apparent difference in ion density is due to the fact that the potential-biased tube fills much more rapidly, thereby depleting the ions near the opening of the tube faster than they can be replaced by the surrounding bath.

section. These structures were then relaxed by grounding the tube (V = 0) to yield the ordinary INTs after about 50 ps of dynamics. The fraction of tubes forming each of the INT morphologies from these three growth procedures were tabulated and can be found in Table 6.3. The results unequivocally demonstrate that the INT fractionation is heavily influenced by kinetic factors. Alternative paths to the same final values of the control parameters (V = 0, T = 800 K) yield different distributions, showing that equilibrium free energy considerations cannot fully describe the results of such simulations.

6.6.1 Concentric INTs in a voltage-biased CNT

Simulations of the filling of (11,11) CNTs with a +10 V bias reveal that the ions crystallize into concentric INTs. The CNT wall is wetted by an tube composed entirely of anions, which are spaced on a hexagonal lattice so as to minimize the repulsions between like ions. This lattice is folded into a tube, which can be characterized in exactly the same manner as the charge-neutral MX tubes. The only difference is that the tube contains only one of the two hexagonal sublattices. Efficient charge-ordering causes a smaller cationic INT to form within the anionic INT, which again can be thought of structurally as the other hexagonal sublattice folded into a tube. The two INTs are not, however commensurate, a fact made clear by the fact that there are unequal numbers of anions and cations in the CNT. Within the cationic INT forms an ordered chain of anions and cations. These structures are all depicted for a single filling simulation in Figure 6.8. The outer tube is a (6,1) structure while the inner is a (5,1). This lack of commensurability comes at an energetic cost since it prevents a one-to-one pairing of anions and cations. The cost is balanced by the benefit of placing as many anions as possible near the positively charged CNT wall. The INTs result from mutual repulsion of all the anions forcing themselves into a packed, symmetric structure. Future work will focus on further characterizing these concentric INTs. It will also be interesting to probe the impact of these structures on the electrochemical response of a nanotube immersed in a molten salt. The relative stability of these INT may well introduce significant hysteresis into the capacitance, something which has been briefly explored using molecular dynamics simulation by Vatamanu *et al.* [64] and which could also be probed experimentally.

In floating tubes the total charge on the CNT was zero, introducing a symmetry between positive and negative induced charges. This symmetry ensures that for any given site, the expectation value for the charge is zero (baring some AFLRO as described earlier). The variance of these charges does, however, depend on the



(a) Cross-sectional view of simulation cell (b) X^- (6,1) tube, M^+ (5,1) tube, and MX_2 chain

Figure 6.8: Molecular dynamics snapshot of filled +10 V CNT. The systems were equilibrated for 45 ps then the ion positions were averaged for 15 ps of additional dynamics to yield the equilibrium internal structures shown in (b). The anions, which are strongly attracted to the excess positive charge on the CNT, wet the inner wall. In repelling each other, the anions adopt a highly symmetric configuration - that of a (6,1) INT. The cations form the same type of structure inside the anion tube, but as the cations have a smaller surface area to wet, they form a smaller (5,1) tube. These two tubes are not commensurate, indicating that the radial charge ordering is more important than the pairing between INT anions and cations. Inside the cationic INT rests an MX_2 chain. Early work suggests that unlike the charge-neutral INTs of Figure 6.2, defects in these INTs are common. The defects could arise from the competition between radial charge ordering induced by the CNT charge and the energetic benefit of local commensurability of the M^+ and X^- tubes. More work is required to determine if these defects anneal out.

position of the metal site along the CNT axis. Due to their smaller connectivity, edge sites tend to support greater charge magnitudes, explaining the greater charge variance. In voltage-biased tubes the charge symmetry is broken, so the reduced edge site connectivity causes extra charge to build up on the tube ends. These effects, displayed in Figure 6.9, are greatly reduced in physical CNTs, which are much longer than the simulated variety. In the short, simulated CNTs the edge effects take on greater importance. Still, Figure 6.9 shows that the anomalously large charge building up on the tube ends decays rapidly along the tube length. INT structures in the middle of the CNT can therefore reasonably be treated as good approximations for structures within longer, more physically realistic tubes.

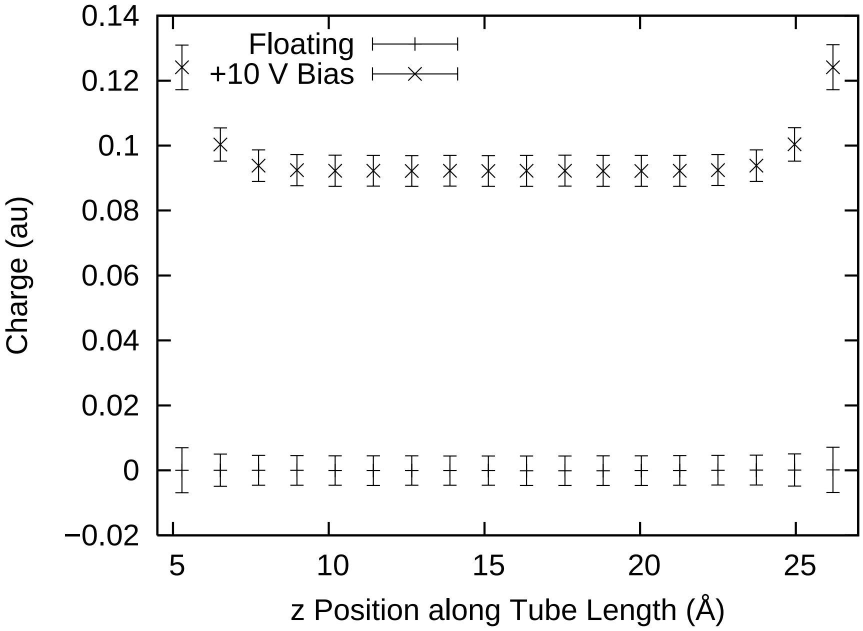


Figure 6.9: Induced charge as a function of position along the CNT axis. Floating CNTs have zero average charge since the symmetry makes any given site as likely to be positive as negative. The standard deviations of these charge distribution, however, have a weak dependence on the z position, with the end sites more likely to build up charges with larger than average magnitudes. The same broadening of the standard deviations occurs at the ends of the voltage-biased tube, but additionally the average charges deviate significantly since the bias breaks the symmetry.

6.7 Conclusions and Discussion

The simulations presented herein, though simple in nature, reveal the fundamental features of INT formation. The distinct INTs are local minima of free energy, each of which can trap the trajectory for times at least on the order of 100 ps (the computationally accessible timescale). The significant rearrangement required to interconvert between INTs suggests that the lifetime of the local minima could be effectively infinite such the multiple structures could all be experimentally observed. We have shown, however, that predicting which INT forms from a given trajectory is not possible due to the heavily chaotic nature of the dynamics (see for example Table 6.4). This forced us to consider only probability distributions of INT morphologies, which were shown to be functions of the path, not just the thermodynamic state. As path functions, these distributions could be altered by changing the trajectory of the control parameters.[4] From a technological viewpoint, it would be optimal to have

---

[4] The trajectory of the control parameters is given by the time-dependence of the macroscopic physically tunable objects like temperature, pressure, voltage bias, etc.

paths that result in the synthesis of pure INTs, but we have argued that the rugged energy landscape makes that possibility exceedingly unlikely. Baring the ability for pure synthesis, it is useful to be able to control the distribution of products so as to simplify the subsequent purification. For example, if one wanted to produce a (4,3) INT, the floating procedure is clearly favorable, whereas generating (6,0) INTs would more wisely be accomplished via the voltage-biased procedure.

The sensitivity to path also significantly challenges the theoretical understanding of these systems. It is far more difficult to adequately sample trajectory space as compared to configuration space. For equilibrium systems the relevant free energies could be computed via a metadynamics approach, but this is not an option for our path-dependent distributions. As such, experimental procedures should be most useful for understanding the effect of control parameter paths on the INT distributions. The major contribution that theory can make to this problem is merely to observe that the path will matter and that this fact could be utilized for technological gain.

In closing, let us identify an area of future work which may not suffer from the same computational and theoretical difficulties. Chaotic dynamics and incomplete sampling complicate the study of long-time events such as CNT filling, but the voltage-biased studies present some potentially fruitful areas of future work that could bypass these challenges. Rather than study the filling of the CNT, the response of an already filled, voltage-biased CNT electrode to modulations in the applied potential could be of great interest. Many of the interesting phenomena would occur on more rapid timescales, perhaps allowing for more mechanistic descriptions of the electrolyte response. These experiments - both physical and computational - would be of great interest since the CNT electrode is sufficiently small that a full double layer of the electrolyte cannot even be contained in the tube. Novel capacitive behavior may well result, particularly given the hysteresis that could be expected to arise from the relative stability of the INTs described in Section 6.6.1. However interesting, the future work is indeed future and will have to wait for another thesis and another time.

Appendix A

Gaussian Charge Ewald with the Convergence Factor Included

We now revert back to Equation (3.28) to deal with the convergence factor fully.[1] Application of Equation (3.12) with $\mathbf{r} = \mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}'$ gives

$$\begin{aligned} U(s) = \frac{1}{2\left(2\pi\right)^6 \sqrt{\pi}} \int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w}\ t^{-1/2} \sum_{i,j} Q_i Q_j \\ \times \exp\left[-\frac{st\left|\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}'\right|^2}{t+s}\right] \sum_{\mathbf{n}} \exp\left[-\left(s+t\right)\left|\mathbf{n} + \frac{t\left(\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}'\right)}{t+s}\right|^2\right] \\ \times \exp\left[-\frac{\left|\mathbf{v}\right|^2}{4\eta^2} + i\mathbf{v}\cdot\mathbf{r}'\right] \exp\left[-\frac{\left|\mathbf{w}\right|^2}{4\eta^2} + i\mathbf{w}\cdot\mathbf{r}''\right] \end{aligned} \tag{A.1}$$

Now the Jacobi imaginary transformation [Equation (3.11)] can be applied to give

$$\begin{aligned} U(s) = \frac{\pi}{2(2\pi)^6 abc} \int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w}\ t^{-1/2}\left(s+t\right)^{-3/2} \sum_{i,j} Q_i Q_j \\ \times \exp\left[-\frac{st\left|\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}'\right|^2}{s+t}\right] \exp\left[-\frac{\left|\mathbf{v}\right|^2}{4\eta^2} + i\mathbf{v}\cdot\mathbf{r}'\right] \exp\left[-\frac{\left|\mathbf{w}\right|^2}{4\eta^2} + i\mathbf{w}\cdot\mathbf{r}''\right] \\ \times \sum_{\mathbf{k}} \exp\left[-\frac{\left|\mathbf{k}\right|^2}{4\left(s+t\right)}\right] \exp\left[\frac{it\mathbf{k}\cdot\left(\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}'\right)}{t+s}\right] \end{aligned} \tag{A.2}$$

---

[1] A more compact version of this derivation is performed in a paper that has been submitted to *Chemical Physics Letters*

We first will integrate over $\mathbf{r}'$ and $\mathbf{r}''$. The x, y, and z parts of these integrals are separable.

$$U(s) = \frac{\pi}{2(2\pi)^6 abc} \int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w}\; t^{-1/2}(s+t)^{-3/2} \sum_{i,j} Q_i Q_j \exp\left[-\frac{|\mathbf{v}|^2 + |\mathbf{w}|^2}{4\eta^2}\right] \\ \times \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4(s+t)}\right] I_x I_y I_z, \tag{A.3}$$

with

$$I_x = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dx'' \exp\left[-\frac{st}{t+s}\left(x_{ij} + x'' - x'\right)^2\right] \\ \times \exp\left[i\left\{\left(v_x x' + w_x x''\right) + \frac{k_x t}{t+s}\left(x_{ij} + x'' - x'\right)\right\}\right] \tag{A.4}$$

This integral is fairly standard and can be taken by transforming into the combined and relative positions. We let $x'' - x' = x_\alpha$ and $x'' + x' = x_\beta$. Equivalently $x' = \frac{x_\beta - x_\alpha}{2}$ and $x'' = \frac{x_\beta + x_\alpha}{2}$. In terms of these new variables,

$$I_x = \int_{-\infty}^{\infty} dx_\alpha \int_{-\infty}^{\infty} dx_\beta \exp\left[-\frac{st}{s+t}\left(x_{ij} + x_\alpha\right)^2 + i\left(v_x\left(\frac{x_\beta - x_\alpha}{2}\right) + w_x\left(\frac{x_\beta + x_\alpha}{2}\right)\right)\right] \\ \times \exp\left[i\left(\frac{k_x t}{s+t}\left(x_{ij} + x_\alpha\right)\right)\right] \tag{A.5}$$

Integrating over dx_β first returns a δ-function by (3.30).

$$\begin{aligned} I_x &= \int_{-\infty}^{\infty} dx_\alpha 2\pi\delta\left(\frac{v_x + w_x}{2}\right) \exp\left[-\frac{st}{s+t}\left(x_{ij} + x_\alpha\right)^2\right] \\ &\quad \times \exp\left[i\left\{\frac{x_\alpha\left(w_x - v_x\right)}{2} + \frac{k_x t}{t+s}\left(x_{ij} + x_\alpha\right)\right\}\right] \\ &= 2\pi\delta\left(v_x + w_x\right) \int_{-\infty}^{\infty} dx_\alpha \exp\left[-\frac{st}{t+s}\left(x_{ij} + x_\alpha\right)^2\right] \\ &\quad \times \exp\left[i\left\{\frac{\left(x_{ij} + x_\alpha\right)\left(w_x - v_x\right)}{2} + \frac{k_x t}{t+s}\left(x_{ij} + x_\alpha\right)\right\}\right] e^{-\frac{i}{2}x_{ij}(w_x - v_x)} \end{aligned} \tag{A.6}$$

Now let $x = x_{ij} + x_\alpha$ to give a standard Gaussian integral, which can be computed with the normal trick of completing the square.

$$\begin{aligned} I_x &= 2\pi\delta\left(v_x + w_x\right) \exp\left[-\frac{i x_{ij}\left(w_x - v_x\right)}{2}\right] \int_{-\infty}^{\infty} dx \exp\left[-\frac{stx^2}{s+t} + i\left(\frac{w_x - v_x}{2} + \frac{k_x t}{t+s}\right)x\right] \\ &= \sqrt{\frac{\pi(s+t)}{st}} 2\pi\delta\left(v_x + w_x\right) \exp\left[-\frac{i x_{ij}\left(w_x - v_x\right)}{2} - \frac{\left[(s+t)\left(v_x - w_x\right) - 2tk_x\right]^2}{16st\left(s+t\right)}\right] \end{aligned} \tag{A.7}$$

Because I_y and I_z are of the identical form, the three integrals recombine neatly as the vector relation

$$I_x I_y I_z = \left(\frac{\pi\left(s+t\right)}{st}\right)^{3/2} \left(2\pi\right)^3 \delta\left(\mathbf{v}+\mathbf{w}\right) \exp\left[-\frac{i\mathbf{r}_{ij}\cdot\left(\mathbf{w}-\mathbf{v}\right)}{2} - \frac{\left|\left(s+t\right)\left(\mathbf{v}-\mathbf{w}\right)-2t\mathbf{k}\right|^2}{16st\left(s+t\right)}\right] \tag{A.8}$$

Inserting this expression back into Equation (A.3) and using the three dimensional δ-function to eliminate the integral over $\mathbf{w}$,

$$\begin{aligned}
U(s) &= \frac{\pi}{2\left(2\pi\right)^6 abc}\int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w}\; t^{-1/2}(s+t)^{-3/2} \sum_{i,j} Q_i Q_j \exp\left[-\frac{\left|\mathbf{v}\right|^2+\left|\mathbf{w}\right|^2}{4\eta^2}\right] \\
&\quad \times \sum_{\mathbf{k}} \exp\left[-\frac{\left|\mathbf{k}\right|^2}{4\left(s+t\right)}\right] \left(\frac{\pi\left(s+t\right)}{st}\right)^{3/2} \left(2\pi\right)^3 \delta\left(\mathbf{v}+\mathbf{w}\right) \exp\left[-\frac{i\mathbf{r}_{ij}\cdot\left(\mathbf{w}-\mathbf{v}\right)}{2}\right] \\
&\quad \times \exp\left[-\frac{\left|\left(s+t\right)\left(\mathbf{v}-\mathbf{w}\right)-2t\mathbf{k}\right|^2}{16st\left(s+t\right)}\right] \\
&= \frac{\pi^{5/2}}{2\left(2\pi\right)^3 abc}\int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{v}\; t^{-2} s^{-3/2} \sum_{i,j} Q_i Q_j \exp\left[-\frac{\left|\mathbf{v}\right|^2}{2\eta^2} + i\mathbf{r}_{ij}\cdot\mathbf{v}\right] \\
&\quad \times \sum_{\mathbf{k}} \exp\left[-\frac{\left|\mathbf{k}\right|^2}{4\left(s+t\right)}\right] \exp\left[-\frac{\left|\left(s+t\right)\mathbf{v}-t\mathbf{k}\right|^2}{4st\left(s+t\right)}\right]
\end{aligned} \tag{A.9}$$

The remaining integral is a three dimensional Gaussian integral over $\mathbf{v}$. For clarity, it can also be split up into three separable integrals.

$$U(s) = \frac{\pi^{5/2}}{2\left(2\pi\right)^3 abc}\int_0^\infty dt\; t^{-2} s^{-3/2} \sum_{i,j} Q_i Q_j \sum_{\mathbf{k}} \exp\left[-\frac{\left|\mathbf{k}\right|^2}{4(s+t)}\right] I'_x I'_y I'_z, \tag{A.10}$$

where

$$\begin{aligned}
I'_x &= \int_{-\infty}^{\infty} dv_x \exp\left[-\frac{v_x^2}{2\eta^2} + i x_{ij} v_x\right] \exp\left[-\frac{\left(\left(s+t\right)v_x - t k_x\right)^2}{4st\left(s+t\right)}\right] \\
&= \left(\frac{\pi st}{2st+\left(s+t\right)\eta^2}\right)^{1/2} 2\eta \exp\left[\frac{2\left(s+t\right)\eta^2 i t k_x x_{ij} - t^2 k_x^2 - 2st\left(s+t\right)x_{ij}^2\eta^2}{2\left(s+t\right)\left(2st+\left(s+t\right)\eta^2\right)}\right]
\end{aligned} \tag{A.11}$$

As before this can be expressed compactly in vector form.

$$\begin{aligned}
I'_x I'_y I'_z &= \left(\frac{\pi st}{2st+\left(s+t\right)\eta^2}\right)^{3/2} 8\eta^3 \exp\left[-\frac{t^2\left|\mathbf{k}\right|^2}{2\left(s+t\right)\left(2st+\left(s+t\right)\eta^2\right)}\right] \\
&\quad \times \exp\left[\frac{i\eta^2 t\mathbf{k}\cdot\mathbf{r}_{ij}}{2st+\left(s+t\right)\eta^2}\right] \exp\left[-\frac{st\left|\mathbf{r}_{ij}\right|^2\eta^2}{2st+\left(s+t\right)\eta^2}\right]
\end{aligned} \tag{A.12}$$

Inserting this expression back into Equation (A.10) gives

$$
\begin{aligned}
U(s) = {} & \frac{\pi\eta^3}{2abc}\int_0^\infty dt\ t^{-1/2}\left(2t+(s+t)\,\eta^2\right)^{-3/2}\sum_{i,j}Q_iQ_j\sum_{\mathbf{k}}\exp\left[-\frac{|\mathbf{k}|^2}{4\,(s+t)}\right] \\
& \exp\left[-\frac{t^2\,|\mathbf{k}|^2}{2\,(s+t)\,(2st+(s+t)\,\eta^2)}+\frac{i\eta^2 t\mathbf{k}\cdot\mathbf{r}_{ij}}{2st+(s+t)\,\eta^2}-\frac{st\,|\mathbf{r}_{ij}|^2\,\eta^2}{2st+(s+t)\,\eta^2}\right] \\
= {} & \frac{\pi\eta^3}{2abc}\int_0^\infty dt\ t^{-1/2}\left(2s+(s+t)\,\eta^2\right)^{-3/2}\sum_{i,j,\mathbf{k}}Q_iQ_j\exp\left[-\frac{|\mathbf{k}|^2\,(2t+\eta^2)}{4\,(2st+(s+t)\,\eta^2)}\right] \\
& \exp\left[\frac{i\eta^2 t\mathbf{k}\cdot\mathbf{r}_{ij}}{2st+(s+t)\,\eta^2}\right]\exp\left[\frac{-st\,|\mathbf{r}_{ij}|^2\,\eta^2}{2st+(s+t)\,\eta^2}\right]
\end{aligned}
\tag{A.13}
$$

Now consider the change of variables $t' = \eta^2 t/\left(2t+\eta^2\right)$, inspired by the $s=0$ transformation used in Section 3.4. When the transformation was first invoked, it was observed that $t^{-2}dt = t'^{-2}dt'$. It is useful to develop a few other properties of this particular transformation. Note first that the transformation can be inverted to give $t = \eta^2 t'/\left(\eta^2-2t'\right)$. Algebraic manipulation also yields

$$
2st+(s+t)\eta^2 = \frac{2s\eta^2 t'}{\eta^2-2t'}+\frac{s\eta^4-2s\eta^2 t'}{\eta^2-2t'}+\frac{\eta^4 t'}{\eta^2-2t'} = \frac{(s+t')\eta^4}{\eta^2-2t'}. \tag{A.14}
$$

Using (A.14) and some more algebra, gives three more useful relations.

$$
\begin{aligned}
t^{-1/2}\left(2st+(s+t)\,\eta^2\right)^{-3/2}dt &= \left[\frac{t'^{-1/2}\left(\eta^2-2t'\right)^{1/2}}{\eta}\right]\left[\frac{\left(\eta^2-2t'\right)}{(s+t')\,\eta^4}\right]^{3/2}\left(\frac{t}{t'}\right)^2 dt' \\
&= t'^{-1/2}\left(s+t'\right)^{-3/2}\eta^{-7}\left(\eta^2-2t'\right)^2\left(\frac{t}{t'}\right)^2 dt' \\
&= t'^{-1/2}\left(s+t'\right)^{-3/2}\eta^{-7}\left(\eta^2 t'\right)^2 t'^{-2}dt' \\
&= t'^{-1/2}\left(s+t'\right)^{-3/2}\eta^{-3}dt'
\end{aligned}
\tag{A.15}
$$

$$
\frac{\eta^2 t}{2st+(s+t)\,\eta^2} = \eta^2\left(\frac{\eta^2 t'}{\eta^2-2t'}\right)\left(\frac{\eta^2-2t'}{(s+t')\,\eta^4}\right) = \frac{t'}{s+t'} \tag{A.16}
$$

$$
\begin{aligned}
\frac{2t+\eta^2}{4\,(2st+(s+t)\,\eta^2)} &= \frac{\left(2t+\eta^2\right)\left(\eta^2-2t'\right)}{4\,(s+t')\,\eta^4} \\
&= \frac{\eta^2 t\left(\eta^2-2t'\right)}{4t'\,(s+t')\,\eta^4} \\
&= \left(\frac{t}{4\,(s+t')}\right)\left(\frac{\eta^2-2t'}{t'\eta^2}\right) \\
&= \left(\frac{t}{4(s+t')}\right)\left(\frac{1}{t}\right) = \frac{1}{4(s+t')}
\end{aligned}
\tag{A.17}
$$

Using these properties to convert the integration from t to t', one obtains

$$\begin{aligned} U(s) = \frac{\pi}{2abc}\int_0^{\eta^2/2} dt'\; t'^{-1/2}(s+t')^{-3/2}\sum_{i,j} Q_iQ_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4\,(s+t')}\right] \exp\left[\frac{i\mathbf{k}\cdot\mathbf{r}_{ij}t'}{s+t'}\right] \\ \exp\left[-\frac{s\,|\mathbf{r}_{ij}|^2\,t'}{s+t'}\right] \end{aligned} \tag{A.18}$$

The integrand is identical to that of Equation (3.14) combined with Equation (3.19) (the $\mathbf{k} = 0$ term). The only difference is the upper bound of integration, which has changed from α^2 to $\eta^2/2$. Equation (A.18) can then be computed in the $s \to 0$ limit with the results of equations (3.17) and (3.21).

$$\lim_{s\to 0} U(s) = \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}\exp\left(-\frac{k^2}{2\eta^2}\right)|S\,(\mathbf{k})|^2 - \frac{\pi}{3V}\sum_{i,j} Q_iQ_j\,|\mathbf{r}_{ij}|^2\,, \tag{A.19}$$

which is precisely the result of Equation (3.33) from the naive $s \to 0$ approach. As in Equation (3.33), this energy is entirely in reciprocal space, so we introduce a partition of the t' integration at α^2. Since the integrand of Equation (A.18) is of the same form as the integrand of equations (3.14) and (3.19), it follows that inverting the Jacobi imaginary transformation will yield an integrand of the form of Equation (3.8).

$$\begin{aligned} &\frac{\pi}{2abc}\int_{\alpha^2}^{\eta^2/2} dt'\; t'^{-1/2}(s+t')^{-3/2}\sum_{i,j} Q_iQ_j \sum_{\mathbf{k}} \exp\left[-\frac{|\mathbf{k}|^2}{4\,(s+t')} + \frac{i\mathbf{k}\cdot\mathbf{r}_{ij}t'}{s+t'} - \frac{s\,|\mathbf{r}_{ij}|^2\,t'}{s+t'}\right] \\ &\quad = \frac{1}{2}\int_{\alpha^2}^{\eta^2/2} dt'\; t'^{-1/2}\sum_{i,j} Q_iQ_j \sum_{\mathbf{n}} \exp\left(-t'\,|\mathbf{r}_{ij}+\mathbf{n}|^2\right)\exp\left(-s\,|\mathbf{n}|^2\right) \end{aligned} \tag{A.20}$$

The $s \to 0$ limit is trivial since it is independent of the integration. Hence Equation (3.35) is confirmed to be the correct $s \to 0$ limit. In summary,

$$\begin{aligned} U = \lim_{s\to 0} U(s) = &\frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}\exp\left(-\frac{k^2}{4\alpha}\right)|S\,(\mathbf{k})|^2 + \left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}}\right)\sum_i Q_i^2 \\ &+ \frac{1}{2}\sum_{\mathbf{n}}\sum_{i\neq j} Q_iQ_j \left(\frac{\operatorname{erfc}\left(\alpha\,|\mathbf{r}_{ij}+\mathbf{n}|\right) - \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}}\,|\mathbf{r}_{ij}+\mathbf{n}|\right)}{|\mathbf{r}_{ij}+\mathbf{n}|}\right) \\ &- \frac{\pi}{3V}\sum_{i,j} Q_iQ_j\,|\mathbf{r}_{ij}|^2\,, \end{aligned} \tag{A.21}$$

exactly as found with the naive $s \to 0$ calculation.

Appendix B

Gaussian Charge Ewald with 2d Periodic Replication

Study of interfaces, particularly charged interfaces, requires the simulation cell to adopt a slab geometry. The calculation of the electrostatic energy of a slab was reported by de Leeuw and Perram assuming point charges. [14] The calculation was extended to Gaussian charges by Reed *et al.* [40] Reed's derivation is nearly correct, but a critical algebraic error causes his reported energy to actually just be the energy of a point charge system plus the self-interaction energy of each Gaussian charge. Here the entire derivation is reproduced, corrected, and (hopefully) made more clear. For greatest clarity, the reader is strongly urged to first review the Gaussian calculation with three periodically replicated directions of Sections 3.3, 3.4, and 3.5.

The Coulomb energy can be written down in a straightforward manner including a Gaussian normalization and a factor of 1/2 to prevent double counting.

$$U = \frac{\eta^6}{2\pi^3} \sum_{i,j} \sum_{k,l=-\infty}^{\infty} \int d\mathbf{r}' \int d\mathbf{r}'' \frac{q_i \exp\left(\left|\mathbf{r}' - \mathbf{r}_i\right|^2 \eta^2\right) q_j \exp\left(\left|\mathbf{r}'' - \mathbf{r}_j\right| \eta^2\right)}{\left|\mathbf{r}_{ij} + \mathbf{r}'' - \mathbf{r}' + k\mathbf{a} + l\mathbf{b}\right|} \tag{B.1}$$

Proceeding in the same manner as for the system periodic in three dimensions,

$$\begin{aligned} U = & \frac{\pi^{3/2}}{2\left(2\pi\right)^6} \int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w}\ t^{-1/2} \sum_{i,j} \sum_{k,l=-\infty}^{\infty} q_i q_j \\ & \times \exp\left[-t\left(x_{ij} + x'' - x' + ka\right)^2 - t\left(y_{ij} + y'' - y' + lb\right)^2 - t\left(z_{ij} + z'' - z'\right)^2\right] \\ & \times \exp\left[-\frac{\left|\mathbf{v}\right|^2}{4\eta^2} + i\mathbf{v}\cdot\mathbf{r}'\right] \exp\left[-\frac{\left|\mathbf{w}\right|^2}{4\eta^2} + i\mathbf{w}\cdot\mathbf{r}''\right], \end{aligned} \tag{B.2}$$

which is (3.28) with $s = 0$ and no periodic replicas in the z direction. By completing the square, it can be observed that

$$\frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} du \exp\left(-\frac{u^2}{4t}+iuz\right) = \frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} du \exp\left[-\left(\frac{u}{2\sqrt{t}}-iz\sqrt{t}\right)^2\right]\exp\left(z^2 t\right)$$
$$= \exp\left(z^2 t\right) \tag{B.3}$$

This identity is applied to the exponential of the z coordinate in Equation (B.2).

$$\begin{aligned}U =& \frac{\pi}{4\left(2\pi\right)^6}\int_0^{\infty} dt \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w} \int_{-\infty}^{\infty} du\; t^{-1}\sum_{i,j}\sum_{k,l=-\infty}^{\infty} q_i q_j \\ &\times \exp\left[-t\left(x_{ij}+x''-x'+ka\right)^2 - t\left(y_{ij}+y''-y'+lb\right)^2 - \frac{u^2}{4t} + iu\left(z_{ij}+z''-z'\right)\right] \\ &\times \exp\left[-\frac{\left|\mathbf{v}\right|^2}{4\eta^2}+i\mathbf{v}\cdot\mathbf{r}'\right]\exp\left[-\frac{\left|\mathbf{w}\right|^2}{4\eta^2}+i\mathbf{w}\cdot\mathbf{r}''\right]\end{aligned} \tag{B.4}$$

Application of the Jacobi imaginary transformation, Equation (3.11), to the x and y exponentials gives (A7) of Reed *et al.* [40]

$$\begin{aligned}U =& \frac{1}{4ab\left(2\pi\right)^6}\int_0^{\infty} dt \int_{\mathbb{R}^3} d\mathbf{r}' \int_{\mathbb{R}^3} d\mathbf{r}'' \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w} \int_{-\infty}^{\infty} du\; t^{-2}\sum_{i,j}\sum_{k,l=-\infty}^{\infty} q_i q_j \\ &\times \exp\left[-\frac{\pi^2 k^2}{a^2 t}+\frac{i2\pi k}{a}\left(x_{ij}+x''-x'+ka\right)-\frac{\pi^2 l^2}{b^2 t}+\frac{i2\pi l}{b}\left(y_{ij}+y''-y'+lb\right)\right] \\ &\times \exp\left[-\frac{u^2}{4t}+iu\left(z_{ij}+z''-z'\right)\right]\exp\left[-\frac{\left|\mathbf{v}\right|^2+\left|\mathbf{w}\right|^2}{4\eta^2}+i\left(\mathbf{v}\cdot\mathbf{r}'+\mathbf{w}\cdot\mathbf{r}''\right)\right]\end{aligned} \tag{B.5}$$

As in the case of three dimensional replication, we use Equation (3.30) to integrate over $\mathbf{r}'$ and $\mathbf{r}''$.

$$\begin{aligned}U =& \frac{1}{4ab}\int_0^{\infty} dt \int_{\mathbb{R}^3} d\mathbf{v} \int_{\mathbb{R}^3} d\mathbf{w} \int_{-\infty}^{\infty} du\; t^{-2}\sum_{i,j} q_i q_j \sum_{k,l=-\infty}^{\infty} exp\left[-\frac{\left|\mathbf{v}\right|^2+\left|\mathbf{w}\right|^2}{4\eta^2}\right] \\ &\times \exp\left[-\frac{\pi^2 k^2}{a^2 t}+\frac{i2\pi k x_{ij}}{a}\right]\exp\left[-\frac{\pi^2 l^2}{b^2 t}+\frac{i2\pi l y_{ij}}{b}\right]\exp\left[-\frac{u^2}{4t}+iuz_{ij}\right] \\ &\times \delta\left(v_x-\frac{2\pi k}{a}\right)\delta\left(v_y-\frac{2\pi l}{b}\right)\delta\left(v_z-u\right)\delta\left(w_x+\frac{2\pi k}{a}\right)\delta\left(w_y+\frac{2\pi l}{b}\right)\delta\left(w_z+u\right)\end{aligned} \tag{B.6}$$

Integration over $\mathbf{v}$ and $\mathbf{w}$ is now trivial on account of the delta functions.

$$\begin{aligned}U =& \frac{1}{4ab}\int_0^{\infty} dt \int_{-\infty}^{\infty} du\; t^{-2}\sum_{i,j} q_i q_j \sum_{k,l=-\infty}^{\infty} \exp\left[-\frac{\pi^2 k^2}{a^2}\left(\frac{1}{t}+\frac{2}{\eta^2}\right)+\frac{i2\pi k x_{ij}}{a}\right] \\ &\times \exp\left[-\frac{\pi^2 l^2}{b^2}\left(\frac{1}{t}+\frac{2}{\eta^2}\right)+\frac{i2\pi l y_{ij}}{b}\right]\exp\left[-\frac{u^2}{4}\left(\frac{1}{t}+\frac{2}{\eta^2}\right)+iuz_{ij}\right],\end{aligned} \tag{B.7}$$

which is just (A10) of Reed. It is at this point that we depart slightly from Reed's approach. Reed partitioned the integral over t into an integral over $[0, \alpha^2]$ and another over $[\alpha^2, \infty)$. Rather than partitioning immediately, we first change variables from t to t' as in Equation (3.31).

$$U = \frac{1}{4ab} \int_0^{\eta^2/2} dt' \int_{-\infty}^{\infty} du\ t'^{-2} \sum_{ij} q_i q_j \sum_{k,l=-\infty}^{\infty} \exp\left[-\frac{\pi^2}{t'}\left(\frac{k^2}{a^2} + \frac{l^2}{b^2}\right)\right] \exp\left[i2\pi\left(\frac{kx_{ij}}{a} + \frac{ly_{ij}}{b}\right) - \frac{u^2}{4t'} + iuz_{ij}\right] \tag{B.8}$$

Now we partition the t' integral at α^2 so that the $[0, \alpha^2]$ and $\left[\alpha^2, \frac{\eta^2}{2}\right]$ terms can be treated in reciprocal and real space respectively.

$$\begin{aligned} U = {} & \frac{1}{4ab} \int_0^{\alpha^2} dt' \int_{-\infty}^{\infty} du\ t'^{-2} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa} \neq 0} \exp\left[-\frac{\left(|\boldsymbol{\kappa}|^2 + u^2\right)}{4t'} + i\left(\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij} + uz_{ij}\right)\right] \\ & + \frac{1}{4ab} \int_0^{\alpha^2} dt' \int_{-\infty}^{\infty} du\ t'^{-2} \sum_{i,j} q_i q_j \exp\left[-\frac{u^2}{4t'} + iuz_{ij}\right] \\ & + \frac{1}{4ab} \int_{\alpha^2}^{\eta^2/2} dt' \int_{-\infty}^{\infty} du\ t'^{-2} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa}} \exp\left[-\frac{\left(|\boldsymbol{\kappa}|^2 + u^2\right)}{4t'} + i\left(\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij} + uz_{ij}\right)\right] \\ \equiv {} & U_a + U_b + U_c, \end{aligned} \tag{B.9}$$

where $\boldsymbol{\kappa} = \left(\frac{2\pi k}{a}, \frac{2\pi l}{b}\right)$ and $\boldsymbol{\rho}_{ij} = (x_{ij}, y_{ij})$. Because the $\kappa = 0$ term diverges, it is treated separately.

At this point there are two clear differences from Reed. First, the partition is different, but this partition is merely a computational tool. If the sums are not truncated, then the energy is entirely independent of the partition. One is free to shift the partition as desired such that the optimal convergence of the sums can be achieved. Later it will be clear why partitioning of t' at α^2 is advantageous. The second difference is the upper limit of integration. Reed's result is simply incorrect. An algebraic error was made in the transformation from t to t', and this error significantly affects the form of the final result.[1] In fact, the error serves to remove the Gaussian nature of the charge, treating everything instead as point charges.

---

[1] The degree to which the error affects numerical results depends on the parameters of the simulation cell. It is quite possible that in some practical situations the error's impact is negligible.

B.1 $U_a(s)$

First we consider the $\boldsymbol{\kappa} \neq 0$ contribution of the reciprocal space term.

$$U_a = \frac{1}{4ab} \int_0^{\alpha^2} dt' \int_{-\infty}^{\infty} du\ t'^{-2} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa} \neq 0} \exp\left[-\frac{|\boldsymbol{\kappa}|^2 + u^2}{4t'} + i\left(\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij} + u z_{ij}\right) \right]$$

The integral over t' is trivial.

$$\begin{aligned}
U_a &= \frac{1}{4ab} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa} \neq 0} \exp\left(i\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij}\right) \exp\left(-\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right) \int_{-\infty}^{\infty} du \frac{\exp\left[-\left(\frac{u^2}{4\alpha^2}\right) + iuz_{ij}\right]}{\left[\frac{|\boldsymbol{\kappa}|^2}{4} + \frac{u^2}{4}\right]} \\
&= \frac{1}{ab} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa} \neq 0} \exp\left(i\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij}\right) \exp\left(-\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right) \int_{-\infty}^{\infty} du \frac{\exp\left[-\left(\frac{u^2}{4\alpha^2}\right) + iuz_{ij}\right]}{|\boldsymbol{\kappa}|^2 + u^2} \\
&= \frac{1}{ab} \sum_{\boldsymbol{\kappa} \neq 0} \int_{-\infty}^{\infty} \frac{du}{|\boldsymbol{\kappa}|^2 + u^2} \exp\left(-\frac{u^2 + |\boldsymbol{\kappa}|^2}{4\alpha^2}\right) \left| \sum_i q_i \exp\left[i\left(\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij} + uz_i\right)\right] \right|^2 \\
&= \frac{1}{ab} \sum_{\boldsymbol{\kappa} \neq 0} \int_{-\infty}^{\infty} \frac{du}{|\boldsymbol{\kappa}|^2 + u^2} \exp\left(-\frac{u^2 + |\boldsymbol{\kappa}|^2}{4\alpha^2}\right) \left|S\left(\boldsymbol{\kappa}, u\right)\right|^2, \qquad \text{(B.10)}
\end{aligned}$$

with $S\left(\boldsymbol{\kappa}, u\right) \equiv \sum_i q_i \exp\left[i\left(\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij} + uz_i\right)\right]$. This is exactly the form of the energy one would expect with a point charge system periodic in two directions. Of course the energy of the Gaussian system is not the same as that of a point charge system, but this difference is entirely handled in the real space part part of the energy. This makes sense because it is only at short range that point charges look different from Gaussians.

One can also compute the integral over u of Equation (B.10), but doing so will require knowing all of the $\mathbf{r}_{ij}$. Therefore the computation scales like N^2. For this reason Kawata and Mikami have suggested that numerical integration of the $\mathcal{O}(N)$ structure factor will be advantageous. For completeness, we now compute the final integral to get a closed form $\mathcal{O}(N^2)$ solution.

$$U_a = \frac{1}{ab} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa} \neq 0} \exp\left(i\boldsymbol{\kappa} \cdot \boldsymbol{\rho}_{ij}\right) \exp\left(-\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right) \int_{-\infty}^{\infty} du \frac{\exp\left[-\left(\frac{u^2}{4\alpha^2}\right) + iuz_{ij}\right]}{|\boldsymbol{\kappa}|^2 + u^2}$$

But as shown in Equation (B.3),

$$\exp\left(-\frac{u^2}{4\alpha^2}\right) = \frac{\alpha}{\sqrt{\pi}} \int_{-\infty}^{\infty} dt \exp\left(-\alpha^2 t^2 + itu\right).$$

Using this alternative representation of the exponential,

$$U_a = \frac{\alpha}{ab\sqrt{\pi}} \sum_{i,j,\boldsymbol{\kappa}\neq 0} q_i q_j \exp\left(i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_{ij}\right) \exp\left(-\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right) \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} du \frac{\exp\left[-\alpha^2 t^2 + iu\left(z_{ij}+t\right)\right]}{|\boldsymbol{\kappa}|^2 + u^2}$$

The integral over u is performed first using a semicircular contour of infinite radius. The integrand has simple poles at $u = \pm i\,|\boldsymbol{\kappa}|$. Whether the upper half plane (UHP) or lower half plane (LHP) contour is used depends on the exponential in the integrand, which will vanish in one half-plane and explode in the other. Observe that if $z_{ij}+t > 0$ the integrand vanishes for u in the UHP, whereas if $z_{ij}+t < 0$ the integrand vanishes in the LHP. This means that the integral over t must be split into these two cases. The Cauchy Residue Theorem is now simply applied as the radial contour's contribution vanishes in the $R \to \infty$ limit.

$$\begin{aligned} U_a =& \frac{\alpha}{ab\sqrt{\pi}} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa}\neq 0} \exp\left(-\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2} + i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_{ij}\right) \\ &\times \left[-\int_{-\infty}^{-z_{ij}} dt\ e^{-\alpha^2 t^2} \frac{2\pi i \exp\left[|\boldsymbol{\kappa}|\left(z_{ij}+t\right)\right]}{-i\,|\boldsymbol{\kappa}| - i\,|\boldsymbol{\kappa}|} + \int_{-z_{ij}}^{\infty} dt\ e^{-\alpha^2 t^2} \frac{2\pi i \exp\left[-|\boldsymbol{\kappa}|\left(z_{ij}+t\right)\right]}{i\,|\boldsymbol{\kappa}| + i\,|\boldsymbol{\kappa}|}\right] \end{aligned}$$

We consider the first of these integrals.

$$\begin{aligned} I_1 &\equiv \int_{-\infty}^{-z_{ij}} dt \frac{\alpha}{\sqrt{\pi}} \exp\left[-\alpha^2 t^2\right] \frac{2\pi \exp\left(|\boldsymbol{\kappa}|\left(z_{ij}+t\right)\right)}{2\,|\boldsymbol{\kappa}|} \\ &= \frac{\pi\alpha \exp\left(|\boldsymbol{\kappa}|\, z_{ij}\right)}{2\,|\boldsymbol{\kappa}|} \left(\frac{2}{\sqrt{\pi}}\right) \int_{-\infty}^{-z_{ij}} dt \exp\left[-\alpha^2 t^2 + |\boldsymbol{\kappa}|\, t\right] \\ &= \frac{\pi\alpha \exp\left(|\boldsymbol{\kappa}|\, z_{ij}\right)}{2\,|\boldsymbol{\kappa}|} \left(\frac{2}{\sqrt{\pi}}\right) \int_{-\infty}^{-z_{ij}} dt \exp\left[-\left(\alpha t - \frac{|\boldsymbol{\kappa}|}{2\alpha}\right)^2\right] \exp\left[\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right] \\ &= \frac{\pi \exp\left(|\boldsymbol{\kappa}|\, z_{ij}\right)}{2\,|\boldsymbol{\kappa}|} \exp\left[\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right] \left(\frac{2}{\sqrt{\pi}}\right) \int_{-\infty}^{-\alpha z_{ij} - \frac{|\boldsymbol{\kappa}|}{2\alpha}} dv \exp\left[-v^2\right] \\ &= \frac{\pi \exp\left(|\boldsymbol{\kappa}|\, z_{ij}\right)}{2\,|\boldsymbol{\kappa}|} \exp\left[\frac{|\boldsymbol{\kappa}|^2}{4\alpha^2}\right] \operatorname{erfc}\left[\alpha z_{ij} + \frac{|\boldsymbol{\kappa}|}{2\alpha}\right] \end{aligned}$$

The second integral can be handled analogously to give

$$\begin{aligned} U_a = \frac{\pi}{2ab} \sum_{i,j} q_i q_j \sum_{\boldsymbol{\kappa}\neq 0} \frac{\exp\left(i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_{ij}\right)}{|\boldsymbol{\kappa}|} \Bigg[&\exp\left(-|\boldsymbol{\kappa}|\, z_{ij}\right) \operatorname{erfc}\left(\frac{|\boldsymbol{\kappa}|}{2\alpha} - \alpha z_{ij}\right) \\ &+ \exp\left(|\boldsymbol{\kappa}|\, z_{ij}\right) \operatorname{erfc}\left(\frac{|\boldsymbol{\kappa}|}{2\alpha} + \alpha z_{ij}\right)\Bigg] \end{aligned} \tag{B.11}$$

B.2 $U_b(s)$

Now we consider the divergent $\boldsymbol{\kappa} = 0$ term. This really should be done carefully with a convergence factor imposing a radial order of summation, but for now the slightly sloppy approach of de Leeuw and Perram suffices. [14] The complete convergence factor calculation for a point charge system is carried out by Bródka *et al.*, justifying the result of this lax derivation. [26]

$$\begin{aligned} U_b &= \frac{1}{4ab}\int_0^{\alpha^2} dt' \int_{-\infty}^{\infty} du\ t'^{-2}\sum_{i,j} q_i q_j \exp\left[-\frac{u^2}{4t'} + iuz_{ij}\right] \\ &= \frac{\sqrt{\pi}}{2ab}\int_0^{\alpha^2} dt'\ t'^{-3/2}\sum_{i,j} q_i q_j \exp\left(-z_{ij}^2 t\right) \\ &= \frac{\sqrt{\pi}}{2ab}\sum_{i,j} q_i q_j \left[\int_0^{\alpha^2} dt'\ t'^{-3/2}\left[\exp\left(-z_{ij}^2 t\right) - 1\right] + \int_0^{\alpha^2} dt'\ t'^{-3/2}\right] \end{aligned} \tag{B.12}$$

While the second term diverges, it is also multiplied by $\sum_i q_i$, which is zero for a neutral system. The use of convergence factors can formally justify the fact that this term will vanish in a charge neutral system. The argument is analogous to the one used to compute U_3 in Section 3.2.3.

$$U_b = \frac{\sqrt{\pi}}{2ab}\int_{i,j} q_i q_j \int_0^{\alpha^2} dt'\ t'^{-3/2}\left[\exp\left(-z_{ij}^2 t\right) - 1\right]$$

Integrating by parts with $u = \left[\exp\left(-z_{ij}^2 t'\right) - 1\right]$ and $dv = dt\ t'^{-3/2}$ yields

$$\begin{aligned} U_b &= \frac{\sqrt{\pi}}{2ab}\sum_{i,j} q_i q_j \left[-2t'^{-1/2}\left[\exp\left(-z_{ij}^2 t'\right) - 1\right]\Big|_0^{\alpha^2} - 2\int_0^{\alpha^2} dt'\ t'^{-1/2} z_{ij}^2 \exp\left(-z_{ij}^2 t\right)\right] \\ &= \frac{\sqrt{\pi}}{2ab}\sum_{i,j} q_i q_j \left[-\frac{2\left[\exp\left(-z_{ij}^2\alpha^2\right) - 1\right]}{\alpha} - 2z_{ij}\int_0^{\alpha^2} dt'\ t'^{-1/2} z_{ij} \exp\left(-z_{ij}^2 t\right)\right] \\ &= \frac{\sqrt{\pi}}{2ab}\sum_{i,j} q_i q_j \left[\frac{-2\left[\exp\left(-z_{ij}^2\alpha^2\right) - 1\right]}{\alpha} - 4z_{ij}\int_0^{z_{ij}\alpha} du \exp\left(-u^2\right)\right] \\ &= -\frac{\sqrt{\pi}}{ab}\sum_{i,j} q_i q_j \left[\frac{\exp\left(-z_{ij}^2\alpha^2\right)}{\alpha} + z_{ij}\sqrt{\pi}\operatorname{erf}\left(z_{ij}\alpha\right)\right] \end{aligned} \tag{B.13}$$

This is (A18) of Reed *et al.* with two small exceptions. Reed's sum excludes the $i = j$ term, an error that would prevent the charge neutrality cancellations. Bródka reports the same expression derived here. The other difference is that Reed's (A18) has

$\beta = \eta\alpha/\sqrt{\eta^2 + 2\alpha^2}$ in place of α. This difference stems from the different partitioning of the t integral. The effect of the difference is to shift more of the computation from real space to reciprocal space. This computational shift is not desirable because it will slow the convergence of the reciprocal space sum. One may worry that the method presented here will merely slow the convergence of the real space sum. As will be seen, the real space sum contains $\mathrm{erfc}(\alpha z_{ij})$ terms, the exact same type of terms that would appear in a point charge calculation. So if α is chosen to make the real space contributions drop off sufficiently rapidly for a minimum image convention, then this value of α will also result in a sufficiently rapidly converging real space sum for Gaussian charges.

B.3 $U_c(s)$

Finally we handle the real space term, U_c as defined in Equation (B.9). Inverting both the Jacobi imaginary transformation and the identity of Equation (B.3) gives

$$\begin{aligned}
U_c &= \frac{1}{2\sqrt{\pi}} \int_{\alpha^2}^{\eta^2/2} dt'\ t'^{-1/2} \sum_{i,j} q_i q_j \sum_{\mathbf{n}} \exp\left(-t' \left|\boldsymbol{\rho}_{ij} + \mathbf{n}\right|^2\right) \exp\left(-z_{ij}^2 t'\right) \\
&= \frac{1}{2} \left\{ \int_{\alpha^2}^{\infty} dt'\ t'^{-1/2} {\sum_{i,j,\mathbf{n}}}' q_i q_j \exp\left[-t'\left(\left|\boldsymbol{\rho}_{ij} + \mathbf{n}\right|^2 + z_{ij}^2\right)\right] \right. \\
&\quad \left. - \int_{\eta^2/2}^{\infty} dt'\ t'^{-1/2} {\sum_{i,j,\mathbf{n}}}' q_i q_j \exp\left[-t'\left(\left|\boldsymbol{\rho}_{ij} + \mathbf{n}\right|^2 + z_{ij}^2\right)\right] + \int_{\alpha^2}^{\eta^2/2} dt'\ t'^{-1/2} \sum_i q_i \right\} \\
&= \frac{1}{2} \left\{ \sum_{i \neq j} q_i q_j \left[\frac{\mathrm{erfc}\left(\alpha \left|\mathbf{r}_{ij}\right|\right) - \mathrm{erfc}\left(\frac{\eta}{\sqrt{2}} \left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} \right] + \sum_i q_i^2 \left(\frac{\eta}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{\pi}} \right) \right\} \qquad \text{(B.14)}
\end{aligned}$$

To obtain the final result, Equation (3.9) was used for $i \neq j$ terms assuming a minimum image convention in the x and y directions.

Appendix C

Gradient of Coulomb Energy

The Coulomb energy was derived in the main text, but it is often necessary to have analytic forms of the gradient. For reference, we collect these expressions here.

$$\begin{aligned}
U_T = {} & \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi e^{-k^2/4\alpha^2}}{k^2}\left|S_1\left(\mathbf{k}\right)+S_2\left(\mathbf{k}\right)\right|^2 + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} q_i q_j \frac{\mathrm{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} \\
& + \sum_{i=1}^{n}\sum_{j=1}^{N} q_i Q_j \frac{\mathrm{erfc}\left(\alpha\left|\mathbf{r}_i-\mathbf{R}_j\right|\right)-\mathrm{erfc}\left(\eta\left|\mathbf{r}_i-\mathbf{R}_j\right|\right)}{\left|\mathbf{r}_i-\mathbf{R}_j\right|} \\
& + \frac{1}{2}\sum_{i=1}^{N}{\sum_{j=1}^{N}}' Q_i Q_j \frac{\mathrm{erfc}\left(\alpha\left|\mathbf{R}_{ij}\right|\right)-\mathrm{erfc}\left(\frac{\eta}{\sqrt{2}}\left|\mathbf{R}_{ij}\right|\right)}{\left|\mathbf{R}_{ij}\right|} \\
& + \left(\frac{\eta}{\sqrt{2\pi}}-\frac{\alpha}{\sqrt{\pi}}\right)\sum_{i=1}^{N} Q_i^2 - \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{n} q_i^2 - V_0\sum_{i} Q_i
\end{aligned} \tag{C.1}$$

The gradient of U_T can be broken up into contributions from the erfc terms and from the Fourier series terms. All erfc terms dependent on $\mathbf{r}_i$ will contribute a term to the gradient of the form

$$\begin{aligned}
\frac{\partial}{\partial\mathbf{r}_i}\left(\frac{\mathrm{erfc}(\alpha\left|\mathbf{r}_i-\mathbf{r}_j\right|)}{\left|\mathbf{r}_i-\mathbf{r}_j\right|}\right) &= \frac{\left[\frac{\partial}{\partial\mathbf{r}_i}\mathrm{erfc}(\alpha\left|\mathbf{r}_i-\mathbf{r}_j\right|)\right]\left|\mathbf{r}_i-\mathbf{r}_j\right|-\mathrm{erfc}(\alpha\left|\mathbf{r}_i-\mathbf{r}_j\right|)}{\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}\frac{\mathbf{r}_i-\mathbf{r}_j}{\left|\mathbf{r}_i-\mathbf{r}_j\right|} \\
&= \frac{-\frac{2\alpha}{\sqrt{\pi}}e^{-\alpha^2\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}\left|\mathbf{r}_i-\mathbf{r}_j\right|-\mathrm{erfc}(\alpha\left|\mathbf{r}_i-\mathbf{r}_j\right|)}{\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}\frac{\mathbf{r}_i-\mathbf{r}_j}{\left|\mathbf{r}_i-\mathbf{r}_j\right|} \\
&= -\left(\frac{2\alpha}{\sqrt{\pi}}\frac{e^{-\alpha^2\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}}{\left|\mathbf{r}_i-\mathbf{r}_j\right|}+\frac{\mathrm{erfc}(\alpha\left|\mathbf{r}_i-\mathbf{r}_j\right|)}{\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}\right)\frac{\mathbf{r}_i-\mathbf{r}_j}{\left|\mathbf{r}_i-\mathbf{r}_j\right|}
\end{aligned} \tag{C.2}$$

The erfc term's partials with respect to charge coordinates are trivial.

$$\frac{\partial U_{erfc}}{\partial Q_i} = \sum_{i=1}^{n} q_i \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{r}_i - \mathbf{r}_j\right|\right) - \operatorname{erfc}\left(\eta \left|\mathbf{r}_i - \mathbf{r}_j\right|\right)}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} + \sum_{i=1}^{N} Q_i \frac{\operatorname{erfc}\left(\alpha \left|\mathbf{R}_{ij}\right|\right) - \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}} \left|\mathbf{R}_{ij}\right|\right)}{\left|\mathbf{R}_{ij}\right|} \tag{C.3}$$

The Fourier contribution to the energy is given by

$$U_{FT} = \frac{1}{2V} \sum_{\mathbf{k} \neq 0} \left|S\left(\mathbf{k}\right)\right|^2 \exp\left(-\frac{k^2}{4\alpha^2}\right),$$

where $S(\mathbf{k}) = S_1(\mathbf{k}) + S_2(\mathbf{k})$. As is clear from the definition of the structure factors, $S_1(-\mathbf{k}) = S_1^*(\mathbf{k})$ and $S_2(-\mathbf{k}) = S_2^*(\mathbf{k})$. Hence $S(-\mathbf{k}) = S^*(\mathbf{k})$, implying that $|S(\mathbf{k})| = |S(-\mathbf{k})|$. This symmetry allows the energy to be computed by summing over half of the $\mathbf{k}$ vectors. We choose the half with $k_1 > 0$.

$$U_{FT} = \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi}{k^2} \left|S(\mathbf{k})\right|^2 \exp\left(-\frac{k^2}{\alpha^2}\right)$$

Finding the gradient of this energy with respect to the ion positions yields

$$\begin{aligned}
\frac{\partial U_{FT}}{\partial \mathbf{r}_i} &= \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi e^{-k^2/4\alpha^2}}{k^2} \left[\frac{\partial S^*\left(\mathbf{k}\right)}{\partial \mathbf{r}_i} S\left(\mathbf{k}\right) + S^*\left(\mathbf{k}\right) \frac{\partial S\left(\mathbf{k}\right)}{\partial \mathbf{r}_i}\right] \\
&= \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi e^{-k^2/4\alpha^2}}{k^2} \left[\left(-iq_i \mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r}_i}\right) S(\mathbf{k}) + S^*(\mathbf{k}) \left(iq_i \mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}_i}\right)\right] \\
&= \frac{q_i}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi \mathbf{k} e^{-k^2/4\alpha^2}}{k^2} \left[i \cos\left(\mathbf{k}\cdot\mathbf{r}_i\right) \left(S^*\left(\mathbf{k}\right) - S\left(\mathbf{k}\right)\right) - \sin\left(\mathbf{k}\cdot\mathbf{r}_i\right) \left(S^*\left(\mathbf{k}\right) + S\left(\mathbf{k}\right)\right)\right] \\
&= \frac{q_i}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{8\pi \mathbf{k} e^{-k^2/4\alpha^2}}{k^2} \left[\cos\left(\mathbf{k}\cdot\mathbf{r}_i\right) \operatorname{Im}\left(S\left(\mathbf{k}\right)\right) - \sin\left(\mathbf{k}\cdot\mathbf{r}_i\right) \operatorname{Re}\left(S\left(\mathbf{k}\right)\right)\right]
\end{aligned} \tag{C.4}$$

The gradient with respect to Gaussian charge magnitudes is

$$\begin{aligned}
\frac{\partial U_{FT}}{\partial Q_i} &= \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi e^{-k^2/4\alpha^2}}{k^2} \left[\frac{\partial S^*\left(\mathbf{k}\right)}{\partial Q_i} S\left(\mathbf{k}\right) + S^*\left(\mathbf{k}\right) \frac{\partial S\left(\mathbf{k}\right)}{\partial Q_i}\right] \\
&= \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi e^{-k^2/4\alpha^2}}{k^2} \left[e^{-i\mathbf{k}\cdot\mathbf{r}_i} S(\mathbf{k}) + S^*(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}_i}\right] \\
&= \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{4\pi e^{-k^2/4\alpha^2}}{k^2} \left[\cos\left(\mathbf{k}\cdot\mathbf{r}_i\right) \left(S^*\left(\mathbf{k}\right) + S\left(\mathbf{k}\right)\right) + i \sin\left(\mathbf{k}\cdot\mathbf{r}_i\right) \left(S^*\left(\mathbf{k}\right) - S\left(\mathbf{k}\right)\right)\right] \\
&= \frac{1}{V} \sum_{k_1 \geq 0} \sum_{k_2, k_3}{}' \frac{8\pi e^{-k^2/4\alpha^2}}{k^2} \left[\cos\left(\mathbf{k}\cdot\mathbf{r}_i\right) \operatorname{Re}\left(S\left(\mathbf{k}\right)\right) + \sin\left(\mathbf{k}\cdot\mathbf{r}_i\right) \operatorname{Im}\left(S\left(\mathbf{k}\right)\right)\right]
\end{aligned} \tag{C.5}$$

Putting it all together,

$$\begin{aligned}\frac{\partial U_T}{\partial \mathbf{r}_i} =&\frac{q_i}{V}\sum_{k_1\geq 0}\sum_{k_2,k_3}{}^{'}\frac{8\pi\mathbf{k}e^{-k^2/4\alpha^2}}{k^2}\left[\cos\left(\mathbf{k}\cdot\mathbf{r}_i\right)\operatorname{Im}\left(S\left(\mathbf{k}\right)\right)-\sin\left(\mathbf{k}\cdot\mathbf{r}_i\right)\operatorname{Re}\left(S\left(\mathbf{k}\right)\right)\right]\\ &-\sum_{j=1}^{N}q_iQ_j\left(\frac{2\alpha}{\sqrt{\pi}}\frac{e^{-\alpha^2|\mathbf{r}_i-\mathbf{R}_j|^2}}{|\mathbf{r}_i-\mathbf{R}_j|}+\frac{\operatorname{erfc}(\alpha\left|\mathbf{r}_i-\mathbf{R}_j\right|)}{\left|\mathbf{r}_i-\mathbf{R}_j\right|^2}\right)\frac{\mathbf{r}_i-\mathbf{R}_j}{\left|\mathbf{r}_i-\mathbf{R}_j\right|}\\ &+\sum_{j=1}^{N}q_iQ_j\left(\frac{2\eta}{\sqrt{\pi}}\frac{e^{-\eta^2|\mathbf{r}_i-\mathbf{R}_j|^2}}{|\mathbf{r}_i-\mathbf{R}_j|}+\frac{\operatorname{erfc}(\eta\left|\mathbf{r}_i-\mathbf{R}_j\right|)}{\left|\mathbf{r}_i-\mathbf{R}_j\right|^2}\right)\frac{\mathbf{r}_i-\mathbf{R}_j}{\left|\mathbf{r}_i-\mathbf{R}_j\right|}\\ &-\sum_{j=1}^{n}q_iq_j\left(\frac{2\alpha}{\sqrt{\pi}}\frac{e^{-\alpha^2|\mathbf{r}_{ij}|^2}}{|\mathbf{r}_{ij}|}+\frac{\operatorname{erfc}(\alpha\left|\mathbf{r}_{ij}\right|)}{\left|\mathbf{r}_{ij}\right|^2}\right)\frac{\mathbf{r}_{ij}}{\left|\mathbf{r}_{ij}\right|}\end{aligned} \tag{C.6}$$

and

$$\begin{aligned}\frac{\partial U_T}{\partial Q_i} =&\sum_{j=1}^{n}q_i\frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_i-\mathbf{r}_j\right|\right)-\operatorname{erfc}\left(\eta\left|\mathbf{r}_i-\mathbf{r}_j\right|\right)}{\left|\mathbf{r}_i-\mathbf{r}_j\right|}+\sum_{j=1}^{N}Q_i\frac{\operatorname{erfc}\left(\alpha\left|\mathbf{R}_{ij}\right|\right)-\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}}\left|\mathbf{R}_{ij}\right|\right)}{\left|\mathbf{R}_{ij}\right|}\\ &+\frac{1}{V}\sum_{k_1\geq 0}\sum_{k_2,k_3}{}^{'}\frac{8\pi e^{-k^2/4\alpha^2}}{k^2}\left[\cos\left(\mathbf{k}\cdot\mathbf{r}_i\right)\operatorname{Re}\left(S\left(\mathbf{k}\right)\right)+\sin\left(\mathbf{k}\cdot\mathbf{r}_i\right)\operatorname{Im}\left(S\left(\mathbf{k}\right)\right)\right]\\ &+2\left(\frac{\eta}{\sqrt{2\pi}}-\frac{\alpha}{\sqrt{\pi}}\right)Q_i-V_0\end{aligned} \tag{C.7}$$

Appendix D

Lagrange Multipliers

Here we review the use of Lagrange multipliers for constrained optimization. To build intuition, we first consider a low-dimensional example. Suppose $f(x, y)$ is a function defined over $\mathbb{R}^2$ and $g(x, y) = 0$ is a curve in $\mathbb{R}^2$. For example we could have a situation as shown in Figure D.1, where level curves of $f(x, y)$ are plotted along with $g(x, y) = 0$ for some particular but unimportant choice of f and g. The constrained extremum is the point on $g(x, y) = 0$ which extremizes $f(x, y)$, but from the contour plot it is clear that the constraint curve does not run through the unconstrained extremum of f. The figure reveals that, at least in two dimensions, the constrained extrema occur at points where the level curves of $f(x, y)$ are parallel to $g(x, y) = 0$. If these two curves were not parallel, moving infinitesimally along the curve $g(x, y) = 0$ would cross a level curve of $f(x, y)$. Hence moving (in one direction) along $g(x, y) = 0$ would decrease the value of $f(x, y)$. Along the opposite direction $f(x, y)$ would of course increase. Recalling that a gradient is perpendicular to level curves, it must then be the case that the gradient of $f(x, y)$ and the gradient of $g(x, y)$ are parallel at the critical points. Hence,

$$\nabla f \propto \nabla g \text{ or equivalently } \nabla f = \lambda \nabla g$$

This can be generalized to arbitrary dimensions, as shown by the following proof.

Claim: If $\mathbf{x}^*$ is a critical point then $\nabla f(\mathbf{x}^*) \propto \nabla g(\mathbf{x}^*)$

Proof. Let $\left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial x_i}{\partial g}\right) = \lambda_i$ and take $\lambda = \min\{\lambda_i\}$. Consider an infinitesimal change from $\mathbf{x}^* \to \mathbf{x}^* + \delta\mathbf{x}$ on the hypersurface $g(\mathbf{x}) = 0$. Since this move remains on the constrained hypersurface,

$$\partial g = \sum_i \frac{\partial g}{\partial x_i} \delta x_i = 0$$

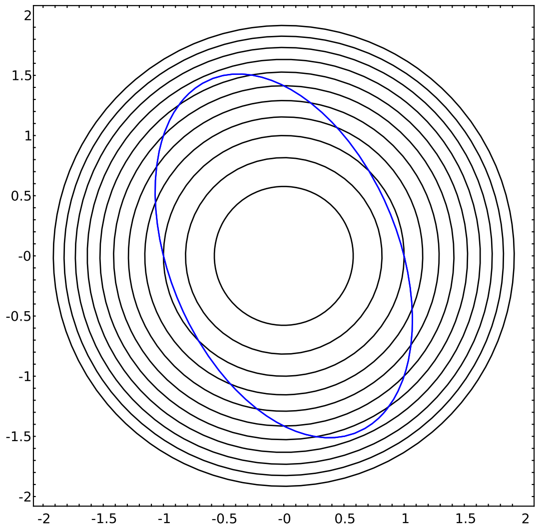


Figure D.1: Level curves of $f(x, y) = x^2 + y^2$ plotted along with the constraint $g(x, y) = 2x^2 + y^2 + xy - 2 = 0$. The function $f(x, y)$ is extremized over $g(x, y) = 0$ at points where the gradients of f and g are parallel.

Now consider the variation in f subject to this constraint.

$$\begin{aligned}
\partial f &= \sum_i \frac{\partial f}{\partial x_i}\delta x_i = \sum_i \lambda_i \frac{\partial g}{\partial x_i}\delta x_i \\
&= \lambda \sum_i \frac{\partial g}{\partial x_i}\delta x_i + \sum_i \left(\lambda_i - \lambda\right) \frac{\partial g}{\partial x_i}\delta x_i \\
&= \lambda \partial g + \sum_i \left(\lambda_i - \lambda\right) \frac{\partial g}{\partial x_i}\delta x_i \\
&= \sum_i \left(\lambda_i - \lambda\right) \frac{\partial g}{\partial x_i}\delta x_i
\end{aligned}$$

As δx_i is arbitrary, $\forall\ i$ with $\lambda_i \neq \lambda$ choose $\mathrm{sgn}(\delta x_i) = \mathrm{sgn}\left(-\frac{\partial g}{\partial x_i}\right)$. By construction, $(\lambda_i - \lambda)$ is guaranteed to be positive, so each term in the sum is negative $\Rightarrow \partial f < 0$ along the specified direction, $\delta \mathbf{x} = \sum_i \delta x_i$. Only in the case that $\lambda_i = \lambda\ \forall\ i$ will there be no direction that gives rise to a nonzero variation in f. Hence at the critical points, $\lambda_i = \lambda\ \forall\ i$. Using our definition of λ_i, this implies that $\frac{\partial f}{\partial x_i} = \lambda \frac{\partial g}{\partial x_i}$. Since this holds for all i, $\nabla f = \lambda \nabla g$. $\square$

D.1 Constrained Optimization as an Effective Unconstrained Problem

Let's revert back to two dimensions temporarily. Our condition for constrained extrema, $\nabla f = \lambda \nabla g$, can be written with respect to each coordinate. These conditions are combined with the constraint to give the following system with three equations and three unknowns (x, y, λ).

$$\begin{cases} \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y} \\ g(x, y) = 0 \end{cases}$$

Suppose we instead considered the function of x, y, and λ given by

$$K(x, y, \lambda) = f(x, y) - \lambda g(x, y) \tag{D.1}$$

The unconstrained optimization of K is straightforward. Critical points require that $\nabla K = 0$. Hence variation in K with respect to x, y, and λ must all be zero. Performing these partial derivatives on Equation (D.1) yields exactly the same system that we found in the case of our constrained optimization. The generalization is natural. Consider the optimization of $f\left(\mathbf{x}\right)$ subject to the k constraints $\{g_k\left(\mathbf{x}\right) = 0\}$. This is equivalent to solving the unconstrained optimization problem of

$$K(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_k \lambda_k g_k(\mathbf{x})$$

Classical mechanics in the Lagrangian form is formally just a constrained optimization problem. Since the constraints must hold for all values of time, rather than indexing these additional multipliers by a discrete number, we index them by the continuous variable t. This means each λ_k is actually a function $\lambda_k(t)$. Hence the action to be minimized for a constrained Lagrangian is

$$S[\mathbf{x}(t)] = \int dt \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) - \sum_k \int dt \lambda_k(t) g_k(\mathbf{x})$$

Equivalently,

$$S_{eff}[\mathbf{x}(t)] = \int dt \mathcal{L}_{eff}(\mathbf{x}, \dot{\mathbf{x}}),$$

where $\mathcal{L}_{eff}(\mathbf{x}, \dot{\mathbf{x}}) \equiv \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) - \sum_k \lambda_k(t) g_k(\mathbf{x})$.

Appendix E

Conjugate Gradient Pseudocode

As discussed in Sections 4.3.4 and 4.4.4, the conjugate gradient method can be applied to electrostatic minimization. Here we give pseudocode for the basic algorithm. The pseudocode for unconstrained minimization is the same as that provided by Shewchuk. [48] Adding a constraint only slightly alters the algorithm.

Algorithm 1 Unconstrained Conjugate Gradient for Constant Potential Calculations

```
$r \leftarrow b - AQ$
$d \leftarrow r$
$\delta_{new} \leftarrow r^T r$
$i \leftarrow 0$
while $i < i_{max}$ and $\delta_{new} > tol$ do
    $y \leftarrow Ad$
    $\alpha \leftarrow \frac{\delta_{new}}{d^T y}$
    $Q \leftarrow Q + \alpha d$
    $\delta_{old} \leftarrow \delta_{new}$
    if $i$ is divisible by 50 then                    ▷ Avoid drifting residuals
        $r \leftarrow b - AQ$
    else
        $r \leftarrow r - \alpha y$
    end if
    $\delta_{new} \leftarrow r^T r$
    $\beta \leftarrow \frac{\delta_{new}}{\delta_{old}}$
    $d \leftarrow r + \beta d$
    $i \leftarrow i + 1$
end while
```

Note that we must be able to compute Ad as well as $(b - AQ)$. We make use of the fact that $(b - AQ) = -\nabla U(Q)$ for an arbitrary vector Q.[1] In fact, this implies

---

[1] This follows most clearly from Equation (4.4).

that

$$Ad = b + \nabla U(d).$$

It does not make physical sense to discuss the gradient of the energy, U, evaluated at the search vector d. Nevertheless, this gradient is precisely the information necessary to deduce the optimal step size, α. Replacing Q_i by d_i everywhere in Equation (C.7) allows one to compute $\nabla U(d)$ and therefore Ad in $\mathcal{O}(N)$ Fourier sums.

Algorithm 2 Constrained Conjugate Gradient for Constant Charge Calculations

```
$r \leftarrow b - AQ$
$p \leftarrow Hr$          ▷ Project $r$ onto the feasible search space. $H$ given in Section 4.4.4
$d \leftarrow p$
$\delta_{new} \leftarrow r^T d$
$\gamma_{new} \leftarrow \delta_{new}$
$i \leftarrow 0$
while $i < i_{max}$ and $\delta_{new} > tol$ do
    $y \leftarrow Ad$
    $\alpha \leftarrow \frac{\delta_{new}}{d^T y}$
    $Q \leftarrow Q + \alpha d$
    $\gamma_{old} \leftarrow \gamma_{new}$
    if $i$ is divisible by 50 then          ▷ Avoid drifting residuals
        $r \leftarrow b - AQ$
    else
        $r \leftarrow r - \alpha y$
    end if
    $p \leftarrow Hr$
    $\gamma_{new} \leftarrow r^T p$
    $\beta \leftarrow \frac{\gamma_{new}}{\gamma_{old}}$
    $d \leftarrow p + \beta d$
    $\delta_{new} \leftarrow r^T d$
    $i \leftarrow i + 1$
end while
```

References

[1] P.M. Ajayan, T.W. Ebbesen, T. Ichihashi, S. Iijima, K. Tanigaki, and H. Hiura. Opening carbon nanotubes with oxygen and implications for filling. *Nature*, 362:522–525, 1993.

[2] P.M. Ajayan, T. Ichihashi, and S. Iijima. Distribution of pentagons and shapes in carbon nano-tubes and nano-particles. *Chemical Physics Letters*, 202(5):384–388, 1993.

[3] P.M. Ajayan, O. Stephan, P. Redlich, and C. Colliex. Carbon nanotubes as removable templates for metal oxide nanocomposites and nanostructures. *Nature*, 375(6532):564–567, 1995.

[4] F. Bertaut. L'énergie électrostatique de réseaux ioniques. *Journal de Physique et le Radium*, 13(11):499–505, 1952.

[5] C.L. Bishop and M. Wilson. The energetics of inorganic nanotubes. *Molecular Physics*, 106(12):1665–1674, 2008.

[6] C.L. Bishop and M. Wilson. The filling of flexible carbon nanotubes by molten salts. *Journal of Materials Chemistry*, 19(19):2929–2939, 2009.

[7] C.L. Bishop and M. Wilson. The mechanisms for filling carbon nanotubes with molten salts: carbon nanotubes as energy landscape filters. *Journal of Physics: Condensed Matter*, 21(11):115301–115306, 2009.

[8] G. Brown, S.R. Bailey, J. Sloan, C. Xu, S. Friedrichs, E. Flahaut, K.S. Coleman, M.L.H. Green, J.L. Hutchison, and R.E. Dunin-Borkowski. Electron beam induced in situ clusterisation of 1D $ZrCl_4$ chains within single-walled carbon nanotubes. *Chemical Communications*, 2001(9):845–846, 2001.

[9] J.M. Caillol. Comments on the numerical simulations of electrolytes in periodic boundary conditions. *The Journal of Chemical Physics*, 101:6080, 1994.

[10] R. Carter, J. Sloan, A.I. Kirkland, R.R. Meyer, P.J.D. Lindan, G. Lin, M.L.H. Green, A. Vlandas, J.L. Hutchison, and J. Harding. Correlation of structural and electronic properties in a new low-dimensional form of mercury telluride. *Physical Review Letters*, 96(21):215501, 2006.

[11] Y.K. Chen, M.L.H. Green, and S.C. Tsang. Synthesis of carbon nanotubes filled with long continuous crystals of molybdenum oxides. *Chemical Communications*, 1996(21):2489–2490, 1996.

[12] B.J. Cox and J.M. Hill. Exact and approximate geometric parameters for carbon nanotubes incorporating curvature. *Carbon*, 45(7):1453–1462, 2007.

[13] S.T. Cui. Electrostatic potential in cylindrical dielectric media using the image charge method. *Molecular Physics*, 104(19):2993–3001, 2006.

[14] S.W. de Leeuw and J.W. Perram. Electrostatic lattice sums for semi-infinite lattices. *Molecular Physics*, 37(4):1313–1322, 1979.

[15] S.W. de Leeuw, J.W. Perram, and E.R. Smith. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proceedings of The Royal Society of London. Series A, Mathematical and Physical Sciences*, 373(1752):27–56, 1980.

[16] M.S. Dresselhaus, R. Saito, and A. Jorio. Semiconducting Carbon Nanotubes. *Physics of Semiconductors; Part A*, 772:25–31, 2005.

[17] P.P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369:253–287, 1921.

[18] M.W. Finnis. Metal-ceramic cohesion and the image interaction. *Acta Metallurgica et Materialia*, 40:S25–S37, 1992.

[19] E. Flahaut, J. Sloan, S. Friedrichs, A.I. Kirkland, K.S. Coleman, V.C. Williams, N. Hanson, J.L. Hutchison, and M.L.H. Green. Crystallization of 2H and 4H PbI_2 in carbon nanotubes of varying diameters and morphologies. *Chemistry of Materials*, 18(8):2059–2069, 2006.

[20] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2nd edition, 2002.

[21] S. Friedrichs, U. Falke, and M.L.H. Green. Phase Separation of LaI_3 inside Single-Walled Carbon Nanotubes. *ChemPhysChem*, 6(2):300–305, 2005.

[22] S. Friedrichs and M.L.H. Green. 1-dimensional lanthanide halide crystals encapsulated within single-walled carbon nanotubes: a brief review. *Zeitschrift für Metallkunde*, 96(5):419–426, 2005.

[23] S. Friedrichs, E. Philp, R.R. Meyer, J. Sloan, A.I. Kirkland, J.L. Hutchison, and M.L.H. Green. LaI_2@(18, 3) SWNT: the crystallisation behaviour of a LaI_2 fragment, confined within a single-walled carbon nanotube. In *Electron microscopy and analysis 2003: proceedings of the Institute of Physics Electron Microscopy and Analysis Group Conference, University of Oxford, 3-5 September 2003*, page 455. Taylor & Francis, 2004.

[24] S. Friedrichs, J. Sloan, M.L.H. Green, R.R. Meyer, A.I. Kirkland, and J.L. Hutchison. Complete characterisation of a Sb_2O_3/(21,8) SWNT inclusion composite. *Chemical Communications*, 2001(10):929–930, 2001.

[25] M. García-Hernández, P.S. Bagus, and F. Illas. A new analysis of image charge theory. *Surface Science*, 409(1):69–80, 1998.

[26] A. Grzybowski, E. Gwóźdź, and A. Bródka. Ewald summation of electrostatic interactions in molecular dynamics of a three-dimensional system with periodicity in two directions. *Physical Review B*, 61(10):6706–6712, 2000.

[27] Y.R. Hacohen, E. Grunbaum, R. Tenne, J. Sloan, and J.L. Hutchison. Cage structures and nanotubes of $NiCl_2$. *Nature*, 395(6700):336–337, 1998.

[28] H.D. Herce, A.E. Garcia, and T. Darden. The electrostatic surface term:(I) Periodic systems. *The Journal of Chemical Physics*, 126:124106, 2007.

[29] M. Hulman, H. Kuzmany, P.M.F.J. Costa, S. Friedrichs, and M.L.H. Green. Light-induced instability of PbO-filled single-wall carbon nanotubes. *Applied Physics Letters*, 85:2068, 2004.

[30] J.D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, Inc, 3rd edition, 1999.

[31] R.O. Jones, P.J. Jennings, and O. Jepsen. Surface-barrier in metals - A new model with application to W(001). *Physical Review B*, 29(12):6474 – 6480, 1984.

[32] A. Loiseau, F. Willaime, N. Demoncy, G. Hug, and H. Pascard. Boron nitride nanotubes with reduced numbers of layers synthesized by arc discharge. *Physical Review Letters*, 76(25):4737–4740, 1996.

[33] V. Meunier, S.V. Kalinin, J. Shin, A.P. Baddorf, and R.J. Harrison. Quantitative Analysis of Electronic Properties of Carbon Nanotubes by Scanning Probe Microscopy: From Atomic to Mesoscopic Length Scales. *Physical Review Letters*, 93(24):246801, 2004.

[34] R.R. Meyer, J. Sloan, R.E. Dunin-Borkowski, A.I. Kirkland, M.C. Novotny, S.R. Bailey, J.L. Hutchison, and M.L.H. Green. Discrete atom imaging of one-dimensional crystals formed within single-walled carbon nanotubes. *Science*, 289(5483):1324, 2000.

[35] M. Neumann. Dipole moment fluctuation formulas in computer simulations of polar systems. *Molecular Physics*, 50(4):841–858, 1983.

[36] T.W. Odom, J.L. Huang, P. Kim, and C.M. Lieber. Structure and electronic properties of carbon nanotubes. *Journal of Physical Chemistry B*, 104(13):2794–2809, 2000.

[37] E. Philp, J. Sloan, A.I. Kirkland, R.R. Meyer, S. Friedrichs, J.L. Hutchison, and M.L.H. Green. An encapsulated helical one-dimensional cobalt iodide nanostructure. *Nature Materials*, 2(12):788–791, 2003.

[38] V.N. Popov. Curvature effects on the structural, electronic and optical properties of isolated single-walled carbon nanotubes within a symmetry-adapted non-orthogonal tight-binding model. *New Journal of Physics*, 6:17, 2004.

[39] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge Press, 3rd edition, 2007.

[40] S.K. Reed, O.J. Lanning, and P.A. Madden. Electrochemical interface between an ionic liquid and a model metallic electrode. *The Journal of Chemical Physics*, 126:084704, 2007.

[41] S.K. Reed, P.A. Madden, and A. Papadopoulos. Electrochemical charge transfer at a metallic electrode: A simulation study. *The Journal of Chemical Physics*, 128:124701, 2008.

[42] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejón. Tight-binding description of graphene. *Physical Review B*, 66(3):35412, 2002.

[43] D.K. Remler and P.A. Madden. Molecular dynamics without effective potentials via the Car-Parrinello approach. *Molecular Physics*, 70(6):921–966, 1990.

[44] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, 3rd edition, 1976.

[45] E.B. Saff and A.B.J. Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelligencer*, 19(1):5–11, 1997.

[46] R. Saito, G. Dresselhaus, and M.S. Dresselhaus. *Physical Properties of Carbon Nanotubes.* Imperial College Press, 1998.

[47] M. Shariff. A constrained conjugate gradient method and the solution of linear equations. *Computers and Mathematics with Applications*, 30(11):25–37, 1995.

[48] J.R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Published privately on web. `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`, August 1994.

[49] J.I. Siepmann and M. Sprik. Influence of surface topology and electrostatic potential on water/electrode systems. *The Journal of Chemical Physics*, 102:511, 1995.

[50] J. Sloan, J. Cook, A. Chu, M. Zwiefka-Sibley, M.L.H. Green, and J.L. Hutchison. Selective deposition of UCl_4 and (KCl) x (UCl_4) inside carbon nanotubes using eutectic and noneutectic mixtures of UCl_4 with KCl. *Journal of Solid State Chemistry*, 140(1):83–90, 1998.

[51] J. Sloan, S.J. Grosvenor, S. Friedrichs, A.I. Kirkland, J.L. Hutchison, and M.L.H. Green. A one-dimensional BaI_2 chain with five-and six-coordination, formed within a single-walled carbon nanotube. *Angewandte Chemie International Edition*, 41(7):1156–1159, 2002.

[52] J. Sloan, A.I. Kirkland, J.L. Hutchison, and M.L.H. Green. Aspects of crystal growth within carbon nanotubes. *Comptes Rendus Physique*, 4(9):1063–1074, 2003.

[53] J. Sloan, P. Mauricio Terrones, S. Nufer, S. Friedrichs, S.R. Bailey, H.G. Woo, M. Ruhle, J.L. Hutchison, and M.L.H. Green. Metastable one-dimensional $\mathrm{AgCl}_{1-x}\mathrm{I}_x$ solid-solution wurzite "tunnel" crystals formed within single-walled carbon nanotubes. *Journal of the American Chemical Society*, 124(10):2116–2117, 2002.

[54] J. Sloan, M.C. Novotny, S.R. Bailey, G. Brown, C. Xu, V.C. Williams, S. Friedrichs, E. Flahaut, RL Callender, A.P.E. York, et al. Two layer 4:4 coordinated KI crystals grown within single walled carbon nanotubes. *Chemical Physics Letters*, 329(1-2):61–65, 2000.

[55] J. Sloan, D.M. Wright, S. Bailey, G. Brown, A.P.E. York, K.S. Coleman, M.L.H. Green, D.M. Wright, J.L. Hutchison, and H.G. Woo. Capillarity and silver nanowire formation observed in single walled carbon nanotubes. *Chemical Communications*, 1999(8):699–700, 1999.

[56] E.R. Smith. Electrostatic energy in ionic crystals. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 375(1763):475–505, 1981.

[57] E.R. Smith. Calculating the pressure in simulations using periodic boundary conditions. *Journal of Statistical Physics*, 77(1):449–472, 1994.

[58] K. Suenaga, C. Colliex, N. Demoncy, A. Loiseau, H. Pascard, and F. Willaime. Synthesis of nanoparticles and nanotubes with well-separated layers of boron nitride and carbon. *Science*, 278(5338):653, 1997.

[59] M.M. Taddei, T.N.C. Mendes, and C. Farina. Subtleties in energy calculations in the image method. *European Journal of Physics*, 30:965–972, 2009.

[60] R. Tenne, M. Homyonfer, and Y. Feldman. Nanoparticles of Layered Compounds with Hollow Cage Structures (Inorganic Fullerene-Like Structures). *Chemistry of Materials*, 10(11):3225–3238, 1998.

[61] S.C. Tsang, Y.K. Chen, P.J.F. Harris, and M.L.H. Green. A simple chemical method of opening and filling carbon nanotubes. *Nature*, 1994.

[62] M.E. Tuckerman, Y. Liu, G. Ciccotti, and G.J. Martyna. Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems. *The Journal of Chemical Physics*, 115:1678, 2001.

[63] M.E. Tuckerman, C.J. Mundy, and G.J. Martyna. On the classical statistical mechanics of non-Hamiltonian systems. *Europhysics Letters*, 45:149–155, 1999.

[64] J. Vatamanu, O. Borodin, and G.D. Smith. Molecular dynamics simulations of atomically flat and nanoporous electrodes with a molten salt electrolyte. *Physical Chemistry Chemical Physics*, 12(1):170 – 182, 2010.

[65] A.P. Willard, S.K. Reed, P.A. Madden, and D. Chandler. Water at an electrochemical interface - A simulation study. *Faraday Discussions*, 141:423 – 441, 2009.

[66] M. Wilson. Structure and phase stability of novel 'twisted' crystal structures in carbon nanotubes. *Chemical Physics Letters*, 366(5-6):504–509, 2002.

[67] M. Wilson. The formation of low-dimensional ionic crystallites in carbon nanotubes. *The Journal of Chemical Physics*, 116:3027, 2002.

[68] M. Wilson. The formation of inorganic nanotubular structures in carbon nanotubes. *Nano Letters*, 4(2):299–302, 2004.

[69] M. Wilson. The formation of low-dimensional chiral inorganic nanotubes by filling single-walled carbon nanotubes. *Chemical Physics Letters*, 397(4-6):340–343, 2004.

[70] M. Wilson. The formation of low-dimensional inorganic nanotube crystallites in carbon nanotubes. *The Journal of Chemical Physics*, 124:124706, 2006.

[71] M. Wilson. Formation of, and ion-transport in, low-dimensional crystallites in carbon nanotubes. *Faraday Discussions*, 134:283–295, 2007.

[72] M. Wilson. The dynamic formation of low-dimensional inorganic nanotubes by filling carbon nanotubes. *The Journal of Chemical Physics*, 131:214507, 2009.

[73] M. Wilson and P.A. Madden. Growth of ionic crystals in carbon nanotubes. *Journal of the American Chemical Society*, 123(9):2101–2102, 2001.

[74] M. Wilson and P.A. Madden. Transformations between tetrahedrally and octahedrally coordinated crystals: the wurtzite → rocksalt and blende → rocksalt mechanisms. *Journal of Physics: Condensed Matter*, 14:4629, 2002.