On the Hardness of Energy Minimisation
for Crystal Structure Prediction

Duncan Adamson[1], Argyrios Deligkas[2], Vladimir Gusev[3], and Igor Potapov[4]

[1,4]Department of Computer Science, The University of Liverpool, United Kingdom
[1,2,3]Leverhulme Research Centre for Functional Materials Design, The University of Liverpool, United Kingdom

October 29, 2019

Abstract

Crystal Structure Prediction (CSP) is one of the central and most challenging problems in materials science and computational chemistry. In CSP, the goal is to find a configuration of ions in 3D space that yields the lowest potential energy. Finding an efficient procedure to solve this complex optimisation question is a well known open problem in computational chemistry. Due to the exponentially large search space, the problem has been referred in several materials-science papers as "NP-Hard and very challenging" without any formal proof though. This paper fills a gap in the literature providing the first set of formally proven NP-Hardness results for a variant of CSP with various realistic constraints. In particular, we focus on the problem of *removal*: the goal is to find a substructure with minimal potential energy, by removing a subset of the ions from a given initial structure. Our main contributions are NP-Hardness results for the CSP *removal* problem, new embeddings of combinatorial graph problems into geometrical settings, and a more systematic exploration of the energy function to reveal the complexity of CSP. In a wider context, our results contribute to the analysis of computational problems for weighted graphs embedded into the three-dimensional Euclidean space.

1 Introduction

One of the central and most challenging problems in materials science and computational chemistry is the problem of predicting the structure of a crystal given the set of ions composing it. At a high level, the goal there is to find a configuration structure of ions in a three-dimensional box that achieves the lowest energy. This problem, termed *Crystal Structure Prediction* (CSP), has remained open due to the complexity of solving it optimally [18] and the combinatorial explosion following a brute-force approach. There are many previous approaches to this problem, largely based on heuristic techniques [10, 12, 14, 16], which have shown some promise, however they still lack the ability to guarantee optimality and moreover they are computationally very demanding.

In the most generic formulation of CSP there are many degrees of freedom due to the numerous parameters of the model: the number of ions to place, their positions, and the unique interactions between each type of ion. Furthermore, real crystals are based on periodic tessellations of 3D space with unit cells whose size and shape may also be changed. The search space remains exponential in size even for greatly simplified versions of CSP. Due to this, CSP has, incorrectly, been referred to in several computational-chemistry papers as "NP-Hard and very challenging" [13]. However, from the computational-theory viewpoint the argument that the search must be done in a set of exponential size does not imply NP-Hardness.

On the other hand, the two results which are often mentioned in context of the NP-Hardness of CSP are [3] and [17]. In [3], within the context of the Ising model, the authors show NP-Hardness in the model of placing ± 1 charges on a graph taking into account only interactions between connected vertices. The reduction works on a grid where each vertex has degree at most 6, making the interaction very local. In [17], it was shown that the problem of placing ions for some given positions is in NP by reducing the problem to TSP. However the reduction goes only one way and does not imply the NP-Hardness of the problem.

In our work, we consider several special variants of CSP and provide a few alternative reasons for the hardness of closely related problems. We take inspiration from hard combinatorial problems in graph theory and propose several new embeddings of NP-Hard graph problems into numerical versions of CSP which can be seen as an optimisation problem for weighted geometric graphs with a non-linear objective function. We focus on the problem of *removal*. Here, the input is a configuration of the ions, and the goal is to remove a subset of the ions such that the interaction energy among the atoms is minimised. The problem of removing vertices of a graph whose deletion results in a subgraph satisfying some specific property have been intensively studied in the combinatorial graph theory. In [9], it was shown that for a large class of properties this problem is NP-Complete. In [19] and [20] this was extended to further properties showing NP-Completeness for bipartite graphs and for non-trivial hereditary properties.

The *removal* problem can be seen as a variant of combinatorial CSP problem, where the positions of the ions correspond to points in discrete grid/lattice. For example, we may find an optimal structure for CSP by placing many copies of the ions that we wish to use to build a new structure in unrealistic positions in the discrete space. This may involve ions "overlapping", i.e., being unrealistically close. Due to the nature of the energy function, when the goal is to minimise the potential energy, the overlapping ions must be removed. In our variant of *removal* problem for which we show NP-Hardness results, the initial configuration (from where we remove ions) is part of the input and has only vacant positions or positions with a single ions in the discrete three-dimensional-Euclidean space.

Our contributions. We give the first NP-Hardness results with more realistic constraints for CSP [7], we provide new embeddings of combinatorial graph problems in geometrical settings, and we explore in a more systematic way the energy function that could reveal the computational complexity of CSP. Moreover, our results can be seen as part of a more general problem of removing vertices from a weighted graph embedded into 3D Euclidean space.

The main challenges for the Euclidean graphs we consider is that they are complete and that the edges are weighted proportional to the distance between their vertices. As such, many classical NP-Hard problems are much harder to embed into this setting. Even for some existing hardness results, in both the geometric and more restricted Euclidean setting, to bring these problems into a bounded number of dimensions often require non-trivial technical proofs as dimension often is part of the input [1, 11]. In some of our constructions we utilise the results on geometric graphs embedded into the plane [5, 6], however many problems in this field remain open. We also study the optimisation problems where we change parameters in the energy function, weights of the nodes, and also restricting their ranges. In this paper we consider three specific versions of the *removal* problem in Euclidean setting:

- k-**Charge Removal:** Remove exactly k charges to minimise the total energy;

- **Minimal-At-Least-k-Charge Removal:** A generalisation of k-charge removal where the removed set is a *minimal* set of at least k charges minimising the total energy;

- **At-Least-k-Charge Removal:** A generalisation of minimal-at-least-k-charge removal where the removed set is of least charges but not necessarily minimal, minimising the total energy.

In all cases we require the sum of the positively weighted vertices to equal the magnitude of the sum of negatively weighted vertices in the set that we are removing. The first version is *k-charge removal*, where for a given k we must remove a set of vertices such that the sum of vertices with positive weights is equal to k. We note in Corollary 1 that determining if there is a solution to this problem

when we have an unbounded number of possible charges for the ions is NP-Hard. One generalisation that will get around this problem is *minimal-at-least-k-charge removal*, where the sum of the weights of positively weighted vertices is greater than or equal to k, and the set is *minimal*, meaning that no strict subset of the removed vertices satisfies the required properties. While every instance of this problem will have a minimal set, we show in Proposition 3 that in the case of unbounded charges it is NP-Hard to verify if any given solution is minimal. We may remove the requirement for a solution to be minimal to avoid this, getting the *at-least-k-charge removal* problem.

We will also consider a variety of settings for these problems, varying the energy function for the crystal, restrictions on the charges and restrictions on the number of ion species. We summarise our results in Table 1. Regarding energy functions, we will primarily consider the class of *controllable* energy functions, $\mathcal{F}$, which we define in Section 2, and the Buckingham-Coulomb potential function, which is popular in computational chemistry. We show in Proposition 1 that Buckingham-Coulomb belongs to the class $\mathcal{F}$. One further energy function we will consider is the Coulomb potential, which is used to calculate the electrostatic potential. We show that depending on the energy function used, and the restrictions on the ion species and charges, we are able to reduce several different combinatorial problems to our problems.

The remainder of this work will be organised as follows: in Section 2 we discuss the preliminaries of these problems, providing relevant notation and definitions. In Section 3 we present our results for the general case of the problems, claiming NP-Hardness with Theorems 1 and 2 for energy functions in $\mathcal{F}$ and an unbounded number of ion species. We also consider some natural restrictions to this problem. In Section 4 we consider the restriction of having only two species of ion under the *Buckingham-Coulomb* potential as our energy function, showing that the problem remains NP-Hard under these restrictions in Theorem 4. In Section 5 we consider the restriction to only the *Coulomb* potential, this time with no restrictions on the number of species of ions or the charges of the ions. In Theorem 5 we show that under these restrictions the problem remains NP-Hard.

Theorem	Summary	Setting
Theorem 1	NP-Completeness by reduction from the clique problem.	All problems, under any energy function in $\mathcal{F}$, charges of $\pm c$ for a given c and an unbounded number of ion species.
Theorem 2	NP-Completeness by extension of Theorem 1.	All problems, under any energy function in $\mathcal{F}$, any bounded set of charges and an unbounded number of ion species.
Theorem 3	Reduction to max-weight-k-clique.	k-charge removal or minimal-at-least-k-charge removal under any computable energy function, charges of $\pm c$ for a given c, and a unbounded number of ion species.
Theorem 4	NP-Completeness by reduction from independent set on penny graphs.	All problems, under the Buckingham-Coulomb potential energy function, charges of ± 1, and two species of ion.
Theorem 5	NP-Completeness by reduction from the knapsack problem.	Minimal-at-least-k-charge removal and at-least-k-charge removal, under the Coulomb potential energy function, unbounded number of charges and unbounded number of ion species.

Table 1: Summary of our results and their corresponding settings.

2 Preliminaries

Unit Cell. A *crystal* is a solid material whose ions, are arranged in a highly ordered arrangement, forming a *crystal structure* that extends in all directions. A crystal structure is described by its *unit cell*; a region of three dimensional space within a parallelepiped representing a period that contains ions in a specific arrangement. The unit cells are stacked in three dimensional space tiling the whole space forming a crystal. The unit cell is a parallelepiped alongside the *arrangement* of ions with their specie. Each unit cell contains a set of n *ions* within the parallelepiped. Each ion, i, has a *specie*, e.g. Ti or Sr, and a non-zero charge q_i. The specie for an ion i will be denoted $S(i)$. In every crystal the unit cell is neutrally charged, i.e., $\sum_{1 \leq i \leq n} q_i = 0$. An arrangement defines a position for every ion in the unit cell, i.e. the positions within $\mathbb{R}^3$.

Energy. The most frequently used technique to compute the energy of a crystal is by summing the pairwise interactions between all pairs of ions. A positive value for the pairwise interaction means the two ions are repelling, while a negative value means they are attracting each other. Formally, for each pair of species there is a unique set of parameters - called *force fields* - which are applied to the common energy function U alongside the Euclidean distance between the ions. In general, energy is defined via series as a crystal is infinite.

In this paper interaction will be restricted with respect to the energy function to a single unit cell only. The primary reason is that the energy between ions in different unit cells quickly converges, making the energy within a single unit cell a good approximation of the total. A second reason is that it is much quicker to compute the energy for a finite set of ions than it is to compute the convergence over an infinite series.

Each arrangement has n ions and a corresponding *potential energy* PE, calculated with respect to the given common energy function U. The goal is to minimise the potential energy. i.e. maximise the magnitude of a negative PE. The pairwise interaction between two ions i and j with respect to the energy function U is $U(i,j)$, denoted U_{ij} when it is clear from the context. The value of U_{ij} is defined by the force field of the ions and the Euclidean distance between them, which is included as one of the parameters. The total potential energy for an arrangement of n ions is given by $PE = \sum_{1 \leq i,j \leq n, i \neq j} U_{ij}$.

This paper will consider a general class of energy functions, called the *controllable* potential functions, denoted by $\mathcal{F}$. All functions in $\mathcal{F}$ are required to be computable in polynomial time for any input. Intuitively, for every $f \in \mathcal{F}$ there exists a set of force field parameters that counteract the effect of the distance parameter r. Formally, a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ belongs to $\mathcal{F}$ if and only if for any given $a \in \mathbb{R}$ and any fixed $r \in \mathbb{R}^+$ there exists a set $\{x_1 \ldots x_{n-1}\} \in \mathbb{R}^{n-1}$ such that $f(x_1, \ldots, x_{n-1}, r) = a$.

The most popular function for crystal structure prediction, which will be focused on in this paper, is the *Buckingham-Coulomb* potential [4], which is the sum of the Buckingham and Coulomb potentials. The Coulomb potential for a pair of ions i, j is defined as $U_{ij}^C = \frac{q_i q_j}{r_{ij}}$, where r_{ij} is the Euclidean distance between the ions. The Buckingham potential for a pair of ions i, j, U_{ij}^B, is defined by four parameters. These are the distance and the three force field parameters, $A_{S(i),S(j)}$, $B_{S(i),S(j)}$, $C_{S(i),S(j)}$, which are dependent on the specie of the ions. It should be noted that all three parameters are positive values. The energy is calculated as $U_{ij}^B = \frac{A_{S(i),S(j)}}{e^{B_{S(i),S(j)}r_{ij}}} - \frac{C_{S(i),S(j)}}{r_{ij}^6}$. Therefore the Buckingham-Coulomb potential is given by

$$U_{ij}^{BC} = U_{ij}^B + U_{ij}^C = \frac{A_{S(i),S(j)}}{e^{B_{S(i),S(j)}r_{ij}}} - \frac{C_{S(i),S(j)}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}}.$$

Proposition 1. *There exists a set of parameters for the Buckingham-Coulomb function such that it is in $\mathcal{F}$.*

Proof. Given a value a and a pair of ions, i and j, at a distance of r_{ij} with arbitrary charges q_i and q_j, the parameters may be set so that the potential at a distance of r_{ij} is a. $B_{S(i),S(j)}$ is set to 0 and the values of $A_{S(i),S(j)}$ and $C_{S(i),S(j)}$ are set as follows. If $a > 0$

$$A_{S(i),S(j)} = \begin{cases} a, & \text{if } q_i q_j > 0; \\ a + \frac{|q_i q_j|}{r_{ij}}, & \text{otherwise.} \end{cases} \quad C_{S(i),S(j)} = \begin{cases} q_i q_j r_{ij}^5, & \text{if } q_i q_j > 0; \\ 0, & \text{otherwise.} \end{cases}$$

If $a \le 0$

$$A_{S(i),S(j)} = \begin{cases} 0, & \text{if } q_i q_j > 0; \\ \frac{|q_i q_j|}{r_{ij}}, & \text{otherwise.} \end{cases} \quad C_{S(i),S(j)} = \begin{cases} |a| r_{ij}^6 + q_i q_j r_{ij}^5, & \text{if } q_i q_j > 0; \\ |a| r_{ij}^6, & \text{otherwise.} \end{cases}$$

Substituting 0 for $B_{S(i),S(j)}$ into the Buckingham-Coulomb potential, the equation becomes

$$U_{ij}^{BC} = A_{S(i),S(j)} - \frac{C_{S(i),S(j)}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}}.$$

Cancelling out the coulomb potential either by adding $q_i q_j r_{ij}^5$ to $C_{S(i),S(j)}$ in the case $q_i q_j > 0$ or $\frac{|q_i q_j|}{r_{ij}}$ to $A_{S(i),S(j)}$ in the case $q_i q_j \le 0$. In the first case the energy added by the Coulomb potential will be $\frac{|q_i q_j|}{r_{ij}}$, which will be cancelled by the addition of $q_i q_j r_{ij}^5$ when multiplied by the $\frac{-1}{r_{ij}^6}$ term applied to $C_{S(i),S(j)}$. Otherwise the Coulomb energy will be $\frac{-|q_i q_j|}{r_{ij}}$, which will be cancelled out by the relevant addition from $A_{S(i),S(j)}$.

By the same arguments the addition of an a term to either $A_{S(i),S(j)}$ or $C_{S(i),S(j)}$ with the appropriate multiplier, will leave just the value of a. Note that in the case that a is negative, the magnitude must be added to $C_{S(i),S(j)}$ so that once the multiplier has been applied it will act as a negative value. $\qquad\square$

Crystals as geometric graphs. Using the above definitions, it can be shown how crystals may be viewed as geometric graphs. Recall that each ion corresponds to a charged point in $\mathbb{R}^3$. Each ion is represented with a weighted vertex, also placed into $\mathbb{R}^3$ at the same position as the ion, giving a total of n vertices. The vertex corresponding to the ion i, denoted v_i, is assigned a weight of q_i. $wt(v_i)$ will denote the weight of a given vertex v_i, i.e. $wt(v_i) = q_i$. For notation, V^+ will denote the set of vertices with a positive weight in V, and V^- for the set of vertices with a negative weight in V. This can be extended to a set of ions A, using A^+ for the positively charged ions and A^- for the negatively charged ions.

Between each pair of vertices there is an edge, weighted by the pairwise interaction of the corresponding ions U_{ij}. Note that from its definition U_{ij} will be determined in part by the length of the edge, which will be drawn as a straight line in the space. The energy of a crystal graph $G = \{V, E\}$ can be computed as $PE = \sum\limits_{\{v_i, v_j\} \in E} U_{ij}$.

Geometric graphs created from a unit cell will be referred to as *crystal graphs*. In the remainder of this work crystals will be described in terms of their physical structure where it makes sense to be considering the ions, and as a graph otherwise.

The *k-Charge Removal* Problem. The k-charge removal problem, henceforth K-CHARGE REMOVAL, will take as input a crystal graph G corresponding to a "dense" initial arrangement of ions, with the goal of removing some vertices in order to minimise the energy of the new subgraph $G' \subset G$. It will be assumed that the initial graph is *charge-neutral*, as defined in Definition 1. As G' must also be neutral, any set of vertices which is removed must therefore be neutral. Using intuition from chemistry regarding the number of ions within a realistic unit cell, a natural number k of charges to remove is chosen, as defined in Definitions 2 and 3.

Definition 1. *A set of vertices R is **neutral** if $\sum\limits_{v_i \in R} wt(v_i) = 0$.*

Definition 2. *A **set of k-Charges** R from a crystal graph $\{V, E\}$ is a neutral subset where $R \subseteq V$ and $\left| \sum\limits_{v_i \in R^+} wt(v_i) \right| = \left| \sum\limits_{v_j \in R^-} wt(v_j) \right| = k$.*

Informally, a set of k-charges is a set of vertices with a total weight is 0, while the magnitudes of the sums of all positively weighted vertices, and of all negatively weighted vertices is k

Definition 3. *A **removal** of a set of vertices R from a graph $\{V, E\}$ is the graph $\{V', E'\}$ where $V' = V \setminus R$ and E' is the set of edges in E with no endpoint in R.*

Informally, a removal is the result of removing a set of vertices and the incident edges from a given graph.

Problem 1. *k-Charge Removal (*K-CHARGE REMOVAL*)*

> *Instance:* A crystal graph G, with edges weighted by a given common energy function U, and a natural number k.
>
> *Goal:* The set of k-charges R from G where $G' = \{V', E'\}$ created by the removal of R from G which minimises $\sum_{\{v_i, v_j\} \in E'} U_{ij}$.

A decision problem can be derived from K-CHARGE REMOVAL by asking if there exists a removal that leaves G' with no-more total energy than some goal g, i.e. $\sum_{v_i, v_j \in V', i \neq j} U_{ij} \leq g$.

The *At-Least-k-Charge Removal* Problem. One generalisation of K-CHARGE REMOVAL is the *at-least-k-charge removal problem*, denoted AT-LEAST-K-CHARGE REMOVAL. This problem takes the same input as in K-CHARGE REMOVAL, however rather than looking to remove a set of exactly k-charges, it is instead sufficient to remove a neutral set of at-least-k-charges. The motivation for this comes from the case that it is not be possible to remove exactly k for some given k. In this generalisation more than k charges may be removed, provided the cell remains neutral. Note that any removal of exactly k will also be valid for this generalisation.

Definition 4. *A set of **At-Least-k-Charges** R from a crystal graph $\{V, E\}$ is a neutral subset where $R \subseteq V$ and $\sum_{v_i \in R^+} wt(v_i) \geq k$.*

Problem 2. *At-Least-k-Charge Removal (*AT-LEAST-K-CHARGE REMOVAL*)*

> *Instance:* A crystal graph G, with edges weighted by a given common energy function U, and a natural number k.
>
> *Goal:* The set of at-least-k-charges R from G where $G' = \{V', E'\}$ created by the removal of R from G which minimises $\sum_{\{v_i, v_j\} \in E'} U_{ij}$.

Proposition 2. *A solution to* K-CHARGE REMOVAL *or* AT-LEAST-K-CHARGE REMOVAL *can be verified in polynomial time.*

Proof. A solution to K-CHARGE REMOVAL contains the set of vertices R that are removed. This can be verified as a set of at-least-k-charges by simply summing up the positive and negative weights, checking that the set is neutral and that $\left| \sum_{v_i \in R^+} wt(v_i) \right| = k$ for K-CHARGE REMOVAL or $\left| \sum_{v_i \in R^+} wt(v_i) \right| \geq k$ for AT-LEAST-K-CHARGE REMOVAL. This will take time of the order of $O(|R|)$. Similarly the sum of the edges in the original graph G that do not have an endpoint in R can be checked against the goal value g. This can be done in $O(|V|^2)$ time, as the graph is complete. Therefore as no step will take more than $O(|V|^2)$ time, a solution to AT-LEAST-K-CHARGE REMOVAL can be verified in polynomial time. Hence K-CHARGE REMOVAL and AT-LEAST-K-CHARGE REMOVAL fall into the class of NP problems. $\square$

The *Minimal-At-Least-k-Charge Removal* Problem An alternative restriction of AT-LEAST-K-CHARGE REMOVAL is the *minimal-at-least-k-charge removal* problem, denoted MINIMAL-AT-LEAST-K-CHARGE REMOVAL. This also serves as a generalisation of K-CHARGE REMOVAL, where the goal is to get close to a set of k-charges, however accepting that it may not be possible to reach the exact value. In this problem a *minimal* set of at-least-k-charges is removed.

Definition 5. *A set R of at-least-k-charges is **minimal** if there exists no subset $R' \subset R$ such that*

$$\left| \sum_{v_i \in R'^+} wt(v_i) \right| \geq k \text{ and } \left| \sum_{v_i \in R'^+} wt(v_i) \right| = \left| \sum_{v_j \in R'^-} wt(v_j) \right|.$$

Informally, Definition 5 means that there is no way of getting closer a set of k-charges from the set, without having fewer than k charges. It follows that for a given crystal graph, there may be multiple minimal at-least-k-charge sets for a given k. A removal of at-least-k-charges is minimal if the set of at-least-k-charges is minimal. It may be noted that a set of k-charges is always a minimal set of at-least-k-charges.

Problem 3. *Minimal-at-least-k-Charge Removal (*K-CHARGE REMOVAL*)*

  Instance: A crystal graph G, with edges weighted by a common energy function U, and a natural number k.

  Goal: The minimal set of k-charges R from G where for $G' = \{V', E'\}$ created by the removal of R from G which minimises $\sum_{\{v_i, v_j\} \in E'} U_{ij}$.

Proposition 3. *It is NP-Hard to verifying if a set of at-least-k-charges is minimal when no bounds are given on the value of the charges.*

Proof. This can be shown by a reduction from the *subset-sum* problem. In the subset-sum problem there is a set of values S, and a goal k. The goal is to choose some subset $S' \subseteq S$ such that $\sum_{i \in S'} i = k$. Note that this problem remains NP-complete in the case the input is only positive integers.

   Given an instance of subset sum $I = \{S, k\}$, a crystal graph is created as follows as follows. For each integer $i \in S$ a new vertex with a weight of i is created, note these will correspond to the set V^+. Two further ions are created, the first having a charge of $-k$ and the second having a charge of $-\left(\left(\sum_{v_i \in V^+} wt(v_i) \right) - k \right)$, these will correspond to V^-. The value k' is chosen as the greater of k and $\left(\sum_{v_i \in V^+} wt(v_i) \right) - k$.

   Given this instance it can be claimed that the only minimal k'-charge removal from I is $R = V$. To disprove this is minimal there must be some subset $R' \subset R$ that is also a set of at-least-k-charges. As k' charges must be removed, any such V'^- must only contain the vertex in V^- with a charge of $-k'$. Therefore if this claim is false, there must be a set $R'^+ \subseteq R^+$ such that $\left| \sum_{v_i \in R'^+} wt(v_i) \right| = k'$. If there is such a R' then there is also have a solution to the subset sum instance as either R'^+ or $R^+ \setminus R'^+$. This can be shown as if $k' = k$, the the values in R'^+ must sum to k, satisfying I. Conversely if $k' = \left| \sum_{v_i \in R'^+} wt(v_i) \right|$, then the ions in $R^+ \setminus R'^+$ must sum to k satisfying I. If there is no such subset then there must be no solution to I, as such a solution would allow the existence of a R'^+ for this set.

   In the other direction, if there is a solution to I then trivially there must be exist such a R'^+ that would make R' non-minimal. Similarly if there is no valid solution to I then the only minimal set of k-charges is the complete set of ions. Therefore it can not be determined if the a solution is minimal in polynomial time. Subsequently as a minimal set of charges for AT-LEAST-K-CHARGE REMOVAL is

required, a solution can not be verified in polynomial time, therefore it is not in NP in the general case. □

Corollary 1. *It is NP-Hard to determine if an instance of* K-CHARGE REMOVAL *has a valid solution in the case there are no bounds on the value of the charges.*

Proof. It follows from the arguments of Proposition 3 that an instance of K-CHARGE REMOVAL may be constructed for a subset sum instance $I = \{S, k\}$ such that it is only satisfiable if the subset sum instance is. □

Corollary 2. *A set of k-charges may be verified as minimal in polynomial time for charge values bounded by a polynomial size.*

Proof. In the case the charges are bounded a solution to the subset sum may be found in polynomial time, for example relative to either the upper limit on the value due to Pisinger [15], or for the number of distinct weights relative and the goal value due to Axiotis and Tzamos [2]. Using these a set of k-charge R can be verified a minimal. This is done by, for every value $1 \leq t \leq \left| \sum_{v_i \in R^+} wt(v_i) \right| - k$ checking

if there a subset of charges $R'^+ \in R^+$ and $R'^- \in R^-$ such that $t = \left| \sum_{v_i \in R'^+} wt(v_i) \right| = \left| \sum_{v_i \in R'^-} wt(v_i) \right|$.

If there exists such a solution for any t then R is not minimal.

The claimed energy may also be verified by checking the sum of pairwise interactions relative to U, with may trivially be done in Polynomial time by the definition of U. Therefore under these restrictions AT-LEAST-K-CHARGE REMOVAL is in NP. □

3 NP-Hardness for an unbounded number of ion species

This section will focus on the class of potential functions $\mathcal{F}$. It will be assumed that the energy function for all cases is an arbitrary function in $\mathcal{F}$ for which the parameters required by the ions to result in the energy from their pairwise interaction to be any arbitrary a are known. NP-completeness for K-CHARGE REMOVAL as well as for the generalisations to MINIMAL-AT-LEAST-K-CHARGE REMOVAL and AT-LEAST-K-CHARGE REMOVAL will be shown when there are bounds on value of the charges (either quantity of charges, or the maximum value). It may be noted that in the case the charges are not bounded, MINIMAL-AT-LEAST-K-CHARGE REMOVAL will remain NP-Hard, however as it will not be in NP it will not be complete. Moreover K-CHARGE REMOVAL and MINIMAL-AT-LEAST-K-CHARGE REMOVAL when all vertices have a weight of $\pm c$ for a given c can be reduced to max-weight-k-clique.

Theorem 1. K-CHARGE REMOVAL, MINIMAL-AT-LEAST-K-CHARGE REMOVAL *and* AT-LEAST-K-CHARGE REMOVAL *are NP-Complete for energy functions in* $\mathcal{F}$ *for charges of* $\pm c$, *for any natural number* c.

Proof. K-CHARGE REMOVAL and AT-LEAST-K-CHARGE REMOVAL are in NP by Proposition 2, and as the charges are bounded, MINIMAL-AT-LEAST-K-CHARGE REMOVAL will be in NP by Corollary 2. Hardness is established via a reduction from CLIQUE. This is shown by reduction to K-CHARGE REMOVAL, noting that any satisfying solution to K-CHARGE REMOVAL will also satisfy AT-LEAST-K-CHARGE REMOVAL and MINIMAL-AT-LEAST-K-CHARGE REMOVAL.

In the Clique problem, henceforth CLIQUE, the input is a graph, G, and a natural number, k. The goal is to find a clique of size k in G, or report that no such clique exists. A clique is a set of vertices in a graph such that all vertices in the set are adjacent to each other.

Given an instance of CLIQUE, $I = (G, k) = ((E, V), k)$ where $n = |V|$, an instance, I', of K-CHARGE REMOVAL is constructed as follows. A unit cell of arbitrary size is chosen. Within this cell $2n$ unique positions are created at arbitrary places. In the first n positive ions are placed and in the last n negative ions are placed. Each ion has its own unique specie. Every vertex $v_i \in V$ corresponds to two

ions, i^+ and i^- with charges c and $-c$ respectively. For two ions i and j associated with v_i and v_j respectively the parameters are set so as to satisfy the following:

$$U_{ij} = \begin{cases} -1 & v_i = v_j \text{ or } \{v_i, v_j\} \in E \\ \infty & \text{otherwise.} \end{cases}$$

The definition of $\mathcal{F}$ guarantees that there exists parameters satisfying these conditions irrespective of the positions, and thus the distance r_{ij}, of the ions. The goal vale is chosen as $g = k(2k-1)$, noting that there are $k(2k-1)$ edges in a clique of size $2k$. Let $k' = n - k$, noting that to remove k' positive and k' negative ions ck' charges must be removed.

From this the corresponding crystal graph $G' = \{V', E'\}$ is constructed as described in the preliminaries. Let the vertices $v_i^+, v_i^- \in V'$ represent the ions corresponding to $v_i \in V$. $v_i^{\pm}$ will be used to denote either v_i^+ or v_i^- where the charge doesn't matter i.e. we are only concerned with the vertex in G that $v_i^{\pm}$ corresponds to. From the definition of the energy function, $wt(v_i^{\pm}, v_j^{\pm}) = -1$ if $i = j$ or $\{v_i, v_j\} \in E$, and ∞ otherwise.

It may now be claimed that I will be satisfiable if and only if I' is satisfiable. First consider the case that I is satisfiable. In this case ck' charges may be removed from I', leaving only the vertices corresponding to the clique in I, denoted A. As all vertices in A correspond to adjacent vertices in G, the energy will simply be -1 multiplied by the number of edges, giving a total energy of $-k(2k-1)$, satisfying the K-CHARGE REMOVAL instance. Conversely if there does not exist a clique of size k in G then any subset of vertices $A \subseteq V'$ of cardinality k clearly must contain at least one edge with a weight of ∞, making I unsatisfiable.

Assume now I' is satisfiable, where R is the set of removed vertices and $A = V' \setminus R$ are the remaining vertices. Note thate the energy must be at most $-k(2k-1)$, and there have exactly $2k$ vertices in A. As the only negative weight edges in G' are those between two vertices representing either the same or adjacent vertices in G. Clearly this can only be achieved by having a set of vertices in G' representing vertices in a clique of size k in G. Similarly if I' is unsatisfiable then there can not be a clique of size k, as this would imply I' is satisfiable. Therefore this problem is NP-Complete in the case that all ions have charges of $\pm c$ for any natural number c for K-CHARGE REMOVAL. The problem is also NP-Complete for AT-LEAST-K-CHARGE REMOVAL and MINIMAL-AT-LEAST-K-CHARGE REMOVAL, as the minimum energy for any at-least-k-charge removal will be $-k(2k-1)$, which can only be achieved by a k-charge removal. $\qquad\square$

It can be noted that this may be extended to other graph problems relatively easily. One example of this would be the MAX-WEIGHT K-CLIQUE problem. In MAX-WEIGHT K-CLIQUE takes as input a weighted graph G, a natural number k, and a goal value v. The problem is to report if a clique of size k where the sum of the weights of the edges is at least v exists. Using the above construction, a crystal graph G' may be created from G. From this the weights on the edges may be adjust as follows:

$$U_{ij} = \begin{cases} -wt(v_i, v_j) & v_i \neq v_j \text{ an } \{v_i, v_j\} \in E \\ -c & v_i = v_j \\ \infty & \text{otherwise.} \end{cases}$$

Where $wt(v_i, v_j)$ denotes the weight between vertices v_i and v_j in G and c is some constant such that $\nexists v_i, v_j \in E$ s.t. $wt(v_i, v_j) \geq c$. The goal value for the K-CHARGE REMOVAL instance is chosen as $-k \cdot c - v$. The correctness of this reduction follows from the arguments in Theorem 1.

Theorem 2. K-CHARGE REMOVAL *remains NP-Hard for set of allowed charges with unique magnitude and an energy function within $\mathcal{F}$.*

Proof. The construction of Theorem 1 may be extended to the case the set of charges is limited to any set of allowed charges. Two charges are chosen from this set, c and d such that $|c| > |d|$ and $cd < 0$ such that $|c| - |d|$ is minimised. The same steps as in Theorem 1 are followed for the construction to

get an initial crystal graph $G = \{V, E\}$ and k'.. Note that I will have a charge *deficiency* of $n(|c|-|d|)$, meaning that some set of vertices must be added to make the cell neutral. To handle the deficiency two sets of dummy vertices with charges of c and d will be created.

The first set will be to deal with the deficiency that would be left from a clique of size k. To construct these, a natural number t is chosen such that there exists a pair of natural numbers t_c and t_d such that $t_c|c| = t$ and $t_d|d| = k(|c-d|)+t$. Using these, t_c vertices with a charge of c and t_d vertices with a charge of c and t_d with a charge of d are added. From the definition of $\mathcal{F}$, the energy between them and all ions in G and between each other can be set as 0.

The second set of dummy vertices will be to counteract the overall deficiency in the initial unit cell. A natural number u is chosen such that there exists a pair of natural numbers u_c and u_d where $u = |c|(u_c + t_c)$ and $u + n(|c| - |d|) = |d|(u_d + t_d)$. u_c vertices with a charge of c and u_d vertices with a charge of d are added. The potential energy between between them and all other vertices, including the set of previously added dummy vertices, is ∞.

To ensure that the best set of ions to be left with will be a clique of size k as well as all of the dummy vertices added in the first step, the following is done. The goal energy remains the same as from Theorem 1. It can be seen that the only way to achieve this is to leave vertices corresponding to a clique of size at least k. As there are $c(n + u_c + t_c)$ charges for one set, and the goal is to be left with $c(k + t_c)$, a value k' is chosen to remove as $c(n + u_c - k)$. In the case that exactly k'-charges are removed, either the dummy vertices or some other vertices corresponding to a clique of size greater than k, ensuring the set remains neutral will be left. In the at-least-k' case, some dummy vertices may also be removed provided the cell remains neutral.

From the arguments in Theorem 1 it can be seen that this will be sufficient to ensure the new instance is satisfiable if and only if the original CLIQUE instance is. Therefore these problems are NP-Hard, even in the case that there are distinct charges c and d, $|c| \neq |d|$. □

While there are simple reductions to other NP-Complete problems such as Integer Programming, embedding this problem into many classical problems is made difficult due to the problem of maintaining the neutrality of the unit cell. To this end, Theorem 3 shows how a restricted version of K-CHARGE REMOVAL may be embedded into MAX-WEIGHT K-CLIQUE.

Theorem 3. K-CHARGE REMOVAL *can be reduced to* MAX-WEIGHT K-CLIQUE *in polynomial time, under the restriction that charges are limited $\pm c$ and the energy function is computable within polynomial time.*

Proof. Note that, given charges of $\pm c$, a valid solution to MINIMAL-AT-LEAST-K-CHARGE REMOVAL will either be valid for K-CHARGE REMOVAL, or there will be no valid solutions to K-CHARGE REMOVAL. Taking as input an instance of MINIMAL-AT-LEAST-K-CHARGE REMOVAL with charges of $\pm c$ with the corresponding crystal graph, it will be claimed that this instance may be represented as an instance of the weighted generalisation of CLIQUE.

In weighted k-clique, denoted WEIGHTED-K-CLIQUE, the input is a weighted graph, a goal value v, and a natural number k. An instance of WEIGHTED-K-CLIQUE is satisfiable if and only if there exist a clique of size k such the the sums of the weight of the edges in the clique is at least v.

Given an instance of MINIMAL-AT-LEAST-K-CHARGE REMOVAL $I = \{G, k\} = \{\{V, E\}, k\}$, an instance I' of WEIGHTED-K-CLIQUE is created as follows. A value k' is chosen as $\frac{|V^+| - k}{c}$ rounded down to the nearest natural number. Note that if $(|V^+| - k)(\mod c) \not\equiv 0$, then there is no valid solution to K-CHARGE REMOVAL, however there may still be some valid solution to MINIMAL-AT-LEAST-K-CHARGE REMOVAL. A new graph $G' = \{V', E'\}$ is created which will initially be empty. For each pair of vertices with different charges a new associated vertex in V' is created. An edge is created between each new vertex if and only if the corresponding vertices are all unique, i.e. given the set of vertices $V^+ = \{v_i, v_j\}$, and $v^- = \{v_k, v_l\}$ an edge would be placed between the new vertex representing (v_i, v_k) and the one representing (v_j, v_l), but not from either to the vertex representing (v_i, v_l). Give two connected vertices corresponding to vertices (v_i, v_k) and (v_j, v_l) the edged is assigned a value of
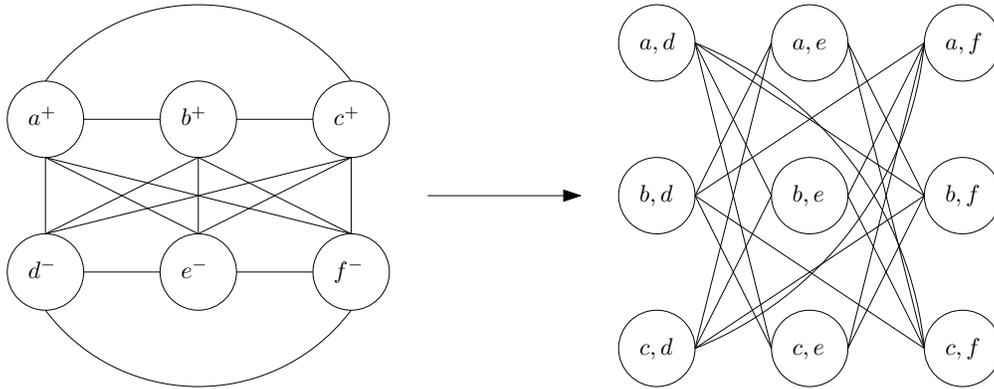
Figure 1: Example of the construction from AT-LEAST-K-CHARGE REMOVAL to clique. Note that vertices $a, b,$ and c have a positive weight, while $d, e,$ and f have a negative weight. Also note that any clique of size 3 will correspond to the original graph.

$-\left(U_{ij} + U_{il} + U_{jk} + U_{kl} + \frac{U_{ij}+U_{jl}}{k'-1}\right)$ to the edge between them. The intuition behind this is for the edge to maintain the weights of the edges in G. $\frac{U_{ik}+U_{jl}}{k'-1}$ is added to this so that within a clique of size k', the edge between the two vertices is fully represented. An example of this construction is shown in Figure 1, omitting weights for legibility.

It may now be claimed that any clique of size m will correspond to a neutrally weighted subset $A \subseteq V$ where $\left| \sum_{v_i \in A^+} wt(v_i) \right| = mc$. This is shown by noting that vertices are only connected if they do not represent a common vertex. As such a clique of size m must contain m unique positively weighted and m unique negatively weighted vertices for the corresponding vertices to be connected as a clique. Therefore by selecting any clique of size k' in this graph, it can be seen that there will be a valid structure left with exactly k' unique positively weighted and k' unique negatively weighted vertices. From the definition of k' this will correspond to an subgraph of G after a minimal removal of k.

It may now be claimed that a maximum weight clique of size k' will correspond to the best subset of ions after a k charge removal. Note that given a clique with total weight w corresponds to a set of ions with total energy $-w$. It is a straightforward extension to see that a maximum weight clique will correspond to a minimum energy subset of ions. This can be seen by noting that by choosing k' as the size of the clique, the corresponding arrangement $A \subseteq V$ will have $\left| \sum_{v_i \in A^+} wt(v_i) \right| = ck'$.

From the definition of k', this requires $\left| \sum_{v_i \in A^+} wt(v_i) \right| \leq \left| \sum_{v_i \in V^+} wt(v_i) \right| - k$, which will satisfy the requirements for a k charge removal. Conversely the definition of k' ensures that the removal must be minimal. Therefore the optimal solution to the WEIGHTED-K-CLIQUE instance must correspond to an optimal solution to the MINIMAL-AT-LEAST-K-CHARGE REMOVALK-CHARGE REMOVAL instance. Similarly any valid solution to the WEIGHTED-K-CLIQUE instance will correspond to some solution to the MINIMAL-AT-LEAST-K-CHARGE REMOVAL instance. □

4 Bounded number of species with Buckingham-Coulomb potential

In Section 3 NP-Hardness was shown for the case that there was an unbounded number of species, and NP-completeness in the case that there is a bounded number of charges. This will now be

strengthened by considering instances with only two unique species. Only the Buckingham-Coulomb potential function with charges of ± 1 will be considered in this section. All three problems will again be considered, noting that for charges of ± 1 K-CHARGE REMOVAL is equivalent to MINIMAL-AT-LEAST-K-CHARGE REMOVAL. NP-Hardness will be shown by a reduction from INDEPENDENT-SET on penny graphs adapting it to the Euclidean settings of crystal graph of ions within a unit cell. The Independent Set problem, denoted INDEPENDENT-SET, takes as input a graph, G, and a natural number k. The goal is to find an *independent set*, i.e. a set of vertices such that no two are adjacent, of size k in G, or report that one does not exist. Penny graphs are the class of graphs where each vertex may be drawn as a unit circle such that no two circles overlap, and an edge between two vertices exist if and only if the corresponding circles are tangent, i.e. they intersect at only a single point. Finding an independent set on this class of graphs was shown to be NP-Hard by Cerioli et al. [5]. The NP-Hardness result for this problem comes from a reduction from max-degree 3 Planar Vertex cover, which was shown to be NP-Complete by Garey and Johnson [8].

Construction of the K-CHARGE REMOVAL **instance:** Starting with an instance of INDEPENDENT-SET on a maximum degree 3 planar graph, containing the graph and a natural number k an instance of K-CHARGE REMOVAL is created as follows. Using Theorem 1.2 from Cerioli et al. to create a new penny graph realisation, G, and a new natural number k. The class of graphs created by this process will be denoted as the *long orthogonal penny graphs*. For the realisation the radius of each circle is chosen as $\frac{n}{2}$.

A region of space in $\mathbb{R}^3$ with a height of at least 1 and a width and length allowing G may be drawn is created. This space will be the parallelepiped for the unit cell. In this space, two copies of G are drawn such that one is has a height 1 heigher than the other. For every circle in G two ions are created, one in the lower copy of G and the other in the higher copy. Each ion is labelled with the vertex from G it corresponds to. In this context *pair* refers to the two ions in the new crystal graph G', labelled with the same vertex from G. Two pairs are *neighbouring* if they represent vertices that are adjacent in G. The lower ions are assigned the positive specie and the upper ions the negative. An example of this arrangement is provided in Figure 2. Note that the minimum distance between two pairs in the same plane that are non-adjacent for circles with a radius of n will be $\sqrt{2}n$, as shown in Figure 3.

The positive and negative species are assigned charges of $+1$ and -1 respectively. From these species there are parameters for the interaction between two ions of the positive specie, two ions of the negative specie, and between one ion of the positive specie and one of the negative specie. For brevity, 1 and 2 will be used to denote the positive and negative specie respectively. Under this construction, the interaction between the two ions of the positive specie is the same as between two ions of the negative specie. Therefore the parameters that may be set are $A_{11}, B_{11}, C_{11}, A_{12}, B_{12}$, and C_{12}.

Let $k' = n - k$, being the number of charges that are required to be removed to be left with an independent set of size k. Note that as the charge of each ion has a magnitude of one, a removal of k' can only be achieved by removing k' positive and k' negative ions. The goal energy for the construction is set as $g = (k-1)(\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1)$. To simplify the equations regarding the interaction between planes, $\hat{r}$ will be used to denote $\sqrt{r^2 + 1}$.

An independent set will be said to be *left* if the ions left after a removal of k' charges have labels corresponding to an independent set in G. To ensure that an independent set is left of size k if and only if one exists, the following three inequalities must be satisfied:

$$\frac{A_{11}}{e^{B_{11}n}} - \frac{C_{11}}{n^6} + \frac{1}{n} + \frac{A_{12}}{e^{B_{12}\widehat{n}}} - \frac{C_{12}}{\widehat{n}^6} - \frac{1}{\widehat{n}} \geq \left| \frac{A_{12}}{e^{B_{12}}} - C_{12} - 1 \right| \tag{1}$$

$$n^2 \left| \frac{A_{11}}{e^{B_{11}r}} - \frac{C_{11}}{r^6} + \frac{1}{r} + \frac{A_{12}}{e^{B_{12}\widehat{r}}} - \frac{C_{12}}{\widehat{r}^6} - \frac{1}{\widehat{r}} \right| \leq \left| \frac{A_{12}}{e^{B_{12}}} - C_{12} - 1 \right|, \qquad r \geq \sqrt{2}n \tag{2}$$

$$\frac{A_{11}}{e^{B_{11}r}} - \frac{C_{11}}{r^6} + \frac{1}{r} + \frac{A_{12}}{e^{B_{12}\widehat{r}}} - \frac{C_{12}}{\widehat{r}^6} - \frac{1}{\widehat{r}} > 0, \qquad r \geq \sqrt{2}n \tag{3}$$

In Lemmas 1, 2 and 3 it will be assumed that when removing k' charges, it will be preferable to choose k' pairs over any other set of charges in the arrangement. This will be proven in Lemma 4.

Lemma 1. *Inequalities (1) and (2) are sufficient to ensure that an independent set is left if one exists.*

Proof. It can be seen that inequality (1) will ensure that if there are two pairs corresponding to points that intersect, the total energy will always decrease the total energy by removing one of the pairs. Inequality (2) compliments this by ensuring that given a pair corresponding to a vertex with no adjacent neighbours, the total energy would increase by removing it. This holds even in the case that all other pairs are at a distance of $\sqrt{2}n$. It can be seen that these two inequalities combined will mean that the global minimum total energy for any subset will be the maximum independent set. Note that the total energy will decrease with the cardinality of the given independent set. $\square$

Lemma 2. *There exists, for any structure created from a long orthogonal penny graph, some parameters such that Inequalities (1, 2) and (3) are satisfied.*

Proof. Values are chosen for A_{12}, B_{12} and C_{12} such that the energy for any pair of ions of opposite charge at a distance of 1 is -1. This is achieved by choosing a value of $\frac{1}{2n^2}$ for A_{12}, 0 for B_{12}, and $\frac{1}{2n^2}$ for C_{12}. Note that these values for A_{12} and B_{12} will lead to a constant interaction of $\frac{1}{n^2}$ between ions on separate planes. This simplifies the energy equation to

$$U^{BC}(r) = \frac{A_{11}}{e^{B_{11}r}} - \frac{C_{11}}{r^6} + \frac{1}{r} + \frac{1}{2n^2} - \frac{1}{2n^2\widehat{r}^6} - \frac{1}{\widehat{r}}.$$

To satisfy Inequality 1, $U^{BC}(n) > 1$. This may be satisfied by choosing values for A_{11}, B_{11}, and C_{11} such that $\frac{A_{11}}{e^{B_{11}n}} - \frac{C_{11}}{n^6} = 1$, noting that $\frac{1}{2n^2} - \frac{1}{2n^2\widehat{r}^6} + \frac{1}{r} - \frac{1}{\widehat{r}} > 0$ for all positive distances greater than 1. This can be satisfied by solving the equation $\frac{A_{11}}{e^{B_{11}n}} - \frac{C_{11}}{n^6} = 1$, choosing $A_{11} = \frac{C_{11}e^{Bn}}{n^6} + e^{Bn}$.

Inequality (2) requires that at a distance of at least $\sqrt{2}n$ the total energy is no more than $\frac{1}{n^2}$. This can be satisfied at a distance of $\sqrt{2}n$ by ensuring that the $\frac{A_{11}}{e^{B_{11}\sqrt{2}n}} - \frac{C_{11}}{8n^6} = 0$, which can be satisfied after substituting in the appropriate value for A_{11} with $C_{11} = \dfrac{e^{Bn}}{e^{B\sqrt{2}n}\left(\frac{1}{(\sqrt{2}n)^6} - \frac{e^{Bn}}{n^6 e^{B\sqrt{2}n}} \right)}$, which

simplifies to $\frac{n^6}{e^{(\sqrt{2}-1)Bn} - 8}$. Finally, consider the value of B_{11}. Note that the value of both A_{11} and C_{11} depend greatly on B_{11}, with a small increase in B_{11} leading to a very rapid increase in the value of A_{11} and a rapid decrease in C_{11}. Similarly, the value of the energy given by $\frac{A_{11}}{e^{B_{11}r}} - \frac{C_{11}}{r^6}$ rapidly decreases initially before converging at approximately 0, which can be seen by noting that the first derivative with respect to r of this equation for a given n is $-B\frac{A_{11}}{e^{Br}} + \frac{6C_{11}}{r^7}$. As such by choosing a suitably large B_{11} Inequality (2) can be easily satisfied, one obvious choice for this would be $B_{11} = n$.

Note that both $\frac{A_{11}}{e^{Br}}$ and $\frac{C_{11}}{r^6}$ will strictly decrease, therefore if $\left| \frac{A_{11}}{e^{Br}} \right| \leq \frac{1}{2n^2} - \frac{1}{r} + \frac{1}{\widehat{r}}$ and $\left| \frac{C_{11}}{r^6} \right| \leq \frac{1}{r} - \frac{1}{\widehat{r}}$ both Inequalities (2) and (3) are satisfied. From the value of B_{11} this becomes $\frac{n^6}{e^{(\sqrt{2}-1)n^2} - 8}$, which will be positive and less than $\frac{1}{n^2}$ for any $n \geq 6$. Note that $\frac{1}{r} - \frac{1}{\widehat{r}} \geq \frac{1}{r^6}$ for any $r \geq 1.3$, hence it is clear that $\frac{C_{11}}{r^6} \leq \frac{1}{r} - \frac{1}{\widehat{r}}$ for $n \geq 6$. Considering $\frac{A_{11}}{e^{Br}}$ at a distance of $\sqrt{2}n$, the equation becomes $\frac{A_{11}}{e^{n^2\sqrt{2}}} = \frac{C_{11}}{n^6 e^{n^2\sqrt{2}}} + \frac{1}{e^{(\sqrt{2}-1)n^2}}$, from the previous arguments it follows that this will be considerably less than $\leq \frac{1}{2n^2} - \frac{1}{r} + \frac{1}{\widehat{r}}$ for $n \geq 6$.

Finally it can be noted that due to the constant $\frac{1}{2n^2}$ term there is a positive value for any distance greater than $\sqrt{2}n$, satisfying Inequality (3). Using these values, it is now shown how to choose a set of parameters which satisfy the Inequalities. $\square$

Lemma 3. *Given k pairs, the energy will be less than $(k-1)(\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1)$ if and only if the pairs correspond to an independent set of size k, for $\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1 < 0$.*

Proof. Given k pairs, the energy between the ions in each pair will be $\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1$, for a total of $k((\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1))$. Inequality (2) ensures that the maximum energy gained from pairs of ions corresponding to non-intersecting circles will be at most $|((\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1))|$. Inequality (3) ensures that having charges on the same plane will lead to a slight positive charge. From this it follows that the maximum energy a set of ions corresponding to an independent set will be $(k-1)(\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1)$. Conversely, from Inequality (1) it is known that if there is a pair of intersecting circles the total energy must be greater than $(k-1)(\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1)$.

It can be noted for AT-LEAST-K-CHARGE REMOVAL that if greater than k' pairs were removed this energy could not be achieved as the minimum energy would be $(k-1)(\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1)$ for the interaction within pairs. As there is a positive interaction between pairs, the total energy must be slightly greater than this for any $k > 1$. Therefore it can be seen that the total energy will be less than $(k-1)(\frac{A_{12}}{e^{B_{12}}} - C_{12} - 1)$ if and only if there is an independent set of size k left. It can be noted that under the choice of variables from Lemma 2, the upper bound of will be $-(k-1)$. $\square$

Lemma 4. *When removing k' charges from the construction from a long orthogonal penny graph, it is always preferable to remove pairs provided that Inequalities (1- 3) hold.*

Proof. Assume that this statement is false, there must be some assignment where it is preferable to remove some set of at least two vertices, v_i^+ and v_j^-, that do not form a pair with any ions that have be removed. Assume that there are t positive and t negative vertices in the graph. If instead v_j^- was left in, while v_i^+ was removed, the remaining energy would change by at least

$-1 + t \left(\frac{A_{11}}{e^{B_{11}\sqrt{2}n}} - \frac{C_{11}}{(\sqrt{2}n)^6} + \frac{1}{\sqrt{2}n} + \frac{A_{12}}{e^{B_{12}\sqrt{2}n}} - \frac{C_{12}}{\widehat{\sqrt{2}n}^6} - \frac{1}{\widehat{\sqrt{2}n}} \right)$. From the arguments in Lemma 1 and the construction in Lemma 2 it can be seen that this will lead to a decrease in total energy, making it preferable and therefore contradicting the assumption. Note that given a positively weighted vertex of the maximum degree, in this case 3, it could contribute at most $\frac{3}{2n^2} - \frac{3}{2n^8} - \frac{3}{n}$ which will have a magnitude less than 1 for any $n \geq 3$. Therefore, by contradiction this holds. $\square$

Theorem 4. K-CHARGE REMOVAL, MINIMAL-AT-LEAST-K-CHARGE REMOVAL *and* AT-LEAST-K-CHARGE REMOVAL *are NP-Complete when limited to only two species of ion and restricted to the Buckingham-Coulomb potential energy function.*

Proof. Building on the results from Lemmas 1, 2, 3, and 4, NP-Completeness will now be shown. Lemma 1 shows that, provided Inequalities (1) and (2) hold, the optimal solution will be to leave an independent set. From Lemma 2 it follows that these inequalities are satisfiable for any graph under the given construction, noting that the assignment of parameters gives an energy of -1 within pairs. Lemma 3 shows that the upper bound is reachable if and only if an independent set has been left. It follows from Lemma 4 that the assumption that it is preferable to remove a set of pairs over any other set of charges holds when the inequalities also do.

Therefore there will be a satisfiable instance of K-CHARGE REMOVAL or any generalisation if and only if the instance of INDEPENDENT SET for the maximum degree 3 planar graph instance is satisfiable. Conversely if the INDEPENDENT SET instance is satisfiable, the corresponding K-CHARGE REMOVAL instance can be satisfied by leaving the vertices corresponding to the independent set in the long orthogonal penny graph construction. Hence under these restriction all three problems will be NP-complete. Note that this may be extended to charges of $\pm c$ for any given c. $\square$
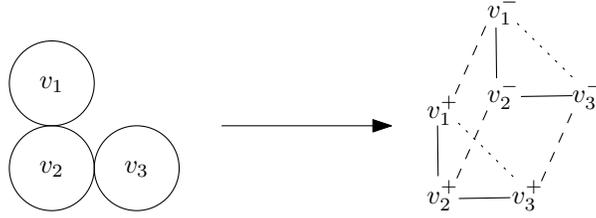
Figure 2: Example of the construction of an arrangement from a penny graph, om this example v_1 and v_2 are adjacent, as are v_2 and v_3, but not v_1 and v_3
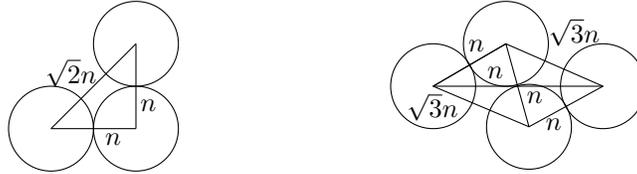

Figure 3: Illustration of the distances between the centre of non-adjacent pennies using the construction of Cerioli et al. [5].

5 Restriction to the Coulomb potential with unbounded charges

The final case that will be considered in this work is when the energy function is the Coulomb potential. NP-Hardness for this case will be shown by a reduction from KNAPSACK to AT-LEAST-K-CHARGE REMOVAL. Note that with an unbounded number charges this problem is not in NP for MINIMAL-AT-LEAST-K-CHARGE REMOVAL due to Proposition 3 and is trivially NP-Hard for K-CHARGE REMOVAL due to Corollary 1. This reduction requires using an unbounded number of charges, thus it follows from proposition 3 that it is NP-Hard to verify if a solution to an instance of AT-LEAST-K-CHARGE REMOVAL is minimal.

In this reduction it will be shown that an instance of AT-LEAST-K-CHARGE REMOVAL such that the set of ions left will correspond to the items for the knapsack instance if and only if there is a set satisfying the knapsack instance.

Theorem 5. AT-LEAST-K-CHARGE REMOVAL *and* MINIMAL-AT-LEAST-K-CHARGE REMOVAL *remains NP-Hard when the energy function is limited to the Coulomb potential.*

Proof. In the knapsack problem, henceforth KNAPSACK, the input is a bag with capacity C, and a set of items S. Each item $i \in S$ has a weight w_i, and a value p_i. In this problem the goal it to find the subset $S' \subseteq S$ such that $\sum_{i \in S'} p_i$ is maximised conditional on $\sum_{i \in S'} w_i \leq C$. Alternatively this may phrased as a decision problem by taking some goal value g and asking if there is an S' such that $\sum_{i \in S'} p_i \geq g$.

NP-completeness for AT-LEAST-K-CHARGE REMOVAL and MINIMAL-AT-LEAST-K-CHARGE REMOVAL will be shown by a reduction from the knapsack problem. Given an instance, I, of the knapsack problem as described above, an instance, I', of AT-LEAST-K-CHARGE REMOVAL is created as follows. For every $i \in S$, two vertices are created denoted v_i^+ and v_i^- and label with the corresponding item. These are assigned a weight of w_i to v_i^+ and $-w_i$ to v_i^-.

The values u and α will be defined such that u is some value such that there does not exit any pair of items, i and j, such that $p_i > p_j$ but $p_j + u \geq p_i$, and u is less than the smallest *unit of precision* for the value of the items. Using this, α is defined as some value satisfying the inequality $u > \frac{4n^2 w_{max}^2}{\alpha}$, where w_{max} is the weight of the heaviest item. This ensures that α is some distance such that if all

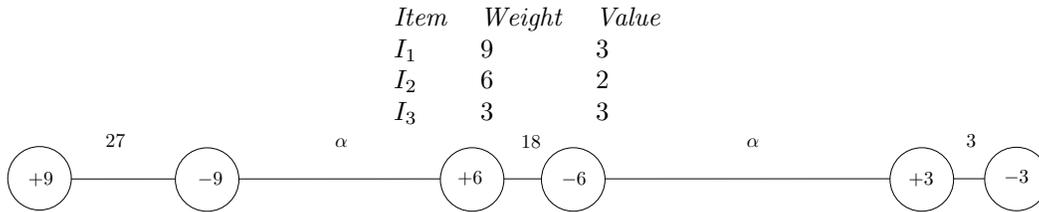|	Item	Weight	Value
	I_1	9	3
	I_2	6	2
	I_3	3	3

Figure 4: Example of construction of the structure from the knapsack instance. In this $u < 1$ and $\alpha > 2916$

vertices are at least α away from each other there will be a difference of no more than u in energy, which will be sufficient to ensure that vertices at that distance may be safely ignored.

These vertices are now placed such that for each item such that the distance between the two vertices v_i^+ and v_i^- will give a potential energy of $-p_i$. Recall that $U_{ij}^C = \frac{q_i q_j}{r_{ij}}$, therefore this can be achieved by placing them at a distance of $\frac{w_i^2}{p_i}$. Each of the pair of vertices representing an item is placed in a line so that the distance between any two pairs is no less than α. An example of this construction is provided in Figure 4.

The value k is chosen as $k = \left(\sum_{i \in I} w_i \right) - c$, ensuring that there will be no more than c charges left after removing k, corresponding to a valid assignment for the knapsack instance. Finally, the goal value will be chosen as $g' = -g + u$.

It follows from this construction that any removal of at-least-k charges will be a valid packing in terms of the capacity.

If the AT-LEAST-K-CHARGE REMOVAL instance is satisfiable then there must be some valid packing of no more than g' energy. As the interaction between vertices corresponding to different items is trivially small, the only way to achieve this is to choose a set of vertex pairs with an energy between them no more than g'. As the energy between pairs is equal to the value of the items, the only way this can be achieved this is to have items corresponding to a packing with value at least g. Conversely if the AT-LEAST-K-CHARGE REMOVAL instance is not satisfiable, there does not exist a packing of value g by the same arguments.

Similarly if the KNAPSACK instance is satisfiable then the AT-LEAST-K-CHARGE REMOVAL instance may be satisfied by removing all vertices not corresponding to a satisfying packing of the KNAPSACK instance. Finally if the KNAPSACK instance is not satisfiable then by the previous arguments the KCR instance also can not be satisfied. Therefore this problem is NP-Complete. Note that as the weights on all items will be positive, with a corresponding negative energy in I', given a non-minimal satisfying solution there will exist some minimal satisfying solution. Therefore MINIMAL-AT-LEAST-K-CHARGE REMOVAL will also be NP-Hard. □

6 Conclusions and future work

In this work we have presented the new problem of k-charge removal, and a class of functions for which the general case is NP-Complete. We have also shown that the problem remains NP-Complete under both the restriction that we have only two species of ions and the Buckingham-Coulomb energy function and the restriction we only use the Coulomb potential on an unbounded number of ion species.

One obvious question would be if approximation results can be gained for this problem. We would submit that while it seems likely that the general case is APX-hard, under a bound on the number of ions this problem may well be approximable within a reasonable factor.

From a chemistry stand point, while we have made progress towards physical constructions there is still a lot that could be done in this regard. As such investigation into the restrictions of having more realistic physical values remains an important unexplored direction. Another question would be if we can investigate the convergence of these interactions, particularly the Coulomb potential, over a periodic structure to more fully understand the energy function.

References

[1] AA Ageev, AV Kelmanov, and AV Pyatkin. Np-hardness of the euclidean max-cut problem. In *Doklady Mathematics*, volume 89, pages 343–345. Pleiades Publishing, 2014.

[2] Kyriakos Axiotis and Christos Tzamos. Capacitated dynamic programming: Faster knapsack and graph algorithms. *arXiv preprint arXiv:1802.06440*, 2018.

[3] F Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241–3253, oct 1982.

[4] Richard A Buckingham. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences.*

[5] Marcia R. Cerioli, Luerbio Faria, Talita O. Ferreira, and Fábio Protti. A note on maximum independent sets and minimum clique partitions in unit disk graphs and penny graphs: complexity and approximation. *RAIRO - Theoretical Informatics and Applications - Informatique Théorique et Applications*, 45(3):331–346, 2011.

[6] Brent N. Clark, Charles J. Colbourn, and David S. Johnson. Unit disk graphs. *Discrete Mathematics*, 86(1):165 – 177, 1990.

[7] C Collins, MS Dyer, MJ Pitcher, GFS Whitehead, M Zanella, P Mandal, JB Claridge, GR Darling, and MJ Rosseinsky. Accelerated discovery of two crystal structure types in a complex inorganic phase field. *Nature*, 546(7657):280, 2017.

[8] M. Garey and D. Johnson. The rectilinear steiner tree problem is np-complete. *SIAM Journal on Applied Mathematics*, 32(4):826–834, 1977.

[9] M. Krishnamoorthy and N. Deo. Node-deletion np-complete problems. *SIAM Journal on Computing*, 8(4):619–625, 1979.

[10] Andriy O. Lyakhov, Artem R. Oganov, and Mario Valle. How to predict very large and complex crystal structures. *Computer Physics Communications*, 181(9):1623 – 1632, 2010.

[11] Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar k-means problem is np-hard. In Sandip Das and Ryuhei Uehara, editors, *WALCOM: Algorithms and Computation*, pages 274–285, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[12] Caroline Mellot-Draznieks, Stphanie Girard, Grard Frey, J. Christian Schn, Zeljko Cancarevic, and Martin Jansen. Computational design and prediction of interesting not-yet-synthesized structures of inorganic materials by using building unit concepts. *Chemistry A European Journal*, 8(18):4102–4113, 2002.

[13] Artem R. Oganov. Crystal structure prediction: reflections on present status and challenges. *Faraday Discuss.*, 211:643–660, 2018.

[14] Artem R Oganov and Colin W Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of chemical physics*, 124(24), 2006.

[15] David Pisinger. Linear time algorithms for knapsack problems with bounded weights. *Journal of Algorithms*, 33(1):1–14, 1999.

[16] Yanchao Wang, Jian Lv, Li Zhu, Shaohua Lu, Ketao Yin, Quan Li, Hui Wang, Lijun Zhang, and Yassie ma. Materials discovery via calypso methodology. *Journal of physics. Condensed matter : an Institute of Physics journal*, 27:203203, 04 2015.

[17] L T Wille and J Vennik. Computational complexity of the ground-state determination of atomic clusters. *Journal of Physics A: Mathematical and General*, 18(8):L419–L422, jun 1985.

[18] Scott M Woodley and Richard Catlow. Crystal structure prediction from first principles. *Nature materials*, 7(12):937, 2008.

[19] M. Yannakakis. Node-deletion problems on bipartite graphs. *SIAM Journal on Computing*, 10(2):310–327, 1981.

[20] Mihalis Yannakakis. Node-and edge-deletion np-complete problems. In *CONFERENCE RECORD OF THE TENTH ANNUAL ACM SYMPOSIUM ON THEORY OF COMPUTING (SAN DIEGO, CALIF., 1978), ACM*, pages 253–264, 1978.