Unterschrift Betreuer

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DISSERTATION

Confinement-Driven Self-Assembly of Charged Particles

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften

unter der Leitung von

Ao. Univ. Prof. Dr. techn. Gerhard Kahl
Institut für Theoretische Physik (E136)
Technische Universität Wien

eingereicht an der Technischen Universität Wien
Fakultät für Physik

von

Dipl.-Ing. Benedikt Hartl
Matrikelnummer 01026773

Wien, 19. Oktober 2020

Unterschrift Student

Kurzfassung

Das Phänomen der Selbstorganisation ist allgegenwärtig in der Natur. Unter bestimmten externen Bedingungen können sich beispielsweise atomare oder molekulare Bausteine selbstständig in komplexe, geordnete Nano-Materialien anordnen, welche dann neuartige Eigenschaften aufweisen. Dieser Prozess kann technologisch genutzt werden, etwa um integrierte Schaltkreise herzustellen, deren logische Elemente nur einige Nanometer groß sind. Selbst Technologien wie Nanomaschinen oder künstliche Rezeptoren scheinen greifbar. Maßgeblich für die physikalischen Eigenschaften von Materialien sind nicht nur die Eigenschaften der jeweiligen Materialbestandteile, sondern auch die vorherrschende mikroskopische Struktur des Materials. Es ist daher von grundlegender Bedeutung zu verstehen, welche Struktur ein Material unter bestimmten äußeren Bedingungen einnehmen wird.

Diese Dissertation umfasst eine umfangreiche computergestützte Abhandlung der Strukturbildung zweier interessanter physikalischer Systeme: Zuerst untersuchen wir das asymmetrische Wigner-Doppelschicht-System, bestehend aus Punktladungen, welche an die Oberflächen zweier paralleler, entgegengesetzt geladener Platten gebunden sind. Des Weiteren behandeln wir ein supramolekulares System von geladenen, polyaromatischen Molekülen, die sich an der Grenzschicht einer Metalloberfläche und einer Flüssigkeit unter elektrochemischer Kontrolle anordnen. Als grundlegende Vorgehensweise verfolgen wir die Strategie, die innere Energie der Modelle der jeweiligen Systeme mithilfe von eigens entwickelten, hoch spezialisierten numerischen Methoden zu minimieren, um geordnete Grundzustände abhängig von den jeweiligen externen Systemparametern vorherzusagen.

Im Speziellen suchen wir im asymmetrischen Wigner-Doppelschicht-System nach geordneten, aber aperiodischen Strukturen dieses Systems, sogenannten Quasikristallen. Bestimmte Kombinationen der Systemparameter erlauben die Ausbildung von zwölfeckigen Clustern und Super-Clustern der Ladungen. Diese lokalen Strukturen stellen wichtige Vorstufen von quasikristalliner Ordnung dar. Im Allgemeinen weisen Quasikristalle sowohl konzeptionell als auch technologisch faszinierende Eigenschaften auf.

Für unsere Untersuchungen zur Selbstorganisation des supramolekularen Systems präsentieren wir eine numerisch effiziente und zuverlässige Methode zur Vorhersage geordneter Konfigurationen der Moleküle. Unsere numerischen Ergebnisse sind in semiquantitativer Übereinstimmung mit experimentell beobachteten zwei- und dreidimensionalen supramolekularen Gitterstrukturen: Abhängig von den gewählten Systemparametern beobachten wir flache, poröse Gitter, Auto-Wirt—Gast Muster und stratifizierte Doppelschichten von den im Experiment verwendeten organischen Molekülen an der Metalloberfläche. Die Struktur und dadurch auch die Eigenschaften einer Oberfläche durch die Wahl der Systemparameter gezielt zu verändern, wäre technologisch höchst interessant.

Abstract

The field of bottom-up self-assembly focuses on the self-organization principles of basic atomic or molecular building blocks into well-ordered functional nano–materials with tailored properties. Assisted by bottom-up processes, nano–electronic devices can be realized and technological applications such as more efficient electrical energy storage devices, nano–machines and molecular drug delivery systems have come within reach. The physical properties of a material are usually determined not only by the related properties of its constituents but also by the microscopic structure of the material. Crucial to the understanding of the corresponding properties of a material is thus to first understand the structure a material will assume under certain external conditions.

This thesis is dedicated to a comprehensive computational study of the structure formation processes of two different systems: first, we consider classical point-charges confined to the surface of two parallel, oppositely charged plates, the so-called asymmetric Wigner bilayer system. Second, we investigate a supramolecular system of charged, polycyclic aromatic molecules, deposited on a metal–liquid-interface under electrochemical conditions. We study the structure formation processes of both systems by minimizing the internal energy of the related models of the systems for different sets of the associated system parameters by employing sophisticated numerical tools.

We systematically search in the asymmetric Wigner bilayer system for possible quasicrystalline ordering: quasicrystals exhibit orientationally long-range ordered yet spatially aperiodic particle arrangements and have intriguing conceptual and technological implications. For selected combinations of the system parameters, we observe the formation of dodecagonal clusters and super-clusters of the point-charges, structures that are important precursors of aperiodic quasicrystalline ordering.

For the study of the self-assembly processes of the supramolecular system, we propose a computationally lean approach to treat this problem reliably with elaborated numerical tools. In a semi-quantitative agreement with experimental data, the target molecules are seen to self-organize into two- and three-dimensional supramolecular lattices for different sets of the system parameters: the molecules form an open porous structure, an auto-host–guest pattern and a stratified bilayer.

Contents

1 Introduction **1**

2 Methods **7**

2.1 Long-Ranged Coulomb Interaction in Quasi-Two and Three Dimensions 8

2.1.1 What is a Long-Ranged Potential? . . . 10

2.1.2 Ewald Summation of a System of Point-Charges in Three Dimensions . . . 13

2.1.3 Anisotropic Directions: Ewald Summation in Quasi–Two Dimensions . . . 19

2.1.4 Convergence Criteria for Evaluating Ewald Sums Numerically . . 22

2.2 Ground State Exploration: Objective, Methods and Algorithms . . . 24

2.2.1 Physical Energy Landscapes can be Complex . . . 24

2.2.2 Navigating towards the Global Minimum . . . 26

2.2.3 Monte Carlo Methods: Simulated Annealing and Replica Exchange Monte Carlo . . . 30

2.2.4 Evolutionary Algorithm . . . 34

2.3 Order Parameters of Structural Data . . . 47

2.3.1 Local Environments and Neighbors: Voronoi Construction . . . 47

2.3.2 Bond Orientational Order Parameters . . . 48

2.3.3 Orientational Order Parameters . . . 50

2.3.4 Global Symmetries: Reciprocal Lattice and Structure Factor . . . 50

2.4 Categorizing Structural Data with Unsupervised Clustering Algorithms . 53

2.4.1 Dimensional Reduction with Principal Components Analysis . . . 55

2.4.2 k-means Clustering . . . 57

2.4.3 Density-Based Clustering: DBSCAN . . . 58

2.4.4 t-SNE Clustering and High-Dimensional Visualization . . . 59

2.4.5 (Adjusted) Mutual Information . . . 61

3 Systems **63**

3.1 Towards Quasicrystalline Order in the Asymmetric Wigner Bilayer System 65

3.1.1 Introduction . . . 65

3.1.2 Terminology . . . 68

3.1.3 The Model of the Asymmetric Wigner Bilayer System . . . 70

3.1.4 Revisiting the Complex Phase Diagram with Unsupervised Clustering Algorithms . . . 73

3.1.5 Towards Dodecagonal Order: Metastable Super-Structures . . . 105

3.1.6 Towards a Bilayer-Modulated Hexagonal Wigner Quasicrystal . . 135

3.1.7 Miscellaneous . . . 168

3.1.8 Conclusions and Outlook . . . 176

3.2 Supramolecular Ordering of Complex Molecules under Electrochemical Conditions . . . 180
3.2.1 Introduction . . . 180
3.2.2 The System and its Representations . . . 183
3.2.3 Identifications of Self-Assembly Scenarios . . . 193
3.2.4 Results for the PQP-ClO_4 System . . . 196
3.2.5 Conclusions and Outlook . . . 203

4 Conclusions and Outlook **205**
4.1 General Conclusions . . . 205
4.2 Outlook and Future Work . . . 208

Acknowledgments **213**

A Supplementary Theory and Information **215**
A.1 The Asymmetric Wigner Bilayer System . . . 216
A.1.1 Unsupervised Clustering of Order Parameters . . . 216
A.1.2 Ground State Symmetries of Clustering Results: Order Parameters and Principal Components . . . 217
A.1.3 Mechanical Stability of Super Square-Triangle Tilings . . . 225
A.1.4 Rectangle-Large–Triangle-Small–Triangle decoration . . . 226
A.1.5 Towards Constructive, Defect-Free Inflation . . . 229
A.2 Predictive Supramolecular Self-Assembly . . . 233
A.2.1 The System and its Representations . . . 233
A.2.2 Identifications of Self-Assembly Scenarios . . . 240
A.2.3 Results . . . 242

Bibliography **245**

1 Introduction

Self-assembly is a ubiquitous phenomenon in nature. Under certain external conditions, rather simple, basic building blocks such as atoms, molecules or nano–particles are observed to organize themselves without external guidance into well-ordered functional nano–materials with novel, tailored properties [1]. For instance, the molecular building blocks of biological systems (such as lipids or amino acid residues) self-assemble into membranes, proteins or DNA, which then, in further hierarchical steps of self-assembly, allow the formation of viruses, cells, tissues, blood vessels or nervous systems [2–5].

More than sixty years ago Richard P. Feynman suggested to study these organizing principles of nature more intensively in order to make them technologically available [6]. Via his famous lecture entitled "There's Plenty of Room at the Bottom." [7] he advertised the field of *bottom-up* self-assembly in 1959: in this research area effort is dedicated to designing the basic, microscopic building blocks of a material such as molecules or nano–particles which then self-assemble, even from an initially disordered configuration, into the desired, well-ordered functional nano–structures [3, 5]. In contrast, a more conservative approach to nanofabrication is the so-called *top-down* method, where nano–structures are synthesized by selectively modifying the surface of an existing crystalline structure on a larger length-scale compared to the size of the material's constituents (e.g. with lithography techniques, now also reaching 10 nm of resolution [3]).

By now, directed self-assembly has become an indispensable tool in numerous (nano-) technological applications in order to create functional materials with tailored properties [8]: by utilizing the morphology of block co-polymers it is possible to self-assemble templates which facilitate accurate control over the structure formation of micro and nano devices and materials, ranging from nano–electronics [9] over more efficient electrical energy storage devices [10, 11] to nano–machines [12] and microscopic robots [13].

Another, rapidly growing field of research is supramolecular chemistry, which is dedicated to the structure formation phenomena of chemical building blocks into supramolecular structures beyond individual molecules, and lies as such at the basis of many nano- and mesoscopic structures found in biology (at length scales of $\sim 10^2$ nm to 1 μm) [14, 15]. Here, the predominant intermolecular interactions are typically of non-covalent nature, such as ion-pairing, hydrogen bonding, van der Waals forces, π–π-stacking or electrostatic contributions, to name a few [16]. Through supramolecular chemistry concepts such as molecular recognition are studied, which are relevant, for instance, for selective binding mechanisms of biological or artificial receptors, or for molecular host-guest mechanisms where selected molecules or ions can be trapped within (and released from) a supramolecular host matrix; such processes may be exploited for molecular drug delivery systems [17], for gene regulation and oncogenesis [18] or for gas-storage, gas-separation or catalysis [19]. Ideally, the well-defined supramolecular architectures

required for these applications self-assemble spontaneously but in a controlled manner under given external conditions [14]. Many different molecular building blocks have been used in supramolecular chemistry, among those are cryptands or crown ethers [14, 20], polycyclic aromatic compounds [21, 22], fullerens [23, 24], dendrimers [25, 26] or nanoparticles and nanorods [27–30], to name a few. Notably, also the solvent, in which the molecular units are immersed, can have a critical impact on the self-organization strategies of supramolecular systems [21, 22].

Related, promising examples for highly flexible and versatile building blocks for self-assembly are colloidal particles (whose sizes range typically from nm to μm), which show a high capacity and propensity to arrange themselves without external guidance into well-ordered structures such as lattices. Colloidal particles can be synthesized in the lab with characteristic features, and there seems to be no limit in tailoring properties such as their shape [31–36], their surface decoration [37–39], or their interaction properties. As for the shape [31], colloidal particles of ellipsoidal, cuboidal, or polyhedral shape [32–36] can meanwhile be synthesized or they can even emerge as octapods [40]. As for the surface decoration, many strategies are applied, such as DNA-grafting [37, 41] (exploiting the selective binding mechanisms of DNA-strands) or treating selected regions of the particles via chemical or physical processes [38, 39]. The preceding list of references is far from being complete.

Some self-assembled nano–materials may even display quasicrystalline ordering [32, 35, 42–50] where the microscopic constituents of the structures exhibit long-range orientational order but no translational symmetry (in contrast to crystalline structures) [51]. Such an ordered, yet aperiodic arrangement of the building blocks of a material has interesting conceptual and technological implications: quasicrystals typically show, for instance, low friction coefficients owed to their inherent structural incommensurability to crystalline materials at the microscopic level. On the other hand, metallic quasicrystals often display low thermal conductivity and corrosion resistance while being exceptionally hard (some quasicrystalline materials are estimated to be roughly ten times as hard as steel) [52]. These material properties render quasicrystals as ideal candidates for "non-stick" coatings, for instance for frying pans [53]. Furthermore, quasicrystalline materials may be used for bone repair and prosthetic applications, where biocompatibility, corrosion resistance and low friction properties are required [52].

In general, the thermal, mechanical, optical or electronical properties of a material – which are relevant for technological applications – are usually determined by the specific microscopic arrangement of the material's constituents, *i.e.*, by the structure formed by the atomic or molecular building blocks (or by even more complex units such as nano–particles or macro-molecules) [54]. So-called *metamaterials* [55], for instance, show novel, unconventional properties (which do not occur naturally) purely based on the specific, highly ordered arrangements of the building blocks of the materials. In that way, special waveguides for electromagnetic-, but also for acoustic- [56] and elastic-waves [57] can be designed; notably, the respective patterns of the particle arrangements must be of a smaller length-scale than the wavelength of the phenomenon which is influenced. Photonic crystals [58, 59] may exhibit unconventional optical properties such as a negative refraction index [60] and special type of electromagnetic wave guides have been proposed, which might act as a "black hole" for radiation of

certain frequencies [61]. Thus, understanding the structure a material will attain (under certain experimental conditions) is key to understanding the corresponding properties of the material [54].

However, in bottom-up processes it is not clear *a priori* which polymorph a material will assume given its constituents and the relevant system parameters (such as temperature, pressure, external electric or magnetic fields, etc.), which additionally govern the structure formation process. The field of computational material science is dedicated to predicting the structure of materials based on modeling of the relevant physics and on computer simulations and optimization algorithms to study the corresponding structure formation process. In that way, the properties of materials may ideally be predicted even before the materials are synthesized in the lab [54].

In this thesis, we are concerned with computational studies of the structure formation processes of two systems composed of (i) negative, classical point-charges confined to the surfaces of two parallel, uniformly but differently charged plates and (ii) of charged, polycyclic aromatic molecules confined to the volume near a metal–liquid-interface under electrochemical conditions.

In Section 3.1, we numerically investigate (i) the asymmetric Wigner bilayer system [62–64] for possible quasicrystalline ordering. In this system, negative, classical point-charges are confined to the surface of two oppositely charged, parallel plates of uniform but different charge densities. The point-charges thereby form Wigner crystals [65] on each plate, whose structures strongly depend on the system parameters (given by the plate separation distance and the charge ratio of the plates). In general, such bilayer model systems may be used to describe systems of charged particles that are located at the interface between two fluids as, for instance, electrons trapped at the surface of liquid Helium [66, 67] or ions which are confined by electromagnetic fields of Paul- and Penning-traps [68–70]. Furthermore, junctions between metals and semiconductors in solid state physics [71–75] or even very simplified representations for biological cell-membranes [76–80] might be described by such bilayer systems.

In Section 3.2, we study (ii) the self-assembly behavior of charged polycyclic aromatic molecules on a metal–liquid interface under electrochemical conditions [21, 22, 81–86]. Here, we deal with intermolecular interactions and structure formation into supramolecular lattices, which originate from a delicate interplay between the typically non-covalent intermolecular and molecule–surface interactions. Such polycyclic aromatic compounds, in particular, are promising candidates that are able to self-assemble into active, functional materials (be it photoactive, electroactive, ionoactive, thermoactive, or chemoactive) that may find applications in supramolecular electronic and opto–electronic devices [14, 22]. Furthermore, the size and topology of the flat conjugated systems, as well as functional groups of such polycyclic aromatic compounds can synthetically be modified in order to control their electronic properties [81–85]. The self-assembly of such complex organic molecules on inorganic substrates often displays a particularly rich and intricate polymorphism of the supramolecular structures [54, 87–91]. In recent experiments [83] the spontaneous and reversible transition between two- and three-dimensional self-assembly scenarios of a supramolecular system ($PQPClO_4$) at a solid–liquid interface under electrochemical conditions (*i.e.*, Au(111) in aqueous 0.1M perchloric acid) has been studied and thoroughly discussed; such

controllable self-assembly scenarios are both of fundamental and of technological relevance, especially in fields such as corrosion, catalysis and nano–electronics [21, 22]. By a simple variation of the interfacial potential it was possible to selectively organize the disc-shaped PQP^+ cations in an open porous pattern, to fill these pores to form an auto-host-guest structure or to stack the building blocks in a stratified bilayer. In this thesis, complementary theoretical investigations are dedicated to rationalize these experimental observations [83].

From a thermodynamic point of view, the specific equilibrium arrangement of the constituent particles of a physical system minimizes the thermodynamic potential as given by the experimental conditions. Such thermodynamic potentials may be the entropy, S, of a system for the NVE-ensembles (constant particle number, N, constant volume, V, and constant energy, E), the Helmholtz free-energy, $F = E - T\,S$, for the NVT-ensemble (constant particle number, N, constant volume, V, and constant temperature, T), and so on.

From a computational point of view, identifying particle configurations can be considered as a sampling problem of the space of all possible configurations with the objective of minimizing the corresponding thermodynamic potential of a system. In practice, the sampling of this configuration space is performed under the variation of all microscopic degrees of freedom (such as the positions and the orientations of the particles or molecules, as well as the shape of the simulation box containing all the particles). In this thesis, we restrict ourselves to investigations of ordered equilibrium structures of the NVT-ensemble of the systems introduced above at vanishing temperature, $T = 0\,\mathrm{K}$. Identifying the particular particle configuration which is indeed realized for a specific system then reduces to the problem of finding the ordered ground state structure which minimizes the internal energy of the system; since the entropic contribution to the free energy vanishes due to $T = 0\,\mathrm{K}$.

Systems (i) and (ii) described in Sections 3.1 and 3.2 fall into a large class of quasi–two dimensional (quasi-2D) systems where the spatial extension in one direction (typically the one perpendicular to the surface) is very small compared to the spatial extent into the other two in-plane directions [92]. Furthermore, we assume that the particles in our (classical) model representations of the two systems carry Coulomb charges, such that the corresponding interparticle potentials are of long-ranged nature. Particular caution needs to be taken to correctly account for the long-ranged contributions to the internal energy of a system when trying to predict ordered ground state configurations in anisotropic quasi-2D geometries with computer simulations: on the one hand, different polymorphs of such systems may exhibit only tiny energy differences at the same system parameters [62–64], complicating the sampling procedure. On the other hand, computationally efficient implementations [93] of techniques used to evaluate the internal energy of systems with long-ranged interactions (such as the method of Ewald summation [94]) typically rely on a spatially fully-periodic treatment of a finite-sized simulation box of the system. We here employ specialized numerical tools to treat long-ranged Coulomb interactions in quasi-2D systems with high accuracy in a numerically efficient manner [62, 92, 95].

In this context, we want to emphasize that, in general, optimization problems faced in computational material science are highly non-trivial since the number of possible

local minima in the potential energy surface (embedded in a high-dimensional parameter space) increases exponentially with the number of particles (and their degrees-of-freedom) of a system [96]. Exhaustive search strategies very quickly hit the computational limits or even exceed the capacities of present-day supercomputers [89, 96].

Over the past decades, a variety of techniques has been used in literature to tackle related optimization problems in computational materials science; among those are: Monte Carlo or molecular dynamics-based techniques such as simulated annealing [97, 98], basin-hopping [99–101], minima hopping [102, 103], and heuristic approaches based on the ideas of natural evolution such as genetic algorithms [63, 104–115].

In recent years, very promising numerical techniques based on the ideas of artificial intelligence and machine learning have emerged to tackle material science problems (or optimization problems in general) [116–119] ranging from statistical learning [54, 89–91, 120–122] over reinforcement learning [123–126] to generative models [127–132].

Here we utilize, extend and employ specialized numerical tools based on simulated annealing [133, 134], replica exchange Monte Carlo [135–138] and memetic evolutionary algorithms [62, 139, 140] to study the ground state self-assembly strategies of the two different physical systems discussed in Sections 3.1 and 3.2. All of these methods are, in principle, capable of determining global ground state configurations in a rugged energy landscape. The practical numerical bottleneck when employing a particular method, however, is always the corresponding computational cost related to a specific problem [141–143].

This thesis is organized as follows: in Chapter 2 we present the methods we use in this work. In Chapter 3 we investigate different physical systems which are studied in the course of this thesis, namely (i) the possible emergence of quasicrystalline order in the asymmetric Wigner bilayer system in Section 3.1 and (ii) the computational prediction of supramolecular ordering of complex molecules under electrochemical conditions in Section 3.2. Eventually, in Chapter 4 we conclude our studies and discuss possible future directions of investigations related to the topics presented in this thesis. Supplementary theory and information can be found in Appendix A.

2 Methods

In this Chapter, we present and discuss methods which we use in this thesis to study the confinement-driven self-assembly of different physical systems and the related emerging structural properties.

We first introduce Ewald summation techniques in order to faithfully handle numerical calculations of the potential energy of a system featuring long-ranged Coulomb interactions in quasi-two and in three dimensions in Section 2.1. We then continue by discussing methods and algorithms used and extended in this thesis to predict global ground state structures – with respect to the internal energy of a system at vanishing temperature – and finite temperature configurations of atomic or molecular lattices in Section 2.2. In Section 2.3 we introduce the concept and explicit implementations of order parameters that allow us to quantify and classify ordered structures. Finally, we present unsupervised clustering techniques in Section 2.4 which we employ in an effort to identify trends in the structure formation processes in the large amount of structural data that we produced through computer experiments of the physical systems investigated in the scope of this thesis.

2.1 Long-Ranged Coulomb Interaction in Quasi-Two and Three Dimensions

In this thesis, we are mostly concerned with numerical studies of systems featuring long-ranged interactions at interfaces of solid surfaces. We consider macroscopically large systems composed of particles such as point-charges, atoms or molecules which interact with each other and with a surface either solely via long-ranged Coulomb interactions in Section 3.1 or via a combination of long-ranged Coulomb and short-ranged Lennard-Jones type of interactions in Section 3.2.

Numerically, we can only treat a computationally manageable number of particles within a simulation box of finite volume. The presence of a macroscopically large bulk system can then be modeled by treating the simulation box as the primitive cell – or, equivalently, unit cell – of an infinite periodic lattice of identical cells [144], as schematically illustrated in Fig. 2.1(a). In that way, the particles within the unit cell not only interact with other particles within the unit cell but also with the periodic images of all particles across the boundaries of the simulation box. To achieve this, particles in the unit cell are usually subjected to periodic boundary conditions with respect to the spatial extent of the simulation box [92, 144].

Interfaces of solid surfaces are intrinsically three-dimensional (3D) systems which exhibit anisotropies in their spatial extensions [92]: the spatial extent of such systems parallel to the surface are considered macroscopically large while the extent into the direction perpendicular to the surface usually is of the same or of a similar order of magnitude as the size of the involved particles (*i.e.*, of microscopic rather than of macroscopic scale [92]). Such systems differ in their physical properties strongly from bulk systems and the numerical treatment of such quasi–two-dimensional (quasi-2D) slab systems is more complicated as compared to 3D bulk systems [92]. To model interfaces, often quasi-2D lattice geometries are used where, again, a finite-sized unit cell can be defined. However, in contrast to 3D bulk systems the unit cell is now only periodically repeated in a two-dimensional lattice of infinite extent in the directions parallel to the surface (which we consider, without the loss of generality, as horizontal) and the height of the unit cell (perpendicular to the surface) represents the finite vertical extent of the interface, see Fig. 2.1(b). Thus, only the in-plane directions of the unit cell are assigned with periodic boundary conditions, while the perpendicular direction is considered to be non-periodic.

If, in addition, the interactions of particles of the system are long-ranged, the situation is even more complicated. Usually, in systems where long-ranged interactions are involved special numerical tools such as Ewald summation techniques [94, 145] (which we heavily use in this thesis) are required to accurately evaluate the internal energy of a system [92]. Conceptually, Ewald summation techniques explicitly rely on the mathematical description of an infinitely large system of particles via periodic images of the particles in a lattice [92]. The problem of evaluating the internal energy of such an infinitely large, periodic system can then be split into two separate, numerically manageable energy-summations in real space and in reciprocal space (which is related to the real space lattice via Fourier transformation, see below). In quasi-2D systems not all directions are periodic but rather anisotropic in their spatial extent, which makes

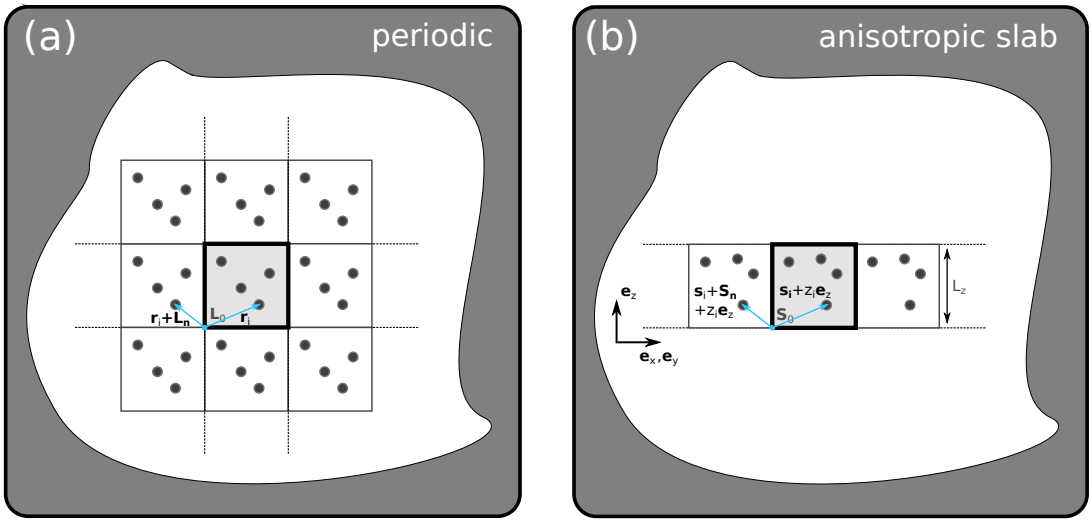


Figure 2.1: (a) Schematic representation of a periodic system of particles of infinite extent, being subject to macroscopic boundary conditions (gray surrounding medium). A simulation box, or unit cell (central, light gray box which is framed by thick black borders indicating the boundaries of the unit cell) is periodically repeated in all spatial directions and images of particle positions within the unit cell, $\mathbf{r}_i = (x_i, y_i, z_i)$, are addressed by $\mathbf{r}_i + \mathbf{L}_\mathbf{n}$. $\mathbf{L}_\mathbf{n} = (n_x L_x, n_y L_y, n_z L_z)$ represents the origin of periodic images of the unit cell which we uniquely label (here in 3D) by three integers $\mathbf{n} = (n_x, n_y, n_z)$ such that $\mathbf{L_0} = (0, 0, 0)$ represents the origin; L_x, L_y and L_z are the lattice constants of the simulation box (here considered as orthogonal box) along the x, y and z directions. (b) Schematic representation of a quasi-2D system labeled "anisotropic slab" of infinite extent in the periodic directions, which we chose without the loss of generality as $\mathbf{e}_x$ and $\mathbf{e}_y$, and of finite extent in the $\mathbf{e}_z$ direction. Images of particle positions of the unit cell of the quasi-2D system are now addressed by $\mathbf{s}_i + \mathbf{S}_\mathbf{n} + z_i\mathbf{e}_z$, where $\mathbf{s}_i = (x_i, y_i, 0)$ describes the periodic part of the position vector and z_i the non-periodic one. $\mathbf{S}_\mathbf{n} = (n_x S_x, n_y S_y, 0)$ represents the origin of the periodic images of the unit cell in the xy-plane which we uniquely label by two integers $\mathbf{n} = (n_x, n_y)$ such that $\mathbf{S_0} = (0, 0, 0)$ represents the origin; S_x and S_y are the lattice constants of the simulation box (here considered as orthogonal box) along the x and y directions and the unit cell height, L_z, along $\mathbf{e}_z$ is referred to as the slab-width. Note that the schematic visualization of the spatial extent of the surrounding medium and of the periodic images of the simulation box are not drawn to scale which, in fact, differ by many orders of magnitude.

the mathematical description of the internal energy of such systems more involved as compared to 3D bulk systems (also see below).

Furthermore, a precise numerical evaluation of the internal energy of long-ranged quasi-2D system via quasi-2D Ewald summation techniques [92] can computationally be very expensive. It is hence of major importance to exploit special properties or symmetries of a given system in the mathematical description of the internal energy such as, for instance, the discrete vertical locations of particles in bilayer systems [62–64, 146, 147] (see also Section 3.1). Otherwise, energy calculations of long-ranged interactions in quasi-2D systems often render numerical studies essentially infeasible due to unacceptably long computation times and often so-called slab-correction schemes [148–151] of 3D Ewald summation techniques are employed.

Section 2.1 is organized as follows: in Subsection 2.1.1 we will define the term "long-ranged" interaction and continue in Subsection 2.1.2 with a brief introduction to Ewald

summation techniques in three dimensions. In Subsection 2.1.3 we introduce Ewald summation techniques for quasi-2D systems and related slab-correction schemes of 3D Ewald summation techniques. We conclude this Section in Subsection 2.1.4 by discussing convergence criteria for accurate evaluations of the internal energy of long-ranged systems using Ewald summation techniques.

2.1.1 What is a Long-Ranged Potential?

In computational physics calculating the internal energy of a system needs to be done as accurate as possible. However, in practice the numerical accuracy with which the internal energy should (or can) be evaluated needs to be carefully put into relation with the computational cost and difficulties that are related to implementing a method [62]. The accuracy of evaluating the internal energy of a system may also strongly depend on the type of interaction of its constituents. Some interactions decay rather quickly with increasing distance, such as a square-well potential or Lennard-Jones type of interactions. For these kinds of interactions the contribution to the internal energy of particles at great distances are negligible compared to those between neighboring particles [62]. If a large enough cutoff radius, r_{cut}, is chosen, *i.e.*, the distance up to which spatially separated, interacting particles are considered to contribute to the internal energy of a system, the internal energy can be evaluated rather accurately by direct summation. However, for other types of interactions, such as Coulomb or dipole potentials which only slowly decay with distance r as $\propto 1/r$ or $\propto 1/r^3$, respectively, the contributions to the internal energy arising from interactions at large distances cannot be neglected.

To demonstrate this we closely follow Ref. [62] and consider a system of N particles, whose number particle density, $\rho(\mathbf{r})$, is given by

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i), \tag{2.1}$$

i.e., by a sum of delta peaks where $\mathbf{r}_i$ specifies the position of particle i. We assume that different particles, i and j, interact via a pair potential $V(\mathbf{r}_i - \mathbf{r}_j)$. The total internal energy of a macroscopic sample of volume V of this system is given by

$$U = \frac{1}{2} \iint_{\mathrm{V}}^{*} d\mathbf{r}\, d\mathbf{r}'\, \rho(\mathbf{r}) V(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}') \tag{2.2}$$

$$= \frac{1}{2} \iint_{\mathrm{V}}^{*} d\mathbf{r}\, d\mathbf{r}' \left[\sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \right] V(\mathbf{r} - \mathbf{r}') \left[\sum_{j=1}^{N} \delta(\mathbf{r}' - \mathbf{r}_j) \right] \tag{2.3}$$

$$= \frac{1}{2} \iint_{\mathrm{V}} d\mathbf{r}\, d\mathbf{r}' \sum_{i=1}^{N} \sum_{j \neq i}^{N} \delta(\mathbf{r} - \mathbf{r}_i) V(\mathbf{r} - \mathbf{r}') \delta(\mathbf{r}' - \mathbf{r}_j) \tag{2.4}$$

where "$\iint^{*}$" in the first two lines indicate that we do not consider self-interaction terms of the form $\delta(\mathbf{r} - \mathbf{r}_i) V(\mathrm{r} - \mathrm{r}') \delta(\mathbf{r}' - \mathbf{r}_i)$ in the double integration.

Transforming into the frame of reference of particle i, *i.e.*, $\mathbf{r}_i = \mathbf{0}$, allows us to extract its respective contribution, U_i, to the total internal energy, $U = \sum_{i=1}^{N} U_i$, by

$$U_i = 2 \times \frac{1}{2} \iint_{\mathrm{V}} d\mathbf{r}\, d\mathbf{r}' \sum_{j \neq i}^{N} \delta(\mathbf{r} - \mathbf{r}_j) V(\mathbf{r} - \mathbf{r}') \delta(\mathbf{r}') \tag{2.5}$$

$$= \int_{\mathrm{V}} d\mathbf{r} \sum_{j \neq i}^{N} \delta(\mathbf{r} - \mathbf{r}_j) V(\mathbf{r}), \tag{2.6}$$

where we performed the integration over $\mathbf{r}'$ in the last step.

We here consider systems of macroscopic scale from a microscopic point of view. Although the sample volume may be macroscopically small the volume, V, and the number of particles, N, involved in the integration of the internal energy in Eq. (2.4) are both large, *i.e.*, N is usually of the order of magnitude of the Avogadro constant $N_{\mathrm{A}} \approx 10^{23}\mathrm{mol}^{-1}$ [62]. It is therefore conceptually easier, from a numerical point of view, to take the thermodynamic limit, $N \to \infty$, and extend the considered sample volume to an infinite extent, *i.e.*, we approximate the integration over a macroscopically large, but finite sized volume by an integration over an infinite extent [92]. As discussed above such macroscopic systems can be considered using the concept of numerically feasible simulation boxes with periodic boundary conditions corresponding to a computationally manageable system of infinite extent.

Assuming an isotropic potential, $V(\mathbf{r}) = V(r)$, in a three-dimensional volume, we can rewrite Eq. (2.6) as

$$U_i = \int_0^{\infty} dr\, r^2 V(r) \int_0^{\pi} d\theta \int_0^{2\pi} d\phi\, \rho(\mathbf{r}) \tag{2.7}$$

$$= 4\pi \int_0^{\infty} dr\, r^2 \rho(r) V(r), \tag{2.8}$$

having introduce the radial number particle density[1], $\rho(r) = \sum_{j=1}^{N} \frac{1}{4\pi r_j^2} \delta(r_j)$, in the last step.

Eq. (2.8) describes an integral which we cannot evaluate directly for an arbitrary macroscopic system with unknown density of particles and an unknown pair potential: it is not guaranteed that the infinite integral converges. In practice (and in case the integrand is well-behaved), the integration in Eq. (2.8) is approximated by defining a cutoff radius, r_{cut}, up to which the integration is performed:

$$U \approx U_{\mathrm{cut}} = 4\pi \int_0^{r_{\mathrm{cut}}} dr r^2 \rho(r) V(r). \tag{2.9}$$

[1] Since $N = \int_{\infty} d\mathbf{r}\, \rho(\mathbf{r}) = 4\pi \int_0^{\infty} dr\, r^2 \rho(r) = \int_0^{\infty} dr\, \rho(r)$ we find that $\rho(r) = \sum_{j=1}^{N} \frac{1}{4\pi r_j^2} \delta(r_j)$.

This causes a relative error, δ_{cut}, in the numerical evaluation of the internal energy, which is defined by

$$\delta_{\text{cut}} = \frac{U_{\text{cut}}}{U}. \tag{2.10}$$

For potentials which decay sufficiently fast with distance we can define a cutoff distance, r_{cut}, in the integration given by Eq. (2.9) such that the relative error made in the numerical evaluation of the internal energy U_{cut} has an upper bound of δ_{cut} with respect to U [62, 145]. Notably, the particular value of the cutoff radius – in order to achieve a desired accuracy in the evaluation of the internal energy – can strongly depend on the actual functional form of $V(r)$ but can always be defined for sufficiently fast decaying, short-ranged potentials [145].

Long-ranged potentials are often defined as potentials which decay no faster with distance than $V(r) \propto r^{-d}$ for $r \to \infty$, where d is the dimensionality of the system [145]. Coulomb potentials with $V(r) \propto r^{-1}$ and dipole interactions with $V(r) \propto r^{-3}$ in three dimensions fall into this category and are thus considered long-ranged. For such types of interactions the involved integrals in Eq. (2.8), when evaluating the internal energy of a system, do not converge [62, 92, 145]. Moreover, using a cutoff radius to numerically evaluate the internal energy via Eq. (2.9) leads to an inestimable relative error with respect to the "correct" internal energy of the system [62].

The convergence of the internal energy given by Eq. (2.6) also depends on the macroscopic boundary conditions for a specific system such as the shape of the macroscopic sample that separates the system from the surrounding medium [92, 152–157] (cf. Fig. 2.1). For electrodynamic interactions also the electrostatic or magnetic properties of the surrounding medium affect the calculation of the internal energy and, in general, a proper treatment of the system with the Maxwell equations [158] is required. If the surrounding medium is vacuum, *i.e.*, with a dielectric constant of $\epsilon_r = 1$, a net polarization would be induced on the surface of a macroscopic sample of point-charges (*i.e.*, on the macroscopic boundaries) which corresponds to the total dipole moment of the system; this induced surface polarization contributes to the internal energy of the system and needs to be considered in the energy calculations [92, 152]. In this thesis, we usually assume metallic (tinfoil) boundary conditions, *i.e.*, we chose a dielectric constant of $\epsilon_r = \infty$ for the medium surrounding the macroscopic system such that the surface contributions to the internal energy stemming from surface polarization effects can be neglected (see for instance Refs. [92, 152] for details).

Also for long-ranged interactions we usually approximate the internal energy of a macroscopically large, yet finite system with a system of infinite extent while still subjecting this infinitely large system to the macroscopic boundary conditions, see Fig. 2.1. Continuing with our example of Coulomb interaction, we consider an infinitely large periodic lattice of images of a finite sized unit cell (*i.e.*, a simulation box) which contains N point-charges q_i. To warrant that the internal energy per particle, U_i, of such an infinite, periodic lattice of identical cells is finite the laws of electrodynamic suggest that the system needs to be electroneutral as a whole [62, 158]. Electroneutrality can either be achieved by considering positive and negative charges in the simulation

box that directly sum up to zero, $\sum_{i=1}^{N} q_i = 0$, or by adding a neutralizing background charge density, $\rho_0(\mathbf{r})$, to the system, satisfying $\sum_{i=1}^{N} q_i + \int_{\mathbf{r}} d\mathbf{r}\, \rho_0(\mathbf{r}) = 0$ [62, 92, 158].

Thus, U_i defined by Eq. (2.6) is conditionally convergent and can be evaluated under certain circumstances [92]. However, the internal electrostatic energy of an infinitely large periodic system of point-charges can still not simply by evaluated by truncating a direct lattice sum of the corresponding pairwise Coulomb energies without making an inestimable error with respect to the correct internal energy of the system.

2.1.2 Ewald Summation of a System of Point-Charges in Three Dimensions

Evaluating the internal energy via Eq. (2.6) for Coulomb interactions is a task far from being trivial and has been addressed by several mathematical models so far. Some of these, such as Ewald- [94] and Lekner sums [159], rearrange the lattice summation by splitting it into manageable, fast converging expressions in real and reciprocal space. Numerical adaptations of Ewald summation techniques, such as the particle-mesh Ewald method [160, 161] or particle-particle-particle-mesh Ewald method [162–165] rely on interpolating particle position onto a discrete grid in real space and calculate the internal energy of the modified, discretized system with the help of numerical fast Fourier transformation (FFT) techniques [92]. Other, related approaches such as Nijboer-de Wette and fast multipole approaches are discussed in Ref. [92].

In this thesis, we exclusively use the method of Ewald summation [94] for quasi-2D bilayer systems [92] and slab-correction techniques of 3D Ewald summations [149, 151] to evaluate long-ranged Coulomb interactions in quasi-2D slab geometries in a numerically accurate yet computationally affordable manner. The basic concept of Ewald summation is to split the slowly converging infinite real space summation into two more rapidly converging parts: one performed in real space, the other performed in reciprocal space. This step is possible for systems featuring electrostatic interactions due to the fact that an electrostatic field potential needs to fulfill the Poisson equation [158].

Closely following Refs. [92, 166] and considering a three-dimensional bulk system with electrostatic interactions and periodic boundary conditions this approach is demonstrated below. The corresponding electrostatic field potential[2] (or in short, the field potential), $\phi(\mathbf{r})$, at the position $\mathbf{r}$ within the simulation box satisfies the Poisson equation [158], given by

$$\nabla^2 \phi(\mathbf{r}) = -4\pi \sum_{\mathbf{L_n}} \rho(\mathbf{r} + \mathbf{L_n}), \tag{2.11}$$

[2] The field potential, $\phi(\mathbf{r}')$, is defined as the electrostatic energy corresponding to a unit charge, $q = 1$, located at the position $\mathbf{r}'$ within a volume V which exhibits a charge distribution $\rho(\mathbf{r})$. Thus, the (internal) potential energy corresponding to a point-charge q_i located at $\mathbf{r}_i$ is related to the field potential via $U_i = q_i\, \phi(\mathbf{r}_i)$ (self-interaction of particle i needs to be avoided in $\phi(\mathbf{r}_i)$) [158].

where $\rho(\mathbf{r})$ is the charge distribution within the simulation box (or equivalently, the unit cell) and $\mathbf{L_n}$ is a symbolic notation for addressing the periodic images of the unit cell within the lattice geometry[3].

We define the charge density for a system of N point-charges q_i located in the simulation box (*i.e.*, in the unit cell of a periodic lattice of cells), similar to Eq. (2.1), as

$$\rho(\mathbf{r}) = \rho_0 + \sum_{i=1}^{N} q_i \, \delta(\mathbf{r} - \mathbf{r}_i); \tag{2.12}$$

ρ_0 is a constant, neutralizing background charge density which assures electroneutrality of the unit cell, *i.e.*, $(V\rho_0 + \sum_{i=1}^{N} q_i) = 0$, V being the unit cell volume. Henceforward, we assume electroneutrality of the form $\sum_{i=1}^{N} q_i = 0$, such that $\rho_0 = 0$ [92].

In order to derive the internal energy of a periodic lattice of point-charges we start with the following considerations: for a single point-charge q_i the electrostatic field potential, $\phi_i(\mathbf{r})$, is defined at an arbitrary position $\mathbf{r}$ in real space via [158]

$$\phi_i(\mathbf{r}) = \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}. \tag{2.13}$$

The total electrostatic field potential, $\phi(\mathbf{r})$, of a periodic lattice of N point-charges q_j within the simulation box is given by

$$\phi(\mathbf{r}) = \sum_{j=1}^{N} \sum_{\mathbf{L_n}} \phi_j(\mathbf{r} + \mathbf{L_n}) = \sum_{j=1}^{N} \sum_{\mathbf{L_n}} \frac{q_j}{|\mathbf{r} + \mathbf{L_n} - \mathbf{r}_j|}, \tag{2.14}$$

with $\phi_j(\mathbf{r} + \mathbf{L_n})$ being defined by Eq. (2.13). The corresponding electrostatic field potential $\phi_{i*}(\mathbf{r})$ of such a system – excluding particle i – can thus be written as

$$\phi_{i*}(\mathbf{r}) = \phi(\mathbf{r}) - \phi_i(\mathbf{r}) = \sum_{j=1}^{N} \sum_{\mathbf{L_n}}^{*} \frac{q_j}{|\mathbf{r} + \mathbf{L_n} - \mathbf{r}_j|}, \tag{2.15}$$

where the notations "$i*$" and "$\sum^{*}$" (used in Eq. (2.15) and henceforward) indicate, that we do not consider contributions $i = j$ for $\mathbf{L_0} = (0, 0, 0)$, *i.e.*, we omit the self-interactions of point-charges within the unit cell. With $\phi_{i*}(\mathbf{r})$ given by Eq. (2.15) we can write the internal energy, U, of the unit cell of the system as

$$U = \frac{1}{2} \sum_{i=1}^{N} q_i \, \phi_{i*}(\mathbf{r}_i). \tag{2.16}$$

The electrostatic field potential as given by Eq. (2.15) represents an infinite, diverging sum considering the infinite extent of the lattice and the long-ranged nature of the

[3] In three dimensions a triclinic simulation box with periodic boundary conditions, or equivalently the unit cell of a lattice, is spanned by the lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The origin of periodic images of the unit cell can thus be addressed by a vector of integers, $\mathbf{n} = (n_1, n_2, n_3)$, via $\mathbf{L_n} = n_1 \, \mathbf{a}_1 + n_2 \, \mathbf{a}_2 + n_3 \, \mathbf{a}_3$. Periodic images of particle positions can then be written as $\mathbf{r} + \mathbf{L_n}$; the original simulation box, with a volume $V = \mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3$, is addressed by $\mathbf{L}_0 = (0, 0, 0)$.

Coulomb interaction (cf. Subsection 2.1.1). Thus, we cannot reliably evaluate the internal energy defined by Eq. (2.16) of the system by employing Eq. (2.15).

The Ewald method for evaluating the internal energy of such a system can be obtained by formally modifying the charge density $\rho(\mathbf{r})$ given by Eq. (2.12) without changing $\rho(\mathbf{r})$ physically: through adding and subtracting smoothly varying charge densities $\pm q_i\, g(\mathbf{r};\mathbf{r}_i,\alpha)$, which are centered at the locations $\mathbf{r}_i$ of the point-charges and which decay over a distance of $\approx 1/\alpha$, we redefine $\rho(\mathbf{r})$ as

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} q_i\, [\delta(\mathbf{r}-\mathbf{r}_i) - g(\mathbf{r};\mathbf{r}_i,\alpha)] + \sum_{i=1}^{N} q_i\, g(\mathbf{r};\mathbf{r}_i,\alpha). \tag{2.17}$$

In that way, each point-charge i is effectively screened by an oppositely charged, smoothly decaying charge density in real space, $\rho_i^{(\mathrm{S})}(\mathbf{r}) = q_i\, [\delta(\mathbf{r}-\mathbf{r}_i) - g(\mathbf{r};\mathbf{r}_i,\alpha)]$, thereby transforming the corresponding Coulomb interaction into a screened short-ranged interaction. In turn, an artificial lattice of charge densities $\rho_i^{(\mathrm{L})}(\mathbf{r}) = q_i\, g(\mathbf{r};\mathbf{r}_i,\alpha)$ of the opposite sign is introduced, which can be treated independently from the now short-ranged screened Coulomb interaction.

Due to linearity, we can split the electrostatic field potential defined in Eq. (2.13) into a corresponding short-ranged part, $\phi_i^{(\mathrm{S})}(\mathbf{r})$, and a long-ranged part, $\phi_i^{(\mathrm{L})}(\mathbf{r}_i)$, given by

$$\phi_i^{(\mathrm{S})}(\mathbf{r}) = \int d\mathbf{r}'\, \frac{\rho_i^{(\mathrm{S})}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} = q_i \int d\mathbf{r}'\, \frac{\delta(\mathbf{r}'-\mathbf{r}_i) - g(\mathbf{r}';\mathbf{r}_i,\alpha)}{|\mathbf{r}-\mathbf{r}'|}, \tag{2.18}$$

$$\phi_i^{(\mathrm{L})}(\mathbf{r}) = \int d\mathbf{r}'\, \frac{\rho_i^{(\mathrm{L})}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} = q_i \int d\mathbf{r}'\, \frac{g(\mathbf{r}';\mathbf{r}_i,\alpha)}{|\mathbf{r}-\mathbf{r}'|}, \tag{2.19}$$

respectively, which both independently satisfy the Poisson equation given by Eq. (2.11). Analogously, we can split $\phi_{i^*}(\mathbf{r}_i) = \phi_{i^*}^{(\mathrm{S})}(\mathbf{r}_i) + \phi_{i^*}^{(\mathrm{L})}(\mathbf{r}_i)$ defined by Eq. (2.15). Consequently, also the corresponding contribution to the internal energy $U_i = U_i^{(\mathrm{S})} + U_i^{(\mathrm{L})}$ can be separated into a short-ranged term $U_i^{(\mathrm{S})} = q_i\, \phi_{i^*}^{(\mathrm{S})}(\mathbf{r}_i)$ and a long-ranged term $U_i^{(\mathrm{L})} = q_i\, \phi_{i^*}^{(\mathrm{L})}(\mathbf{r}_i)$. It is now convenient to reintroduce the self-interaction term for the long-ranged part of the electrostatic field potentials, $\phi_{i^*}^{(\mathrm{L})}(\mathbf{r}_i) = \phi^{(\mathrm{L})}(\mathbf{r}_i) - \phi_i^{(\mathrm{L})}(\mathbf{r}_i)$, such that the internal energy $U \equiv U_{\mathrm{Ewald}}^{(\mathrm{3D})}$ of the system is given by

$$U_{\mathrm{Ewald}}^{(\mathrm{3D})} = \frac{1}{2}\sum_{i=1}^{N} q_i\, \phi_{i^*}^{(\mathrm{S})}(\mathbf{r}_i) + \frac{1}{2}\sum_{i=1}^{N} q_i\, \phi^{(\mathrm{L})}(\mathbf{r}_i) - \frac{1}{2}\sum_{i=1}^{N} q_i\, \phi_i^{(\mathrm{L})}(\mathbf{r}_i) \tag{2.20}$$

$$= U^{(\mathrm{S})} + U^{(\mathrm{L})} + U^{(\mathrm{Self})}, \tag{2.21}$$

where we defined $U^{(\mathrm{S})} = \frac{1}{2}\sum_{i=1}^{N} q_i\, \phi_{i^*}^{(\mathrm{S})}(\mathbf{r}_i)$, $U^{(\mathrm{L})} = \frac{1}{2}\sum_{i=1}^{N} q_i\, \phi^{(\mathrm{L})}(\mathbf{r}_i)$ and $U^{(\mathrm{Self})} = \frac{1}{2}\sum_{i=1}^{N} q_i\, \phi_i^{(\mathrm{L})}(\mathbf{r}_i)$ in the last step.

Henceforward, we explicitly use Gaussian charge densities $g(\mathbf{r};\mathbf{r}_i,\alpha)$ of the form

$$g(\mathbf{r};\mathbf{r}_i,\alpha) = \left(\frac{\alpha^2}{\pi}\right)^{3/2} \exp\left(-\alpha^2(\mathbf{r}-\mathbf{r}_i)^2\right), \tag{2.22}$$

with a standard deviation of $\sqrt{2}/\alpha$ in real space (although other functional forms would also be possible in what follows) [92].

As is well-known from literature [92], the electrostatic field potential of a Gaussian charge density $g(\mathbf{r};\mathbf{r}_i,\alpha)$ can be solved using the error function, defined as $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z dt\, \exp(-t^2)$, such that $\phi_i^{(\mathrm{S})}(\mathbf{r})$ and $\phi_i^{(\mathrm{L})}(\mathbf{r})$ given by Eqs. (2.18) and (2.19) yield

$$\phi_i^{(\mathrm{S})}(\mathbf{r}) = \frac{q_i}{|\mathbf{r}-\mathbf{r}_i|}\mathrm{erfc}(\alpha|\mathbf{r}-\mathbf{r}_i|), \tag{2.23}$$

$$\phi_i^{(\mathrm{L})}(\mathbf{r}) = \frac{q_i}{|\mathbf{r}-\mathbf{r}_i|}\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{r}_i|), \tag{2.24}$$

respectively, where we also introduced the complementary error function $\mathrm{erfc}(z) = 1-\mathrm{erf}(z)$. In comparison to the field potential of a "bare" Coulomb point-charge which is long-ranged and singular, the expression $\phi_i^{(\mathrm{S})}(\mathbf{r})$ describes a short-ranged singular- and $\phi_i^{(\mathrm{L})}(\mathbf{r})$ a long-ranged non-singular electrostatic field potential. Moreover, we can identify the short-ranged electrostatic field potential $\phi_{i^*}^{(\mathrm{S})}(\mathbf{r}_i)$ which all other charges q_j in the system generate at the position of charge q_i as

$$\phi_{i^*}^{(\mathrm{S})}(\mathbf{r}) = \sum_{j=1}^{N}{\sum_{\mathbf{L_n}}}^{*}\frac{q_j}{|\mathbf{r}+\mathbf{L_n}-\mathbf{r}_j|}\mathrm{erfc}(\alpha|\mathbf{r}+\mathbf{L_n}-\mathbf{r}_j|). \tag{2.25}$$

The corresponding short-ranged energy $U^{(\mathrm{S})}$ defined in Eq. (2.21) then becomes

$$U^{(\mathrm{S})} = \frac{1}{2}\sum_{i,j=1}^{N}{\sum_{\mathbf{L_n}}}^{*}\frac{q_i\,q_j}{|\mathbf{r}_i+\mathbf{L_n}-\mathbf{r}_j|}\mathrm{erfc}(\alpha|\mathbf{r}_i+\mathbf{L_n}-\mathbf{r}_j|). \tag{2.26}$$

Notably, $U^{(\mathrm{S})}$ is similar to the usual Coulomb energy but every term in the sum is strongly suppressed at distances $r \gtrsim \sqrt{2}/\alpha$ via the complementary error function term. Consequently, we can evaluate $U^{(\mathrm{S})}$ in a numerically faithful manner by truncating the sum at large enough distances.

However, we have yet to define the long-ranged part, $U^{(\mathrm{L})}$, and the self-energy term, $U^{(\mathrm{Self})}$, of the total internal energy $U_{\mathrm{Ewald}}^{(\mathrm{3D})} = U^{(\mathrm{S})}+U^{(\mathrm{L})}+U^{(\mathrm{Self})}$ defined in Eq. (2.21).

We start with the self-energy term: with the identity $\lim_{z\to 0}\mathrm{erf}(z) = 2z/\sqrt{\pi}$ we find that the long-ranged electrostatic field potential $\phi_i^{(\mathrm{L})}(\mathbf{r})$ given by Eq. (2.24) takes on the value $\phi_i^{(\mathrm{L})}(\mathbf{r}_i) = q_i\,2\alpha/\sqrt{\pi}$ as $\mathbf{r}\to\mathbf{r}_i$, and thus

$$U^{(\mathrm{Self})} = \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{N}q_i^2. \tag{2.27}$$

In order to derive a computationally manageable expression for the long-ranged part $U^{(\mathrm{L})}$ of the internal energy (cf. Eq. (2.21)) we exploit the fact that we are dealing with a periodic charge density $\rho^{(\mathrm{L})}(\mathbf{r})$ given by

$$\rho^{(\mathrm{L})}(\mathbf{r}) = \sum_{\mathbf{L_n}}\sum_{j=1}^{N}\rho_j^{(\mathrm{L})}(\mathbf{r}+\mathbf{L_n}). \tag{2.28}$$

Consequently, also the corresponding electrostatic field potential $\phi^{(\mathrm{L})}(\mathbf{r})$ is a periodic function and can be defined as

$$\phi^{(\mathrm{L})}(\mathbf{r}) = \sum_{\mathbf{L_n}} \sum_{j=1}^{N} \phi_j^{(\mathrm{L})}(\mathbf{r} + \mathbf{L_n}), \tag{2.29}$$

using $\phi_j^{(\mathrm{L})}(\mathbf{r})$ given by Eq. (2.24).

Hence, it is convenient to transform both $\rho^{(\mathrm{L})}(\mathbf{r})$ and $\phi^{(\mathrm{L})}(\mathbf{r})$ into the reciprocal space[4]. We can expand the periodic functions $\rho^{(\mathrm{L})}(\mathbf{r})$ and $\phi^{(\mathrm{L})}(\mathbf{r})$ into a Fourier series as

$$\rho^{(\mathrm{L})}(\mathbf{r}) = \sum_{\mathbf{k}} \rho_{\mathbf{k}}^{(\mathrm{L})} \exp(\imath \mathbf{k} \cdot \mathbf{r}), \tag{2.30}$$

$$\phi^{(\mathrm{L})}(\mathbf{r}) = \sum_{\mathbf{k}} \phi_{\mathbf{k}}^{(\mathrm{L})} \exp(\imath \mathbf{k} \cdot \mathbf{r}), \tag{2.31}$$

respectively, where the Fourier coefficients $\rho_{\mathbf{k}}^{(\mathrm{L})}$ and $\phi_{\mathbf{k}}^{(\mathrm{L})}$ are associated with the wave vectors, $\mathbf{k}$, representing the vertices of the reciprocal lattice[5]. The Fourier coefficients $\rho_{\mathbf{k}}^{(\mathrm{L})}$ and $\phi_{\mathbf{k}}^{(\mathrm{L})}$ of the periodic charge density $\rho^{(\mathrm{L})}(\mathbf{r})$ and of the long-ranged electrostatic field potential $\phi^{(\mathrm{L})}(\mathbf{r})$ can then be determined via

$$\rho_{\mathbf{k}}^{(\mathrm{L})} = \frac{1}{V} \int_V d\mathbf{r}\, \rho^{(\mathrm{L})}(\mathbf{r}) \exp(-\imath \mathbf{k} \cdot \mathbf{r}), \tag{2.32}$$

$$\phi_{\mathbf{k}}^{(\mathrm{L})} = \frac{1}{V} \int_V d\mathbf{r}\, \phi^{(\mathrm{L})}(\mathbf{r}) \exp(-\imath \mathbf{k} \cdot \mathbf{r}), \tag{2.33}$$

respectively, where the integration is performed over the volume V of the simulation box. With the Gaussian charge densities $\rho_i^{(\mathrm{L})}(\mathbf{r}) = g(\mathbf{r}; \mathbf{r}_i, \alpha)$ given by Eq. (2.22) we find that $\rho_{\mathbf{k}}^{(\mathrm{L})}$ reads [166]

$$\rho_{\mathbf{k}}^{(\mathrm{L})} = \frac{1}{V} \sum_{i=1}^{N} \exp(k^2/4\alpha^2) \exp(-\imath \mathbf{k} \cdot \mathbf{r}_i), \tag{2.34}$$

with $k = |\mathbf{k}|$. By following the Poisson equation in reciprocal space, defined as

$$k^2 \phi_{\mathbf{k}}^{(\mathrm{L})} = 4\pi \rho_{\mathbf{k}}^{(\mathrm{L})}, \tag{2.35}$$

we find with $\rho_{\mathbf{k}}^{(\mathrm{L})}$ given by Eq. (2.34) the following expression for the Fourier coefficients $\phi_{\mathbf{k}}^{(\mathrm{L})}$ of the long-ranged field potential:

$$\phi_{\mathbf{k}}^{(\mathrm{L})} = \frac{4\pi}{V} \sum_{j=1}^{N} q_j \frac{\exp(-k^2/4\alpha^2)}{k^2} \exp(-\imath \mathbf{k} \cdot \mathbf{r}_j). \tag{2.36}$$

[4] The reciprocal space, conjugate to the real space lattice, is spanned by the reciprocal lattice vectors $\mathbf{a}_1^*$, $\mathbf{a}_2^*$ and $\mathbf{a}_3^*$ which are defined by the relation $\mathbf{a}_k^* \mathbf{a}_i = 2\pi\delta_{ki}$ with respect to the real space lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ [92].

[5] The wave vectors $\mathbf{k} = h\,\mathbf{a}_1^* + k\,\mathbf{a}_2^* + l\,\mathbf{a}_3^*$ define the vertices of the reciprocal lattice with reciprocal lattice vectors $\mathbf{a}_1^*$, $\mathbf{a}_2^*$ and $\mathbf{a}_3^*$ and with integers h, k and l such that $\exp(\pm\imath \mathbf{k} \cdot \mathbf{L_n}) = 1$ [92].

Substituting Eq. (2.36) into Eq. (2.31) yields the following expression for the long-ranged field potential $\phi^{(\mathrm{L})}(\mathbf{r})$ in real space

$$\phi^{(\mathrm{L})}(\mathbf{r}) = \frac{4\pi}{V} \sum_{\mathbf{k}\neq 0} \sum_{i=1}^{N} q_i \, \frac{\exp\left(-k^2/4\alpha^2\right)}{k^2} \exp\left(\imath \mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_i)\right) \tag{2.37}$$

for $\mathbf{k} \neq 0$; the contributions to Eq. (2.37) with $\mathbf{k} = 0$ include diverging terms (not shown here) which eventually cancel in the evaluation of the internal energy $U^{(\mathrm{3D})}_{\mathrm{Ewald}}$ if the system is electroneutral, *i.e.*, if $(V\,\rho_0 + \sum_{i=1}^{N} q_i) = 0$ [92]. In that case, the total long-ranged contribution $U^{(\mathrm{L})}$ to the internal energy of the system, defined in Eqs. (2.16) and (2.21), can be written as

$$U^{(\mathrm{L})} = \frac{2\pi}{V} \sum_{\mathbf{k}\neq 0} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j \, \frac{\exp\left(-k^2/4\alpha^2\right)}{k^2} \exp\left(\imath \mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)\right) \tag{2.38}$$

$$= \frac{2\pi}{V} \sum_{\mathbf{k}\neq 0} \frac{\exp\left(-k^2/4\alpha^2\right)}{k^2} \left| \sum_{i=1}^{N} q_i \exp\left(\imath \mathbf{k} \cdot \mathbf{r}_i\right) \right|^2 . \tag{2.39}$$

Via Eq. (2.39) we see that also the long-ranged term of the internal energy can be expressed via a quickly converging sum in reciprocal space since each term in the summation over the reciprocal lattice is suppressed for large values of $k \gtrsim \alpha$.

Summarizing, the internal energy $U^{(\mathrm{3D})}_{\mathrm{Ewald}} = U^{(\mathrm{S})} + U^{(\mathrm{L})} + U^{(\mathrm{Self})}$ defined by Eq. (2.21) of an infinitely large three-dimensional periodic lattice of cells with N point-charges in the simulation box can indeed be evaluated using 3D-Ewald summation techniques by combining Eqs. (2.26), (2.27) and (2.39); $U^{(\mathrm{3D})}_{\mathrm{Ewald}}$ is explicitly given by [92]

$$U^{(\mathrm{3D})}_{\mathrm{Ewald}} = \frac{1}{2} \sum_{i,j=1}^{N} {\sum_{\mathbf{L_n}}}^{*} q_i \, q_j \frac{\mathrm{erfc}(\alpha |\mathbf{r}_i + \mathbf{L_n} - \mathbf{r}_j|)}{|\mathbf{r}_i + \mathbf{L_n} - \mathbf{r}_j|}$$
$$+ \frac{2\pi}{V} \sum_{\mathbf{k}\neq 0} \frac{\exp\left(-k^2/4\alpha^2\right)}{k^2} \left| \sum_{i=1}^{N} q_i \exp\left(\imath \mathbf{k} \cdot \mathbf{r}_i\right) \right|^2 - \frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 , \tag{2.40}$$

which is – irrespective of the particular numerical value of the Ewald α parameter – a mathematically exact but numerically manageable representation of the internal energy of the system given by Eq. (2.16). Since Eq. (2.40) still contains an infinite number of terms, the real space and the reciprocal space summations have to be (carefully) truncated for any practical purposes, which will be discussed below. The corresponding time complexity of the three-dimensional Ewald summation is $\mathcal{O}(N^2)$, *i.e.*, the time necessary to evaluate $U^{(\mathrm{3D})}_{\mathrm{Ewald}}$ via Eq. (2.40) scales with the square of the number of charges in the simulation box. The time complexity can be improved further to $\mathcal{O}(N^{3/2})$ by a proper choice of α (see Ref. [92]).

2.1.3 Anisotropic Directions: Ewald Summation in Quasi–Two Dimensions

Ewald summations of quasi-2D systems [92] can be seen as a limiting case of Ewald summations in 3D: one of the spatial directions (here chosen without the loss of generality as $\hat{\mathbf{e}}_z$=(0,0,1)) is non-periodic but is assumed to be finite in its spatial extent.

The unit cell (or the simulation box) of such a quasi-2D system is defined by two lattice vectors $\mathbf{a_1} = (a_{11}, 0, 0)$ and $\mathbf{a_2} = (a_{21}, a_{22}, 0)$ in the xy-plane and by the non-periodic extent into the z-direction, $L_z\hat{\mathbf{e}}_z$, which defines the unit cell area (in the xy-direction) as $A = |\mathbf{a}_1 \times \mathbf{a}_2|$ and the volume as $V = A\,L_z$. The finite extent, L_z, of the unit cell into the z-direction is often referred to as *slab width* and we use the term quasi-2D geometry synonymously to quasi-2D slab- or simply slab geometry. The origin of the periodic images of the unit cell in the xy-plane are described symbolically by the vectors $\mathbf{S_n} = n_1\mathbf{a_1} + n_2\mathbf{a_2}$ which are uniquely labeled by the two integers $\mathbf{n} = (n_1, n_2)$. Also the wave vectors of the reciprocal lattice, $\mathbf{k} = \mathbf{G} + k_z\hat{\mathbf{e}}_z$, are treated separately for the periodic and non-periodic directions. The component of the wave vector along $\hat{\mathbf{e}}_z$ is defined by $k_z = 2\pi n_l/L_z$, while the wave vectors of the reciprocal lattice – associated with the 2D lattice with lattice vectors $\mathbf{a_1}$ and $\mathbf{a_2}$ – is given by $\mathbf{G} = n_h\mathbf{a}_1^* + n_k\mathbf{a}_2^*$ with $\mathbf{a}_k^* \cdot \mathbf{a}_i = \delta_{ki}$, where n_h, n_k and n_l are integers [92].

In the definition of the position vectors of point-charges within the unit cell of such quasi-2D systems, $\mathbf{r}_i = \mathbf{s}_i + z_i\hat{\mathbf{e}}_z$, we treat the periodic part of the vectors (*i.e.*, $\mathbf{s}_i = (x_i, y_i, 0)$ in the xy-plane) and the non-periodic part (*i.e.*, $z_i\hat{\mathbf{e}}_z = (0, 0, z_i)$ in the z-direction) separately. Periodic images of particle positions in quasi-2D geometries can thus be written as $\mathbf{r}_i + \mathbf{S_n} = \mathbf{s}_i + \mathbf{S_n} + z_i\hat{\mathbf{e}}_z$.

The spatial anisotropy of periodic directions in the plane and the non-periodic directions out of the plane greatly complicates the calculation of the internal energy of point-charges in quasi-2D slab geometries via Ewald summation techniques: many steps during the derivation leading to Eq. (2.40) rely on the periodicity of the particle positions in all spatial directions. Nevertheless, for $L_z \to \infty$ an expression for the electrostatic energy in quasi-2D (Q2D) can be given as [92]

$$\begin{aligned} U_{\text{Ewald}}^{(\text{Q2D})} = \frac{1}{2}\sum_{i,j=1}^{N}\sum_{\mathbf{S_n}}^{*} q_i\,q_j\,\frac{\operatorname{erfc}(\alpha|\mathbf{r}_{ij}+\mathbf{S_n}|)}{|\mathbf{r}_{ij}+\mathbf{S_n}|} \\ + \frac{\pi}{2A}\sum_{i,j=1}^{N} q_i\,q_j \sum_{\mathbf{G}\neq 0} F(G,\alpha,z_{ij})\,\exp\left(\imath\mathbf{G}\cdot\mathbf{s}_{ij}\right) \\ - \frac{\pi}{A}\sum_{i,j=1}^{N} q_i\,q_j\left(z_{ij}\operatorname{erf}(\alpha\,z_{ij}) + \frac{\exp\left(-\alpha^2 z_{ij}^2\right)}{\alpha\sqrt{\pi}}\right) - \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2, \end{aligned} \tag{2.41}$$

with $F(G, \alpha; z)$ defined as

$$F(G,\alpha;z) = \frac{1}{G}\left[\exp(Gz)\operatorname{erfc}\left(\frac{G}{2\alpha}+\alpha z\right)+\exp(-Gz)\operatorname{erfc}\left(\frac{G}{2\alpha}-\alpha z\right)\right], \quad (2.42)$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, $\mathbf{s}_{ij} = \mathbf{s}_j - \mathbf{s}_i$, $z_{ij} = z_j - z_i$, A is the surface area of the unit cell and $G = |\mathbf{G}|$; $\operatorname{erf}(z)$ is the Gauss error function, $\operatorname{erfc}(z)$ is the complementary error function and the asterisk in $\sum_{\mathbf{S_n}}{}^*$ indicates, that we do not consider contributions $i = j$ for $\mathbf{S_0} = (0,0,0)$, *i.e.*, we omit self-interaction of point-charges within the unit cell.

Analytically exact energy evaluations of systems of point-charges in quasi-2D slab geometries achieved via Eq. (2.41) are for many practical applications prohibitively slow (as for instance to study the ground state assembly of complex, molecular systems at a solid–liquid interface as presented in Section 3.2). This origins in the second term of Eq. (2.41): the function $F(G,\alpha;z)$ is asymmetric with respect to the periodic and non-periodic directions such that the sums in the second term in Eq. (2.41) cannot be factorized anymore, this was the case for 3D periodic systems (cf. step from Eq. (2.38) to Eq. (2.39)). The associated time complexity of the second term in Eq. (2.41) is of the order $\mathcal{O}(N^2)$ for each wave vector, $\mathbf{G}$, while the equivalent term in the 3D case, given by Eq. (2.40), is a one-particle sum with a complexity of $\mathcal{O}(N)$.

Slab Correction

An alternative approach to analytically exact quasi-2D Ewald summation techniques is to use so called slab correction techniques [148, 149, 154, 155]. The basic idea is to make the quasi-2D system fully periodic again by introducing periodic boundary conditions along $\mathbf{e}_z$ and additionally separate the (now) vertically periodic images of the slabs in the z direction by introducing wide layers of vacuum of width L_v between them, see Fig. 2.2(a) for a schematic representation. These vacuum regions are not accessible to particles but the 3D periodic treatment allows employing numerically efficient 3D Ewald summation techniques, given by Eq. (2.40), to the modified system. To correct the effects caused by introducing periodic boundary conditions and vacuum layers in the z direction, the 3D Ewald energy, given by Eq. (2.40), is modified by a correction term, $J(\mathbf{M},P)$, as [148, 149]

$$U_{\text{Ewald}}^{(\text{slab})} = U_{\text{Ewald}}^{(\text{3D})} + J(\mathbf{M},P) \quad (2.43)$$

accounting for the induced dipole-dipole interaction across the (now) periodic slabs in the z direction by the total dipole moment of the entire simulation box, $\mathbf{M} = \sum_{i=1}^{N} q_i \mathbf{r}_i$; the parameter P in Eq. (2.43) symbolically specifies the geometry (as defined by the macroscopic boundary conditions) of the system and different, P-dependent correction terms need to be applied in Eq. (2.43). For a spherical geometry, $P = S$, with a surrounding medium with dielectric constant ϵ_r the correction term $J(\mathbf{M},S)$ is given by [149]

$$J(\mathbf{M},S) = \frac{2\pi}{(2\epsilon_r+1)V}|\mathbf{M}|^2, \quad (2.44)$$

that is $J(\mathrm{M}, S) = 0$ for tinfoil boundary conditions with $\epsilon_r = \infty$ and $J(\mathrm{M}, S) = \frac{2\pi}{3V}|\mathbf{M}|^2$ for surrounding vacuum with $\epsilon_r = 1$. A similar approach can also be applied for systems composed of point dipoles [150] and for non-neutral quasi-2D slab systems [151].

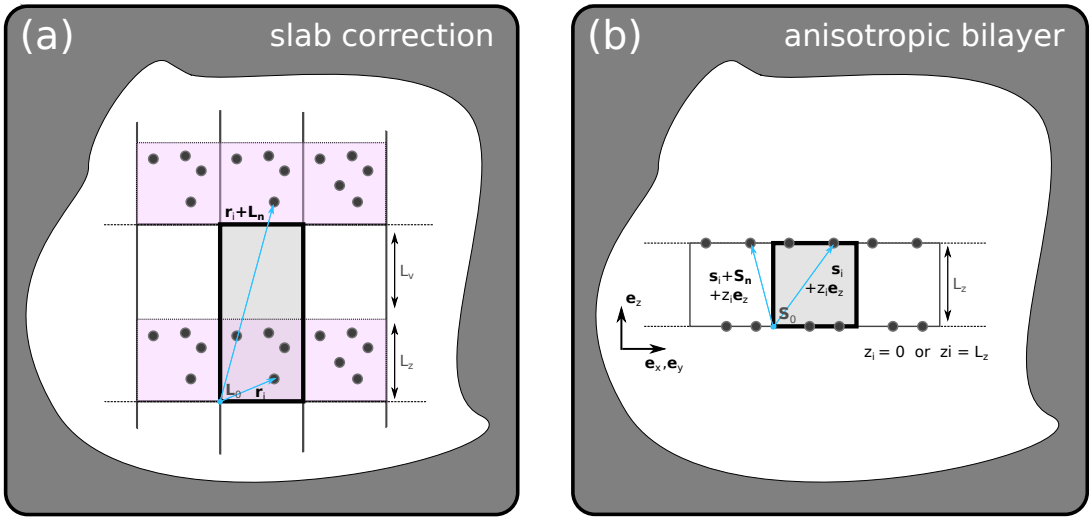


Figure 2.2: Similar to Fig. 2.1(b), but here panel (a) shows the approach of slab correction techniques where periodic images of the simulation cell (emphasized by pink areas) are separated by vacuum layers of width L_v which are not accessible to particles. In panel (b) a quasi-2D bilayer geometry is presented where particles are restricted to occupy only the boundaries of the unit cell in the non-periodic direction, *i.e.*, $z_i = 0$ or $z_i = L_z$.

Bilayer System

In case of a further restricted slab geometry where particles are only allowed to occupy the horizontal slab boundaries, *i.e.*, $z_i = 0$ or $z_i = L_z$ (see Fig. 2.2(b) for an illustration), Eq. (2.41) can be further simplified; equivalently we can define $z_i = \pm d/2$ with d being the slab-width or the "layer separation distance". The discrete spacing in the $\hat{\mathbf{e}}_z$ direction permits us to split the energy summation, $U_{\mathrm{Ewald}}^{(\mathrm{Q2D})} = U_{\mathrm{Ewald}}^{(\mathrm{inter})} + U_{\mathrm{Ewald}}^{(\mathrm{intra})}$, into an inter-layer term $U_{\mathrm{Ewald}}^{(\mathrm{inter})}$, where $z_{ij} = 0$, and into an intra-layer term $U_{\mathrm{Ewald}}^{(\mathrm{intra})}$ with $z_{ij} = \pm d$, which can symbolically be written as [92]

$$\sum_{i,j}^{N} \bullet = \sum_{i_1}^{N_1}\sum_{j_1}^{N_1} \bullet + \sum_{i_1}^{N_1}\sum_{j_2}^{N_2} \bullet + \sum_{i_2}^{N_2}\sum_{j_1}^{N_1} \bullet + \sum_{i_2}^{N_2}\sum_{j_2}^{N_2} \bullet = \sum_{k,l=1}^{2}\sum_{i_k}^{N_k}\sum_{j_l}^{N_l} \bullet . \tag{2.45}$$

The symbolic notations i_k and j_l indicate that particles i_k and j_l belong to layer $k = 1, 2$ and $l = 1, 2$, respectively, and the respective summations over i_k and j_l cover all N_k particles in layer k and all N_l particles in layer l. As already mentioned, in every sum on the right-hand side of Eq. (2.45) the vertical spacing between particles is constant[6] and we define, without the loss of generality, $d_{kl} = z_{i_k j_l} = (l - k)\, d$ with $k, l = 1, 2$. With this discrete set of possible vertical spacings of the particles, $d_{kl} = 0$ for $k = l$

[6] This also holds true more generally for multilayer systems [62, 92].

and $\pm d$ otherwise, we can rewrite and further simplify the second term in Eq. (2.41) – which is computationally especially costly – as

$$U^{\text{Q2D}}_{\text{Ewald,II}} = \frac{\pi}{2A} \sum_{i,j=1}^{N} q_i\, q_j \sum_{\mathbf{G}\neq 0} F(G,\alpha,z_{ij})\, \exp\left(\imath \mathbf{G}\cdot\mathbf{s}_{ij}\right) \tag{2.46}$$

$$= \frac{\pi}{2A} \sum_{k,l=1}^{2} \sum_{i_k=1}^{N_k} \sum_{j_l=1}^{N_l} q_{i_k}\, q_{j_l} \sum_{\mathbf{G}\neq 0} F(G,\alpha,d_{kl})\, \exp\left(\imath \mathbf{G}\cdot\mathbf{s}_{i_k j_l}\right) \tag{2.47}$$

$$= \frac{\pi}{2A} \sum_{k,l=1}^{2} \sum_{\mathbf{G}\neq 0} F(G,\alpha,d_{kl}) \sum_{i_k=1}^{N_k} q_{i_k}\, \exp\left(-\imath \mathbf{G}\cdot\mathbf{s}_{i_k}\right) \sum_{j_l=1}^{N_l} q_{j_l}\, \exp\left(\imath \mathbf{G}\cdot\mathbf{s}_{j_l}\right), \tag{2.48}$$

with $\mathbf{s}_{i_k j_l} = \mathbf{s}_{j_l} - \mathbf{s}_{i_k}$ and $d_{kl} = (l-k)\,d$. The corresponding inter-layer term, $U^{(\text{inter})}_{\text{Ewald,II}}$, with $k = l$ can be written as

$$U^{(\text{inter})}_{\text{Ewald,II}} = \frac{\pi}{2A} \sum_{k=1}^{2} \sum_{\mathbf{G}\neq 0} F(G,\alpha,0) \left| \sum_{i_k=1}^{N_k} q_{i_k}\, \exp\left(\imath \mathbf{G}\cdot\mathbf{s}_{i_k}\right) \right|^2 \tag{2.49}$$

and the intra-layer term, $U^{(\text{intra})}_{\text{Ewald,II}}$, with $k \neq l$ as

$$U^{(\text{intra})}_{\text{Ewald,II}} = \frac{\pi}{2A} \sum_{(k\neq l)=1}^{2} \sum_{\mathbf{G}\neq 0} F(G,\alpha,d_{kl}) \sum_{i_k=1}^{N_k} q_{i_k}\, \exp\left(-\imath \mathbf{G}\cdot\mathbf{s}_{i_k}\right) \sum_{j_l=1}^{N_l} q_{j_l}\, \exp\left(\imath \mathbf{G}\cdot\mathbf{s}_{j_l}\right) \tag{2.50}$$

such that $U^{\text{Q2D}}_{\text{Ewald,II}} = U^{(\text{inter})}_{\text{Ewald,II}} + U^{(\text{intra})}_{\text{Ewald,II}}$. The corresponding time complexity for numerical evaluations of Eqs. (2.49) and (2.50) then becomes comparable to the time complexity of the corresponding term in the three-dimensional case (cf. second sum in Eq. (2.40)).

Hence, the discrete set of possible vertical particle spacings permits us to study bilayer systems of point-charges [62–64, 146, 147] with quasi-2D Ewald summation methods with great accuracy and with great efficiency[7] [62, 92].

2.1.4 Convergence Criteria for Evaluating Ewald Sums Numerically

The Ewald summations given by Eqs. (2.40) and (2.41) are explicitly defined for systems of point-charges in three and in quasi–two dimensions. Other long-ranged interaction potentials may lead to variations of the particular form of the Ewald summation, but many basic features remain the same [62, 92]. We want to stress, that the Ewald summations given by Eqs. (2.40) and (2.41) are mathematically exact (irrespective of the numerical value of the Ewald α parameter, *i.e.*, of the (reciprocal) width of the Gaussian charge distributions) and still contain an infinite number of terms.

[7] In this thesis, we rely on the efficient implementation of quasi-2D Ewald summation methods for bilayer systems of Coulombic point-charges of M. Antlanger [62].

For any practical purposes when evaluating Ewald summations numerically, cutoff distances in both the real space, r_{cut}, and the reciprocal space, k_{cut}, need to be introduced, up to which the summations in Eqs. (2.40) and (2.41) are carried out. Introducing cutoff distances and truncating the corresponding sums causes a numerical error in the evaluation of the internal energy.

However, we emphasize that the internal energy of a long-ranged system can very accurately be evaluated numerically via truncated Ewald summations [92]: r_{cut} and k_{cut} can be chosen in accordance with the Ewald summation parameter α such that a desired accuracy, $1 - \delta_{\text{cut}}$, of the internal energy of the system is achieved [62]. In practice, the particular choice of those three parameters, r_{cut}, k_{cut} and α, should not affect the value of the internal energy of the system (up to the specified numerical accuracy). Thus, they must be chosen such that U_{Ewald} exhibits, as a function of r_{cut}, k_{cut} and α, a plateau where the energy evaluation is independent to changes of α in a reasonable large parameter range [92]. A rule of thumb is to chose $k_{\text{cut}} \sim \alpha$ and $r_{\text{cut}} \sim 1/\alpha$, since $\sqrt{2}/\alpha$ defines the width of the Gaussian charge distributions and, more importantly, roughly sets the distance at which the complementary Gauss error functions annihilate the integrands in the real space integrals in Eqs. (2.40) and (2.41); an analogous argument holds for $k_{\text{cut}} \sim \alpha$ in reciprocal space. In general, these parameters need to be carefully adjusted [62] such that the internal energy of arbitrary particle arrangements in a simulation box can faithfully be evaluated.

Choosing the cutoff parameters r_{cut}, k_{cut} and the Ewald parameter α accordingly, such that they do not affect the energy evaluations, can directly be applied for systems of quasi-2D slab geometries discussed in Subsection 2.1.3. For slab correction models the internal energy, cf. Eq. (2.43), should also be independent of the width of the introduced vacuum layers, L_v, which needs to be chose large enough.

2.2 Ground State Exploration: Objective, Methods and Algorithms

In this Section, we will introduce methods and algorithms used in this thesis to predict ground state configurations of particles or molecules arranged in lattices. First we introduce the concept of the potential energy landscape in Subsection 2.2.1 which is followed by a general introduction and background to optimization strategies in Subsection 2.2.2. We then introduce numerical approaches mainly used in this thesis to reliable identify global ground state configurations, namely simulated annealing (SA) and Replica Exchange Monte Carlo (REMC), see Subsection 2.2.3, and (memetic) evolutionary algorithms (EA), see Subsection 2.2.4. We introduce these different numerical tools (*i.e.*, SA, REMC and EA) mostly in a general framework. Whenever it is necessary or helpful for the discussion we address related algorithmic details of particular applications to specific physical system which are studied in this thesis (cf. Chapter 3).

2.2.1 Physical Energy Landscapes can be Complex

The main objective of this thesis is to identify ground state particle or molecule[8] arrangements in lattices, *i.e.*, structural configurations that minimize the internal energy (per particle) of a system at vanishing temperature. In what follows, we will use the term "particle" synonymously for the elemental objects whose properties, such as their positions or – in case of aspherical molecules – their center of mass coordinates and orientations, are variables in the structure optimization procedures.

For simple (yet instructive) problems such as an attractive harmonic potential or Lennard Jones interaction between two particles, the potential energy as a function of the particle distance exhibits a single minimum which can easily be identified. For more complex systems with more than two particles, the situation drastically changes [1]: the potential energy as a function of all particle positions quickly becomes much more complicated, may not be a strictly convex function anymore but rather be of rugged shape with a large number of local minima. In fact, for a simple Lennard-Jones system (or other systems of simple, isotropic pair-wise interactions), the number of possible local minima of the potential energy as a function of particle positions increases exponentially with the number of particles [96], rendering the identification of the global minimum a far from trivial task.

We are ultimately interested in identifying ordered particle arrangements – in lattice geometries – which globally minimize the internal energy (or, in our case equivalently, the potential energy) of different physical systems. In other words, we want to identify the particle arrangement which represents the global ground state configuration for a given system. The challenge of finding the structural ground state configurations of a system of isotropic interactions is to identifying the optimal particle positions, $\mathbf{R}^N = (\mathbf{R}_1, \ldots, \mathbf{R}_N)$, of all N particles in the unit cell. We also might look for the

[8] We consider interactions of point-charges in Section 3.1 as well as interactions between rigid molecules whose atomic constituents exhibit partial charges and a Lennard-Jones type interaction in Section 3.2.

optimal unit cell geometry, *i.e.*, for the best choice of lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, here presented in three dimensions. Furthermore, for anisotropic particles also their orientations, given by $\mathbf{P}^N = (\mathbf{P}_1, \ldots, \mathbf{P}_N)$, need to be considered in the structure optimization procedure.

Although often not all parameters are independent (the unit cell volume might be constant or certain angles of the unit cell might be fixed, etc.), the number of parameters that need to be considered, in our example $\mathbf{R}^N, \mathbf{P}^N, \mathbf{a}_1, \mathbf{a}_2$ and $\mathbf{a}_3$, can be quite large and may reach up to several hundreds. We collect all variable parameters in the set $\mathcal{X}$, *i.e.*, $\mathcal{X} = \{\mathbf{R}^N, \mathbf{P}^N, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ for our example. Thus, we can define the objective function we aim to optimize for different physical systems such as the potential energy per particle[9]

$$\mathcal{U}(\mathcal{X})/N \to \min\,. \tag{2.51}$$

The overall shape of $\mathcal{U}(\mathcal{X})/N$ is called the *potential energy landscape* which is an analytic (yet complicated) function of the set of all variable parameters, $\mathcal{X}$, of the system. The space of all possible realizations of $\mathcal{X}$ is called the *configuration space*.

Furthermore, for interactions operating on vastly different length-scales in geometries with anisotropic directions, as we discuss above and in Section 3.2, both, faithfully evaluating the energy of the system and studying the ground state self-assembly can become rather difficult and may even lead to effects such as frustration[10]. A vast number of other configurations – sitting at minima of the potential energy landscape with energies comparable to the global energy minimum of the system – may also play an important role in the structure formation processes [89]. When experimentally (or numerically, see Subsection 2.2.3 below) cooling a structural configuration in an effort to identify the global ground state of a system different low energy polymorphs can easily be kinetically trapped, and the global ground state may remain hidden [89]. From a numerical point of view, this is equivalent to sampling a configuration space for the global minimum in the presence of many disconnected local minima of comparable energies – which is a quite challenging task. If competing structures can be transformed into each other with negligibly energetic cost even aperiodic quasicrystals can emerge, at least at finite (small) values of temperature [167, 168], which makes a numerical treatment conceptually even more difficult (see Section 3.1). Thus, identifying and classifying such competing structures is of great importance for understanding trends in the structure formation process of physical systems at vanishing temperatures.

In order to cope with the complexity of such global optimization problems exhaustive search strategies are numerically intractable, even for nowadays most powerful supercomputers [89, 96]. To tackle this problem we utilize, combine and deploy within this thesis advanced optimization techniques ranging from deterministic gradient descent solvers over stochastic simulated annealing (SA) and Replica Exchange Monte

[9] The algorithms presented below are not restricted to minimize the potential energy but operate on any (sufficiently well-behaved) objective function $\mathcal{U}(\mathcal{X})$.

[10] Lowering the contribution to the internal energy from one interaction may increase the contribution from another which can potentially trap numerical optimization procedures in local minima.

Carlo samplers (see Subsection 2.2.3) to heuristic, memetic evolutionary algorithms (see Subsection 2.2.4).

2.2.2 Navigating towards the Global Minimum

Finding an (arbitrary) local minimum can be performed in a relative straightforward manner: the region in the configuration space surrounding a local minimum is called an *attraction basin* and following down the gradient, $\partial\mathcal{U}/\partial\mathcal{X} \to 0$, leads to the local minimum. A vast number of local, gradient based optimizer exist for that purpose [169–176]

Identifying the global minimum of a complex function (such as the energy landscape of a physical system) is a far more complicated challenge yet of great importance in a vast variety of fields of research such as protein folding [126, 177–179], chemical engineering [180–183], or – very prominently due to the recent outbreak of the SARS-CoV-2 pandemic – in computational drug-design [184]; the above list is far from being comprehensive.

Such tasks usually fall into the category of *NP* or even *NP-hard* problems, *i.e.*, a set of non-deterministic polynomial acceptable problems that are considered to be not solvable in polynomial time (scaled by their degrees of freedom) with a deterministic turing machine [185]; one very prominent example for the class of NP problems is the traveling salesman problem [186].

Many researchers have been working on strategies to efficiently solve such optimization problems computationally for decades, consequently a large number of algorithms emerged over the years. They can roughly be grouped into *deterministic*, *stochastic* and *heuristic* algorithms.

Deterministic Optimization Strategies

If the problem allows it, deterministic optimization strategies aim at solving the optimization problem in an exact and reproducible manner. Based on interval arithmetic [187, 188] some algorithms try to describe the solution space analytically in order to identify possible solutions deterministically. In other cases the general shape and/or symmetries of solutions are known and the search space can be greatly reduced by focusing on a relevant subspace of configurations. A proper educated guess for the initial candidate may allow suitable, local, gradient descent based optimization techniques[11] to identify the global optimum.

[11] Gradient descent (GD) based optimizers minimize an objective function by following the direction of the steepest gradient in configuration space, *i.e.*, $\mathcal{U}(\mathcal{X}) \to \mathcal{U}(\mathcal{X}) - \eta \cdot \nabla_{\mathcal{X}}\mathcal{U}(\mathcal{X})$ with η being a numerical step size; of course many quite sophisticated GD algorithms exist.

Stochastic Optimization Strategies

A different class of optimization algorithms relies on random variables in order to sample the parameter space for the optimal solution. Monte Carlo [189–192] algorithms are typical realizations of such stochastic techniques. Via random moves, *i.e.*, by randomly changing degrees of freedom, these techniques, on the one hand, are able to improve solutions and, on the other hand, allow jumps between attraction basins of different local minima in order to explore the configuration space more broadly. While the algorithmic parameters – such as the numerical step sizes of random moves of different variables and the acceptance criteria of proposed moves, etc. – need to be tuned carefully stochastic optimization strategies may eventually identify the global minimum of an optimization problem.

An improvement of standard Monte Carlo algorithms for global optimization purposes is, for instance, stochastic tunneling [193, 194]. The original objective function (which is to be optimized) is non-linearly transformed allowing more efficient jumping between regions in the vicinity of local minima. Thereby, convergence rates of the algorithm are improved.

Simulated annealing [133, 134] techniques are inspired by the process of annealing in metallurgy, a technique used to grow a crystal of high purity via heating and cooling of the material. Simulated annealing algorithms mimic this technique by successively modify the acceptance rates of regular Monte Carlo sampling [190, 191] via an artificial temperature: starting from high temperatures (*i.e.*, from high acceptance rates) large jumps in configuration space allow to broadly explore the search space. A configuration (or equivalently a sample) is gradually cooled down and the jumps become successively smaller such that, at lower temperatures, the sample approaches the global optimum[12].

Extending this idea, parallel tempering (or replica exchange Monte Carlo) [135–138] performs sampling of several, different configurations in parallel at different temperatures. Additional *tempering* steps are used to exchange configurations between neighboring temperatures, constantly allowing configurations evaluated at high temperatures to be further optimized at low temperatures (and vice versa), see Subsection 2.2.3 for more details.

Heuristic Optimization Strategies

In contrast to the techniques discusses above, heuristic optimization algorithms try to explore the search space in a more intelligent way by making use of already obtained information on the target function such as problem specific symmetries and boundaries, or by learning good update policies of parameters. Even for very complicated optimization problems they guide the search process towards promising regions in configuration space and explore regions of nearly optimal solutions in a problem specific and more intelligent way as, for instance, stochastic procedures. It is worth noting that usually such algorithms need to be specifically tailored for particular problems

[12] The global optimum of a configuration space may eventually be identified via simulated annealing if the cooling process is performed adiabatically, *i.e.*, slow enough.

since it cannot be expected from an algorithm that can solve one particular class of problem efficiently to work as good for a completely different task [62]: "there ain't no free lunch", neither for search nor for optimization [141–143].

Examples of heuristic optimization strategies are reactive- or *tabu* search [195, 196]: (i) a local search strategy is extended by also considering worsening steps during updates in the minimization process and (ii) aspects from machine learning are applied such that already visited solutions are marked *tabu* in order to avoid them during the continuing optimization process.

Graduated optimization [197] aims at greatly reducing the difficulty of the full optimization problem in the initial phase of the algorithm; the complexity of the simplified search problem is then gradually increased during optimization until, eventually, the original problem complexity is restored and can be treated.

Other techniques rely on ensembles of candidate solutions, *i.e.*, on a pool of configurations which is gradually updated and thereby gather information about the problem. In biologically inspired swarm optimization algorithms, such as particle swarm optimization [198, 199], an ensemble of candidate solutions, *i.e.*, particles, propagates in configuration space according to specific, simple mathematical models [200] for their positions and velocities. The members of the swarm or of even several diverse swarms (as in multi-swarm approaches [201, 202]) can be located and propagate in a single or between different attraction basins while exploring the configuration space. Eventually the entirety of particles is guided towards the global optimum (although convergence cannot be guaranteed). A related technique is differential evolution [203]: Based on a population of different candidate solutions selected configurations within this population are picked and their parameters are combined in a so-called crossover move which is performed according to a specific, simple formulae [203]. The new sample may be better suited to optimize the problem in which case it is accepted and forms a new part of the population, otherwise it is simply discarded.

In this thesis, we use and further develop tools based on evolutionary algorithms [62, 204, 205] which are built upon ideas of biological evolution, *i.e.*, natural (or better artificial) selection, reproduction and mutation. An ensemble of configurations forms a population of previously obtained solutions from which selected *parents* are chosen, combined or modified in order to produce *offsprings* which may be better equipped to optimize the problem [62]. Due to slow convergence rates, especially in continuous optimization problems (as in material science), these tools can be combined with local optimization techniques, such as gradient descent algorithms. This can steer offsprings configurations towards local minima before accepting or discarding samples; such a combination of local and global optimization is referred to as memetic algorithms [139].

Above methods make only few or no assumptions about the problem and are classified as meta-heuristics. However, sometimes prior information about the system is available and can be incorporated into the optimization strategy. Bayesian approaches to global optimization [206], for instance, are sequential design strategies which make use of this property in terms of Bayesian statistics.

Recent Trends in Optimization: Artificial Intelligence

In recent years, very promising numerical techniques have emerged to tackle material science problems (or optimization problems in general) which are based on the ideas of artificial intelligence and machine learning (for very instructive reviews of this topic we refer to by Refs. [116–119]). Current trends in optimization strategies heavily point into the direction of using data driven models and artificial intelligence such as generative models [129] and reinforcement learning [123, 124]. Generative models (such as Boltzmann-machines [130]) can be used to draw high quality samples from an arbitrarily complex configuration space (e.g. from the space of images of human faces) after a model has been trained on a (typically large) data set of representative examples which follow the probability distribution of the configuration space [119]. Reinforcement learning has, for instance, very successfully been applied to find optimal solutions to protein folding problems [126] beyond the capabilities of conventional approaches in a similar way as AlphaGo [125] was able to master the infamous board game.

Even more abstract machine learning models are used in materials science or chemistry which are based on ideas of natural language processing of sequential data (such as chemical structure formulas or written, scientific text). For instance, a deep neural network[13] based model, denoted IBM RXN [182, 183], has successfully been employed to predict chemical reactions based on a text-representation of the involved chemical reactants. It has been shown [182, 183] that this tool is able to outperform professional human scientists in predicting the outcome of chemical reactions. The interesting thing here is, that the IBM RXN model is solely trained on text representations of chemical reactions, *i.e.*, no other physical or chemical models what so ever are involved during the training of the model. Similar tools have been used to screen large data bases of molecules or material science data in order to "discover" new materials with specific target-properties, such as previously unknown thermo electrica [209]; very recently such an approach was employed in the field of computational drug design to identify certain molecules which may be promising candidates to treat a CoV-Sars-2 infection [184].

All of these machine-learning based approaches introduced above aim at identifying optimal candidates from a configuration space of possible solutions to a given optimization problem. The field of artificial intelligence based optimization methods is rapidly growing (not at least owed to the now available computational resources and data), and the above list is far from being comprehensive. It will be fascinating to see, which problems can be tackled with tools from artificial intelligence in the near and more distant future and also to learn about the limitations of these methods.

[13] An artificial neural network (or *neural network* in short) is a set of interconnected artificial neurons which collect weighted signals (either from external sources or from other neurons) and create and redistribute output signals generated by a nonlinear activation function [207]. In that way a neural network can process information in a very efficient and, more importantly, very flexible way: by adjusting the weights and biases of connections between different neurons (or by adjusting the network topology), especially so-called *deep neural networks* can be trained to map an arbitrary input of the network onto desired output signals in a highly non-trivial way thereby realizing task specific operations which are often too complicated to be implemented manually [207, 208]. Usually, a neural network is referred to as "deep", if it is composed of more than three consecutively interconnected layers of neurons [208].

With this we conclude Subsection 2.2.2 and we will proceed with a more in-depth explanation of three algorithms which we heavily used in this thesis, namely simulated annealing, replica exchange Monte Carlo and a memetic evolutionary algorithm. These three algorithms are – in principle – able to find global minima in high-dimensional, rugged energy landscapes.

2.2.3 Monte Carlo Methods: Simulated Annealing and Replica Exchange Monte Carlo

Monte Carlo (MC) methods [189] are scalable, general tools mainly used to draw random samples from a certain probability distribution. They are used for numerical integration or – relevant for our purposes – to numerically solve optimization problems. They are widely used not only in physics and mathematics but also in computer science, finances, industry and many other fields of research [192]. The core idea is to repeat a computer experiment many times based on random sampling, *i.e.*, simulating sufficiently long, to gather sufficient information about the underlying statistics of a random quantity which mimics the behaviour of real-life process; basically, MC methods can be applied to any problem which has a probabilistic interpretation [192].

For parametrized probability distributions, so-called Markov chain Monte Carlo methods (MCMC) can be applied for obtaining sequences of samples from (preferably high-dimensional) probability distributions for which direct sampling is difficult [190, 191]. The Metropolis–Hastings algorithm [191] is probably the most prominent MCMC algorithm. The Boltzmann distribution in statistical physics[14] is one very prominent example of a parametrized probability distribution which takes the form

$$p(\mathcal{X}, T) = \frac{1}{Z} \exp\left(-\mathcal{U}(\mathcal{X})/k_{\mathrm{B}}T\right), \tag{2.52}$$

where $Z = \int_{V_\mathcal{X}} p(\mathcal{X}, T)\, d\mathcal{X}$ is the partition function ensuring $p(\mathcal{X}, T)$ is normalized, k_{B} is Boltzmann's constant and T is the temperature. $\mathcal{X}$ is a particular realization of the random quantity – to which we frequently refer to as sample – from all possible realizations within the volume $V_\mathcal{X}$ of the so-called configuration space. In our case, $\mathcal{X}$ might represent a structural configuration composed of point-charges or molecules.

Metropolis-Hastings Algorithm

The Metropolis-Hastings Algorithm represents an MCMC process which is uniquely defined by its transition probabilities $P(\mathcal{X}'|\mathcal{X}_t)$, *i.e.*, the probability of transition to get from one sample, $\mathcal{X}_t$, to another state, $\mathcal{X}'$. It is assumed that the MCMC process reaches asymptotically a stationary distribution, $p(\mathcal{X})$, such that the detailed balance condition

$$P(\mathcal{X}'|\mathcal{X}_t)p(\mathcal{X}_t) = P(\mathcal{X}_t|\mathcal{X}')p(\mathcal{X}') \tag{2.53}$$

[14] As put forward by Jaynes [119, 127, 128] the Boltzmann distribution, underlying statistical physics, can be derived in a purely information theoretic way when applying the method of maximum entropy [127, 128] to the conservation of expectation values of thermodynamic variables over ensembles of microstates; The maximum entropy method represents a generative model [129] which inspired, for instance, Boltzmann machines [210] in machine learning [119].

is fulfilled: detailed balance implies that transitions from $\mathcal{X}_t$ to $\mathcal{X}'$ are as probable as the inverse process. The transition probability, $P(\mathcal{X}'|\mathcal{X}_t)$, is then separated into a conditional[15] proposal probability, $g(\mathcal{X}'|\mathcal{X}_t)$, and a joint[16] acceptance probability, $A(\mathcal{X}', \mathcal{X}_t)$, *i.e.*, $P(\mathcal{X}'|\mathcal{X}_t) = g(\mathcal{X}'|\mathcal{X}_t)\, A(\mathcal{X}', \mathcal{X}_t)$. Inserting the latter expression for $P(\mathcal{X}'|\mathcal{X}_t)$ into Eq. (2.53) yields

$$\frac{A(\mathcal{X}', \mathcal{X}_t)}{A(\mathcal{X}_t, \mathcal{X}')} = \frac{p(\mathcal{X}')}{p(\mathcal{X}_t)} \frac{g(\mathcal{X}_t|\mathcal{X}')}{g(\mathcal{X}'|\mathcal{X}_t)}. \tag{2.54}$$

Now, an acceptance probability ratio $A(\mathcal{X}', \mathcal{X}_t)/A(\mathcal{X}_t, \mathcal{X}')$ needs to be chosen which satisfies Eq. (2.54); a common choice is the so-called *Metropolis choice* given by

$$A(\mathcal{X}', \mathcal{X}_t) = \min\left(1, \frac{p(\mathcal{X}')}{p(\mathcal{X}_t)} \frac{g(\mathcal{X}_t|\mathcal{X}')}{g(\mathcal{X}'|\mathcal{X}_t)}\right), \tag{2.55}$$

where either $A(\mathcal{X}', \mathcal{X}_t) = 1$ or $A(\mathcal{X}_t, \mathcal{X}') = 1$ such that detailed balance is fulfilled.

In cases of constant proposal probability, $g(\mathcal{X}'|\mathcal{X}_t) = g(\mathcal{X}_t|\mathcal{X}')$, the acceptance probability reduces to $A(\mathcal{X}', \mathcal{X}_t) = \min(1, p(\mathcal{X}')/p(\mathcal{X}_t))$ which yields for the Boltzmann distribution defined in Eq. (2.52) the following expression for the acceptance probability:

$$A(\mathcal{X}', \mathcal{X}_t; T) = \min\left(1, \exp\left(-\Delta\mathcal{U}(\mathcal{X}', \mathcal{X}_t)/k_\mathrm{B}T\right)\right). \tag{2.56}$$

The acceptance probability given by Eq. (2.56) depends on the energy difference, $\Delta\mathcal{U}(\mathcal{X}', \mathcal{X}_t) = \mathcal{U}(\mathcal{X}') - \mathcal{U}(\mathcal{X}_t)$, between the candidate sample, $\mathcal{X}'$, and the current sample, $\mathcal{X}_t$, and additionally on an artificial temperature, T, at which the simulation is conducted.

The Metropolis-Hastings algorithm consists of the following steps:

1. Start from an arbitrary initial sample, $\mathcal{X}_t$ (and set $t = 0$).
2. Generate a new sample, $\mathcal{X}'$, via random updates (or, equivalently, random moves) applied to $\mathcal{X}_t$ following the proposal probability $g(\mathcal{X}'|\mathcal{X}_t)$.
3. Accept $\mathcal{X}'$ if a uniform random number $r \in [0, 1] \leq A(\mathcal{X}', \mathcal{X}_t)$, reject otherwise.
4. If accepted, set $\mathcal{X}_{t+1} = \mathcal{X}'$, otherwise $\mathcal{X}_{t+1} = \mathcal{X}_t$
5. Increment $t \to t + 1$ and continue with step 2.

Over multiple iterations this Markov chain process generates a random walk sequence of realizations from the stationary distribution $p(\mathcal{X})$.

For so-called ergodic systems any state can be transformed into any other from the same configuration space by a sequence of update-moves. For such systems, expectation values of certain functions, $f(\mathcal{X})$, of the random process can be approximated – by the law of large numbers – using the empirical mean, *i.e.*, $\langle f(\mathcal{X})\rangle = \int_{V_\mathcal{X}} f(\mathcal{X})\, p(\mathcal{X})\, d\mathcal{X} \approx (1/N) \sum_{t=0}^{N-1} f(\mathcal{X}_t)$, where $V_\mathcal{X}$ is the configuration space volume [190, 191].

[15] A conditional probability distribution $p(X|Y)$ quantifies the probability that an event X occurs given Y has occurred. In our case, the proposal probability $g(\mathcal{X}'|\mathcal{X}_t)$ quantifies the probability of proposing a state $\mathcal{X}'$ given $\mathcal{X}_t$.

[16] A joint probability distribution $p(X, Y, \dots)$ quantifies the probability that several events $X, Y, \dots$ occur simultaneously. In our case, $A(\mathcal{X}', \mathcal{X}_t)$ represents the probability to accept the proposed state $\mathcal{X}'$ starting from state $\mathcal{X}_t$.

Monte Carlo based Optimization: Simulated Annealing

Continuing our discussion on optimization problems, the inherent randomness of MC algorithms is the key quantity which makes them applicable for global optimization tasks. In contrast to gradient descent algorithms, which are well suited tools to reliably evaluate local minima, MC methods are able to jump – with a certain probability – between different basins in the potential energy surface; given enough time and/or resources MC based optimization algorithms may eventually identify the global minimum. A very prominent example for an MC based optimization algorithm is simulated annealing (SA) [133, 134] which is inspired by the process of annealing in metallurgy, *i.e.*, a technique involving heating and cooling of the material, to grow a crystal of high purity.

SA is an adaptation of the Metropolis-Hastings algorithm. A single sample is constantly subjected to the Metropolis-Hastings MC algorithm while the temperature, T_i, in the acceptance probability $A(\mathcal{X}', \mathcal{X}_t; T_i)$ defined by Eq. (2.56) is successively decreased over time in $i = N_i, N_i - 1, \ldots, 0$ additional annealing steps. The temperature grid, $T_{N_i} > \ldots > T_i > T_{i-1} > \ldots > T_0$, is usually predefined, the initial temperature, T_{N_i}, is typically large and the final temperature $T_0 = 0$ K.

The SA algorithm can be described as follows:

1. Start with an (arbitrary) initial sample, $\mathcal{X}_{t=0}$, and at an initial large value of the temperature, $T_{i=N_i}$, by setting $i = N_i$.

2. Perform N_t iterations of Metropolis-Hastings MC (or use another suitable MCMC sampler) at a given temperature, T_i, starting from the sample $\mathcal{X}_0$.

3. Decrease the temperature $T_i \to T_{i-1}$ by setting $i \to i - 1$.

4. If $i \geq 0$ set $\mathcal{X}_0 = \mathcal{X}_{N_t}$ and continue with step 2., otherwise terminate with the result $\mathcal{X}_{N_t}$.

At the initially large value of the temperature, T_{N_i}, the acceptance probability of $A(\mathcal{X}', \mathcal{X}_t; T_{N_i}) \approx 1$ permits large jumps in configuration space during the N_t Metropolis-Hastings MC steps. Large energy-difference are compensated by large temperatures in the acceptance probability and the initial sample, $\mathcal{X}_0$, is subject to large thermal fluctuations such that $\mathcal{X}_{N_t}$ is essentially randomized.

If SA is performed adiabatically, *i.e.*, if the temperature is decreased slowly enough (via a sufficiently fine T_i grid) and Metropolis-Hastings MC sampling at each temperature, T_i, is given enough time (via a sufficiently large N_t), SA cools down a sample and guides it towards the global optimum (although this process might be computationally infeasible for many difficult problems due to prohibitively long execution times).

For temperatures close to zero the acceptance probability approaches 0 for worsening steps (*i.e.*, if a proposed sample is less favorable with respect to the objective function compared to the current one) and the sample configurations freezes out, ideally into the optimal solution.

Replica Exchange Monte Carlo Optimization

Another approach to tackle global optimization problems with Monte Carlo algorithms is Replica Exchange Monte Carlo (REMC) [135–138], or parallel tempering (we use both phrases synonymously in this thesis). REMC extends the idea of simulated annealing by subjecting several samples in parallel to Metropolis-Hastings MC (or MCMC approaches in general) at different temperatures. The samples of neighboring temperatures are then successively swapped in additional tempering steps instead of annealing a single sample as in SA.

The REMC algorithm can be described as follows:

1. Start from N_T+1 arbitrary initial samples, $\mathcal{X}^{N_T} = \{\mathcal{X}_{T,0}, \mathcal{X}_{T,1}, \ldots, \mathcal{X}_{T,i}, \ldots \mathcal{X}_{T,N_T}\}$ and define an associated, ordered temperature grid $T^{N_T} = \{T_0 < T_1 < \ldots < T_i < \ldots < T_{N_T}\}$.

2. Perform N_t iterations of Metropolis-Hastings MC sampling (or use another suitable MCMC sampler) for each temperature, T_i, respectively starting from the current (associated) sample $\mathcal{X}_{T,i}$. *This can be done in parallel.*

3. Set $\mathcal{X}_{T,i} = \mathcal{X}_{N_t}$ with the final sample, $\mathcal{X}_{N_t}$, of the respective Metropolis-Hastings MC simulation performed at the temperatures, T_i, for all $i = 0, \ldots, N_T$.

4. Exchange samples $\mathcal{X}_{T,k}$ and $\mathcal{X}_{T,l}$ between neighboring temperatures $T_k < T_{l=k+1}$ in the following way:

   4.1 Pick $k = 2j$ and $l = 2j+1$ pairs every even cycle and $k = 2j+1$ and $l = 2j+2$ pairs every odd cycle for every $j = 0, 1, \ldots, N_t/2$; N_t is considered to be an even number and if $l > N_T$ the k, l pair is omitted. *In that way, the samples can diffuse along the temperature direction*

   4.2 A sample swap between neighboring temperatures $T_k < T_l$ is always accepted if $\Delta\mathcal{U}_{kl} = \mathcal{U}(\mathcal{X}_{T,l}) - \mathcal{U}(\mathcal{X}_{T,k}) < 0$ and accepted according to the REMC acceptance criteria, *i.e.*, if a uniform random number $r \in [0, 1] \leq \exp\left(-\Delta\mathcal{U}_{kl}/k_\mathrm{B}\Delta T_{kl}\right)$ with $\Delta T_{kl} = T_l - T_k$.

   4.3 The actual swap, $\mathcal{X}_{T_k} \leftrightarrow \mathcal{X}_{T_l}$, for a particular k, l pair is only performed if the swap between the temperature pair T_k and T_l was accepted.

5. Continue with step 2. until a convergence criterion is met, otherwise terminate with the result $\mathcal{X}_{T,0}$.

At large temperatures the parallel tempering algorithm allows to broadly explore the configuration space and good samples (with low energies) are successively swapped down to lower temperatures. Good samples are continuously further optimized at low temperatures but are always in competition with other low-energy samples which may independently emerge during the execution of the algorithm. REMC can be understood as several interlinked simulated annealing procedures, which are executed in parallel and which are able to exchange good samples in additional tempering steps. Furthermore, REMC also allows good, yet sub-optimal solutions to resolve possible defects when being swapped to higher temperatures.

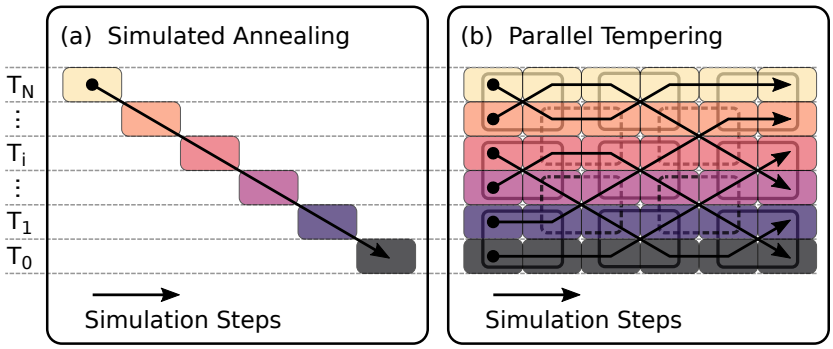


Figure 2.3: Schematic time evolution of (a) simulated annealing (SA) procedure and (b) of parallel tempering or replica exchange Monte Carlo (REMC). Colored boxes represent MCMC simulations at the respective temperatures, $T_0, \cdots, T_N$ (color coded). For SA the optimization is a "single trajectory" (the sample evolution is indicated by the thick black arrow) where the temperature is successively decreased after sufficient MCMC iterations at a single temperature. REMC, in contrast, can be performed in parallel at many different temperatures: samples are exchanged between neighboring temperatures after the respective MCMC simulations (cf. thick black arrows) according to the REMC acceptance criterion. At even swap steps neighboring (T_{2i}, T_{2i+1}) sample pairs are exchanged (indicated by the nine solid-framed boxes in (b)) while at odd swap steps neighboring (T_{2i+1}, T_{2i+2}) sample pairs are exchanged (four dashed-framed boxes in (b)).

REMC can easily be utilized using standard MCMC algorithms and, furthermore, can easily be scaled to run on many parallel cores on supercomputers. Thus, parallel tempering representing a highly efficient, powerful tool for global ground state exploration – particularly for sampling discrete configuration spaces.

In Fig. 2.3 we schematically visualize a time evolution of both, the simulated annealing procedure and the parallel tempering or REMC algorithm. Note that the SA algorithm is essentially a "one shot" trajectory and the annealing needs to be performed very carefully in order to faithfully identity the global minimum. In REMC, on the other hand, an ensemble of different samples is considered in parallel and good solutions at low temperatures are constantly in competition with other potential ground state solutions emerging at larger values of the temperature.

2.2.4 Evolutionary Algorithm

From Natural Selection to Global Ground State Search

Inspired by the principles of Darwin's theory of evolution [211] of biological systems, evolutionary algorithms (EA)[62, 63, 104–109, 139, 140, 205] use concepts of natural (or better artificial) selection to find the best solution to an objective function.

In the theory of evolution a population of genetically diverse individuals, which are called *phenotypes*, populate a shared environment. Based on their genetic blueprint, *i.e.*, based on their *genotype*, they exhibit different physical traits reflecting on how well they are adapted to their natural environment. Individuals of a population (or better of a species) can pass on their (shared) genetic information via reproduction to their offsprings, which may then be better equipped for survival than other members

of the population and take over their place; individuals less adapted will eventually perish and whole species not adequately equipped may become extinct. The rate of adaptation, *i.e.*, how well individuals perform, is called *fitness* and in nature only the fittest may survive. It is more likely for well adapted individuals to pass on their genetic information than others, with lower fitness, since the former have a higher survival probability, *i.e.*, they have been around for a longer time thus having more chances to reproduce. This also suggests that associated *good traits* are passed on to subsequent generations with a higher probability such that, over time, the majority of individuals of a population are gradually adapted to their environment in order to increase their chances of survival. Furthermore, because nothing seems to be perfect in nature, errors may occur during the reproduction stage – at random – when genetic material is passed on over generations. Such *mutations* usually affect how an individual is adapted to its environment and can have either beneficial or disadvantageous effects. However, random mutations may give access to completely new features and strategies of survival and open up new regions in the configuration space of possible solutions to the evolutionary process.

Both the genome representation of individuals and the environment of biological systems are far too complicated to be simulated in all their details by present day technology. Still, evolutionary algorithms take on the heuristic yet flexible concepts of biological evolution and adaptation processes – in a very simplified way – in order to explore the configuration space of an optimization problem for the global minimum. From an algorithmic point of view, individuals can be represented by objects or data structures on a computer (describing, for instance, particle arrangements in a lattice) and further can be serialized to a *genome string*, $\mathcal{X} \in \mathbb{R}^{N_A}$, where N_A is the dimensionality of the search problem, *i.e.*, the number of degrees of freedom to which we refer to as *genetic traits* or *attributes*; these digital individuals, in the language of EAs, are called *candidates* (or here sometimes *configurations*) to make a clear distinction from their biological counterparts. The environment in evolutionary algorithms is typically represented simply by an analytic objective- or target-function, $\mathcal{U}(\mathcal{X})$, which is to be optimized (this might be the internal energy of a system).

In contrast to biological systems, where a population also affects its environment, in most implementations of evolutionary algorithms the target function stays constant over time and is also not affected by the population in any way. This greatly simplifies the task of quantifying the degree of adaptation of a configuration to the environment via a so-called fitness function $F(\mathcal{X}) : \mathbb{R}^{N_A} \to \mathbb{R}$; the value of the fitness function of a candidate may directly be related to the value of the objective function, $F(\mathcal{X}) = F(\mathcal{U}(\mathcal{X}))$. A large value of the fitness function indicates that a candidate is a better solutions for minimizing the objective function compared to candidates with lower fitness values.

The two main ingredients which have inspired evolutionary algorithms are reproduction and mutation processes of biological systems: the former is usually referred to as *crossover* operation (or crossover move), where traits from several parent configurations of the most recent population are heuristically extracted and intercombined to create new offsprings, or candidate solutions. Some possible combinations of different traits of the genetic material of a particular pair of parent configurations may form a

better solution to the optimization problem as the parents themselves (or offsprings of other parents). Over time, the crossover operation of an evolutionary process aims at identifying the best possible combination of genetic material present in the population. With only a finite amount of genetic information being present in a population only a finite (yet potentially huge) number of combinations of different attributes or traits in the genetic material can be achieved. If the "correct" attributes, which would allow the formation of the global ground state via crossover operations, are not present in the population, the global optimum cannot be identified by a purely crossover-based evolutionary process. The purpose of the *mutation* operation in EAs is to explore the configuration space of candidate solutions more broadly for completely new combinations of genetic traits (*i.e.*, for new features and strategies which are, hopefully, favorable for the optimization process). Notably, this operation is of similar nature as random moves in stochastic optimization strategies, namely to jump between disconnected regions in the configuration space, an ingredient which is indispensable for global ground state search strategies. Mutation operations are, of course, only applied rather rarely as opposed to crossover moves, since an evolutionary process also need time to identify good offspring candidates via successive recombinations of the available genetic material.

A candidate solution generated that way is then evaluated by its fitness and either integrated into the population, if it has a sufficiently high fitness value, or discarded otherwise. Over time (and given enough computational resources) this artificial selection procedure will successively bring forth highly adapted candidates with respect to the optimization problem and will hopefully steer the solution towards the global optimum.

For many problems the convergence of such an algorithm governed only by crossover and mutation operators [205], is quite poor. For example, the recombination of two different molecular lattice configurations may frequently lead to offspring configurations with overlapping molecules. Such offsprings have large values of the internal energy, correspondingly low fitness values and are thus frequently rejected by the evolutionary selection process, which renders this approach quite inefficient. So-called *memetic* approaches circumvent this problem by combining global and local search strategies: while the global optimization strategy (here the EA) continuously – heuristically – explores the configuration space for different attraction basins, local optimization (often gradient descent algorithms) rapidly guides newly generated candidate solutions to the nearest local minimum [62, 139, 140].

The memetic evolutionary algorithm [86] which we further developed [62, 140, 205] and which we have heavily applied in this thesis can be outlined as follows (Fig. 2.4 depicts a flow-chart of this procedure):

0. Define a convergence criterion *(e.g., terminate after a maximum number of cycles of the evolutionary algorithm or terminate, if for a certain number of cycles no better solution as the current optimal solution was found, etc.).*

1. Initialize a population of candidate solutions at random or load a predefined one *(potentially from a previous optimization run).*

2. While the algorithm has not converged generate new offsprings from the current population via a *reproduction operation*, otherwise continue with step 7.
3. Locally optimize, *i.e.*, relax, the newly generated offsprings to their nearest local minimum in configuration space *(this is, by far, the most time-consuming task and can be performed in parallel).*
4. Evaluate the fitness of the newly relaxed candidate solutions and either include them into the population in case of sufficiently large fitness (while removing other configurations from the current population, keeping the population size constant), reject them otherwise.
5. In the case a new candidate was inserted into the population the updated population is stored (*i.e.*, exported) to a checkpoint file, which iteratively collects all populations encountered during the execution of the EA. *(Furthermore, the updated population is synchronized among all parallel processes.)*
6. Continue with step two.
7. Export the most recent population to the checkpoint file and export the best encountered candidate solution to the optimization problem, and terminate.

Over all, this memetic evolutionary algorithm may be quite time-consuming – especially the relaxation step represents, by far, the computationally most expensive task – but the overall convergence rate can easily be improved: the crossover and relaxation procedures for generating new candidate configurations can be performed in parallel.

Our implementation of the memetic evolutionary algorithm, which emerged throughout this thesis[17], is completely modular and flexible such that practical applications of the algorithm may be tailored problem-specifically (cf. Sections 3.1 and 3.2 and Ref. [213]). Below we explain several concepts and terms used in our EA implementation such as *population* and *genome representation* of candidate solutions. The *reproduction operation* is split up into several operations, which we all explain in detail below: the *mating*, *crossover*, *mutation*, *invasion* and *insertion*-operation; a flowchart of the entire *reproduction operation* is presented in Fig. 2.5. Furthermore, we discuss the *offspring relaxation* operation and introduce the functional form of the *fitness function* we used in this thesis. We then explain concepts such as *elitism*, *extinction*, *genetic diversity* and *nichening* and eventually provide some details on the inherent *parallelization* of our implementation of the memetic evolutionary algorithm.

Population and Genome Representation

The total number of N_{EA} different candidate configurations at a given iteration step of the EA is referred to as the (evolutionary) *population*, which is constantly exposed to concepts of natural (or, rather, artificial) selection.

The candidates of a population are commonly defined by a genome representation, $\mathcal{X}$, of the object which is subject to the evolutionary optimization [139, 205]; $\mathcal{X}$ usually

[17] Our memetic evolutionary algorithm implementation, introduced in Ref. [86], is available (on request) via Ref. [212] and is based on the ideas of Refs. [62, 140], which we further developed.

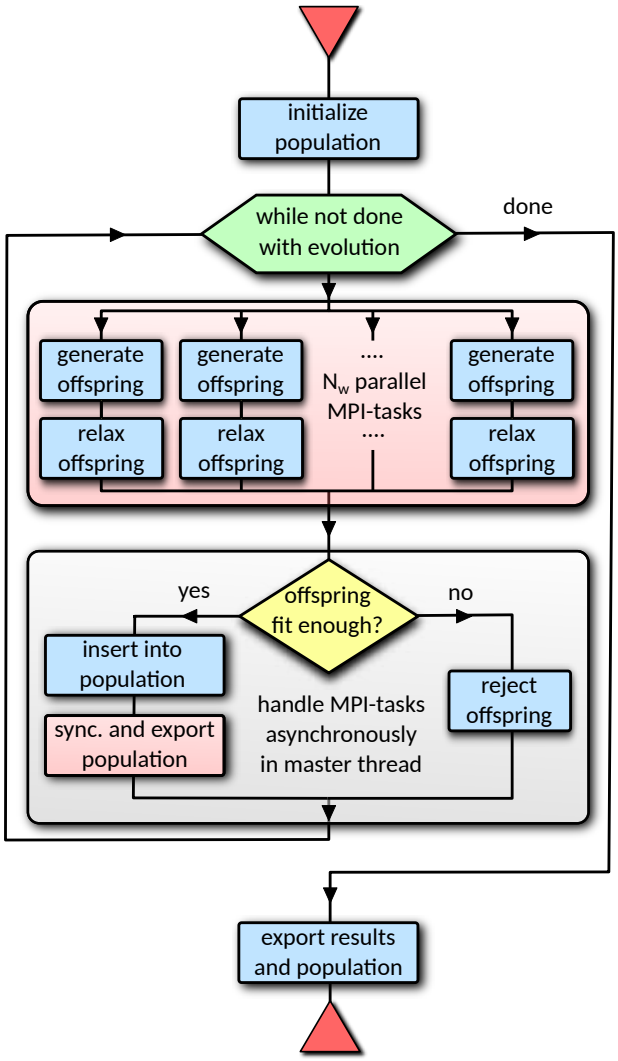


Figure 2.4: Flowchart of the memetic evolutionary algorithm starting at the top and ending at the bottom (see description in the text). Potentially parallelizable blocks (and associated synchronization operations) are colored by bright red background color: the reproduction operation (here labeled "generate offspring") and the subsequent local, gradient descent optimization (LO), here labeled "relax offspring", are performed in parallel on N_w worker tasks. The relaxed offsprings are collected by a "master" thread (see text below) which keeps a synchronous list of the most current population. After updates of the population occurred the updated population has to be synchronized amongst all idle worker processes; this is done asynchronously by the master thread. For parallelization we rely on the message passing interface (MPI) [214, 215] (see text below).

represents a structural configuration in this thesis, whose N_A degrees of freedom – to which we refer to as *genetic traits* or *attributes* – the EA can modify in order to minimize the objective function $\mathcal{U}(\mathcal{X}) \to \min$.

For instance, we might represent an arrangement of molecules within a lattice (cf. Section 3.2) phenotypically (rather than genotypically [205]) by the set $\mathcal{X} = \{\mathbf{R}^N, \mathbf{P}^N, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ of positions, $\mathbf{R}^N = \{\mathbf{R}_1, \ldots, \mathbf{R}_N\}$, and orientations, $\mathbf{P}^N = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$, of

all N molecules within a simulation box, which is spanned by the lattice vectors, $\mathbf{a}_1, \mathbf{a}_2$ and $\mathbf{a}_3$. In the example above the objective would be to find the optimal positions, orientations and lattice vectors which minimize the internal energy of the molecular lattice. This, of course, depends on the particular type of interaction of the molecules and on other system parameters (see Section 3.2). We would like to emphasize, however, that the EA described here is not restricted to this particular system.

Henceforward different candidate configurations are labeled by Latin indices via $\mathcal{X}_i$ and we define the population of candidate configurations as $\mathbf{X}^{N_{\text{EA}}} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{N_{\text{EA}}}\}$.

Reproduction Operation: Mating Operation, Pick Parents

In the first step of generating an offspring, two *parent* configurations, $\mathcal{X}_i$ and $\mathcal{X}_j$, are picked at random or via the "roulette wheel" method from the current population (see below) [62, 104, 105, 140, 216]; this strategy favors parents of high quality hence making them more likely to be used for reproduction than "weaker" configurations, *i.e.*, configurations from the population with less favorable values of the objective function $\mathcal{U}(\mathcal{X})$.

Reproduction Operation: Crossover Operation

After the two parents $\mathcal{X}_i$ and $\mathcal{X}_j$ were chosen, the two configurations are combined via a crossover operation (*i.e.*, a cut-and-splice process) creating thereby an offspring configuration, $\mathcal{X}_{i\oplus j}$, with the subscript '$i \oplus j$' emphasizing the executed crossover operation between $\mathcal{X}_i$ and $\mathcal{X}_j$. The purpose of this operation is to save high quality blocks of the genetic material (*e.g.*, the relative positions and orientations of molecules within the unit cell) in order to efficiently sample the parameter space [62, 104, 105, 139, 140, 205, 216].

The cut-and-splice process can be performed in several ways [62, 140, 205] but boils down to untangling the genomes of two (or in general N_P) parents into their attributes and mix different blocks of genetic material together in order to form a valid offspring candidate. In our implementation, the two parent genomes are cut at N_C randomly chosen positions (at the same positions for all parent genomes) partitioning their attributes into piecewise blocks, always assuring that attributes belonging to the same object (*e.g.*, position and orientation of one molecule for our example of a molecular lattice) belong to the same parent configuration. The partitioned genomes are then zipped together (*i.e.*, the first piece stems from parent one, the second from parent two, the third again from parent one[18], and so on). Special, attributes which can only be taken from one parent (such as the unit cell vectors at constant volume calculation of our molecular lattice example) can be labeled and are then only taken from one parent, *i.e.*, usually parent one.

The here introduced crossover operation has limited applicability to efficiently explore a configuration space for studying ground states arrangements of highly aspherical

[18] The crossover move between parents has been extended to N_P parents in our implementation of the EA, but we usually use $N_P = 2$.

particles or molecules at high packing fractions, (*i.e.*, if the combined volume of all particles or molecules is of the same or of comparable order as the available volume of the unit cell) [213]. The reason for this is, that in such cases with very high probability a crossover operation leads to unphysical overlaps between particles or molecules which is, as a consequence, associated with large numerical values of the internal energy of a configuration. To avoid this problem, we perform so called *ramp-compression*–moves [213] after a crossover operation of the EA: with a certain probability we scale up the volume V of the unit cell by a factor of s_i if the energy of a configuration after a crossover operation exceeds a certain threshold. The unit cell is then successively shrunk in N_i equidistant steps of $\Delta V = (s_i - 1)V/N_i$ until the original volume is restored. For every intermediate unit cell volume we iteratively perform a (small) number, N_r, of gradient descent based optimization steps for the positional and orientational variables of $\mathcal{X}_{i\oplus j}$. Thereby, we could efficiently favor non-overlapping configurations after crossover operations and reduce the rejection rate for offspring candidates[19] [213].

Reproduction Operation: Mutation Operation

The newly generated offspring configuration, $\mathcal{X}_{i\oplus j}$, is then exposed to random mutation moves: particular mutation operations of single attributes of $\mathcal{X}_{i\oplus j}$ need to be specified specifically for a particular optimization problem, either in predefined numerical boundaries of the different attributes or in a more specialized way.

In case of the above example of a molecular lattice, mutations are either translations or rotations of single molecules, swaps of center-of-mass positions or orientations of pairs of molecules or deformations of the unit cell. Each of these mutation moves are performed with a certain probability and within preset numerical boundaries [$\mathcal{X}_\mathrm{min}$, $\mathcal{X}_\mathrm{max}$].

This step of the algorithm has the purpose of exploring disconnected areas in configuration space, a feature which is indispensable in global minimization techniques.

Reproduction Operation: Invasion- and Injection-Operations

We usually perform independent optimization runs simultaneously for several, different sets of system parameters in this thesis. In Section 3.2 such system parameters might be the molecular density or the value of an external control parameters such as the electrostatic field strength. In a related project [213] on the self-assembly of ionic liquid crystals[20] [217, 218] the locations and values of the charges within the anisotropic particles, the aspect ratio and the packing fraction of the particles of ellipsoidal shape are the system parameters. For a particular problem, some of these

[19] Notably, *ramp-compression*-moves are not applied in this thesis but in a related project in the scope of R. Wanzenböck's diploma thesis [213] on the "self-assembly of ionic liquid crystals into smectic phases", *i.e.*, of ellipsoidal shaped, charged molecules, in three dimensions [217, 218].

[20] Ionic liquid crystals have attracted a steadily increasing interest during recent years both, in academic and in industrial research. With their anisotropic shape and their ability to carry charges they combine properties of charged particles and liquid crystals which are, for instance, reflected in their self-assembly capacities, making them technologically highly interesting [217, 218].

sets of parameters are very similar, such that we can expect that the respective evolutionary optimization processes lead to similar results (or at least similar, competing candidate solutions may emerge during the respective evolutionary processes). We define *compatible* sets of parameters as sets of parameters which exhibit the smallest available numerical difference – amongst all considered sets of parameters in different evolutionary optimization procedures – in all predefined "directions" of the parameter space; which particular parameters are considered to be relevant for identifying compatible sets of parameters is an input to the algorithm, *i.e.*, a hyper-parameter. The populations of evolutionary optimization procedures which use a compatible parameter set are labeled as *compatible populations*. To make use of the heuristically evaluated genetic traits of different, independent evolutionary optimization procedures, which use compatible sets of parameters, we introduce a so-called *invasion*-operation:

1. Choose one particular current compatible population – which represents a population of *alien* candidates to the current optimization procedure (with different, yet compatible system parameters).

2. Pick one high fitness candidate from the chosen compatible population according to the "roulette wheel" (or a suitable other) method; the fitness values of the alien candidates from the compatible evolutionary process are used for this purpose.

3. We label the particularly picked alien candidate as *invader configuration*.

The invasion-operation is performed with a certain probability, r_I, and, if applied, replaces the *mating* operation: the invader configuration replaces the offspring in the rest of the reproduction operation, see Fig. 2.5.

This move allows to more efficiently perform evolutionary optimization for a large grid of system parameters since good genetic information can diffuse throughout different parallel optimization runs of compatible (neighboring) sets of system parameter: different, compatible optimization procedures can make use of the heuristics of other compatible populations. In that way, the invasion-operation permits to easily and efficiently embed additional evolutionary optimization procedures for new sets of system parameter into a grid of already converged evolutionary optimization procedure which use different system parameters: a rather new, non-converged population can be quickly guided towards promising regions in the configuration space by employing frequent invasion operations. However, in general the invasion-operation needs to be applied with great caution, especially along phase boundaries in the parameter space, since successive invasions introduce a certain bias to the evolutionary process.

We also introduce a so-called *injection*-operation which allows us to bias an evolutionary ground state search towards selected regions in the configuration space (*i.e.*, towards specific candidate solutions with specific, selected traits). The injection-operation is performed with a certain probability, r_J, and, if applied, replaces the *mating* operation: an *injection population* of specially designed candidates is generated at the beginning of the EA, which are considered to be promising candidates for (or at least valid competitors to) the optimal solution of the optimization problem. The candidates of the injection population are usually designed via an educated guess

or are good solutions to the optimization problem suggested by other sources[21]. One designed candidate configuration is chosen at random from the injection population and replaces the offspring in the rest of the reproduction operation.

With such an injection-operation the convergence of the EA can be biased and analysed: for difficult problems it can be quite time-consuming for the EA to identify promising regions in configuration space starting from a random population. Furthermore, the EA can prematurely converge towards a local minimum of the optimization problem (and *trap* itself there). Via injecting specially designed candidate solutions both, an initially random or a trapped population can intentionally be steered towards other regions in the configuration space. On the other hand, if the EA has, indeed, identified the global optimum, the injection-operation can be used to challenge the solution of the EA.

Offspring Relaxation: Local Optimization

After the *crossover* and *mutation* steps, and assuming the offspring configuration, $\mathcal{X}_{i\oplus j}$, does not yet represent a local minimum with respect to the objective function, a *local, gradient-descent based optimization* (LO) is performed. For such LOs, we mainly rely on the implementations of gradient descent algorithms such as the sequential least squares programming [169] *(SLSQP)* gradient-descent algorithm or the limited memory [170] Boyden-Fletcher-Goldfarb-Shanno gradient-descent algorithm [171–174] in its bounded variant [175] *(L-BFGS-B)* from the open-source software package "*SciPy*" [176]. These algorithms allow us to define numerical boundaries and constraints of the attributes of the genomes during a gradient descent optimization.

In our molecular lattice example, the LO operation has the purpose of minimizing the forces and torques between the molecules and the stress of the unit cell. The above mentioned constraint gradient descent optimizers are very helpful numerical tools, for instance, to keep the unit cell volume fixed for restricted volume calculations and to prevent re-orientations of the molecules which would cause some of their atomic constituents to be transferred into positions outside of the slab geometry (see Section 3.2).

Subsequently we perform several *basin-dropping* (BD) steps, where we further try to lower the energy of the candidate configuration by applying several small random moves to the attributes of the locally optimized or, equivalently, relaxed offspring; from the emerging configurations only the ones improving the value of the objective function are accepted. The BD operation turned out to considerably improve the convergence rate of the local optimization, in particular if multiple and alternating sequences of LO and BD runs are applied.

[21] Often, such designed candidates can be identified via studying related problems (in literature) or by identifying special symmetries of the optimization problem which favors special candidate solutions over others.

Artificial Selection: Survival of the Fittest

After the local search procedure the optimized offspring configuration, $\mathcal{X}_{i\oplus j}$, becomes a new candidate to enter the evolutionary population, $\mathbf{X}^{N_{\mathrm{EA}}}$. The objective of the EA is to retain the best configurations (*i.e.*, the most favorable one with respect to the objective function $\mathcal{U}(\mathcal{X})$) within the population and to include only candidates with values of the objective function, $U_i = \mathcal{U}(\mathcal{X}_i)$, better or comparable to those of the current population[22]. In an effort to quantify the quality of the candidates, their so-called fitness is evaluated [62, 104, 105, 140, 205, 216], for which we have used in this thesis the function:

$$F(U) = \exp\left(-s\frac{U - U_{\min}}{U_{\max} - U_{\min}}\right); \tag{2.57}$$

$F(U)$ is a monotonic function of the evaluated objective function, U, of the candidates, whose value ranges within the interval $0 \leq F(U) \leq F(U_{\min}) = 1$; $U_{\min}$ and $U_{\max}$ are the minimal and maximal values of the objective functions realized in the current population. The selection parameter, s, quantifies the reproduction-rate for configurations within the population in the sense that large values of s tend to exclude configurations with low fitness from reproduction; following Ref. [62] we commonly use $s = 3$.

The aforementioned "roulette wheel" method for choosing suitable parent configurations also relies on the fitness function (and hence the selection parameter): assuming that the configurations within the population $\mathbf{X}^{N_{\mathrm{EA}}}$ are sorted by their respective fitness values in descending order, $F(U_i) > F(U_{i+1})$, the probability, $f(U_i)$, of a configuration, $\mathcal{X}_i$, to be selected for reproduction is given in terms of the relative fitness [62, 104, 140, 216]:

$$f(U_i) = \sum_{j=i}^{N_{\mathrm{EA}}} F(U_j) \cdot \left[\sum_{k=1}^{N_{\mathrm{EA}}}\sum_{l=k}^{N_{\mathrm{EA}}} F(U_l)\right]^{-1}, \tag{2.58}$$

N_{EA} being the total number of configurations within the population.

With a certain probability (commonly in 20% of all crossover moves) we also allow reproduction between randomly chosen configurations.

Elitism and Extinction

Once a new configuration is accepted to enter the population another configuration has to be eliminated. The probability $p(U_i)$ for a configuration, $\mathcal{X}_i$, to be eliminated is given by

[22] As soon as a sufficient number of candidates are exchanged within the population, the later is stored in a checkpoint file with a unique generation label; any generation may be used to initialize a population in subsequent optimization runs.

$$p(U_i) = \exp\left[-sF(U_i)\right] \left[\sum_{j=1}^{N_{\mathrm{EA}}} \exp\left[-sF(U_j)\right]\right]^{-1}, \tag{2.59}$$

a value which is again related to the fitness of the configuration, $F(U_i)$, and the selection parameter s. Thus, configurations with low fitness are more likely to be eliminated.

A certain number, N_E, of the best configurations within the population are retained in an effort to keep the so far best solutions as appropriate parent candidates for the above-mentioned crossover procedures (a strategy referred to in the literature as *elitism* [140]).

It should be emphasized that this strategy does not follow biological selection mechanisms [211], where populations are replaced entirely once that new generations have been formed; however, our strategy ensures to protect the best genetic material from extinction during the entire search procedure [62, 140].

Maintaining Diversity: Nichening

During evolutionary optimization processes it can happen that very few, highly adapted (in our case structurally identical) candidates "overrun" the population and thereby bring the evolutionary process effectively to an end. In an effort to prevent this from happening we aim at maintain the *genetic diversity* within the population, *i.e.*, we aim at maintaining qualitatively different candidates which feature different, yet potentially promising genetic traits for maximizing the fitness value of the objective function. We achieve this by additionally applying a so-called *nichening* operation [112] in the artificial selection process of the EA: locally optimized offspring configurations will be discarded if the values of their objective function are too close to values of any other configuration in the current population.

However, this procedure alone cannot cope with "degenerate" configurations, *i.e.*, if qualitatively different configurations exhibit essentially the same values of the objective function (within the specified nichening tolerance). In our approach we allow configurations to enter the population only if their genetic traits differ significantly from those of the competing, degenerate configurations (with similar values of the objective function). In order to quantify the difference between the traits of configurations we associate a feature vector, $\mathbf{x}_i$, which collects a set of order parameters pertaining to configuration $\mathcal{X}_i$ (see Section 2.3 for details for lattice configurations). The degree of similarity between two configurations, $\mathcal{X}_i$ and $\mathcal{X}_j$, is then evaluated by taking the Euclidean distance between the corresponding feature vectors, *i.e.*, $\Delta_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ (other measures are also possible); similar configurations will have a small distance, while unlike configurations will have a large distance. If Δ_{ij} is above a certain threshold value D_f, which is a parameter to the algorithm and depends on the magnitude of the involved order parameters, the offspring configuration, $\mathcal{X}_{i\oplus j}$, will not be discarded by the objective function nichening operation.

Parallelization

In every cycle of the EA a new configuration, *i.e.*, an *offspring*, is created from existing configurations of the most recent population, via a reproduction operation. This new configuration is then subjected to an LO, an operation which represents by far the most time-consuming task in our algorithm and is therefore performed in parallel using the "*mpi4py*" framework [214, 215, 219]. For an optimal load-balance we additionally spawn a "master" thread on one of the MPI-processes, which asynchronously distributes reproduction and optimization tasks of offspring configurations among all idle MPI-processes. The task of the master thread is represented by the entire flowchart depicted in Fig. 2.4; the sub-task of the worker processes (*i.e.*, the total number of N_w MPI-processes which perform the main computational "workloads" of the EA) are also indicated in Fig. 2.4.

The relaxed configurations are gathered by the master thread, which keeps a synchronous list of the most current population. The master thread then decides – via criteria based on the fitness values of the configurations and the nichening operations with respect to the current population – whether the new relaxed candidates are accepted or rejected. In case of acceptance, the elitism and extinction operations are applied and updates to the current population are made. The updated population is stored into a checkpoint file (*i.e.*, into separate files for every set of system parameters of the evolutionary optimization procedure), which iteratively collects all encountered populations during the evolutionary process. The updated population is then asynchronously distributed amongst all idle MPI-processes.

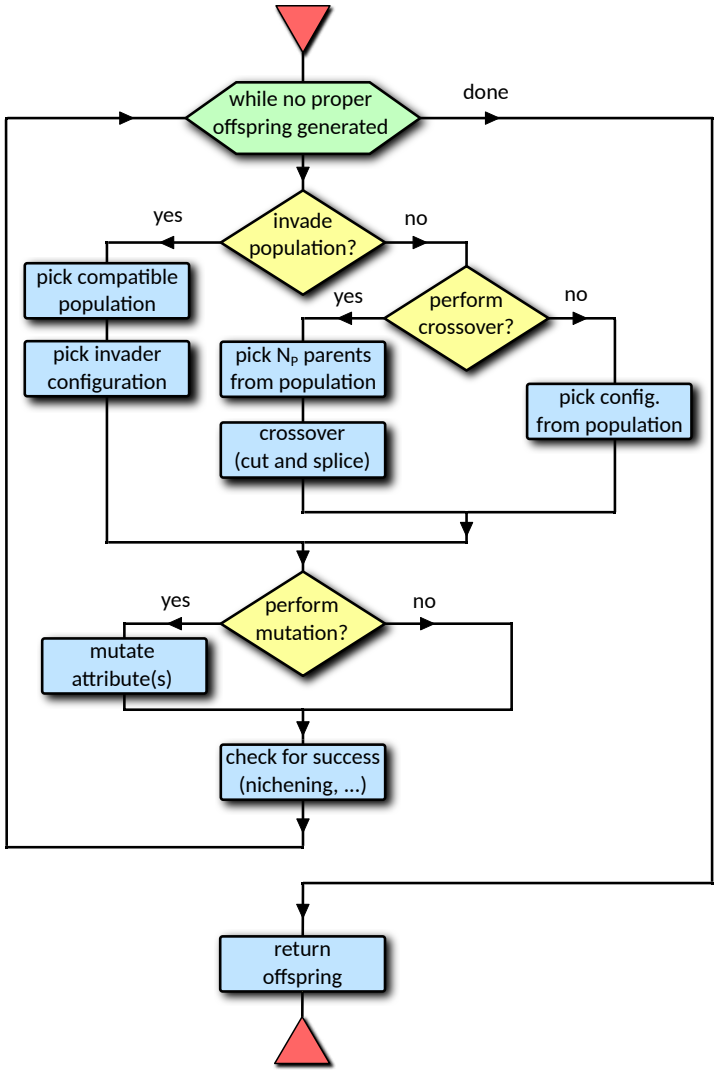


Figure 2.5: Flowchart of the *reproduction-operation* in our (memetic) evolutionary algorithm starting at the top and returning an offspring candidate at the bottom. Until a proper offspring is generated (*i.e.*, an offspring which satisfies the nichening criterion and whose fitness is reasonably large), the following steps are iteratively applied according to the flowchart: first, it is randomly decided if an invasion operation is applied (with a rate r_I, which we typically set $< 1\%$) or, otherwise, if a regular crossover operation is (potentially) performed. In case of being in the "yes" branch of the *invade population* decision, an invasion operation is applied, see text (here also an insertion operation can be performed, not shown here). In case of being in the "no" branch of the *invade population* decision, *i.e.*, in the crossover branch, it is randomly decided (with a rate r_C, which we typically set $\approx 80\%$) whether the crossover operation is applied or if a single candidate is directly picked from the current population as offspring. In case of being in the "yes" branch of the *perform crossover* decision, first, N_P parents (typically two) are picked from the population, following the mating operation, which are then subject to the crossover operation in order to generate an offspring. The offspring outcome after the *invade population* decision, *i.e.*, either an invader configuration or a generated or single offspring, is then potentially (with a rate r_M, which we typically set $\approx 20\%$) subject to mutations of its attributes, *i.e.*, to mutations of the variables of the offspring candidate. Notably, after the *perform mutation* decision we check if the (potentially mutated) offspring fulfills the nichening criterion (and if its fitness is reasonably large). If so, the offspring is considered a *proper* offspring and the reproduction operation is considered to be successful (otherwise the above steps are repeated). A successfully generated offspring is eventually returned.

2.3 Order Parameters of Structural Data

Order parameters are essential tools when it comes to understand structures, especially when categorizing many of them into a phase diagram. Order parameters are functions that map structural data, $\mathcal{X}$, such as coordinates, orientations, etc., to a numerical value[23], $\Psi(\mathcal{X})$. Such a function usually quantifies a particular "fingerprint" in structural data which allows us to relate a structure to an entire family of structures (such as crystalline structures of specific symmetries). Typically, $\mathcal{X}$ can be related to a certain family of structures if the value of the related order parameter is large, *i.e.*, usually close to unity, $\Psi(\mathcal{X}) \approx 1$. On the other hand, low values of an order parameter suggest that $\mathcal{X}$ can be distinguished from the related structural family. A set of order parameters, thus, may provide us with essential information about the symmetries in structural data.

Order parameters, from a thermodynamic point of view, are quantities which specify the type of transition between two thermodynamic phases. Changes of the structure, induced by changing external system parameters, are reflected in changes of the respective (representative) order parameters, which eventually vanish if the order is lost. In case of a discontinuous jump of an order parameter at the transition point we speak of a first order transition, in case of a continuous change between two phases the transition is of second order [220–222].

In this thesis, we mostly quantify the order of structural data by means of the local neighborhood of the constituents of ordered ground state structures. Therefore we start this Section by introducing the concept of local neighborhoods on the basis of Voronoi constructions in Subsection 2.3.1. We then define bond orientational order parameters in Subsection 2.3.2 and orientational order parameters in Subsection 2.3.3. Further, we introduce the concept of diffraction patterns, structure factors and reciprocal lattices in Subsection 2.3.4 in order to quantify characteristic length scales and global rotational symmetries of crystalline structures.

2.3.1 Local Environments and Neighbors: Voronoi Construction

The evaluation of local order parameters strongly depends on the method on how to identify neighbors. In this thesis, we use the well-defined method of Voronoi construction [223, 224]. We are mostly concerned with particles or molecules in quasi–two-dimensional slab or bilayer geometries. We therefore present the basic concept of Voronoi construction in two dimensions (2D) below.

A Voronoi construction describes the partitioning of a space filled with a finite number of objects (also called seeds, vertices or generators) into non-overlapping polygons without leaving voids; each polygon of a Voronoi construction defines the local, "closest" environment of an object. To be more specific, a Voronoi construction defines a tessellation of a space into Voronoi cells around these objects such that all points within such a cell are closer[24] to the central object than to any other vertex [223, 224].

[23] Usually – or often – the value of order parameters are normalized $\Psi(\mathcal{X}) \in [0, 1]$.
[24] Usually, Euclidean distance metrices are used.

In practice, Voronoi constructions are generated with the help of Delaunay triangulation – which is possible in any dimensions – and we present the algorithm in 2D (the bullet points correspond to panels (a) to (c) in Fig. 2.6):

(a) For each vertex i, *i.e.*, for each particle position $\mathbf{r}_i$, we draw connections to all other vertices, $\mathbf{r}_j$.

(b) These connections are then "cut in half" by straight lines, $\mathbf{g}_{ij}$, satisfying $\mathbf{g}_{ij} : (\mathbf{r}_i - \frac{1}{2}(\mathbf{r}_i + \mathbf{r}_j)) \cdot \hat{\mathbf{r}}_{ij}^{\perp} = 0$, where $\hat{\mathbf{r}}_{ij}^{\perp}$ is the unit vector orthogonal to $\mathbf{r}_j - \mathbf{r}_i$, thereby drawing a convex hull around the centering particle, i. The so-called partitioning lines, $\mathbf{g}_{ij}$, thus contain all points which are equidistant to both particles i and j.

(c) The Voronoi cell of a vertex, i, is defined as the smallest polygon built by these partitioning lines, $\mathbf{g}_{ij}$, around the vertex; the corners of this polygon define the vertices of the Delaunay-triangulation. In other words, the Voronoi-cell of vertex i is the region in the vicinity of $\mathbf{r}_i$ that is void of all partitioning lines, $\mathbf{g}_{ij}$.

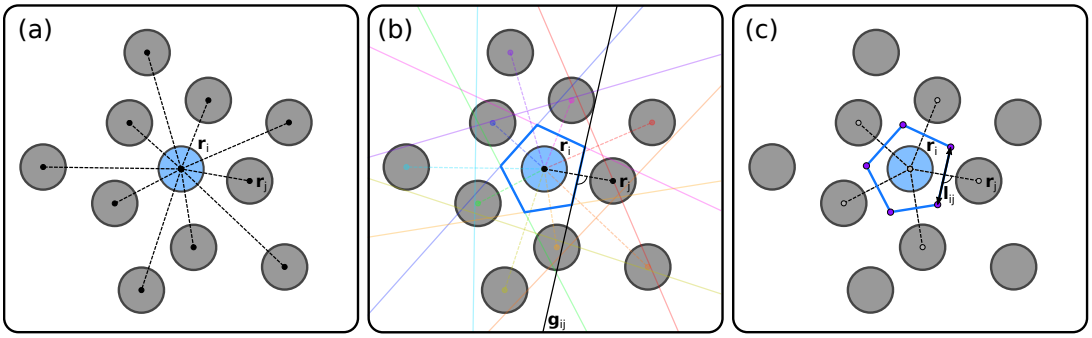


Figure 2.6: Schematic illustration of the Voronoi construction using Delaunay-triangulation following the algorithm described in the text. (a): Drawing connection from a central vertex i to all other vertices j (dotted lines). (b): Drawing partitioning lines, $\mathbf{g}_{ij}$, orthogonal to connection lines $\mathbf{r}_j - \mathbf{r}_i$, *i.e.*, solid, color-coded lines for each vertex connection (dotted lines). The smallest polygon surrounding the central vertex, *i.e.*, its Voronoi cell, is emphasized by thick blue lines. (c): Nearest neighbors to vertex i, according to Voronoi construction, are connected with dashed lines, the vertices of the Delaunay-triangulation are emphasized by small filled circles at the corners of the Voronoi cell (thick blue polygon).

Vertices, or particles, i and j, that share a side, l_{ij}, of a Voronoi polygon (or the area of a polyhedron in higher dimensions) are considered the nearest neighbors of the Voronoi construction (cf. connected vertices in Fig. 2.6(c)).

Many open source software packages [169] provide faithful and numerically optimized implementations of the Voronoi tessellation scheme explained above for any dimensions.

2.3.2 Bond Orientational Order Parameters

In this thesis, we mostly employ order parameters which describe global properties (or symmetries) of an ordered structure in two dimensions based on the local proximity of all of its atomic or molecular constituents. These parameters can be expressed as a

sum (or the average) over local order parameters ψ_i, calculated for all N particles in the system, *i.e.*,

$$\Psi_\nu(\mathcal{X}) \sim \sum_{i=1}^{N} \psi_i(\mathbf{r}_i, \mathbf{r}^{\mathcal{N}_i}; \mathbf{u}_i, \mathbf{u}^{\mathcal{N}_i}; \nu); \tag{2.60}$$

where, here, $\mathcal{X} = \{\mathbf{r}^N, \mathbf{u}^N\}$ represents the set of particle positions, $\mathbf{r}^N = \{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$, and orientations, $\mathbf{u}^N = \{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ of a particular structure, the vectors $\mathbf{r}_i$ and $\mathbf{u}_i$ specify the position and the orientation of particle i, while $\mathbf{r}^{\mathcal{N}_i}$ and $\mathbf{u}^{\mathcal{N}_i}$ are the sets of positions and orientations of the neighboring particles of particle i, respectively; $\mathcal{N}_i$ is the number of neighbors[25] of particle i. The role of the parameter ν will be specified below.

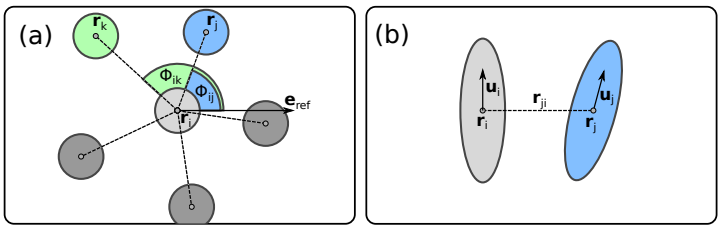


Figure 2.7: (a) Schematic representation of ingredients of bond-orientational order parameters defined by Eqs. (2.61) and (2.62) and (b) of orientational order parameters given by Eqs. (2.63) and (2.64); polygon sides, l_{ij}, are shown in Fig. 2.6(c).

We define the so-called bond orientational order parameters (BOOPs), $\Psi_\nu(\mathcal{X})$, as defined in Refs. [225, 226] and revisited in Ref. [227], as

$$\Psi_\nu(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \exp[\imath\nu\phi_{ij}] \right|; \tag{2.61}$$

who depend only on the angles, ϕ_{ij} (with $\cos\phi_{ij} = \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{e}}_{\mathrm{ref}}$), which is enclosed between the bonds of a central particle, i, to each of its $\mathcal{N}_i$ neighbors. In Eq. (2.61) we introduced the vector $\hat{\mathbf{r}}_{ij} = (\mathbf{r}_j - \mathbf{r}_i)/|\mathbf{r}_j - \mathbf{r}_i|$, *i.e.*, the unit vector between two neighboring particles i and j, the reference axis $\hat{\mathbf{e}}_{\mathrm{ref}}$ (which is of unit length), and the complex unity $\imath$, $\imath^2 = -1$; the situation is depicted in Fig. 2.7(a).

The orientational symmetry, which is related to the mean coordination number, is quantified by the (integer) variable ν: the ν-fold bond orientational order parameter, $\Psi_\nu(\mathcal{X})$, assumes the value one if the angles between neighbors are multiples of $2\pi/\nu$ and attains values close to zero for a disordered particle arrangement or if the ν-fold symmetry is not present.

The numerical value of $\Psi_\nu(\mathcal{X})$ defined by Eq. (2.61) strongly depends on the number of neighbors $\mathcal{N}_i$ and on the complex phase, $\imath\nu\phi_{ij}$, between neighboring particles i and j. Even tiny numerical deviations of the particle positions from an ideal configuration can

[25] Neighbors might be defined according to the Voronoi construction described in Fig. 2.6.

change the number of neighbors and the associated complex phases between neighbors in a way that strongly affects the value of $\Psi_\nu(\mathcal{X})$ [62]. In this thesis, the lattice structures we are dealing with are never perfect and the number of nearest neighbors can differ from the ideal value. In an effort to guarantee better numerical stability in the evaluations of bond orientational order parameters of numerically imperfect configurations we use a modification of the POOPs defined by Eq. (2.61), which was proposed in Ref. [228]: this modified definition includes a weighting factor which is related to the polygon side length, l_{ij}, that neighboring particles share

$$\Psi_\nu(\mathcal{X}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{L_i}\sum_{j=1}^{\mathcal{N}_i} l_{ij}\exp[\imath\nu\phi_{ij}]\right|; \tag{2.62}$$

with $L_i = \sum_{j=1}^{\mathcal{N}_i} l_{ij}$; the polygon side lengths, l_{ij}, are extracted from the Voronoi construction, cf. Fig. 2.6(c).

We usually drop the argument, $\mathcal{X}$, for both, the bond-orientational order parameter defined by Eq. (2.61) and for the orientational order parameters defined by Eq. (2.62), and write $\Psi_\nu \equiv \Psi_\nu(\mathcal{X})$, unless indicating a particular realization of a structure, $\mathcal{X}_i$, is explicitly necessary for the discussion.

2.3.3 Orientational Order Parameters

In case of molecular configurations one often deals with anisotropic shapes and interactions; it is therefore useful to quantify the orientational order of molecular configurations. Similar to Eq. (2.62) we can quantify global *orientational* order including again the above Voronoi nearest-neighbor construction in the following way:

$$\beta = \frac{1}{2N}\sum_{i=1}^{N}\frac{1}{L_i}\sum_{j=1}^{\mathcal{N}_i} l_{ij}\left|\hat{\mathbf{u}}_i\cdot\hat{\mathbf{u}}_j\right|; \tag{2.63}$$

here $\hat{\mathbf{u}}_i$ is a unit-vector which specifies the orientation of particle i, see Fig. 2.7(b).

Finally, we can combine orientational order with positional degrees of freedom, using the unit vector $\hat{\mathbf{r}}_{ij}$ between two neighboring particles i and j:

$$\alpha = \frac{1}{2N}\sum_{i=1}^{N}\frac{1}{L_i}\sum_{j=1}^{\mathcal{N}_i} l_{ij}\left|(\hat{\mathbf{u}}_i\cdot\hat{\mathbf{r}}_{ij})^2 + (\hat{\mathbf{u}}_j\cdot\hat{\mathbf{r}}_{ij})^2\right|, \tag{2.64}$$

suggesting again a modified version of the order parameters with the side lengths of the Voronoi polygons, l_{ij} [229]; $\hat{\mathbf{u}}_\mathbf{i}$ is again a unit-vector defining the orientation of a particle in the lab-frame.

2.3.4 Global Symmetries: Reciprocal Lattice and Structure Factor

X-ray, electron or neutron diffraction experiments are important tools in order to investigate the microscopic structure of materials, be it of crystalline, quasicrystalline or

of disordered nature. Here we are concerned with the static structure factor, which is an important mathematical description of how a material scatters incident radiation. Further, we are only concerned with crystalline structures, *i.e.*, a spatially periodic lattice of copies of a unit cell (which is spanned by the lattice vectors $\mathbf{a}_i$) with N atoms or molecules in its basis. We define the lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ in three dimensions as

$$\mathbf{a}_1 = (a_{11}, 0, 0), \tag{2.65}$$
$$\mathbf{a}_2 = (a_{21}, a_{22}, 0) \text{ and} \tag{2.66}$$
$$\mathbf{a}_3 = (a_{31}, a_{32}, a_{33}). \tag{2.67}$$

In diffraction experiments an incident beam is scattered at scattering centers (*i.e.*, particles) within the lattice; constructive (or destructive) interference of the scattered waves from different scattering centers form a scattering pattern. We assume weak[26] and elastic scattering[27] with an incident wave vector $\mathbf{k}_I$ and a scattered wave vector $\mathbf{k}_S$, whose wave length is given by $\lambda = 2\pi/|\mathbf{k}_I| = 2\pi/|\mathbf{k}_S|$ with $|\mathbf{k}_S - \mathbf{k}_I| = \frac{4\pi}{\lambda}\sin(\theta)$, where θ is the angle between incident and scattered wave vectors [220, 230].

According to Bragg's law [231], *i.e.*, $2d\sin(\theta) = n\lambda$, constructive interference of scattered beams only occurs at angles θ, where the phase-shift between beams that are scattered at different, parallel planes, d, of the periodic scattering centers is an integer multiple of the wave length, λ.

The reciprocal lattice vectors, $\mathbf{a}^*{}_i$, are defined through the condition $\mathbf{a}_i \cdot \mathbf{a}^*{}_j = 2\pi\delta_{ij}$, with $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. For a three-dimensional lattice the reciprocal lattice vectors can be given by

$$\mathbf{a}_1^* = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} = \frac{2\pi}{V}(\mathbf{a}_2 \times \mathbf{a}_3), \tag{2.68}$$
$$\mathbf{a}_2^* = \frac{2\pi}{V}(\mathbf{a}_3 \times \mathbf{a}_1) \text{ and} \tag{2.69}$$
$$\mathbf{a}_3^* = \frac{2\pi}{V}(\mathbf{a}_1 \times \mathbf{a}_2), \tag{2.70}$$

where $V = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = a_{11}\, a_{22}\, a_{33}$ is the unit cell volume; $\mathbf{a} \times \mathbf{b}$ denotes the cross product and $\mathbf{a} \cdot \mathbf{b}$ the inner product between two vectors $\mathbf{a}$ and $\mathbf{b}$, respectively.

The integer Miller indices $\{h, k, l\}$, which define all possible parallel planes, d_{hkl}, in the real space lattice (cf. Bragg's law above) also define the lattice points, $(h\mathbf{a}_1^*, k\mathbf{a}_2^*, l\mathbf{a}_3^*)$, of the reciprocal lattice. In that way, the reciprocal lattice determines the positions (*i.e.*, the angles θ) of constructively interfering diffracted beams $\mathbf{k}_I = (h\mathbf{a}_1^*, k\mathbf{a}_2^*, l\mathbf{a}_3^*)$ by Bragg's law.

A perfect crystal is mathematically defined by a unit cell, *i.e.*, by the lattice, and the atomic or molecular basis, *i.e.*, the arrangements of particles within this unit cell.

[26] The amplitude of the incident beam is assumed to be constant throughout the sample and no multiple scattering shall occur.

[27] Incident wave vector $\mathbf{k}_I$ and scattered wave vector $\mathbf{k}_S$ have the same wave length, $|\mathbf{k}_I| = |\mathbf{k}_S|$.

Analogously, a reciprocal crystal – which is actually observed in scattering experiments – is composed of a reciprocal lattice and a structure factor, the latter being defined by

$$F_{hkl} = \sum_{i=1}^{N} f_i \exp\left(-2\pi\imath(hx_i + ky_i + lz_i)\right) \tag{2.71}$$

where N is the number of particles in the primitive unit cell, x_i, y_i and $z_i \in [0, 1]$ are the coordinates within the unit cell and f_i is the form-factor (which we here usually set one) of particle i. This expression relates the amplitude and phase of the incident beam, $\mathbf{k}_I$, diffracted by the $\{h, k, l\}$ planes of the crystal to that produced by a single scattering center, i, of the unit cell [220, 230].

2.4 Categorizing Structural Data with Unsupervised Clustering Algorithms

Scientists and engineers nowadays are often confronted with huge *data sets*, may it be images or written text of any kind, scattering data from particle detectors, geometric data of lattice structures (*i.e.*, particle arrangements) generated by computer experiments or order parameters of such particle configurations [119]. Analysing such data sets is usually a task far from being trivial, especially in high-dimensional spaces: a data set, $\mathbf{X} = \{\mathbf{x}_1, \dots \mathbf{x}_N\}$, consists of $i = 1, \dots, N$ data elements (or, equivalently, *data points*), $\mathbf{x}_i = \{x_1, \dots, x_{N_f}\}$. Each data element, $\mathbf{x}_i$, also referred to as *feature vector*, may contain a large number, $j = 1, \dots, N_f$, of so-called *features*, x_j, such as the values of different pixels of an image, the different channels of the measurement of a particle collision or the coordinates, orientations and/or order parameters of a particle arrangement of a (lattice) structure. In certain circumstances a few clear signals in the data (*i.e.*, a few characteristic features, x_j, in the feature vectors, $\mathbf{x}_i$, of a data set $\mathbf{X}$) may permit us to categorize the elements of a data set (e.g. into different structural families with clear, characteristic order parameters). However, the shear size of typical data sets and the often immense complexity of the involved features usually render a classification scheme intractable to be manually carried out by a human being. In such cases, and to obtain a more comprehensive picture of the properties of the underlying data in general, one can turn to methods from unsupervised machine learning [119, 232].

In this thesis, we are mostly concerned with organizing large sets of lattice structures, which are generated by different computer experiments for a variety of system parameters (e.g. with algorithms described in Section 2.2), into archetypical families of structures (*i.e.*, into different categories). To achieve this, we rely on a set of order parameters[28,29] (cf. Section 2.3) which we evaluate for every structure and which we then collect in a corresponding data set of order parameters; every element in the order parameter data set represents a feature vector which is related to one specific lattice structure from the computer experiments. The elements, or data points of such an order parameter data set may be arranged in the N_f-dimensional so-called *feature space*, spanned by the N_f features, in a certain geometric way, such that spatially separated clusters (in the different directions of the feature space) of similar data elements[30] may be identified. Methods from unsupervised machine learning [119] can be used to analyse a data set of feature vectors (or of order parameters in our case) for certain similarity measures in the features which may permit us to algorithmically organize the elements of the data set into an initially unknown set of categories [232].

We use maybe simplest (and therefore maybe the most applied) form of unsupervised machine learning [119], namely *clustering algorithms* [119, 235, 236], in order to organize data sets of lattice structures into families of structures. In the language of

[28] Of course, a set of order parameters needs to be defined specifically for a particular problem.

[29] Other techniques may even directly operate on the structural data (*i.e.*, on a data set of coordinates of the particles), see for instance Refs. [233, 234].

[30] One specific cluster is considered to contain similar data elements. See Ref. [232] for an in depth discussion on different, problem specific similarity measures in data science problems.

clustering algorithms, the procedure of categorizing data elements via similarity measures of data points in a feature space into different clusters is usually denoted as *clustering* or *labeling*: each of the $i = 1, \ldots, N$ elements of a data set is labeled by an identifier, k_c, which assigns each element i to one of the $c = 1, \ldots, N_c$ categories (or *clusters*) identified by the clustering algorithm[31].

There exists a vast, ever growing number of clustering algorithms and the decision for choosing a suitable method may depend on several issues [119]:

- The intrinsic geometries forced upon the feature space of a data set by the particular choice of features[32] as well as the specific geometric shapes of clusters in the feature space of a particular data set may influence the outcome of a clustering algorithm[33] [119].
- The size, N, of the data set and the dimensionality, N_f, of the feature space may influence the choice of clustering algorithm used in practice (from a conceptual point of view).
- The data set size and feature space dimensionality are also relevant when considering the time complexity of different clustering algorithms (with respect to N and N_f) and the available resource when applying a specific method.

Depending on the intrinsic properties of the available features and on the particular realization of the data elements of a data sets some algorithms might be favorable over others [119, 232].

Many different clustering algorithms (and many other machine learning methods) are implemented in the *Scikit-Learn* Python package [237] which also provides great tutorials and literature on that topic. A very instructive review on such useful tools and many other helpful machine learning applications in physics or chemistry can be found in Ref. [119].

This Section is organized as follows: in Subsection 2.4.1 we first present some ideas form the literature how to identify and capture characteristic features in a data set and how to reduce the dimensionality, N_f, of the feature space to an N_l-dimensional (representative) *latent space* with $N_l \leq N_f$, using principal component analysis [238]. We then focus on two heavily used standard clustering algorithms, namely *k-means* clustering [239–242] in Subsection 2.4.2 and *density-based* (DB) clustering [243] in Subsection 2.4.3, which both rely on the Euclidean distance as a measure for similarity of data points in the feature space. In Subsection 2.4.4 we introduce *t-distributed stochastic embedding* (t-SNE) [244], a clustering algorithm which does not rely on Euclidean distance measures. In Subsection 2.4.5 we introduce the concept of (adjusted) mutual

[31] Depending on the particularly applied clustering algorithm the number of clusters, N_c, may be a preset parameter to the algorithm or may even be identified by the algorithm during execution.

[32] For instance, to identify spatially separated, spherical clusters of data elements in an N_f dimensional feature space might be conceptually different from identifying clusters of data elements which are arranged on the surface of spheres with different radii but with a common center (in the same feature space).

[33] Different clusters may be of different geometric shape, of different spatial extent, may be separated or may overlap either due to intrinsic properties of the features (which might not be capable of distinguishing between different categories) or due to noise in the features of the data set.

information [245, 246], which quantifies the agreement of different clusterings of the same data set via information theoretic measures.

2.4.1 Dimensional Reduction with Principal Components Analysis

Introducing order parameters (or features) of a system often is not an easy task. Furthermore, interpreting features, especially in high-dimensional spaces, spanned by the number of data elements, N, and the dimensionality, N_l, of the feature space, quickly becomes very difficult. Working in such high, $N \times N_f$-dimensional spaces – with often more than hundred-thousands or millions of data elements and more than thousands of features per data element – can be very challenging and impractical. If all data in this data sets were relevant, *i.e.*, if the information necessary for categorizing the data was uniformly distributed across all captured features, we would have a hard time identifying meaningful results; this is often referred to as the curse of dimensionality. Luckily, in real-world physical problems a vast number of features of raw or processed data may indeed be redundant or simply irrelevant to the problem at hand. In statistical physics, for example, macroscopic systems with a huge number of microscopic degrees of freedom can often be described by very few order parameters such as temperature, magnetisation, density, etc.; a few characteristic, collective quantities, which are based on ensembles of microscopic states of a system, are often sufficient to describe a macroscopic physical system [119].

In this Subsection we present some ideas form the literature on how to identify and capture characteristic features in an arbitrary data set and how to reduce the dimensionality of the N_f-dimensional feature space of a data set to an N_l-dimensional (representative) *latent space* using *principal component analysis* [238] (PCA).

In general, dimensional reduction describes the process of transforming a data set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \in \mathbb{R}^{N \times N_f}$ of N data elements, $\mathbf{x}_i \in \mathbb{R}^{N_f}$, represented in a high, N_f-dimensional feature space, via the operator $\mathbb{P}$ to a low-dimensional latent space representation of the data, $\mathbf{L} = \mathbb{P}(\mathbf{X}) = \{\mathbf{l}_1, \ldots, \mathbf{l}_N\} \in \mathbb{R}^{N \times N_l}$, with $\mathbf{l}_i \in \mathbb{R}^{N_l}$ being the N_l-dimensional latent space representation of $\mathbf{x}_i$, see Fig. 2.8 for an illustration.

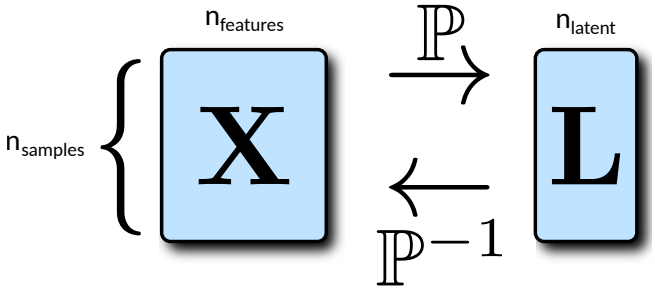


Figure 2.8: Schematic illustration of the transformation of a data set, $\mathbf{X}$, of n_{sample} data elements (or samples) defined in a high, n_{features}-dimensional feature space ($\mathbf{X} \in \mathbb{R}^{n_{\text{sample}} \times n_{\text{feature}}}$) to a lower, n_{latent}-dimensional latent space representation, $\mathbf{L} = \mathbb{P}(\mathbf{X}) \in \mathbb{R}^{n_{\text{sample}} \times n_{\text{latent}}}$, of the data. The transformation from $\mathbf{X}$ to $\mathbf{L}$ is indicated by the operator $\mathbb{P}$, the inverse transformation (which is, in general, not possible without information loss) as $\mathbb{P}^{-1}$ [238, 247, 248].

In most cases such transformations are (necessarily) associated with information loss, *i.e.*, the transformation is not bijective and we have $\mathbb{P}^{-1}(\mathbf{L}) = \mathbf{X}' \neq \mathbf{X}$. It is, however, crucial the low-dimensional representations of the data in the latent space, $\mathbf{L}$, is able to address the essential features of the original data, $\mathbf{X}$, as good as possible: the content of the relevant information in $\mathbf{X}$ should still be contained in $\mathbf{X}'$. For that purpose, it is, for instance, desirable that pairwise distances[34], $D(\mathbf{x}_i, \mathbf{x}_j) = |\mathbf{x}_j - \mathbf{x}_i|$, of data elements $\mathbf{x}_i$ and $\mathbf{x}_j$ of the data set in the feature space representation, $\mathbf{X}$, are conserved, as good as possible, in the latent space representation, $\mathbf{L}$, of the data, *i.e.*, $D(\mathbf{x}_i, \mathbf{x}_j) \approx D(\mathbf{l}_i, \mathbf{l}_j)$. The distances may differ in scale but the important point for our purposes is, that neighboring points, $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{N_f}$, in the feature space should also be neighboring points, $\mathbf{l}_i, \mathbf{l}_j \in \mathbb{R}^{N_l}$, in the latent space [119].

One very frequently used method for dimensional reduction is the above mentioned principal component analysis. PCA can be understood as a projection, $\mathbb{P}_{XL}$, of the N data elements of a data set represented in the high-dimensional feature space, $\mathbf{X} \in \mathbb{R}^{N \times N_f}$, to a low-dimensional latent space representation, $\mathbf{L} = \mathbb{P}_{XL}(\mathbf{X}) \in \mathbb{R}^{N \times N_l}$.

Closely following the derivation in Ref. [119], we start from N data elements (or equivalently, data points), $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, each living in an N_f-dimensional feature space, $\mathbf{x}_i \in \mathbb{R}^{N_f}$. We assume – without the loss of generality – zero empirical mean, $\mathbf{x}_i \to \mathbf{x}_i - E[\mathbf{x}]$, with $E[\mathbf{x}] = N^{-1} \sum_{i=1}^{N} \mathbf{x}_i$ and unit-variance, $\mathrm{Var}[\mathbf{x}] = (N-1)^{-1} \sum_{i=1}^{N} (E[\mathbf{x}] - \mathbf{x}_i)^2 = \mathbf{1}$. The $N \times N_f$ matrix $\mathbf{X}$ is called design matrix (whose rows are the data points and whose columns are the different features). We are interested in a linear transformation that reduces the covariance between different features which can be expressed using the $N_f \times N_f$ symmetric covariance matrix, $\mathbf{\Sigma}(\mathbf{X})$, defined by

$$\mathbf{\Sigma}(\mathbf{X}) = \frac{1}{N-1} \mathbf{X}^\intercal \mathbf{X}, \tag{2.72}$$

whose diagonal elements, $\mathbf{\Sigma}(\mathbf{X})_{jj}$, measure the variance of features and whose off-diagonal elements, $\mathbf{\Sigma}(\mathbf{X})_{ij}$, the covariance between feature i and j; $\mathbf{X}^\intercal$ is the transformed of $\mathbf{X}$. Further, we can perform singular value decomposition (SVD) on the design matrix

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\intercal, \tag{2.73}$$

with $\mathbf{S}$ being the $N_f \times N_f$ diagonal matrix of singular values, s_i, and the $N \times N_f$ matrix $\mathbf{U}$ and the $N_f \times N_f$ matrix $\mathbf{V}$ are orthogonal matrices whose columns represent, respectively, the left and right singular vectors of $\mathbf{X}$. We use Eq. (2.73) to rewrite the covariance matrix given by Eq. (2.72) as follows

$$\mathbf{\Sigma}(\mathbf{X}) = \frac{1}{N-1} \mathbf{V}\mathbf{S}\mathbf{U}^\intercal \mathbf{U}\mathbf{S}\mathbf{V}^\intercal = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\intercal, \tag{2.74}$$

where we introduced the $N_f \times N_f$ diagonal matrix $\mathbf{\Lambda} = \mathbf{S}^2/(N-1)$ representing the eigen-decomposition of the covariance matrix $\mathbf{\Sigma}(\mathbf{X})$. The along the diagonal descending-ordered $i = 1, \ldots, N_f$ eigenvalues, $\lambda_i = s_i^2/(N-1)$, of $\mathbf{\Lambda}$ can be used for dimensional reduction: large values of λ_i label directions along the associated singular vectors $\mathbf{v}_i \in \mathbb{R}^{N_f}$ (collected in the matrix $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_{N_f})$) of high variance in the feature space of the data which are, in many cases, those directions which contain

[34] Here we assume a Euclidean distance metric but other choices of distance metrices are also possible.

the relevant information of the data. In contrast, directions with small variance, *i.e.*, with associated small values of λ_i, are usually associated with noise and can potentially be ignored. The name "principal component analysis" stems exactly from the ordering $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_{N_f}$ of the eigenvalues of $\mathbf{\Lambda}$: the singular vector, $\mathbf{v}_i$, with the largest (second largest, third largest, ...) eigenvalue, λ_i, is called the first (second, third, ...) principal component.

An important measure to quantify the information contained in a principal component is its percentage in the explained variance $\lambda_i^{(\mathrm{e})}$, given by

$$\lambda_i^{(\mathrm{e})} = \left(\sum_{j=1}^{N_f} \lambda_j \right)^{-1} \lambda_i. \tag{2.75}$$

Often, only very few of the singular values, s_i, are significantly larger than zero. Selecting the $N_l = N_f^*$ largest eigenvalues, λ_i, and the associated singular vectors, $\mathbf{V}^* = (\mathbf{v}_1, \ldots, \mathbf{v}_{N_f^*}) \in \mathbb{R}^{N_f \times N_f^*}$, provides us with an effective way to project the data points to a low-dimensional (but representative) latent space $\mathbf{L} = \mathbf{X}\mathbf{V}^* = \mathbb{P}_{XL}(\mathbf{X})$. In that way, the transformation $\mathbb{P}_{XL}$ is simply a linear projection from $\mathbb{R}^{N_f}$ to $\mathbb{R}^{N_f^*}$ for every data point. For a more comprehensive explanation of PCAs see [238, 249, 250].

There are many examples in physics where PCA has been applied, many of which are listed in Ref. [119]. We employ the concept of PCA heavily in Subsection 3.1.4 in order to identify families of structures from a large data set of unlabelled (*i.e.*, not categorized), ordered ground state configurations of the asymmetric Wigner bilayer system [62–64].

2.4.2 k-means Clustering

Probably simplest and yet instructive form of unsupervised learning are clustering algorithms, whose objective is to identify groups in unlabeled data according to similarity or distance measures of one kind or another [119, 235, 236]. Here we introduce the k-means algorithm [239–242] by closely following the derivation in [119].

Starting from N data points, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, in an N_f-dimensional feature space, $\mathbf{x}_i \in \mathbb{R}^{N_f}$, the objective is to distribute a predefined (fixed) number of K cluster centers, called the *cluster means* $\mathbf{K} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_K\}$ with $\boldsymbol{\mu}_k \in \mathbb{R}^{N_f}$, in the feature space, such that data points assigned to the different clusters minimize the following cost function

$$\mathcal{C}(\mathbf{X}, \mathbf{K}) = \sum_{k=1}^{K} \sum_{i=1}^{N} r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k)^2. \tag{2.76}$$

In Eq. (2.76) the assignment of data point i to cluster k is realized via the binary variable $r_{ik} = 1$ (and $r_{ik'} = 0 \, \forall \, k \neq k'$). Consequently, $\sum_{k=1}^{K} r_{ik} = 1 \, \forall \, i$ expresses the fact that data points are assigned exclusively to one cluster only. $\sum_{i=1}^{N} r_{ik} = N_k$ defines the size of cluster k, *i.e.*, the number of data points associated with it. The set of assignments $\mathbf{k} = \{r_{ik}\}$ is also called *labeling* or *clustering* of the data points.

Minimizing Eq. (2.76) can be interpreted as finding the K cluster means, $\mathbf{K}$, and assigning, via r_{ik}, the N data points to different clusters, k, such that the (scaled) variance of each cluster, $\sum_{i=1}^{N} r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^2$, is minimized. This is performed in a two-step procedure:

1. Eq. (2.76) is minimized with respect to $\boldsymbol{\mu}_k$ given a set of assignments $\{r_{ik}\}$, *i.e.*, $(\partial\mathcal{C}/\partial\boldsymbol{\mu}_k)|_{\{r_{ik}\}} = 0$, yielding the update rule for $\boldsymbol{\mu}_k = N_k^{-1}\sum_{i=1}^{N} r_{ik}\mathbf{x}_i$, *i.e.*, $\boldsymbol{\mu}_k$ is assigned as the geometric center of the members $r_{ik}\mathbf{x}_i$ of cluster k.

2. Given the cluster means, $\mathbf{K} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_K\}$, we want to find the assignments $\mathbf{k} = \{r_{ik}\}$ which minimize Eq. (2.76) by assigning each data point to its nearest cluster-mean: $r_{ik} = 1$ if $k = \arg\min_{k'}(\mathbf{x}_i - \boldsymbol{\mu}_{k'})^2$ and $r_{ik} = 0$ otherwise.

These two steps are performed in an alternating way until some convergence criterion is met, for instance if the change of the object function, given by Eq. (2.76), between two iteration steps is smaller than a threshold.

The k-means algorithm scales linearly with the size of the data set, more specifically as $\mathcal{O}(KN)$, and can therefore be used for a large amount of data. However, Eq. (2.76) is in general a non-convex function and the minimization result may largely depend on the initial (random) choice of the means $\mathbf{K}$ and the assignments $\mathbf{k} = \{r_{ik}\}$. In practice, the k-means algorithm is applied multiple times with different (random) initial conditions which may result in different assignments. Then, usually the particular assignment with the minimal value of $\mathcal{C}(\mathbf{X}, \mathbf{K})$, given by Eq. (2.76), compared to all other assignments, is chosen as the "best" solution to the clustering problem.

2.4.3 Density-Based Clustering: DBSCAN

Density-based (DB) clustering is particularly useful in cases where different clusters of data points, $\mathbf{x}_i \in \mathbb{R}^{N_f}$, of a data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N\}$ form high density regions in the N_f-dimensional feature space (*i.e.*, regions with many data points per feature space volume) which are separated by regions of lower density. The core idea behind DB clustering is to assume that a relative local density estimation in the feature space of the data is possible (see below for details) which allows the data to be ordered (and clustered) by their densities [119].

The most prominent DB clustering algorithm is called *density-based spatial clustering of applications with noise* (DBSCAN) [243]. All data points, $\mathbf{x}_n$, that are at a distance, $|\mathbf{x} - \mathbf{x}_n|$, smaller than an ϵ to other data points in $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N\}$ are collected in an ϵ-neighborhood $N_\epsilon(\mathbf{x}_n) = \{\mathbf{x} \in \mathbf{X} : |\mathbf{x} - \mathbf{x}_n| < \epsilon\}$ and are considered as a group. If a minimum number of data points, N_m, (which, besides of ϵ, is one of the two parameter of the algorithm) lie within $N_\epsilon(\mathbf{x}_n)$ the point $\mathbf{x}_n$ is considered to be a so-called *core-point*, otherwise $\mathbf{x}_n$ is considered to be a *noise-point*. Overlapping core points – by means of the ϵ-neighborhood $N_\epsilon(\mathbf{x}_n)$ – are considered to form a cluster.

The main advantage of DB clustering is, that the number and the size of the clusters can be detected by the algorithm. DBSCAN works quite accurately in low-dimensional feature spaces, such as $N_f = 2$ or $N_f = 3$. However, density based measures, especially

in high-dimensions, can easily be affected by noise in the data, which limits the applicability of DB clustering methods to data with low-dimensional feature spaces for many practical purposes [119]. Often, dimensional reduction of a high-dimensional feature space via, for instance, PCA (see Subsection 2.4.1) can be applied prior to DBSCAN (or, in general, to any other clustering algorithm) to transform the clustering problem into an easier one with a more manageable number of degrees of freedom of the latent space representation of the data.

2.4.4 t-SNE Clustering and High-Dimensional Visualization

Another quite useful tool for visually representing a set of data from a high-dimensional feature space in a low-dimensional latent space, *i.e.*, in a few *embedding coordinates*, is t-stochastic neighbor embedding (t-SNE) [244]. To be more specific, t-SNE is a non-linear mapping of a total number of N data points, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{R}^{N_f}$, to low-dimensional embedding coordinates (in two [or three] dimensions), $\mathbf{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_N\}$ with $\mathbf{t}_i = (t_{i1}, t_{i2}[, t_{i3}]) \in \mathbb{R}^{N_l=2[\text{or } 3]}$, while aiming at preserving "local structures" in the original feature space representation, $\mathbf{X}$, of the data[35]. N_f is the number of features in the data set and N_l is the dimension of the latent space.

The idea behind the t-SNE analysis is to identify neighboring data points in the feature space by a proper distance and similarity measure with respect to the different features of the data. Specifically, t-SNE aims at matching the probability distribution p_{ij} (defined below) of two data points, $\mathbf{x}_i$ and $\mathbf{x}_j$, being neighbors in the features space with the probability distribution q_{ij} (also defined below) of the same two data points, $\mathbf{t}_i$ and $\mathbf{t}_j$, being neighbors in the latent space representation of the data. This is achieved by choosing the embedding coordinates $\mathbf{t}_i$ for all $i = 1, \ldots, N$ data points accordingly, as discussed below [119].

The Gaussian likelihood, $p_{i|j}$, that data point $\mathbf{x}_i$ is a neighbor of $\mathbf{x}_j$ in the data set $\mathbf{X}$ of N data points can be written as [244]

$$p_{i|j} = \frac{\exp(-|\mathbf{x}_i - \mathbf{x}_j|^2/(2\sigma_i^2))}{\sum_{j\neq i}^{N} \exp(-|\mathbf{x}_i - \mathbf{x}_j|^2/(2\sigma_i^2))}, \tag{2.77}$$

where σ_i is free a bandwidth parameter which is usually determined by fixing the local entropy $H(p_i) = -\sum_{j=1}^{N} p_{j|i} \log_2(p_{j|i})$ for each data point and where we set $p_{i|i} = 0$. For every $i = 1, \ldots, N$ data point $\mathbf{x}_i$ the local entropy $H(p_i)$ is set to the same constant value Σ, *i.e.*, to the so-called *perplexity* $\Sigma = 2^{H(p_i)}$ [244]. In that way, σ_i is determined for every $i = 1, \ldots, N$ data point $\mathbf{x}_i$ such that the value of σ_i is smaller for data points in high-density regions of the data points $\mathbf{X}$ in the feature space and larger for data points in low-density regions [119].

The main contribution to the Gaussian likelihood $p_{i|j}$ defined by Eq. (2.77) stems from data points that are "nearby" in the feature space (which should also be nearby in the t-SNE latent space). For data points that are "far apart" the contributions to $p_{i|j}$

[35] Usually, t-SNE is combined with a preceding data preparation, such as scaling all features to unit-variance– and zero-mean coordinates and applying principal component analysis or other dimensional reduction tools beforehand, $\mathbf{x}_i \to \mathbf{x}'_i$, see Subsection 2.4.1.

is vanishingly small which, however, can consequently lead to ambiguous embedding coordinates [119, 244]. For that reason, the symmetrized probability distribution p_{ij}, defined by

$$p_{ij} = \frac{1}{2N}(p_{i|j} + p_{j|i}), \tag{2.78}$$

is used, rather than the Gaussian likelihood $p_{i|j}$, to describe the local neighborhood relations of data points in the feature space. Note that $p_{ii} = 0$ (since $p_{i|i} = 0$) and $\sum_{i,j=1}^{N} p_{ij} = 1$.

The t-SNE approach now aims at identifying a *similar* probability distribution, q_{ij}, of two data points, $\mathbf{t}_i$ and $\mathbf{t}_j$, being neighbors in the low-dimensional latent space representation, $\mathbf{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_N\}$, of the data with $\mathbf{t}_i \in \mathbb{R}^{N_l}$; q_{ij} is defined by [244]

$$q_{ij} = \frac{(1 + |\mathbf{t}_i - \mathbf{t}_j|^2)^{-1}}{\sum_{j \neq i}^{N}(1 + |\mathbf{t}_i - \mathbf{t}_j|^2)^{-1}}, \tag{2.79}$$

which is chosen as long tail distribution in contrast to short tail symmetrized Gaussian distribution p_{ij} defined via Eqs. (2.77) and (2.78).

The two probability distributions p_{ij} and q_{ij}, respectively defined by Eqs. (2.78) and (2.79), are then matched by minimizing the *cost function* $\mathcal{C}(\mathbf{T})$, which is defined by the Kullback-Leibler divergence, $D_{\mathrm{KL}}(p||q)$, as

$$\mathcal{C}(\mathbf{T}) = D_{\mathrm{KL}}(p||q) = \sum_{i,j=1}^{N} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right); \tag{2.80}$$

$\mathcal{C}(\mathbf{T})$ is minimized under the variation of the embedding coordinates, $\mathbf{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_N\}$, and the embedding coordinates minimizing Eq. (2.80) represent the latent space representation of the original data.

At this point, it is instructive to explicitly present the gradient, $\partial_{\mathbf{t}_i} C(\mathbf{T})$, of the cost function, defined by Eq. (2.80), with respect to the embedding coordinates $\mathbf{t}_i$, *i.e.*

$$\partial_{\mathbf{t}_i} C(\mathbf{T}) = \sum_{j \neq i}^{N} 4\, p_{ij} q_{ij} Z_i (\mathbf{t}_i - \mathbf{t}_j) - \sum_{j \neq i}^{N} 4 q_{ij}^2 Z_i (\mathbf{t}_i - \mathbf{t}_j) \tag{2.81}$$

$$= \mathbf{F}_i^{(\mathrm{attr.})} - \mathbf{F}_i^{(\mathrm{rep.})}, \tag{2.82}$$

with $Z_i = 1/(\sum_{k \neq i}^{N}(1 + |\mathbf{t}_k - \mathbf{t}_i|^2)^{-1})$. We can see in Eqs. (2.81) and (2.82), that for every data point $\mathbf{t}_i$ in the latent space, the gradient of the cost function can be separated into two contributions: the first sum in Eq. (2.81) can be interpreted as a collective force $\mathbf{F}_i^{(\mathrm{attr.})}$, defined in Eq. (2.82), acting on a data point $\mathbf{t}_i$ in the latent space, which origins in attractive interactions between $\mathbf{t}_i$ and all other data points $\mathbf{t}_{j \neq i}$; analogously, the second sum in Eq. (2.81), represented by $\mathbf{F}_i^{(\mathrm{rep.})}$ in Eq. (2.82), can be interpreted as a collective force, acting on data point $\mathbf{t}_i$, originating from repulsive contributions between the data points [119].

In that way, the t-SNE approach of minimizing the Kullback-Leibler divergence between the feature space distribution p_{ij} and the latent space distribution q_{ij} can be

interpreted as equilibrating the effects of attractive forces $\mathbf{F}_i^{(\text{attr.})}$ and repulsive forces $\mathbf{F}_i^{(\text{rep.})}$ acting on the data points in the latent space: data points which are nearby in the feature space (with large contributions to p_{ij}) experience large attractive forces in the latent space (via large contributions to q_{ij}) and are thereby grouped together in clusters in the latent space representation of the data by minimizing $\mathcal{C}(\mathbf{T})$; in that way, short distance information (*i.e.*, relative neighborhoods) of the data points is preserved. Data points which are far apart in the feature space, on the other hand, are repelled from each other in the latent space due to the long range tails of q_{ij}. The representation of the data in the t-SNE embedding coordinates thus forms clusters of similar data points but different clusters are repelled such that they can more clearly be distinguished from each other in the latent space compared to the original feature space [119].

Often it is easier to visually separate clusters in a few t-SNE dimensions than in the higher-dimensional feature space [119]. One can subsequently use clustering techniques such as DBSCAN [243] (see Subsection 2.4.3) to label the different clusters identified by the t-SNE analysis (see Refs. [86, 119]).

2.4.5 (Adjusted) Mutual Information

In the Subsections above we introduced clustering algorithms, with the objective of partitioning a set, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, of N data elements, into subsets, $\mathbf{U}^R = \{\mathbf{U}_1, \mathbf{U}_2, \ldots,$
$\mathbf{U}_R\}$, such that the union of all clustered elements, $\cup_{i=1}^{R} \mathbf{U}_i = \mathbf{X}$, uniquely gives the set, *i.e.*, they do not overlap, $\mathbf{U}_i \cap \mathbf{U}_j = \emptyset \; \forall i \neq j$ [246].

Commonly used algorithms, such as k-means clustering or DBSCAN, see Subsection 2.4.2 and Subsection 2.4.3, are, on the one hand, applicable in different situations and, on the other hand, are not unique in their behaviour. The final result of such algorithms may strongly depend on the respective initial conditions of the particularly applied clustering algorithm (such as the initial, usually random choice of assigning data points to clusters, etc.) and on the choice of the parameters of the algorithm, or on noise in the data [237].

Here we are interested in comparing the results of different clusterings[36], $\mathbf{U}^R = \{\mathbf{U}_1, \mathbf{U}_2,$ $\ldots, \mathbf{U}_R\}$ and $\mathbf{V}^C = \{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_C\}$, by quantifying their overlap, or in other words, by quantifying the shared information of different clusterings.

A fundamental class of techniques for comparing clusterings is formed by information theoretic measures [246]. In this thesis we used the concept of adjusted mutual information, *i.e.*, an information theoretic approach used for comparing clusterings of a labeled data set [245, 246].

In order ot do so, we first define the $R \times C$ contingency table $M = [n_{ij}]_{j=1\ldots C}^{i=1\ldots R}$ in Table 2.1, whose elements, $n_{ij} = |\mathbf{U}_i \cap \mathbf{V}_j|$, quantify the number of common objects in clustering $\mathbf{U}_i$ and $\mathbf{V}_j$.

[36] We want to compare either the results of different clustering algorithms or compare the results of the same algorithm but with different initial conditions.

$\mathbf{U}^R/\mathbf{V}^C$	$\mathbf{V}_1$	$\mathbf{V}_2$	$\dots$	$\mathbf{V}_C$	Sums
$\mathbf{U}_1$	n_{11}	n_{12}	$\dots$	n_{1C}	a_1
$\mathbf{U}_2$	n_{21}	n_{22}	$\dots$	n_{2C}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\mathbf{U}_R$	n_{R1}	n_{R2}	$\dots$	n_{RC}	a_R
Sums	b_1	b_2	$\dots$	b_C	$\sum_{ij} n_{ij} = N$

Table 2.1: Contingency table between two different clusterings, $\mathbf{U}^R = \{\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_R\}$ and $\mathbf{V}^C = \{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_C\}$, with $n_{ij} = |\mathbf{U}_i \cap \mathbf{V}_j|$ being the number of common objects in clusterings $\mathbf{U}_i$ and $\mathbf{V}_j$; and $a_i = \sum_{j=1}^{C} n_{ij}$ and $b_j = \sum_{i=1}^{R} n_{ij}$.

The mutual information, $I_M(\mathbf{U}, \mathbf{V})$, of two different clusterings, $\mathbf{U} \equiv \mathbf{U}^R$ and $\mathbf{V} \equiv \mathbf{V}^C$, is defined as [245, 246]

$$I_{\mathrm{M}}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{R} \sum_{j=1}^{C} P_{\mathbf{UV}}(i,j) \log \frac{P_{\mathbf{UV}}(i,j)}{P_{\mathbf{U}}(i) P_{\mathbf{V}}(j)} \tag{2.83}$$

where $P_{\mathbf{UV}}(i,j) = |\mathbf{U}_i \cap \mathbf{V}_j|/N$ is the probability that a data point, i, belongs to both clusters $\mathbf{U}_i$ (in $\mathbf{U}$) and $\mathbf{V}_j$ (in $\mathbf{V}$) and $P_{\mathbf{U}}(i) = |\mathbf{U}_i|/N$ and $P_{\mathbf{V}}(j) = |\mathbf{V}_j|/N$ denote the probabilities, that the data points i and j fall into the cluster $\mathbf{U}_i$ and $\mathbf{V}_j$, respectively. In that way, $I_{\mathrm{M}}(\mathbf{U}, \mathbf{V})$ quantifies the information which is shared between two clusterings and thus can be interpreted as a similarity measure for clusterings; notably, the upper bounds of $I_{\mathrm{M}}(\mathbf{U}, \mathbf{V})$ are the entropies $H(\mathbf{U}) = -\sum_{i=1}^{R} P_{\mathbf{U}}(i) \log P_{\mathbf{U}}(i)$ and $H(\mathbf{V}) = -\sum_{j=1}^{C} P_{\mathbf{V}}(j) \log P_{\mathbf{V}}(j)$ [246].

The adjusted mutual information, $I_K(\mathbf{U}, \mathbf{V})$, corrects the information theoretic measures of mutual information agreement of clusterings for chance (see Refs. [245, 246, 251] for details), and can be given by

$$I_K(\mathbf{U}, \mathbf{V}) = \frac{I_{\mathrm{M}}(\mathbf{U}, \mathbf{V}) - E_{\mathrm{MI}}(\mathbf{U}, \mathbf{V})}{\max(H(\mathbf{U}), H(\mathbf{V})) - E_{\mathrm{MI}}(\mathbf{U}, \mathbf{V})}, \tag{2.84}$$

where the expected mutual information, $E_{\mathrm{MI}}(\mathbf{U}, \mathbf{V})$, between two (random) clusterings is defined by

$$E_{\mathrm{MI}}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{n_{ij}=\max(1, a_i+b_j-N)}^{\min(a_i, b_j)} \frac{n_{ij}}{N} \log\left(\frac{N\, n_{ij}}{a_i b_j}\right) \times \frac{a_i\,!\, b_j\,!\, (N-a_i)\,!\, (N-b_j)\,!}{N\,!\, n_{ij}\,!\, (a_i - n_{ij})\,!\, (b_j - n_{ij})\,!\, (N - a_i - b_j + n_{ij})\,!}, \tag{2.85}$$

with $a_i = \sum_{j=1}^{C} n_{ij}$ and $b_j = \sum_{i=1}^{R} n_{ij}$ being the partial sums over the contingency table $M[n_{ij}]_{j=1\dots C}^{i=1\dots R}$ defined in Table 2.1.

A value of $I_K(\mathbf{U}, \mathbf{V}) = 1$ means perfect overlap between two different clusterings (*i.e.*, the two clusterings label the data equivalently but potentially use different numerical values to label the different clusters), a value smaller than one indicates differences in the clusterings.

3 Systems

In this Chapter, we present our numerical investigations of the structural ground state self-assembly of two different systems, which are composed of (i) negative, classical point-charges confined to the surfaces of two parallel, uniformly but differently charged plates or (ii) of charged, polycyclic aromatic molecules confined to the volume near a metal–liquid-interface under electrochemical conditions.

In Section 3.1 we numerically investigate the ordered ground state configurations of the asymmetric Wigner bilayer system [62–64, 146, 147], a system of negative, classical point-charges confined to the surfaces of two oppositely charged, parallel plates of uniform but different charge densities. Based on the studies of Antlanger *et al.* [62–64] we investigate the asymmetric Wigner bilayer system for ordered, yet aperiodic quasicrystalline structures [51] for selected values of the system parameters. To identify the corresponding regions in the parameter space we first employ clustering algorithms to the structural database of currently suggested bilayer ground state configurations from Refs. [62–64]. We then focus on structures related to the snub-square ground state configurations of the system – featuring an equilateral square-triangle tiling on one of the plates, which is thus often considered to be a precursor of a dodecagonal quasicrystal [252–254]. Thereafter, we investigate possibly aperiodic ground states configurations related to the trihexagonal phase of the system which give rise to *pseudo*-dodecagonal traits in the tiling (clusters of rectangles and two different types of equilateral triangles resembling regular dodecagons emerge). In both cases we identify a series of self-similar super-structures which we then compare with ground state candidate configurations of the asymmetric Wigner bilayer system known from literature [62–64].

In Section 3.2 we investigate the ground state self-assembly behavior of complex charged molecules on a metallic surfaces under electrochemical conditions with elaborated numerical tools [21, 22]. We propose a computationally lean, two-stage approach [86] to treat this problem reliably. Stage one uses *ab initio* simulations to provide reference data for the structures of the different molecules as well as of energies (evaluated for archetypical configurations) to fit the parameters of a conceptually much simpler and computationally less expensive force field of the molecules: classical, spherical particles, represent the respective atomic entities while a flat and perfectly conducting wall represents the metallic surface. Stage two feeds the energies that emerge from this force field into highly efficient and reliable optimization techniques to identify via energy minimization the ordered ground state configurations of the molecules. We demonstrate the power of our approach by successfully reproducing on a semi-quantitative level the intricate supramolecular ordering observed experimentally for PQP^+ and ClO_4^- molecules on a Au(111)-electrolyte interface, including the formation of open-porous, auto-hosts–guest, and stratified bilayer phases as a function of the electric field at

the solid–liquid interface. We also discuss the role of the perchlorate ions in the self-assembly process, whose positions could not be identified in the related experimental investigations [83].

3.1 Towards Quasicrystalline Order in the Asymmetric Wigner Bilayer System

3.1.1 Introduction

In a Wigner bilayer system [146, 147] classical point-charges are confined between two parallel plates with uniform, opposite surface charge densities, σ_1 and σ_2, maintaining overall charge neutrality, see Fig. 3.1(a) for a schematic representation. Only recently, analytic results of the zero temperature ground state formation of the charges for the symmetric case, $\sigma_1 = \sigma_2$, could be determined [146, 147].

Detailed investigations [63, 64] on the self-assembly scenarios of the charges on the two layers of the asymmetric case, $\sigma_1 \neq \sigma_2$, have revealed a rich plethora of zero-temperature ground state configurations; the phase diagram is shown in Fig. 3.1(d) and Fig. 3.3. Any state of this system can be characterized by the separation of the plates, d, and the ratio of the surface charge densities of the two plates[1], $A = \sigma_2/\sigma_1$; the setup is illustrated in Fig. 3.1(a) and a detailed explanation of the physics of the system is presented in Subsection 3.1.3.

The results presented in Refs. [63, 64] show ordered zero temperature ground-state structures which can be built by periodic stacking of respective irreducible unit cells (see Fig. 3.3). In well-defined regions of the parameter space of the system (spanned by $\eta \propto d$ and A), ordered structures have been identified which can be considered as precursors of quasicrystals [51]: amongst them are snub-square type of tilings, S_1, *i.e.*, a square-triangle tiling [255] on one plate and squares on the other (cf. right panel of Fig. 3.1(b)). Then there are tilings with pentagonal motives on one plate and rectangle-triangle tilings on the other (cf. left panel of Fig. 3.1(b)) and I_x-Cairo type tilings [255] with rectangles and two different types of equilateral triangles representing the basic tiles of one plate of a bilayer structure which give rise to a hexagonal monolayer structure if the particles of both plates were projected onto the same layer, see Fig. 3.1(c).

In contrast to an ordered crystalline structure, the particles of a quasicrystal – atoms, molecules or in our case point-charges – form ordered but aperiodic (*i.e.*, quasiperiodic [51]) patterns: a finite number of local motives (or tiles) can be used to aperiodically tile the entire space (or plane) without leaving any voids [256, 257]. Quasicrystals exhibit long range orientational order but no translational symmetry [51]. Often quasicrystals exhibit conspicuous five-fold, ten-fold, eight-fold or twelve-fold rotational symmetries apparent in X-ray diffraction patterns which are inconsistent with any kind of periodic crystalline structure with traditionally allowed two-, three-, four- and sixfold rotational symmetry [51, 258, 259]. A prominent example of a quasicrystalline pattern in two dimensions is the Penrose tiling [257, 260] with local pentagonal motives formed by two distinct rhombic tiles with the same side-lengths. The respective angles of the thin rhombus are 36° and 144°, for the thick rhombus we have 72° and 108° (*i.e.*, multiples of 36° = 360°/10). The shapes of the tiles are related to the golden mean, $\Phi = (1+\sqrt{5})/2$: the long diagonal of the thick rhombus is Φ and the short diagonal of the thin rhombus is $1/\Phi$ (in units of the side-lengths of the tiles). When

[1] Following Refs. [62–64] and without the loss of generality we chose $A \in [0, 1]$.

imposing very specific local matching rules of these two basic tiles during tiling, global decagonal rotational symmetry is achieved.

Some physical systems are structurally quite similar to quasicrystals by forming crystalline structures with large unit cells (thereby approximating aperiodic order) such as Frank-Kasper phases [261] or, more generally, some tetrahedrally close-packed structures [262] or Mackay icosahedra [263]; so-called quasicrystalline *approximants* can interpolate between crystalline and quasicrystalline order by forming ever-larger irreducible unit cells[2] which all feature local motives and environments of a certain tile-set that also allows quasicrystalline ordering [264]. Quasicrystalline approximants often show up in the experiment at similar (yet slightly different) values of the system parameters (such as temperature, pressure, etc.) as real quasicrystals [264] and potentially indicate the presence of a quasicrystal in the system [265–268].

Mathematically, quasicrystals can be related to regular, periodic lattices in a higher-dimensional space: it has been shown, for instance, that three-dimensional icosahedral quasicrystals correspond to arrangements of three-dimensional hypersurfaces (or hypersurface segments) in a six-dimensional hpyercubic lattice [259, 269]; two-dimensional aperiodic square-triangle based dodecagonal quasicrystals can be represented as periodic lattices in a four-dimensional space [253, 270, 271]. Thus, the structure of aperiodic quasicrystals can be described via projections of higher-dimensional periodic lattices onto hypersurfaces (*i.e.*, onto the so-called physical space) that exhibit certain angles with the hyperlattice vectors [264].

In this thesis, we focus on specific regions in the parameter space of the asymmetric Wigner bilayer system featuring specific tilings or symmetries such as snub-square tilings or structures with strong signals in the five- or twelve-fold orientational order parameters [225–227], which thereby suggest structures related to quasicrystals.

Below we provide an outline of this Section. We first provide an overview of the basic physics of the asymmetric Wigner bilayer system in Subsection 3.1.3. The results of our investigations are then subdivided into three main topics – schematically depicted in Fig. 3.1(d,c,b) – which can be considered as the precursors of three forthcoming contributions:

In Subsection 3.1.4 we propose a new scheme based on k-means clustering [239–242] of structures based on their order parameters to extend the current picture of the phase-diagram of Refs. [62–64] via automatized (unsupervised) classification of structure families (cf. Fig. 3.1(d)). With this tool at hand, we can gain broader insights into the complex phase behavior of the asymmetric Wigner bilayer system which will help us identifying promising regions in the phase diagram which potentially host quasicrystalline states.

In an effort to find dodecagonal quasicrystalline ground state candidates we specifically explore in Subsection 3.1.5 ever-larger square-triangle tilings which approximate a dodecagonal quasicrystal [252, 254, 273] (cf. right inset in Fig. 3.1(b)) and test these approximants against ground state configurations in the S_1 region. We systematically

[2] An irreducible unit cell is the smallest possible unit cell, with as few particles inscribed as possible, which is sufficient to fully describe a periodic structure.

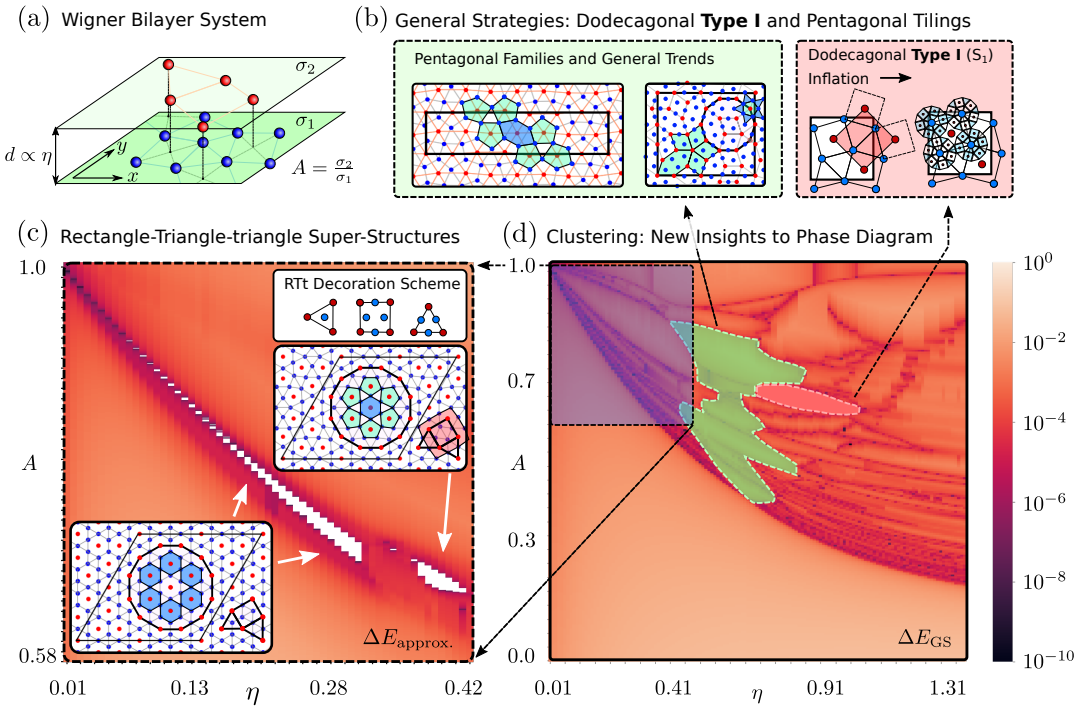


Figure 3.1: Schematic overview of different subdivided studies on quasicrystalline order in the asymmetric Wigner bilayer system performed in this thesis: (a) schematic representation of the here investigated asymmetric Wigner bilayer system [62–64, 146, 147] at $T = 0$ K with blue/red colored charges belonging to the bottom/top plates (layer one/two), respectively (see Subsection 3.1.3). In this thesis, we always color charges of layer one in blue and charges of layer two in red in the visual representations of bilayer structures. (b) General strategies to systematically explore the asymmetric Wigner bilayer system for quasicrystalline ordering related to the S_1 configuration (example structures featured in right, red shaded panel in (b), cf. Subsection 3.1.5) and to structures belonging to *pentagonal* families of bilayer configurations (example structures featured in left, green shaded panel in (b), cf. Subsection 3.1.7) using so-called *inflation rules* [252, 254, 272] and tailored optimization algorithms in the relevant regions of the (η, A)-plane (emphasized by dashed, correspondingly colored areas in (d)). (c) Energy difference (logarithmic scale) of rectangle-large–triangle-small–triangle ($\mathrm{RT}_l\mathrm{t}_s$) super-structures to the ground state energies from Refs. [62–64] in the relevant region of the (η, A)-plane (emphasized by dashed, blue rectangle in (d), cf. Subsection 3.1.6). Insets show the decoration scheme of the rectangular and the two types of triangular tiles and selected, newly discovered $\mathrm{RT}_l\mathrm{t}_s$ ground state candidate super-structures, emerging in the white regions in (c), see Subsection 3.1.6. (d) Energy difference (logarithmic scale) between the ground state– and the (lowest energy) competing structures from literature [62–64] in the (η, A)-plane identified with the help of unsupervised clustering techniques (see Subsection 3.1.4) highlighting phase boundaries of the asymmetric Wigner bilayer system (via dark colors, cf. colorbar). The dashed, shaded regions in (d) emphasizes regions of interest – possibly featuring quasicrystalline ordering in the Wigner bilayer system – which are studied in this thesis.

increase the complexity of the unit cells by a factor of $(2 + \sqrt{3})$ using "Stampfli"-inflation [252] and periodic stacking of the unit cells, covering 6, 19, 22, 76, 82, 88, 306, 1142 and 1224 particles per unit cell. The required structure optimization of these increasingly complex unit cells is performed with specifically implemented simulated annealing and replica exchange Monte-Carlo methods, utilizing so-called *zipper*-moves [254].

In Subsection 3.1.6, we identify a new class of zero temperature ground state candidate structures in the proximity of the phase boundary between the regions of stability of the trihexagonal- (I_x) and the honeycomb structure (H). Elemental rectangular and two different types of equilateral triangular tiles arrange themselves in pseudo-dodecagonal motives – composed of rectangles and triangles rather than squares and equilateral triangles in S_1– which themselves form triangular and rectangular super-tiles. We provide design rules for mapping approximate dodecagonal square-triangle structures via bilayer modulations[3] of the hexagonal lattice which is related to hexagonal symmetric metallic mean quasicrystals [45, 264, 274]. Our analysis – numerically based on replica exchange Monte-Carlo and a careful study of local environments – reveals the emergence of zero temperature super-tiling global ground state candidates of the asymmetric Wigner bilayer system with 25, 49, 56 or 192 particles per unit cell (cf. Fig. 3.1(c)).

In Subsection 3.1.7, we discuss the current state of our research and miscellaneous topics regarding pentagonal ground state structures of the asymmetric Wigner bilayer system, where the trends in the structure formation are less clear at first sight. However, we can identify routes towards understanding the structure formation in some of these pentagonal regions in parameter space (cf. left inset in Fig. 3.1(b) and green region in Fig. 3.1(d)). Also, we briefly discuss the region of stability of V_x structures (*i.e.*, hexagonal bilayer structures with different particle densities per layer) and we report relations between different ground state configurations in the asymmetric Wigner bilayer system and three-dimensional atomistic structures known from intermetallics.

We conclude this Section with summarizing our results and with final remarks on our findings in Subsection 3.1.8.

3.1.2 Terminology

Here we explicitly collect different terms and phrases that are frequently used in the investigations of the structural ground state self-assembly scenarios in Section 3.1.

We refer to a specific arrangement of particles (*i.e.*, charges) in a quasi-2D bilayer lattice geometry as *structure* or, synonymously, as *configuration*. The position of a single charge or of multiple particles forming the structure is referred to as *vertex* or *vertices* of a configuration. The lattice geometry which underlies a structure is referred to as *lattice* which is defined by the *lattice vectors*. A unit cell is the simulation box hosting N particles in a volume of finite spatial extent (given by the lattice vectors) with periodic boundaries in the x and y direction and with a finite slab-width in the z direction. The quasi-2D lattice of 2D-periodic copies of the unit cells (including 2D-periodic images of all particles inscribed to the unit cell) defines a structure (or a configuration). The smallest possible unit cell, *i.e.*, a unit cell with as few inscribed particles as possible with whom a structure can fully be described is called the *irreducible unit cell* of the configuration.

[3] We vertically reassign particles of a hexagonal monolayer to the other plate into rectangular and two different types of triangular tiles.

A *tiling* (or equivalently a *tesselation*) of a flat surface is defined as the covering of a plane with non-overlapping geometric shapes without leaving holes. The different geometric shapes with whom the tiling is generated are called *tiles*. In Section 3.1 we refer to the geometric shapes, *i.e.*, the regular or irregular polygons formed by nearest neighbor vertices of a configuration, as tiles; the corresponding structure is, in this respect, also referred to as a tiling. If all tiles of a structure are of a special geometric shape such as, for instance, of equilateral triangular or square shape, this geometric property might be indicated via writing *triangular* tiling or *square* tiling, respectively. For tilings composed of regular polygons with a unique side length, we also refer to the side length of the tiles as *tiling length*.

We usually refer to the tiling in layer one of a bilayer structure as *layer one tiling* and the tiling in layer two as *layer two tiling*, respectively. Wherever we want to address the structure, the configuration, the vertices or the tiles specifically for one of the two layers we emphasize this through writing *layer one* structure, *layer one* configuration, etc., and analogously for *layer two*. Occasionally, we also address those structures emerging when the planar coordinates of the vertices of both layers are projected to the same layer as *projected structure*, but we especially highlight this in the text whenever necessary. Notably, we use the words *layer* and *plate* synonymously in the context of the Wigner bilayer system.

We here use the phrase "*with respect to the periodic boundary conditions*" to address the toroidal periodicity of a specific configuration or of a family of configurations (*i.e.*, of a configuration space) that has a fixed number of particles per unit cell with fixed lattice vectors. The periodic boundary conditions of a specific realization of a family of structures from a configuration space impose, which tiles can be formed geometrically by the particles in the unit cell and across periodic directions.

We use the term *super-tile* to address tiles, which are formed by clusters of several tiles on a larger length scale by connecting special vertices of the clusters and we refer to the tiling composed of such *super-tiles* as *super-tiling* and to the related structures as *super-structures*.

3.1.3 The Model of the Asymmetric Wigner Bilayer System

A classical Wigner bilayer system [146, 147] consists of two parallel plates of infinite extent[4] with uniform charge densities, σ_1 and σ_2, respectively, with classical, oppositely charged point particles, $q_i = -e$, being confined between the two layers, maintaining overall charge neutrality; e being the elementary charge. At zero temperature, $T = 0$ K, the charges are – as a consequence of the Earnshaw-theorem [275] – prohibited to occupy the interior of the volume but have to be located on its boundaries, *i.e.*, on the plates. The system is sketched in Fig. 3.1(a).

In the asymmetric Wigner bilayer system [62–64], where $\sigma_1 \neq \sigma_2$, any state of the system can be characterized by the separation of the plates, d, or by the dimensionless quantity η defined by

$$\eta = d\sqrt{\frac{\sigma_1 + \sigma_2}{2}}, \tag{3.1}$$

and by the ratio of the surface charge densities of the two plates, *i.e.*,

$$A = \frac{\sigma_2}{\sigma_1} \in [0, 1]. \tag{3.2}$$

We exclude from our investigations the trivial case of $A < 0$, where all particles are located on one plate and, without loss of generality, we chose $\sigma_2 \leq \sigma_1$, since inverting the z axis covers situations where $\sigma_2 > \sigma_1$ [62].

The total number density[5], $\rho = N/S_0 = \rho_1 + \rho_2$, together with the number densities of layer one, $\rho_1 = N_1/S_0$, and layer two, $\rho_2 = N_2/S_0$, specify the total number of charges, $N = N_1 + N_2$, which occupy layer one and two in the quasi 2D unit cell of a bilayer structure with surface area S_0 and height d. Overall charge neutrality is satisfied when

$$\sigma S_0 = N\, e = \rho\, S_0\, e \tag{3.3}$$

which relates the total uniform background charge density of the two plates, $\sigma = (\sigma_1 + \sigma_2)$, and the number charge density of the particles, $-\rho\, e = -(\rho_1 + \rho_2)e$. In other words, the total charge of the particles, $-N\, e = -\rho\, S_0\, e = -Q_0$, within a unit cell of area S_0 is compensated by the total plate charge ot the unit cell, $Q_0 = \sigma S_0$. Overcharging effects of the two layers have been observed for zero temperature ground state structures in Refs. [63, 64], *i.e.*, charge neutrality does not need to be satisfied locally, $-e\rho_i \neq \sigma_i$, as long as Eq. (3.3) is satisfied.

The classical point-charges interact via long range Coulomb interaction with each other and with the uniformly charged plates. Further, there is a distance dependent but otherwise constant plate to plate interaction contributing to the total (internal) electrostatic energy of the unit cell of a bilayer structure, $E(\mathbf{r}^N; A, \eta)$, which is given by

$$E(\mathbf{r}^N; A, \eta) = \sum_{i=1}^{N} \left[\sum_{j=1}^{N} {\sum_{\mathbf{S_n}}}^{*} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{S_n}|} - 2\pi e^2 (\sigma_1 - \sigma_2) z_i \right] + const. \tag{3.4}$$

[4] Which we assume perpendicular to the z axis.
[5] Without loss of generality we fixed $\rho = 1$ in this thesis following Refs. [62–64, 146, 147].

$\mathbf{r}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the set of position vectors, $\mathbf{r}_i = (x_i, y_i, z_i)$, of all N charges in the unit cell, $z_i = 0$ or $z_i = d \propto \eta$ defines if particle i occupies plate one or two and $\mathbf{S_n}$ is a symbolic notation for periodic images of the unit cell in the x and y directions used to carry out the lattice summation[6] [92] (see Subsection 2.1.3 for details). For convenience, we chose the dielectric constant, ϵ, of the dielectric medium in which the particles are immersed, as well as the dielectric constant of the two plates, ϵ_1 and ϵ_2, to be one, *i.e.*, $\epsilon = \epsilon_1 = \epsilon_2 = 1$. Closely following Ref. [62] we employ Ewald summation techniques [94] – specifically implemented for quasi-2D bilayer geometries [62, 92] – to numerically evaluate the long range electrostatic energy of the system in a highly reliable and computationally efficient manner.

Investigating global ground state configurations in the asymmetric Wigner bilayer system for any pair of η and A boils down to identifying simultaneously the correct number of particles, N, per unit cell, to find the optimal particle arrangement, $\mathbf{r}^N$, on both plates and identifying the correct unit cell geometry[7] which, all together, minimize the total unit cell energy per particle, $E(\mathbf{r}^N; A, \eta)/N$, given Eq. (3.4). We collect all of these variational parameters to the structural ground state search in the expression

$$\mathcal{X} = (\mathbf{r}^N, \mathbf{a}_1, \mathbf{a}_2) \tag{3.5}$$

and we also might write $E(\mathcal{X}; A, \eta)/N \equiv E(\mathbf{r}^N; A, \eta)/N$ to parametrize the energy per particle[8]. If the particular values of A and η are not important for the discussion we may simply write $E(\mathcal{X})/N$ or even drop the argument of the energy E/N totally.

The required accuracy with which the energies of competing structures need to be evaluated is tremendously high: relative differences in the sevenths or eighths digit are very relevant when comparing energies of magnitude $E/\left(N\sqrt{\rho}e^2\right) \approx -1$, which sets the energy scale of this system. The high accuracy with which the internal energy of the system needs to be evaluated poses a very delicate and complicated optimization problem in a high dimensional search space, a task which has very successfully been carried out in Ref. [62] with the help of memetic evolutionary algorithms, see Subsection 2.2.4.

The computational cost for exploring the search space in the (η, A)-plane can be reduced significantly as the energy, E/N, defined by Eq. (3.4), can be separated into a (i) structure-dependent but A-independent part and an (ii) A-dependent but structure-independent part: Following Refs. [62–64] we first introduce the reduced energy per particle as

[6] $\sum^*$ in Eq. (3.4) indicates, that for $\mathbf{S_n} = (0,0,0)$ the sum is carried out only for $j > i$ to avoid double counting within the unit cell.

[7] The components a_{11}, a_{21} and a_{22} of the lattice vectors $\mathbf{a}_1 = (a_{11}, 0, 0)$ and $\mathbf{a}_2 = (a_{21}, a_{22}, 0)$ are subject to the structure optimization problem under the constraint of keeping a constant unit cell area, $S_0 = a_{11}\, a_{22}$; the lattice vector $\mathbf{a}_3 = (0, 0, d)$ is fixed by the plate separation distance, d.

[8] Notably, the energy, $E(\mathbf{r}^N; A, \eta)/N$, defined by Eq. (3.4) also depends on the in-plane lattice vectors, $\mathbf{a}_1$ and $\mathbf{a}_2$, via the definition of $\mathbf{S_n}$ (cf. Subsection 2.1.3). We do not explicitly list the lattice vectors in the argument list of $E(\mathbf{r}^N; A, \eta)$ but we emphasize the intrinsic dependency of the energy of a configuration on the lattice vectors in the definition of the variational parameters, $\mathcal{X}$, of a structure defined by Eq. (3.5).

$$\frac{E^*(\mathbf{r}^N;A,\eta)}{N} = \frac{E(\mathbf{r}^N;A,\eta)}{N\sqrt{\rho}e^2} \tag{3.6}$$

and we then define the structure independent term as

$$\frac{E^*_A(A,\eta,x)}{N} = 2^{3/2}\pi\eta\frac{A}{(1+A)^2}[A-2x(1+A)], \tag{3.7}$$

where we introduced the composition, x, defined by

$$x = \frac{N_2}{N},. \tag{3.8}$$

Eq. (3.7) allows us to redefine the total (reduced) unit cell energy per particle as

$$\frac{E^*(\mathbf{r}^N;A,\eta)}{N} = \frac{1}{N}\left[E^*(\mathbf{r}^N;A_0,\eta) - E^*_A(A_0,\eta,x) + E^*_A(A,\eta,x)\right] \tag{3.9}$$

and we use $E^*(\mathbf{r}^N;A,\eta,x) \equiv E^*(\chi;A,\eta)$ occasionally; notably, the composition $x = N_2/N$ can be evaluated by the vertical particle arrangement of a particular configuration, *i.e.*, by the z coordinates in $\mathbf{r}^N$, in Eq. (3.9).

With this the computational overhead of ground state identification in the entire (η, A)-plane can be greatly reduced since the energies of geometrically optimized structures (at fixed plate separation, η, and composition, x) at one specific value of A_0 – usually and without loss of generality at $A_0 = 0$ [62–64] – can be compared for all values of A using Eq. (3.9).

To be more specific, the geometric ground state configurations have been determined in [62] by independent evolutionary searches at a fixed value of $A = 0$ for different, numerically tractable values of the composition[9], $x = N_2/N \in [0, 0.5]$, with N ranging from one to 40 and N_2 from zero to 20 ($N = 40$ represents a numerical limit for the evolutionary algorithm [62]) for different values of the plate separation η. Such a set of structural ground state configurations of different compositions which are identified at $A = 0$ but at a particular value of η provide all necessary information to subsequently identify the ground state configuration for any state point (η, A) and renders the approach very efficient. For a given (η, A)-pair the configuration which minimizes $E^*(\mathbf{r}^N;A,\eta,x)/N$ given by Eq. (3.9) is considered as the ground state and we label the ground state energy as a function of η and A as $E^*_{\mathrm{GS}}(A,\eta)/N$. Henceforward, we usually drop the arguments of the energies for different configurations – unless we want to emphasize the dependency of certain arguments – and we synonymously use E^*/N for Eq. (3.9) and E^*_{GS}/N for the ground state energy in the entire parameter space.

[9] The total number of compositions at each value of η is $N_{\mathrm{tot}} = 1 + \frac{1}{2}\sum_{n=2}^{N}(n - n \bmod 2) = 401$ for $N = 40$, since $N_2 = 0$ for $N_1 = 1$ and $0 < N_2 \leq \frac{N_1}{2}$ for $N_1 > 1$ was chosen.

3.1.4 Revisiting the Complex Phase Diagram with Unsupervised Clustering Algorithms

The structural ground states of the asymmetric Wigner bilayer have intensively been studied with the help of evolutionary algorithms and Monte Carlo simulations[10] in Refs. [62–64]. A cascade of ordered patterns has been observed when changing the system parameters, *i.e.*, the layer distance, η, or the plate charge asymmetry, A, see Fig. 3.2. The structural versatility of the ordered ground states at different system parameters is based on the competition between commensurability of the attained configurations in both layers and charge neutralization of the long range Coulomb interaction [64].

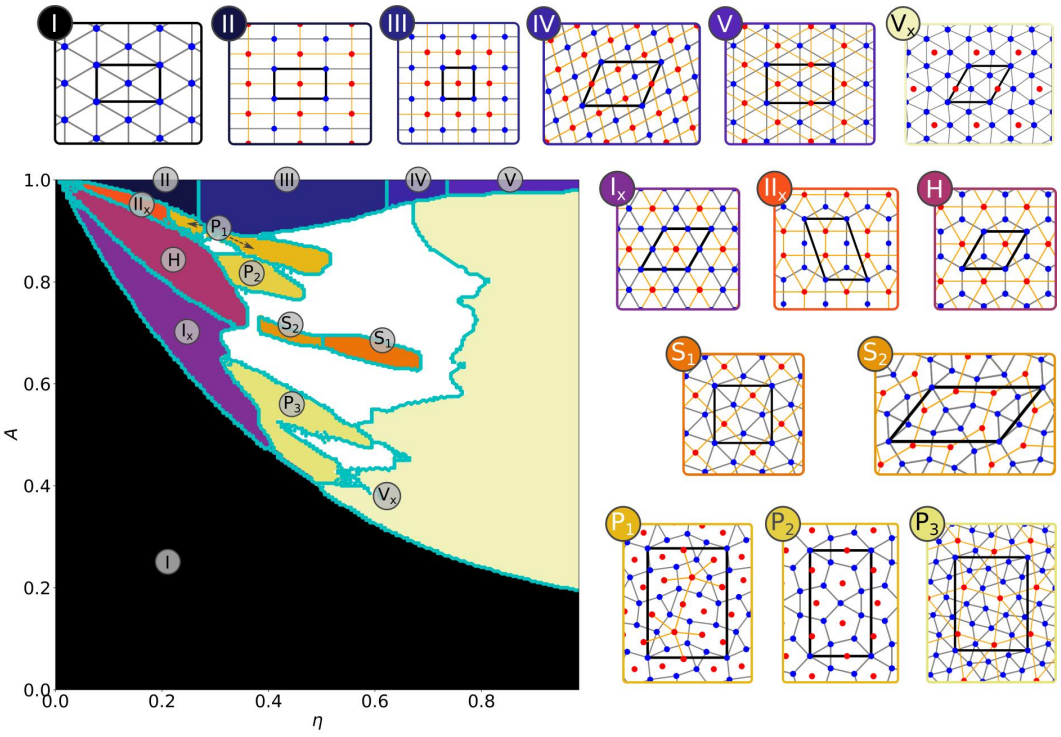


Figure 3.2: Redrawn phase-diagram from Refs. [62–64]. Colored and labeled regions denote identified phases with structures indicated by correspondingly labeled subfigures (layer one particles are colored blue, layer two particles are colored red and the thick black frames indicate the unit cells of the bilayer structures). The cyan lines in the left panel highlight the phase boundaries. Structures within the white region have not been classified yet.

Several phases were classified in Refs. [62–64] in the categories shown in Fig. 3.2: to be more precise, 14 structures, or rather families of structures, can be related to special values, or features of order parameters, such as the composition x defined by Eq. (3.8) and bond-orientational order parameters [225–227] defined in Subsection 2.3.2 – which we will discuss in detail below.

[10] The algorithms are presented in Subsections 2.2.3 and 2.2.4.

Despite giving broad insights into the phase-diagram the phase classification scheme suggested in Refs. [62–64] and depicted in Fig. 3.2 draws not a complete picture of the emerging phases. There is still a large region in the phase-diagram where the respective ground state structures – due to the considerable structural complexity – have not yet been classified (cf. white region in Fig. 3.2). Furthermore, there is strong evidence [276] that other regions in the diagram – such as the I_x region – can be subdivided into more specific structures of their own kind.

In this Subsection we present a route towards (i) systematically identifying families of structures using unsupervised clustering algorithms and (ii) we use this information to highlight sweet-spots in the phase-diagram which hint at possible quasicrystalline order of dodecagonal (and possibly even decagonal) rotational symmetry in the asymmetric Wigner bilayer system.

3.1.4.1 Order Parameters and Phase Relations

As previously mentioned, the ground state phase-diagram of the asymmetric Wigner bilayer system [62] shown in Fig. 3.2 was drawn by relating phases to special features of order parameters. A labeling of bilayer configurations into structure-families has been performed in Refs. [62–64, 146, 147] according to selection rules based on characteristic features of certain order parameters (cf. Table 3.1).

This labeling mainly relies on the composition $x = N_2/N$, defined in Eq. (3.8), and on different variants of so called bond-orientational order parameters (BOOPs) [225–227], Ψ_ν, which we introduced in Subsection 2.3.2 by Eq. (2.62). Four different types of these BOOPs have been defined in Ref. [62] namely

(1) $\Psi_\nu^{(1)}$, quantifying ν-fold bond-orientational order of particles in layer one,

(2) $\Psi_\nu^{(2)}$, quantifying ν-fold bond-orientational order of particles in layer two,

(3) $\Psi_\nu^{(3)}$, quantifying ν-fold bond-orientational order of particles of both layers projected onto the same plate, and

(4) $\Psi_\nu^{(4)}$, quantifying ν-fold bond-orientational order of particles of layer two, considering only layer one particles (projected onto the same layer) as neighbors.

We here introduce a short-hand notation for addressing a set of $u_i = u_1, \ldots, u_n$ different BOOPs, $\Psi_{\nu_j}^{(u_i)}$, of different bond-orientational order, $\nu_j = \nu_1, \ldots, \nu_m$, via $\Psi_{[\nu_1,\ldots,\nu_m]}^{(u_1,\ldots,u_n)} = \{\Psi_{\nu_1}^{(u_1)}, \ldots, \Psi_{\nu_m}^{(u_1)}, \Psi_{\nu_1}^{(u_2)}, \ldots, \Psi_{\nu_m}^{(u_2)}, \ldots, \Psi_{\nu_1}^{(u_n)}, \ldots, \Psi_{\nu_m}^{(u_n)}\}$; if only one lower index is present, e.g. $[\nu]$, the lower brackets may also be omitted and we may write $\Psi_\nu^{(u_1,\ldots,u_n)} = \{\Psi_\nu^{(u_1)}, \ldots, \Psi_\nu^{(u_n)}\}$.

Twelve of these order parameters have been used in Ref. [62], *i.e.*, $\nu_j = 4, 5, 6$-fold BOOPs of all four types, $\Psi_{[4,5,6]}^{(1,2,3,4)}$. The geometrical classification power of these order parameters is demonstrated in the top left panel of Fig. 3.3 which shows the so called "hole" BOOPs $\Psi_{[4,5,6]}^{(4)}$ with $4, 5, 6$-fold order for the currently suggested ground state configurations in the (η, A)-plane in an [R, G, B]-scheme [62–64]: the value of the $\nu_1 = 4$-fold order parameter gives the relative amount of red, the $\nu_2 = 5$-fold order

parameter the relative amount of green and the $\nu_3 = 6$-fold order parameter the relative amount of blue color[11]; every color channel (*i.e.*, the relative amounts of red, green and blue) is normalized to a minimum value of zero and a maximum value of one. Note that in Ref. [62] the specific features in the order parameters which are used to label a certain family of structures (*i.e.*, to label different phases) were carefully chosen according to Table 3.1, such that equally labeled structures exhibit similar values of the bond-orientational order parameters.

Label	Description	Characteristic Features
I	hexagonal monolayer	$x = 0$
II	rectangular bilayer	$x = \frac{1}{2}$, $\Psi_4^{(1,2)} = 1$, $0 < \Psi_6^{(1,2)} < 1$
III	square bilayer	$x = \frac{1}{2}$, $\Psi_4^{(1,2)} = 1$, $\Psi_6^{(1,2)} = 0$
IV	rhombic bilayer	$x = \frac{1}{2}$, $0 < \Psi_4^{(1,2)} < 1$, $0 < \Psi_6^{(1,2)} < 1$
V	hexagonal bilayer	$x = \frac{1}{2}$, $\Psi_4^{(1,2)} = 0$, $\Psi_6^{(1,2)} = 1$
I_x	trihexagonal (layer one)	$0 < x < \frac{1}{3}$, $\Psi_6^{(3)} > 0.9$
H	honeycomb (layer one)	$x = \frac{1}{3}$, $\Psi_6^{(3)} > 0.9$
II_x	modified rectangular bilayer	$\frac{1}{3} < x < \frac{1}{2}$, $\Psi_6^{(3)} > 0.9$
V_x	hexagonal bilayer	$0 < x < \frac{A}{1+A}$, $(1-x)\Psi_6^{(1)} + x\Psi_6^{(2)} > 0.9$
S_1	snub square (layer one)	$x = \frac{2}{6}$, $\Psi_5^{(1)} > 0.7$, $\Psi_4^{(2)} > 0.9$
S_2	snub square like (layer two)	$x = \frac{2}{6}$, $\Psi_5^{(2)} > 0.45$
P_1	pentagonal type two	$\frac{1}{3} < x < \frac{1}{2}$, $\Psi_5^{(2)} > 0.45$
P_2	pentagonal holes	$\frac{1}{3} < x < \frac{1}{2}$, $\Psi_5^{(4)} > 0.9$
P_3	pentagonal holes	$0 < x < \frac{1}{3}$, $\Psi_5^{(4)} > 0.9$

Table 3.1: Structure labels, description and characteristic features of the current phase-diagram of the asymmetric Wigner bilayer system from literature [62–64] shown in Fig. 3.2.

Here we are aiming at automatically identifying phases in the asymmetric Wigner bilayer system using a more systematic classification scheme based on unsupervised clustering (cf. Section 2.4) of a data set of order parameters which corresponds to the database of structural ground state candidates of the system from Refs. [62–64]. We thus introduce additional bond-orientational order parameters with $\nu = 3, 8, 10, 12$-fold symmetry for all four types of BOOPs, which are, in combination with the order parameters of Refs. [62–64], hopefully sufficient for the clustering algorithm to also identify structures with eightfold, tenfold or twelvefold-symmetric traits. (These newly introduced BOOPs can be written in short-hand notation as $\Psi_{[3,8,10,12]}^{(1,2,3,4)}$.) In Fig. 3.3 we present the ground state phase-diagram of the asymmetric Wigner bilayer system in a similar [R, G, B]-scheme as introduced for the $\Psi_{[4,5,6]}^{(4)}$ BOOPs (cf. top left panel in Fig. 3.3) where we use some of the 16 newly introduced BOOPs as [R, G, B] values in the color coding (*i.e.*, $[\Psi_4^{(2)}, \Psi_6^{(2)}, \Psi_{12}^{(2)}]$ in the top right and $[\Psi_{10}^{(2)}, \Psi_5^{(4)}, \Psi_6^{(1)}]$ in the bottom left panels of Fig. 3.3).

A first, visual inspection of the top right panel of Fig. 3.3, showing $[\Psi_4^{(2)}, \Psi_6^{(2)}, \Psi_{12}^{(2)}]$ in an [R, G, B]-scheme, indeed reveals an area of interest with a strong (blue) signal of

[11] The values of any BOOP defined by Eq. (2.62) is bound between zero and one.

the dodecagonal order parameter of plate two, *i.e.*, of $\Psi_{12}^{(2)}$, below the honey-comb, H, and within the I_x-region (cf. Fig. 3.1(c)), while $\Psi_4^{(2)}$ and $\Psi_6^{(2)}$ are not as pronounced; we will continue this discussion in Subsection 3.1.6.

Furthermore, the bright whitish/greenish area in the P_3 region of the bottom left panel of Fig. 3.3 which depicts $[\Psi_{10}^{(2)}, \Psi_5^{(4)}, \Psi_6^{(1)}]$ (also in an [R, G, B]-scheme) indicates decagonal (or at least pentagonal) ordering in layer two via large values of $\Psi_{10}^{(2)}$. The pentagonal "holes" BOOP, $\Psi_5^{(4)}$, and the hexagonal BOOP in layer one, $\Psi_6^{(1)}$ also obtain large values; we will come back to this particular combination of BOOPs in Subsection 3.1.7.

In an effort to quantify the ratio of the average nearest neighbor distance in layer one, $r_{\mathrm{nn}}^{(1)}(\mathcal{X})$, and the average nearest neighbor distances in layer two, $r_{\mathrm{nn}}^{(2)}(\mathcal{X})$, for a certain bilayer configuration, $\mathcal{X}$, we define the *intralayer nearest neighbor ratio* order parameter, $r_g(\mathcal{X})$, as

$$r_g(\mathcal{X}) = \frac{r_{\mathrm{nn}}^{(1)}(\mathcal{X})}{r_{\mathrm{nn}}^{(2)}(\mathcal{X})}, \tag{3.10}$$

where we assume for monolayer structures, $\mathcal{X}_m$, with $N_2(\mathcal{X}_m) = 0$ that $r_{\mathrm{nn}}^{(2)}(\mathcal{X}_m) = \infty$ such that $r_g(\mathcal{X}_m) = 0$. In the bottom right panel of Fig. 3.3 we present $r_g \in [0, 1.07]$ for the ground state configurations of the asymmetric Wigner bilayer system suggested in Refs. [62–64] in the (η, A)-plane. Note that the $r_g(\mathcal{X})$ order parameter is not bound to a maximum value of unity, but we empirically find an upper limit of $\max(r_g(\mathcal{X})) \approx 1.07$ for all considered bilayer ground state configurations from the literature database [62–64] with $x \in [0, 0.5]$.

We define the feature vector, $\mathbf{x}(\mathcal{X})$, by

$$\mathbf{x}(\mathcal{X}) = \{f_1(\mathcal{X}), \ldots, f_{N_f=30}(\mathcal{X})\} \tag{3.11}$$

collecting all $i = 1, \ldots, 30$ features, $f_i(\mathcal{X})$, of the 28 bond-orientational order parameters, $\Psi_{[3,4,5,6,8,10,12]}^{(1,2,3,4)}(\mathcal{X})$ (written in shorthand notation), the value of $r_g(\mathcal{X})$ and the composition $x(\mathcal{X})$ related to the respective structure $\mathcal{X}$.

Henceforward, we may drop the argument of feature vectors, $\mathbf{x}(\mathcal{X})$, and of the different features, $f_i(\mathcal{X})$, and simply write $\mathbf{x}$ and f_i unless emphasizing $\mathcal{X}$ is explicitly necessary for the discussion. Analogously, we may refer to $\Psi_{[3,4,5,6,8,10,12]}^{(1,2,3,4)}(\mathcal{X})$, $r_g(\mathcal{X})$ and $x(\mathcal{X})$ simply as $\Psi_{[3,4,5,6,8,10,12]}^{(1,2,3,4)}$, r_g and x if no particular configuration is addressed.

3.1.4.2 Unsupervised Clustering: Outline

To make the order parameter based classification more rigorous, we will extend the phase-classification scheme of Refs. [62–64] with the help of unsupervised machine-learning techniques in order to automatically identify different families of structures directly from the feature vector, $\mathbf{x}$, given by Eq. (3.11) (*i.e.*, by the set order parameter evaluated for a given bilayer structure). To be more specific, we perform a principal component analysis [238] (PCA) on the feature vectors of all structures from Ref. [62]

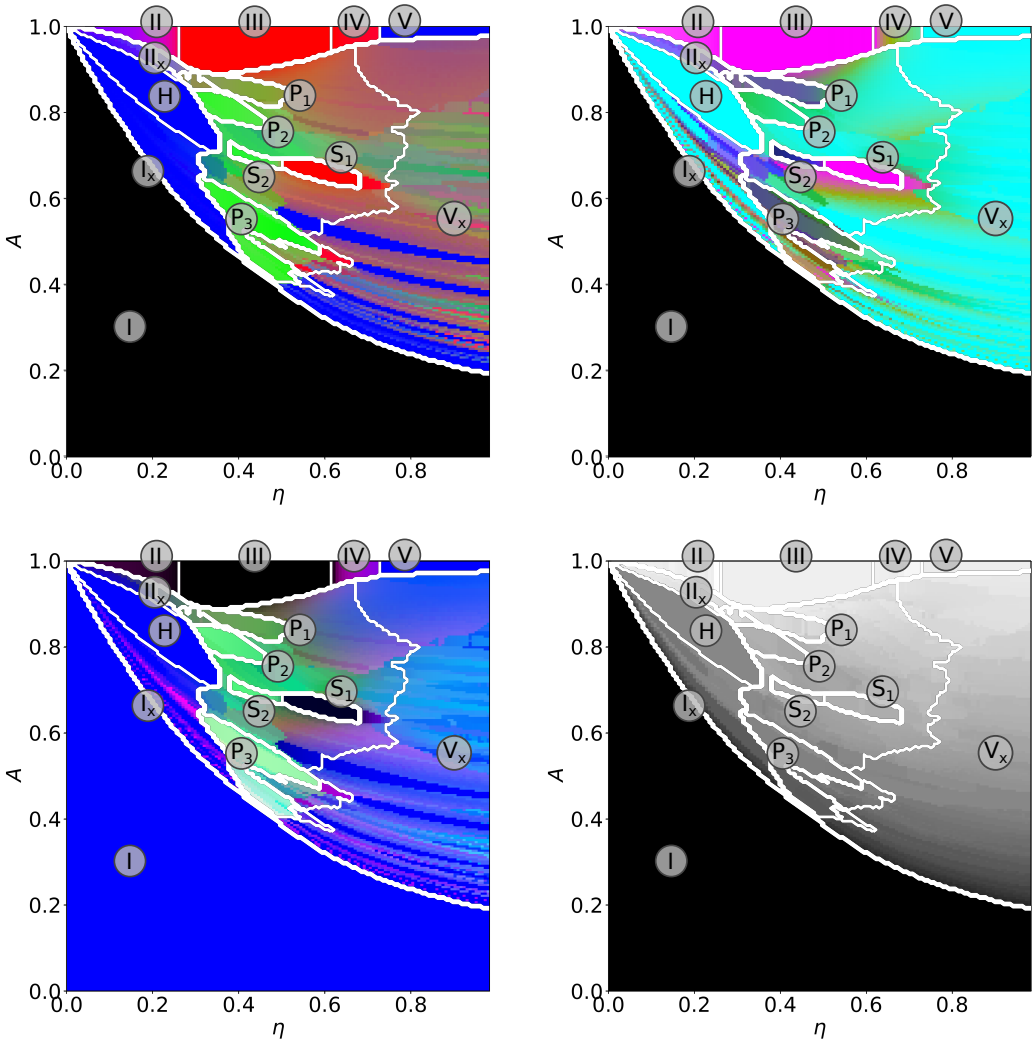


Figure 3.3: Phase diagram of the asymmetric Wigner bilayer system colored by bond-orientational order parameters (BOOPs) in a [R, G, B] color-scheme for $[\Psi_4^{(4)}, \Psi_5^{(4)}, \Psi_6^{(4)}]$ (top left), and the newly introduced order parameters $[\Psi_4^{(2)}, \Psi_6^{(2)}, \Psi_{12}^{(2)}]$ (top right) and $[\Psi_{10}^{(2)}, \Psi_5^{(4)}, \Psi_6^{(1)}]$ (bottom left) and in gray-scale for the intralayer nearest neighbor ratio order parameter r_g defined by Eq. (3.10) (bottom right) of the ground state configurations of the asymmetric Wigner bilayer system in the (η, A)-plane [62–64]. The color-scheme of the top panels and the bottom left panel is given by an [R, G, B]-scheme with respect to the three indicated BOOPs of each panel which define the relative amount of [red, green, blue] for each pixel representing an (η, A)-pair ($\mathrm{red} = \Psi_4^{(4)}$, $\mathrm{green} = \Psi_5^{(4)}$, $\mathrm{blue} = \Psi_6^{(4)}$ for the top left panel and analogously for the other panels showing other BOOP combinations); each color channel [red, green, blue] is normalized to attain values between zero and one. The color-scheme of the bottom right panel is in gray-scale such that $r_g = 0$ is colored black and $r_g = 1.07$ is colored light-gray (to be distinguishable from the white colored phase boundaries). Phase boundaries from literature [62–64] are indicated by thick white lines and the identified phases are labeled according to Table 3.1.

(which defines our data set). This allows us to identify directions of large variance in the data set[12] – holding the most information amongst the different features, f_i, cf. Subsection 2.4.1.

We then apply k-means [239–242] clustering[13,14] to the latent space representation of the data set (spanned by the leading principal components) in order to identify new, previously unclassified phases which are potentially hidden in the large set of the original structural data from Refs. [62–64].

As a benchmark for this approach and for illustrative reasons we start our analysis with the simplest problem in the domain of this Subsection, namely with the symmetric Wigner bilayer system [146, 147] with $\sigma_1 = \sigma_2$ or, equivalently, $A = 1$.

3.1.4.3 k-means Clustering of Structural Data of the Symmetric Wigner Bilayer System

In the symmetric case with $A = 1$ the ground state pattern formation of the Wigner bilayer system can be solved analytically [146, 147] and five emerging phases, *i.e.*, I through V depicted in Fig. 3.4, can be identified: the hexagonal monolayer (I) is stable only at $\eta = 0$ and transforms in a rectangular bilayer (II), which is stable at $0 < \eta \lesssim 0.263$. Increasing η further leads to a second order transition into a square bilayer (III), stable at $0.263 \lesssim \eta \lesssim 0.621$, which transforms – again as second order transition – into a rhombic bilayer phase (IV), stable at $0.616 < \eta \leq 0.728$. For $0.728 < \eta$ a hexagonal bilayer (V) emerges (also see $A = 1$ line in Fig. 3.2).

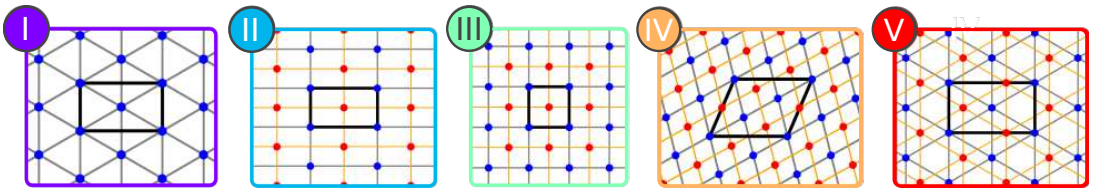


Figure 3.4: Phases I through V (color coded in purple, blue, green, orange and red) representing ground state solutions of the symmetric Wigner bilayer system [146, 147]. The presented structures are archetypical structures of phases I through V identified by the evolutionary algorithm in Ref. [62] and the respective unit cells are indicated by thick black frames.

We now want to apply the clustering approach outlined above (cf. Subsections 2.4.1 and 2.4.2) in order to identify these phases directly from the order parameters of

[12] Actually, we first transform the data set of feature vectors to unit-variance and zero-mean coordinates (for each feature, f_i), a technique called whitening used to unbias PCA from the relative scale of different features [119].

[13] The k-means algorithm identifies groups of data points by iteratively reassigning data points to one of the k groups (the number of groups, k, is fixed during the entire algorithm) with the nearest group average, also called the group mean. In a second state group averages are redefining when data points have been reassigned and both steps are repeated until a convergence criterion is met, see Subsection 2.4.2 for details.

[14] Other approaches, such as DBSCAN of principal components or of t-SNE manifolds [244] similar as in [86] have not been very fruitful approaches in this perspective, see Appendix A.1.1 for details.

all $N_{\text{sym}} = 141$ ground states identified by the memetic evolutionary algorithm (cf. Subsection 2.2.4) in Ref. [62] for different values of $\eta \in [0, 1]$ at $A = 1$.

We first perform a principal component analysis (PCA) [238] on the set of (unit-variance and zero-mean) feature vectors, $\mathbf{X}^{(\text{sym})} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\text{sym}}}\}$ with $\mathbf{x}_i \in \mathbb{R}^{N_f=30}$ defined in Eq. (3.11), of all structures and transform the data to an N_f^*-dimensional latent space representation, $\mathbf{L}^{(\text{sym})} = \{\mathbf{l}_1, \ldots, \mathbf{l}_{N_{\text{sym}}}\}$ with $\mathbf{l}_i = (v_{i1}, v_{i2}, \ldots, v_{iN_f^*}) \in \mathbb{R}^{N_f^*}$ (see Subsection 2.4.1). The numerical value of $N_f^* \leq N_f$ defines how many leading principal components are considered in the latent space representation of the data and by default $N_f^* = N_f$. In the top panels of Fig. 3.5 we visualize all $i = 1, \ldots, N_{\text{sym}}$ data points as a function of the first three principal components, *i.e.*, we present the coordinates, v_{i1}, v_{i2} and v_{i3} corresponding to the data points $\mathbf{x}_i$ in the first three latent space directions $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ (cf. Subsection 2.4.1); we color-code the data points according to the different phases I through V they correspond to, following the color-scheme of Fig. 3.4.

As illustrated in Fig. 3.5 the leading three principal components already allow – to some extent – to distinguish the different phases by eye: data points (*i.e.*, structures) belonging to a certain phase are organized in clusters in the latent space and clusters of different structures are – to some extent – spatially separated from each other. In the bottom panel of Fig. 3.5 we present the explained variance $\lambda_j^{(\text{e})}$, defined by Eq. (2.75), for all $j = 1, \ldots, 30$ principal components. The explained variance quantifies the amount of information encoded in each principal component direction $\mathbf{v}_j$. The value of $\lambda_j^{(\text{e})}$ drops quickly from $\lambda_1^{(\text{e})} \sim 1/3$ to $\lambda_6^{(\text{e})} < 5\%$, and further by several orders of magnitudes, such that the higher principal components are insignificant compared to the leading ones. Thus, we can safely restrict ourselves to the leading five principal components and set $N_l = N_f^*{=}5$ for the here studied symmetric case of the Wigner bilayer system.

We now perform k-means clustering (cf. Subsection 2.4.2) on the $N_l{=}5$ dimensional latent space representation $\mathbf{L}^{(\text{sym})}$ of the data $\mathbf{X}^{(\text{sym})}$ (corresponding to the leading five principal components) to assign all $i = 1, \ldots, N_{\text{sym}}$ data points with a cluster label $c_i \in \{1, \ldots, K\}$ and thereby define the labeling (or clustering) $\mathbf{k}^{(\text{sym})} = \{c_1, \ldots, c_{N_{\text{sym}}}\}$ of the data set. In the particular case of the symmetric Wigner bilayer system we know the numbers of phases and we set $K = 5$. As can be seen in Fig. 3.6 the "$K{=}5$-means" clustering, $\mathbf{k}^{(\text{sym})}$, of the data is in excellent agreement with the phase-assignment from literature [146, 147], $\mathbf{w}^{(\text{sym})} = \{C_1, \ldots, C_{N_{\text{sym}}}\}$, where $C_i = 1$ through 5 label the corresponding phase I through V, respectively for every data point i by following Table 3.1. Notably, the particular numerical values that associate the data points with a certain cluster are unique but usually arbitrarily chosen by the k-means algorithm: in Fig. 3.6 we see that the clusters related to phases I through V are respectively labeled by $c_i = 5, 1, 3, 4$ and 2. However, the clustering of the data points into the different phases is almost perfect and the labels c_i can be redefined to match the numerical values of C_i (not shown here). We report an (adjusted) mutual information score, defined by Eq. (2.84), of $I_K(\mathbf{k}^{(\text{sym})}, \mathbf{w}^{(\text{sym})}) = 94.97\%$ between the clustering $\mathbf{k}^{(\text{sym})}$ and the phase-assigment from literature $\mathbf{w}^{(\text{sym})}$ [146, 147]. Only two data points, which are emphasized in Eq. (2.84) as *outliers*, are assigned with a wrong label by the $K{=}5$-means clustering algorithm (see discussion below).

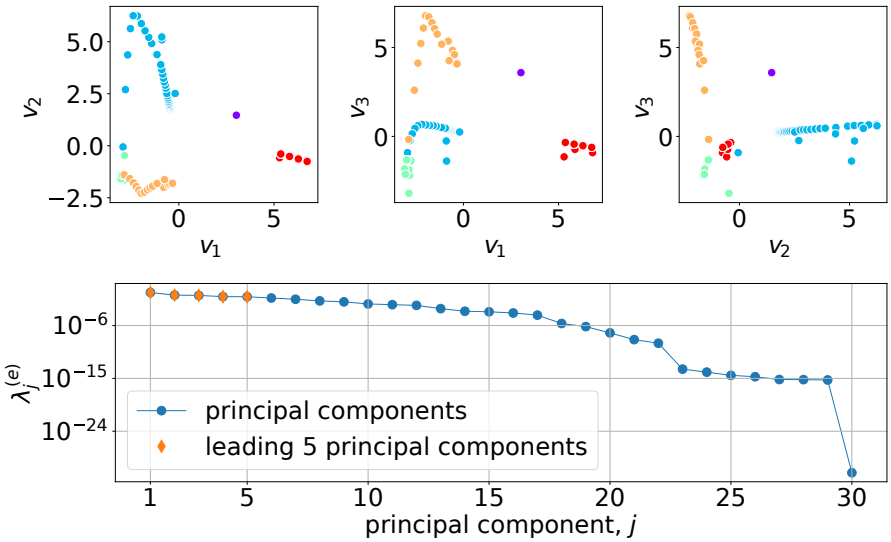


Figure 3.5: Top panel: ground state structures of the symmetric Wigner bilayer system, $A = 1$, for different value sof η represented as a function of the first three principal components, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ of the latent space representation $\mathbf{L}^{(\mathrm{sym})}$ of the data set $\mathbf{X}^{(\mathrm{sym})}$. Each data point represents one symmetric Wigner bilayer structure and we color-code the data points according to the phase they belong to (*i.e.*, I through V, cf. Fig. 3.4). Bottom panel: Explained variance, $\lambda_j^{(\mathrm{e})}$ defined by Eq. (2.75) for all $j = 1, \ldots, 30$ principal components (blue) and for the leading five principal components (orange).

In the right panel of Fig. 3.6 we present a t-SNE [244] analysis (see Subsection 2.4.4 for details) of the five leading principal components of the data points in $\mathbf{X}^{(\mathrm{sym})}$ which are thereby mapped (and visually clustered by similarity) into a two-dimensional t-SNE latent-space manifold, spanned by the axes "t-sne 1" and "t-sne 2". t-SNE is a very useful tool to present high-dimensional data in a low-dimensional latent space (here given by the two axes "t-sne 1" and "t-sne 2"). Indeed, the different phases can be distinguished in this two-dimensional visualization of the data points by spatially well separated clusters of, respectively, similar data points (except[15] maybe for phase I and V)

The results from k-means clustering may depend on the initial conditions of the algorithm such as (i) the initial, usually arbitrary placement of the K different cluster centers in the latent space of the data set $\mathbf{X}^{(\mathrm{sym})}$ and (ii) the initial data point assignments to the clusters. To justify the results shown in Fig. 3.6 we thus perform $l = 1, \ldots, 100$ independent runs of k-means clustering on the leading five principal components of the data set $\mathbf{X}^{(\mathrm{sym})}$ and we find that all corresponding labelings, $\mathbf{k}_l^{(\mathrm{sym})}$, share the same adjusted mutual information, defined by Eq. (2.84), of $I_K(\mathbf{k}_l^{(\mathrm{sym})}, \mathbf{w}^{(\mathrm{sym})}) = 94.57\%$ with the results from literature $\mathbf{w}^{(\mathrm{sym})}$. Furthermore we varied the number of leading principal components from five to 30 without observing significant changes in the results

[15] The proximity of phases I and V in the t-SNE plot shown in the right panel of Fig. 3.6 could be owed to the fact that the hexagonal monolayer,*i.e.*, phase I, only appears once in the data set, *i.e.*, at $\eta = 0$, while there is several structures of the family V, *i.e.*, of hexagonal bilayer structures.

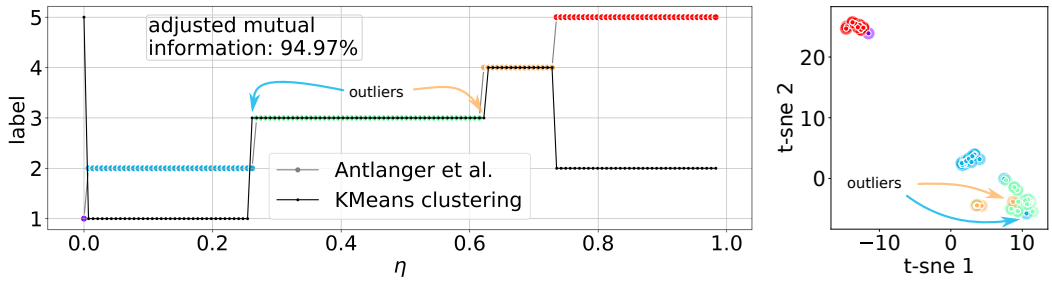


Figure 3.6: Left: Labeling $\mathbf{w}^{(\mathrm{sym})}$ [146, 147] of the ground states of the symmetric Wigner bilayer system, *i.e.*, I: 1, II: 2, III: 3, IV: 4 and V: 5 (color-coding according to Fig. 3.4), and labeling $\mathbf{k}^{(\mathrm{sym})}$ by the k-means clustering (black) for every data point $i = 1, \ldots, N_{\mathrm{sym}}$ of the data set $\mathbf{X}^{(\mathrm{sym})}$ identified in Ref. [62] for different values of $\eta \in [0, 1]$ at $A = 1$. Note that the numerical value of a particular cluster label $c_i = 1$ through 5 assigned by the k-means algorithm to all data points belonging to one particular cluster is unique but arbitrary. An almost perfect mapping between the data points assigned to phases I through V and the respective clusters of data points labeled by $c_i = 5, 1, 3, 4$ and 2, respectively, can be achieved: an adjusted mutual information score, defined by Eq. (2.84), of $I_K(\mathbf{k}^{(\mathrm{sym})}, \mathbf{w}^{(\mathrm{sym})}) = 94.57\%$ is realized and only two outliers (highlighted by arrows) amongst all data points occur in the clustering. Right: Mapping of the N_l=5-dimensional latent space representation, $\mathbf{L}^{(\mathrm{sym})}$, of the data $\mathbf{X}^{(\mathrm{sym})}$ (corresponding to the leading five principal components) onto a two-dimensional t-SNE manifold [244] (t-sne 1 and t-sne 2). Each data point, i, is color-coded according to its label, c_i, assigned by the k-means clustering, $\mathbf{k}^{(\mathrm{sym})} = \{c_1, \ldots, c_{N_{\mathrm{sym}}}\}$, such that clusters labeled by $c_i = 5, 1, 3, 4$ and 2 appear in the same colors as the corresponding phases I through V following the color-scheme in the left panel and in Fig. 3.4. The two outliers are again indicated by arrows.

(when using less than five principal components the results become unreliable). This suggests that the clustering shown in Fig. 3.6 is indeed the optimal k-means clustering to group the data points of $\mathbf{X}^{(\mathrm{sym})}$ into the phases I through V.

Interestingly, there is a small discrepancy in labeling the structures at the transition from phase II to III as well as from III to IV: the k-means algorithm identifies a tiny portion of configurations to be part of phase III instead of phases II and IV to both sides of the respective η range. The bright, green colored circles in the right panel of Fig. 3.6 correspond to phase III and from a visual inspection of the t-SNE plot the two outliers from phase II and IV (small orange and blue dots within the cluster of large green background dots) should rather be associated to phase III. The respective order parameters that characterize the phases of these two outlier structures according to Table 3.1 are given in Table 3.2:

$\mathbf{w}^{(\mathrm{sym})}$ label [146, 147]	$\mathbf{k}^{(\mathrm{sym})}$-means label	x	$\Psi_4^{(1)}$	$\Psi_4^{(2)}$	$\Psi_6^{(1)}$	$\Psi_6^{(2)}$
II	III	0.5	1	1	0.025	0.025
IV	III	0.5	0.949	0.949	0.125	0.125

Table 3.2: Phase classification and order parameters of outliers shown in Fig. 3.6 (see also text).

Clearly, these two data points can respectively be labeled as phase II and IV from the point of view of an unsupervised clustering algorithm as the transition from II

to III and III to IV is of second order [62, 146, 147]: structure III (square bilayer) can be considered as a special case of both phases, II (rectangular bilayer) and IV (rhombic bilayer), and transitions of II →III and of III →IV are continuous. However, the distinction between a rectangular bilayer, a square bilayer or a rhombic bilayer configuration can be performed analytically by clearly associated values and boundaries of some order parameters to certain phases (see Table 3.2) [146, 147]. The clustering results are simply wrong here but increasing the resolution in η could most likely resolve this issue.

In general, small variations in the data, which are often related to artifacts such as noise, can trigger undesired effects in clustering approaches and may lead to artificial partitioning of data in a clustering or classification task. A proper preparation of the data with, for instance, PCA can help to reduce the effects of noise on the outcome of a clustering approach of a particular data set. However, sometimes small variations in the data indeed have physical meaning such as, for instance, in continuous phase transitions and great caution needs to be taken for correctly distinguishing between different clusters of data points.

Nevertheless, clustering tools can be incredibly useful for establishing first insights into a large amount of data as well as to gain a deeper understanding of hidden features in a data set which are often too complicated to identify manually.

3.1.4.4 Principal Component Analysis of Structural Data of the Asymmetric Wigner Bilayer System

In the symmetric Wigner bilayer system discussed above with $\sigma_1 = \sigma_2$ five ground state configurations, I through V, emerge at different values of η. The energies of these structures can be evaluated with analytical tools [146, 147] and the corresponding phase boundaries can subsequently be calculated with high precision. In contrast, identifying ground state structures of the asymmetric case where $\sigma_1 \neq \sigma_2$ (see Fig. 3.2) is much more complicated and one has to rely on global optimization tools such as memetic evolutionary algorithms (cf. Subsection 2.2.4) to investigate the ground states of the system [62]. The suggested phase boundaries of the asymmetric case in Ref. [62] mainly rely on the so-called bond orientational order parameters. These boundaries were identified such as that structures pertaining to certain categories – which are labeled according to Table 3.1 – occupy regions in the phase-diagram of similar symmetry. In the top left panel of Fig. 3.3 we see, for instance, that the $\Psi^{(4)}_{[4,5,6]}$ order parameters are good candidates for describing the symmetries of the ground state configurations of the asymmetric Wigner bilayer system for large, extended regions in the (η, A)-plane [62–64]. Thus, the phase classification scheme given by Table 3.1 was obtained [62] by relating phases of structures with monochrome regions in $[\Psi^{(u_1)}_{\nu_1}, \Psi^{(u_2)}_{\nu_2}, \Psi^{(u_3)}_{\nu_3}]$-based [R, G, B] ground state phase-diagrams similar to the one presented in the top left panel of Fig. 3.3. This classification scheme has high predictive power but is, to some extent, arbitrary.

We have demonstrated in Subsection 3.1.4.3, that a combination of (i) principal component analysis [238] (PCA) and (ii) successive k-means clustering [239–242], $\mathbf{k}^{(\mathrm{sym})}$, of a data set of order parameters of structural data of the symmetric Wigner bilayer system,

$\mathbf{X}^{(\mathrm{sym})}$ [62–64], is in excellent agreement with analytic labeling results, $\mathbf{w}^{(\mathrm{sym})}$ [146, 147]. The phase-diagram in η can (almost perfectly) be recovered, however, correctly identifying second order transitions with the current set of order parameters turned out to be numerically delicate (cf. Fig. 3.6). We would like to emphasize, that we did not only performed this clustering analysis of structural data of the symmetric Wigner bilayer system as a benchmark for the PCA and the k-means clustering algorithm but also for pedagogical reasons to tackle the much more involved asymmetric case below.

Following the approach in Subsection 3.1.4.3, we here perform the same analysis (PCA and k-means clustering) on the set of feature vectors, $\mathbf{X}^{(\mathrm{asym})} = (\mathbf{x}_1, \cdots, \mathbf{x}_{N_{\mathrm{asym}}})$ with $\mathbf{x}_i \in \mathbb{R}^{N_f=30}$ being defined by Eq. (3.11), for the entire database of $i = 1, \ldots, N_{\mathrm{asym}}$ structures considered in the asymmetric Wigner bilayer system [62–64]. The structural database of $N_{\mathrm{asym}} \sim 64000$ configurations was generated by independent evolutionary structure optimization of configurations with $N = 1$ to $N = 40$ particles per unit cell and all possible values of the composition, $x = N_2/N$, at 140 different values of $\eta \in [0, 1.4]$ (with a step size of $\Delta\eta = 0.01/\sqrt{2}$) at $A = 0$ (see Subsection 3.1.3 for further details). At any point in the (η, A)-plane the particular bilayer configuration which minimizes $E^*(\mathcal{X}_i; A, \eta)/N_i$, defined by Eq. (3.9), is assigned as ground state amongst all 401 competing structures for each (η, A)-tuple [62]; $\mathcal{X}_i$ is defined by Eq. (3.5) and represents a particular realization of the set of variable attributes of a configuration with composition x_i and N_i particles per unit cell from the database of $i = 1, \ldots, N_{\mathrm{asym}}$ structures.

Via PCA we transform the data set $\mathbf{X}^{(\mathrm{asym})}$ to a latent space representation $\mathbf{L}^{(\mathrm{asym})} = (\mathbf{l}_1, \ldots, \mathbf{l}_{N_{\mathrm{asym}}})$ of the data, with $\mathbf{l}_i \in \mathbb{R}^{N_f^*}$ being the vector of the data point $\mathbf{x}_i \in \mathbb{R}^{N_f}$ in the latent space of dimension $N_f^* \leq N_f$ (the numerical value of N_f^* is yet to be defined). In Fig. 3.7 we present the explained variance, $\lambda_j^{(\mathrm{e})}$ defined by Eq. (2.75), for each of the $j = 1, \ldots, 30$ principal components of the (unit-variance and zero-mean coordinates of the) data set of feature vectors, $\mathbf{X}^{(\mathrm{asym})}$, of the entire database of structures considered in the asymmetric Wigner bilayer system in Refs. [62–64]. We see that, similar to the symmetric case above (cf. Fig. 3.5), only very few principal components appear to carry relevant information (*i.e.*, have significant values of $\lambda_i^{(\mathrm{e})}$). Thus, we restrict our analysis to the leading nine principal components, whose explained variance is larger than 2%, and we set $N_l = N_f^* = 9$ for the here investigated asymmetric case.

In Fig. 3.8 we present the leading nine principal components $\mathbf{v}_1, \ldots, \mathbf{v}_9 \in \mathbb{R}^{N_f=30}$ in the feature space spanned by $\mathbf{x} = (f_1, \ldots, f_{N_f})$, cf. Eq. (3.11). The $i = 1, \ldots, 30$ elements of the principal component vector, $\mathbf{v}_j$, indicate the direction of the principal component in the feature space; we here refer to the values of the $i = 1, \ldots, 30$ elements of a principal component as *feature weights*. Large positive or large negative values of certain feature weights of a particular principal component indicate important features which quantify information in the data set (*i.e.*, directions of high variance). These characteristic features of principal components can be used to identify important order parameters or, if several feature weights are dominant in a particular principal component, combinations of order parameters. Feature weights close to zero, on the other hand, indicate less important directions.

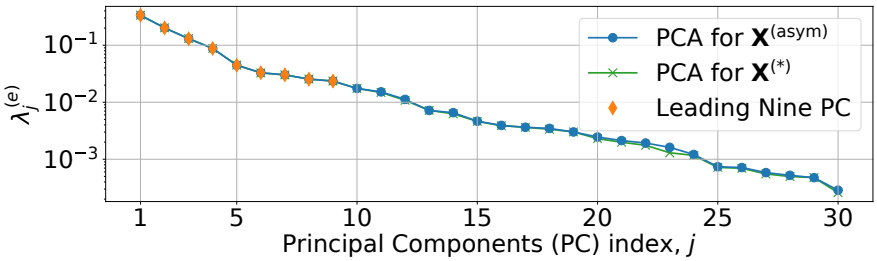


Figure 3.7: Explained variance, $\lambda_j^{(\mathrm{e})}$, of the principal components (PC) of the data set $\mathbf{X}^{(\mathrm{asym})}$ of feature vectors of all $N_{\mathrm{asym}} \sim 64000$ configurations considered in the construction of the phase-diagram in Refs. [62–64] (blue dots) and of the leading nine PC, N_l=9, whose explained variance $\lambda_j^{(\mathrm{e})} > 2\%$ (orange diamonds). Also structures that do not correspond to ground state configurations in the database [62–64] are considered. We define the data set $\mathbf{X}^{(*)}$ which contains all data points of $\mathbf{X}^{(\mathrm{asym})}$ except for those which correspond to hexagonal monolayer configurations where $x = N_2/N = 0$ and we also present the explained variance, $\lambda_j^{(\mathrm{e})}$, of the PC of the data set $\mathbf{X}^{(*)}$ (green crosses).

We see in Fig. 3.8 that the first principal components, $\mathbf{v}_1$, exhibits large positive feature weights from the six and twelvefold order parameters $\Psi^{(1,3,4)}_{[6,12]}$ and medium to large negative weights of $\Psi^{(1,3,4)}_{[4,5,8,10]}$, r_g and x. The principal components $\mathbf{v}_2$ and $\mathbf{v}_3$ exhibit medium to large feature weights distributed over a range of order parameters which makes them more difficult to interpret than $\mathbf{v}_1$. Starting from the fourth principal component, single (or very few) directions in feature space become important ($\Psi^{(2)}_{[10,12]}$ in the case of $\mathbf{v}_4$ and $\Psi^{(1)}_{[10,12]}$ in the case of $\mathbf{v}_5$). We can see in Fig. 3.8 that throughout all principal components, the feature weights for the intralayer nearest neighbor order parameter, r_g, and for the composition, x, are practically the same. Thus, we conjecture the nearest neighbor distances of particles in both layers for low energy configurations of the asymmetric Wigner bilayer system[16] are largely governed by the composition; in turn this suggests, that the particles tend to be distributed as uniform as possible throughout both layers for ground state configurations of the system.

In the left panel of Fig. 3.9 we present the ground state phase-diagram of the asymmetric Wigner bilayer system in the (η, A)-plane in an [R, G, B]-scheme – similar to Fig. 3.3 – based on the leading three principal components[17] of the data set $\mathbf{X}^{(\mathrm{asym})} = (\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\mathrm{asym}}})$: we consider the latent space representations, $\mathbf{l}_g$, of the feature vectors, $\mathbf{x}_g$, which correspond to the suggested ground state configurations, $\mathcal{X}_g$, of the asymmetric Wigner bilayer system from the literature database [62–64] for different values of the system parameters, η and A. For each of these data points, $\mathbf{l}_g = (v_{g1}, \ldots, v_{gN_l})$, we use the first three coordinates, $[v_{g1}, v_{g2}, v_{g3}]$, *i.e.*, the coordinates of $\mathbf{l}_g$ associated to the principal components $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, to define the relative amount of red, green and blue color used in the [R, G, B]-scheme to color every pixel

[16] We consider a constant particle area-density of bilayer configurations in this thesis.
[17] The coordinates of the data points in the principal component directions are not limited to the interval $[0, 1]$ but are, in general, real values.

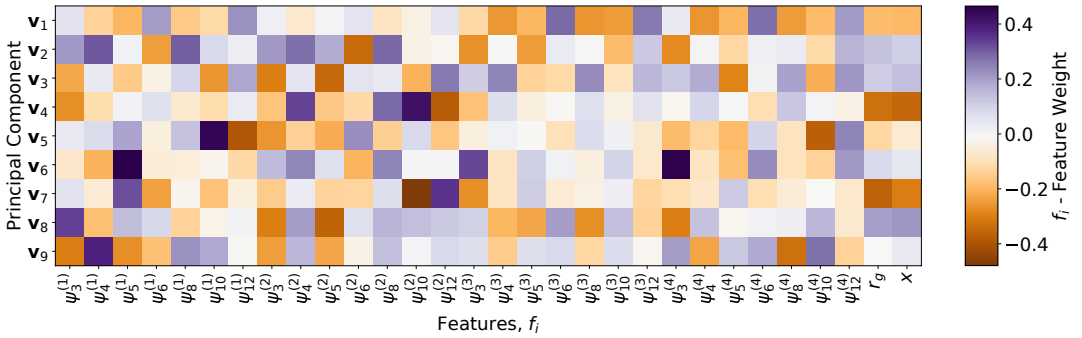


Figure 3.8: Leading nine principal components, $\mathbf{v}_1, \ldots, \mathbf{v}_9 \in \mathbb{R}^{N_f=30}$, of the data set $\mathbf{X}^{(\mathrm{asym})}$ represented in the feature space spanning $\mathbf{x} = (f_1, \ldots, f_{N_f})$, cf. Eq. (3.11). The $f_1, \ldots, f_{30}$ features are listed along the horizontal axis and represent the order parameters used to evaluate $\mathbf{X}^{(\mathrm{asym})}$ from the structural database of the asymmetric Wigner bilayer system from literature [62–64]; the respective $i = 1, \ldots, 30$ elements of the principal components, $\mathbf{v}_j$, related to the directions of f_i (*i.e.*, to which we refer to as *feature weights*) are presented via the color-coding specified by the colorbar: large positive (negative) values of feature weights are emphasized by dark purple (orange) coloring and values close to zero by white coloring.

in the (η, A)-plane of the ground state phase-diagram depicted in the left panel of Fig. 3.9. We can see that the phase boundaries suggested in Ref. [62–64] again nicely correlate with the values of the principal components, but there also appear to be regions in the (η, A)-plane which might require a closer inspection. The I_x region, for instance, might have a more sophisticated phase behaviour than previously suggested as indicated by the different values of $\mathbf{v}_2$ (green) and $\mathbf{v}_3$ (blue), a feature which we will further investigate in the following Subsections.

We can see from the left panel of Fig. 3.9 that phase I can uniquely be identified (via black colors of the [R, G, B]-scheme). Further, the leading three principal components of structures II, II_x, H and I_x are clearly different to those of the pentagonal structures P_1, P_2, P_3 and S_2. The former structures,II, II_x, H and I_x, exhibit large [G, B] values (green and blue) associated to the second and third principal components, *i.e.*, large values of the latent space coordinates into the directions $\mathbf{v}_2$ and $\mathbf{v}_3$; the corresponding bilayer structures have the property that when projecting their particles onto a single plate a hexagonal monolayer is formed, respectively. The pentagonal structures are, in general, more complicated to interpret[18] but we can see that their symmetry (cf. Fig. 3.8) is either dominated by the first principal component $\mathbf{v}_1$ (red) or by combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ (yellow, generated by adding red and green colors in [R, G, B]-notation). Furthermore, structures in the V_x region show strong signals either from the third principal component $\mathbf{v_3}$ (blue) or from combinations of $\mathbf{v}_1$ and $\mathbf{v}_3$ (red and blue becomes magenta) and can in that way be distinguished from the (red and yellow) pentagonal region in the phase-diagram depicted in the left panel of Fig. 3.9.

In the right panel of Fig. 3.9 we present a t-SNE [244] analysis (see Subsection 2.4.4) of the data set $\mathbf{X}^{(\mathrm{asym})}$ represented by the leading nine principal components. The corre-

[18] Notably, the region in the phase-diagram depicted in Fig. 3.9, where the latent space representation of the related ground state structures are dominated by $\mathbf{v}_1$ or by combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$, largely corresponds to the unclassified white region in Fig. 3.2.

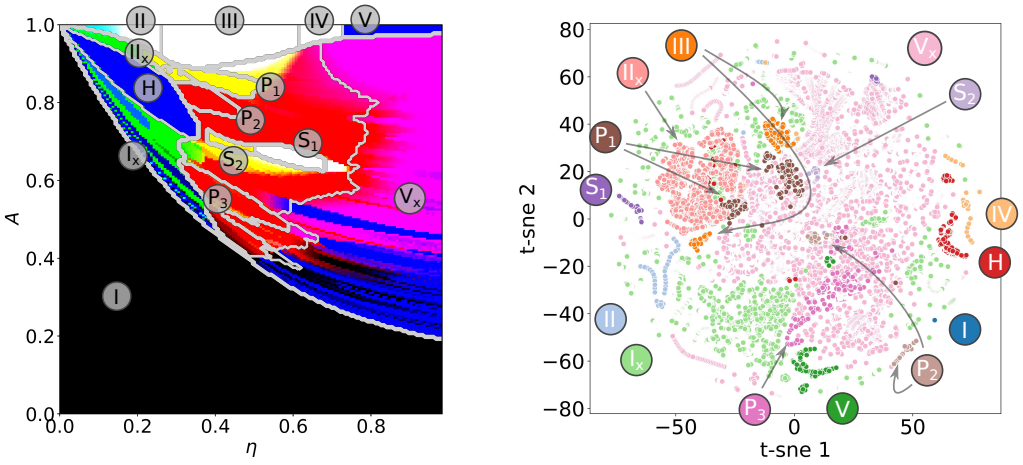


Figure 3.9: Left: Zero temperature ground state phase-diagram in the (η, A)-plane in an [R, G, B]-scheme similar to Fig. 3.3 based on the first three principal components $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ of the data set $\mathbf{X}^{(\mathrm{asym})}$: for every (η, A)-pair we define the relative amount of red, green and blue color, [R, G, B], of the corresponding pixel in the (η, A)-plane respectively by the coordinates, $[v_{g1}, v_{g2}, v_{g3}]$, of the latent space data point, $\mathbf{l}_g = (v_{g1}, \ldots, v_{gN_l})$, of the associated ground state configuration of the asymmetric Wigner bilayer system from the literature database [62–64]; the values of the coordinates in the [R, G, B]-scheme are truncated by $[0, 1]$. The light-gray lines indicate phase boundaries from Ref. [62] as illustrated in Fig. 3.2. Phases from literature [62–64] are labeled according to Table 3.1. Right: t-SNE [244] analysis (see Subsection 2.4.4), mapping the N_l=9-dimensional latent space representation, $\mathbf{L}^{(\mathrm{asym})}$, of all bilayer configurations from the literature database [62–64] (also considering structures not corresponding to ground state configurations) onto a two-dimensional t-SNE manifold spanned by "t-sne 1" and "t-sne 2". Each point represents a structure from the database embedded into the two-dimensional t-SNE plot and is colored according to the labeling given by Table 3.1 (*i.e.*, by rules from literature [62–64] on how to assign a phase label to a structure via the related data point in the feature space, $\mathbf{x}_i$, defined by Eq. (3.11)); the chosen colors are arbitrary.

sponding N_l=9-dimensional latent space representation of the data, $\mathbf{L}^{(\mathrm{asym})}$, is mapped onto a two-dimensional t-SNE manifold, *i.e.*, onto two axes[19] "t-sne 1" and "t-sne 2". In that way, similar data points (*i.e.*, nearby data points in the N_l=9-dimensional latent space $\mathbf{L}^{(\mathrm{asym})}$) are represented in clusters in the two t-SNE dimensions while different clusters (of cluster-internally similar data point) are spatially separated by the t-SNE analysis. We label the data points based on the classification scheme listed in Table 3.1 and color-code the respective phases in an arbitrary color-scheme (data points which cannot be classified with the help of Table 3.1 are omitted here). This enables us to visually distinguish parts of the $N_{\mathrm{asym}} \sim 64000$ structures available in the database in a two-dimensional plot which represents another kind of structural "phase-diagram". Interestingly, some phases which appear to be neighbors in the phase-diagram (cf. Fig. 3.2) appear to be located at very different positions in the t-SNE plot (cf. both panels in Fig. 3.9). This suggests, that the corresponding principal components are sufficiently expressive to distinguish between these clusters and may thereby indicate a discontinuous phase transition between the involved phases.

[19] Note that the scales of the axes are arbitrary.

Via the t-SNE analysis presented in the right panel of Fig. 3.9 we can clearly separate several phases by eye: phase I, for instance, is represented as a single dot in the bottom right corner of Fig. 3.9 which highlights the absence of structural variability within all monolayer structures in the structural database from literature [62–64]. Phases II, II_x, S_1, S_2 and P_2 form well separated clusters in the two-dimensional t-SNE space and thus can clearly be separated from the rest of the structures from the database. Phases III, IV, V, H and P_1 form well separated clusters in the t-SNE plot but each cluster is, on its own, split into two spatially separated clusters suggesting additional (possibly minor) structural variabilities of the structures that have been categorized into these phases, respectively. The splitting of P_1 in the t-SNE plot, for instance, can be related to the splitting of the P_1 phase into two regions in the phase-diagram depicted in Fig. 3.2: besides of the large P_1 region below phase III (*i.e.*, at smaller values of A) there is also a region in the vicinity of phase II_x at slightly larger values of η which is assigned as phase P_1. This splitting of phase P_1 is reflected in the t-SNE analysis shown in Fig. 3.9: one part of the P_1 cluster of data points is embedded in the II_x cluster (roughly at t-sne 1 ~ -35 and t-sne 2 ~ 0) while the other part of the P_1 cluster is located in the proximity of the larger III cluster (roughly at t-sne 1 ~ 0 and t-sne 2 ~ 20). Furthermore, the t-SNE mapping of the assigned phases I_x, P_3 and V_x is more difficult to interpret as compared to the other phases: data points which are assigned to be members of I_x, P_3 and V_x appear to be widely spread out in the t-SNE embedding coordinates t-sne 1 and t-sne 2. When trying to categorize the two-dimensional distribution of the data points in the t-SNE plot (Fig. 3.9) corresponding to the phases I_x, P_3 and V_x by eye, several sub-clusters of the data points appear to be present. We suspect that especially structures assigned as members of I_x and V_x can be further subdivided into sub-families of the respective phases (see below).

The t-SNE analysis is a viable tool when trying to quantify the information content of a data set, especially when the challenge is to categorise the data via unsupervised clustering algorithms. In our case, the t-SNE analysis of the leading nine principal components of $\mathbf{X}^{(\mathrm{asym})}$ suggests, that the information content provided by the PCA is (most likely) sufficient to identify the different ground state phases of the asymmetric Wigner bilayer system suggested in literature [62–64] (and shown in Fig. 3.2). Moreover, we suspect that several new phases (possibly also sub-categories of existing ones) can be related to certain values and boundaries of the order parameters and of the principal component representation of the data.

3.1.4.5 k-means Clustering of Structural Data of the Asymmetric Wigner Bilayer System

Both of the two investigations above based on PCA and t-SNE suggest that an additional, systematic analysis of the phase behaviour of the asymmetric Wigner bilayer system is of high interest since new ground state families might be hidden in the incredibly rich plethora of crystalline bilayer structures from literature [62–64]. In this thesis, we try to achieve such a systematic analysis of the phase behaviour of the asymmetric Wigner bilayer system by means of k-means clustering (see Subsection 2.4.2) of the nine leading principal component representation of the data set $\mathbf{X}^{(\mathrm{asym})} = (\mathbf{x}_1, \ldots, \mathbf{x}_{\mathbf{N}_{\mathrm{asym}}})$

of feature vectors $\mathbf{x}_i$, defined by Eq. (3.11), of the entire structural database of Refs. [62–64] featuring $N_{\text{asym}} \sim 64000$ different configurations.

However, there is one serious complication to the k-means clustering problem compared to the symmetric case discussed in Subsection 3.1.4.3: we do not know the precise number, K, of clusters, which is actually a parameter of the k-means algorithm. Since this number of clusters is not known in advance, we performed independent k-means clusterings, for $K = 14$ to $K = 43$ clusters[20] and we use the concept of adjusted mutual information [245, 246], $I_K(\mathbf{k}_i, \mathbf{k}_j) \in [0, 1]$ defined by Eq. (2.84), to compare the results of different clusterings[21], $\mathbf{k}_i$ and $\mathbf{k}_j$ (see Subsection 2.4.5 for details). Large values of $I_K(\mathbf{k}_i, \mathbf{k}_j)$ (*i.e.*, close to unity) indicate that two clusterings, $\mathbf{k}_i$ and $\mathbf{k}_j$, are similar (or perfectly match in case of $I_K(\mathbf{k}_i, \mathbf{k}_j) = 1$) while smaller values of $I_K(\mathbf{k}_i, \mathbf{k}_j)$ indicate deviations of the clusterings (*i.e.*, that the assignment of data points from a certain data set into specific clusters is performed qualitatively different by $\mathbf{k}_i$ and $\mathbf{k}_j$). Based on comparing commonly occurring information of different clustering results (*i.e.*, common cluster labels of structures) we aim at identifying a reasonable – ideally the correct – number of clusters which represent the different structural families of the system.

At this point another comment is in order: for the k-means algorithm the choice of the initial location of the total number of K different clusters (here in the N_l=9-dimensional latent space of the data $\mathbf{L}^{(\text{asym})}$) is usually arbitrary. However, the final k-means clustering result may depend on the particular choice of the initial cluster coordinates and on the initial assignment of the different data points to the clusters. This choice of the *initial conditions* of the k-means algorithm was not an issue for the symmetric case above since the problem appears to be well-behaved. For more complicated tasks, such as the current one, it is good practise to apply k-means clustering several times with independent initial conditions. The results of these independent clusterings, $\mathbf{k}_1$, $\mathbf{k}_2, \ldots$, can then be analysed for frequently occurring solutions or similarities in the labeling of structures. Also for this purpose we can use the concept of the adjusted mutual information, $I_K(\mathbf{k}_i, \mathbf{k}_j)$, as described in Subsection 2.4.5, in order to identify the most common clustering results, first for a fixed value of the number of clusters, K, and later to compare clustering results with different numbers of clusters.

To simplify the k-means clustering problem of the data set $\mathbf{X}^{(\text{asym})}$ we can make use of the following considerations: there is no structural variability of hexagonal monolayer structures, I, in the database from literature [62–64] and we can clearly identify phase I via the composition $x = N_2/N = 0$. In turn, phase I structures can easily be treated separately from the rest of the structures in the database. We can manually assign all hexagonal monolayer structures, I, in the data set $\mathbf{X}^{(\text{asym})}$ and further only need to consider the remaining structures in the k-means clustering based phase classification scheme which we employ below. We thus define the reduced data set $\mathbf{X}^{(*)}$ which covers all data points of $\mathbf{X}^{(\text{asym})}$ except for the feature vectors of hexagonal monolayer structures with $x = 0$; in what follows, quantities which are based on the data set $\mathbf{X}^{(*)}$ are also emphasized by an asterisk. As shown in Fig. 3.7 performing a PCA on $\mathbf{X}^{(*)}$

[20] $K = 14$ clusters represent the 14 phases listed in Table 3.1 and $K = 43$ is already a quite large number of different phases.

[21] A clustering uniquely maps a structure via its feature vector to a cluster (or to a label) and assigns it thereby to a structural family, see Subsection 2.4.2 and cf. Subsection 3.1.4.3.

instead of $\mathbf{X}^{(\mathrm{asym})}$ has only minor consequences on the explained variance of the principal components of the entire database. However, working with the data set $\mathbf{X}^{(*)}$ has the advantage that the k-means clustering will not be biased by the phase I structures, which represent a large portion of the entire phase-diagram. Henceforward, we refer to the specific k-means clustering of the data set $\mathbf{X}^{(*)}$ as "k^*-means clustering", where we manually label all data points $\mathbf{x}_i$ in $\mathbf{X}^{(\mathrm{asym})}$ which correspond to phase I with $c_i = 1$; the labeling of the remaining data with the cluster labels $2, \ldots, K$ is then performed by k-means clustering of the leading nine principal component representation of the data set $\mathbf{X}^{(*)}$.

In Fig. 3.10 we present the adjusted mutual information, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$, of $i, j = 0, \ldots, 39$ independent clustering results, $\mathbf{k}_i^*$ and $\mathbf{k}_j^*$, of the k^*-means clustering algorithm for a total number of $K = 14$ and $K = 32$ clusters, respectively. Henceforward, we refer to $\mathbf{k}_i$ and $\mathbf{k}_j^*$ also as clustering samples of the k-means and k^*-means algorithm, respectively.

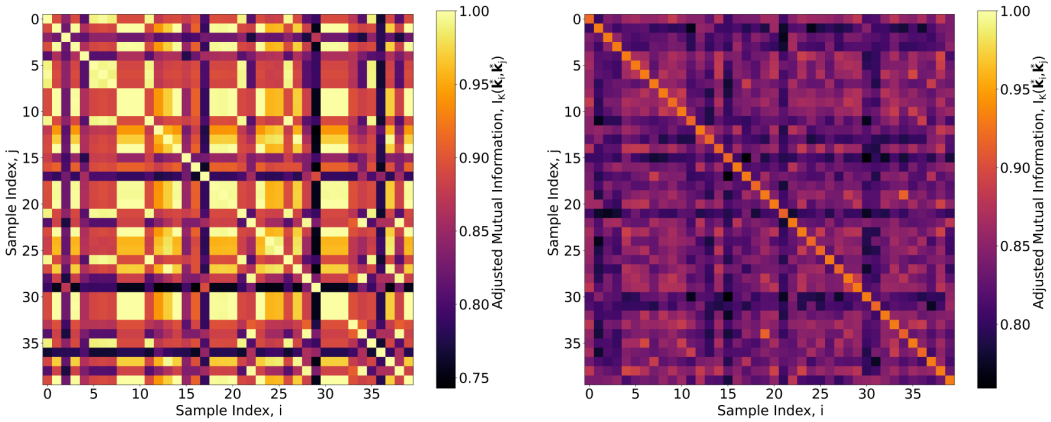


Figure 3.10: Adjusted mutual information score, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$ defined by Eq. (2.84), of $i, j = 0, \ldots, 39$ different, randomly initialized k^*-means clustering results, $\mathbf{k}_i^*, \mathbf{k}_j^*$, with a total number of $K = 14$ (left) and $K = 32$ (right) clusters, respectively. Values of $I(\mathbf{k}_i^*, \mathbf{k}_j^*)$ close to unity (bright, yellow regions) suggest large overlap between the different clusterings, $\mathbf{k}_i^*$ and $\mathbf{k}_j^*$, while smaller values $I(\mathbf{k}_i^*, \mathbf{k}_j^*) \approx 0.8$ (black and purple) indicate different results. Note that $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*) = I_K(\mathbf{k}_j^*, \mathbf{k}_i^*)$ is symmetric.

For smaller numbers of clusters (such as $K = 14$) the algorithm is more stable: many different samples, $\mathbf{k}_i^*$ and $\mathbf{k}_j^*$, exhibit a perfect score of $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*) = 1$ of the adjusted mutual information (cf. yellow pixels in the left panel of Fig. 3.10) which indicates that the algorithm has identified the same results several times. For larger numbers of clusters (such as $K = 32$) the situation is more complicated since the number of possible clustering results grows rapidly with the number of clusters. In general, for large numbers of clusters the adjusted mutual information becomes larger by construction rendering it in principle harder to define a quality measure to compare different results [245, 246].

In both cases, *i.e.*, using $K = 14$ or $K = 32$, there appears to be qualitatively different clustering results to the clustering problem as depicted in Fig. 3.10. In Fig. 3.11 we present histograms of the adjusted mutual information score, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$, depicted in Fig. 3.10 for selected clustering samples, some of which share little information with other clustering results (dark regions in Fig. 3.10 and orange distributions in Fig. 3.11) and others with a more similar clustering (bright regions in Fig. 3.10 and green distributions in Fig. 3.11).

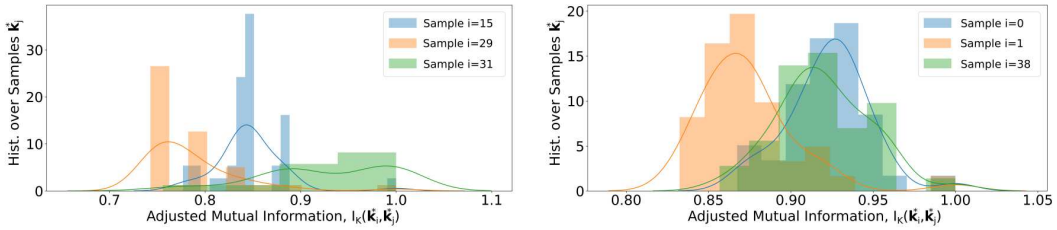


Figure 3.11: Left: Column-wise histogram of the adjusted mutual information score, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$ defined by Eq. (2.84), of clustering samples $\mathbf{k}_{i=15}^*$ (blue), $\mathbf{k}_{i=29}^*$ (orange) and $\mathbf{k}_{i=31}^*$ (green) for the $K = 14$ clustering results shown in the left panel of Fig. 3.10. Right: Column-wise histogram of the adjusted mutual information score, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$, of clustering samples $\mathbf{k}_{i=0}^*$ (blue), $\mathbf{k}_{i=1}^*$ (orange) and $\mathbf{k}_{i=38}^*$ (green) for the $K = 32$ clustering results shown in the right panel of Fig. 3.10.

In order to compare the quality of different clustering results we present in Fig. 3.12 the column-wise average value, *i.e.*, $\langle I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)\rangle_j = \sum_{j=0}^{N_c} I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)/N_c$, and the median (analogously defined) of the adjusted mutual information score of each clustering sample, $\mathbf{k}_i^*$, with all other clusterings, $\mathbf{k}_{j=0,\ldots,39}^*$, and we mark both the maximum of the mean and the median. For $K = 14$ the clustering sample $i = 31$, *i.e.*, $\mathbf{k}_{31}^*$, seems to be a good choice for the final clustering result (cf. left panels of Figs. 3.10 to 3.12). However, for a larger number of clusters, e.g. $K = 32$, it is harder to decide what the optimal clustering could be: both samples, $i = 0$ and $i = 38$, appear to have qualitatively similar traits as can be seen in the right panels of Figs. 3.10 to 3.12.

However, there is another ingredient we can include into our analysis to bias the adjusted mutual information score into a physically motivated direction: the ground state solutions of the symmetric Wigner bilayer system also show up in the phase-diagram of the asymmetric Wigner bilayer system at $A = 1$. We can identify the fraction of the data points in the data sets $\mathbf{X}^{(\mathrm{asym})}$ which correspond to the ground state solutions of the symmetric Wigner bilayer system and collect them in a separate data set $\mathbf{X}^{(\mathrm{sym})}$. We assign all data points in $\mathbf{X}^{(\mathrm{sym})}$ to the phases I through V following Table 3.1 [146, 147] and collect the corresponding phase labels in the set $\mathbf{w}^{(\mathrm{sym})}$. Analogously, we collect in the set $\mathbf{k}_i^{(\mathrm{sym})}$ the particular clustering labels from the clustering result $\mathbf{k}_i$ (performed on the full data set $\mathbf{X}^{(\mathrm{asym})}$ after PCA) which correspond to the data points in $\mathbf{X}^{(\mathrm{sym})}$. Hence, the adjusted mutual information score $I_K(\mathbf{w}^{(\mathrm{sym})}, \mathbf{k}_i^{(\mathrm{sym})})$ quantifies the overlap between the clustering result $\mathbf{k}_i^{(\mathrm{sym})}$ and the analytically known labeling $\mathbf{w}^{(\mathrm{sym})}$ (*i.e.*, the amount of commonly labeled data points) of the data set $\mathbf{X}^{(\mathrm{sym})}$ of the feature vectors of the ground states of the symmetric case. We now define the *biased*

adjusted mutual information score, $S(\mathbf{k}_i, \mathbf{k}_j|\mathbf{w}^{(\mathrm{sym})})$, as

$$S(\mathbf{k}_i, \mathbf{k}_j|\mathbf{w}^{(\mathrm{sym})}) = I_K(\mathbf{k}_i, \mathbf{k}_j) \times \sqrt{I_K(\mathbf{w}^{(\mathrm{sym})}, \mathbf{k}_i^{(\mathrm{sym})})}\sqrt{I_K(\mathbf{w}^{(\mathrm{sym})}, \mathbf{k}_j^{(\mathrm{sym})})}, \quad (3.12)$$

which weights the adjusted mutual information, $I_K(\mathbf{k}_i, \mathbf{k}_j)$, of different k-means (or analogously k^*-means[22]) clustering results, $\mathbf{k}_i$ and $\mathbf{k}_j$, with the square root of the respective adjusted mutual information scores of $\mathbf{k}_i^{(\mathrm{sym})}$ and $\mathbf{k}_j^{(\mathrm{sym})}$ with $\mathbf{w}^{(\mathrm{sym})}$.

In our case, the biased adjusted mutual information score, $S(\mathbf{k}_i, \mathbf{k}_j|\mathbf{w}^{(\mathrm{sym})})$, is an important measure for the quality of the clustering results $\mathbf{k}_i$ and $\mathbf{k}_j$ since we demand of a corresponding labeling to be as accurate as possible, especially for the fraction of the data, $\mathbf{X}^{(\mathrm{sym})}$, that can be labeled analytically via $\mathbf{w}^{(\mathrm{sym})}$. In Fig. 3.13 we present the biased adjusted mutual information score, $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$, of the same selected samples as used in Fig. 3.11 and we also present the corresponding mean and median values of all biased sample scores (*i.e.*, $S(\mathbf{k}_i, \mathbf{k}_j|\mathbf{w}^{(\mathrm{sym})})$) as we have already shown for the unbiased case (*i.e.*, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$) in Fig. 3.12.

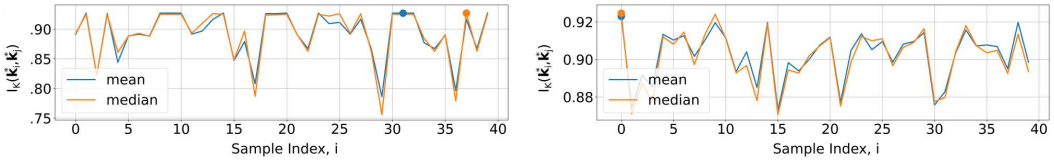


Figure 3.12: Mean (blue) and median (orange) of the adjusted mutual information score, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$, for each of the $i = 0, \ldots, 39$ randomly initialized clustering samples, $\mathbf{k}_i^*$, with respect to all other clustering samples, $\mathbf{k}_{j=1,\ldots,39}^*$ shown in Fig. 3.10 for $K = 14$ (left) and $K = 32$ (right) clusters (*i.e.*, column-wise average and mean of the data shown in Fig. 3.10). Maxima of the mean and median of the adjusted mutual information score, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$, as a function of the sample indices $i = 0, \ldots 39$ are indicated by filled circles which emphasize clustering results which potentially share the most information with other results on average (or represent the maximum median thereof).

By comparing the biased, $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$, and the unbiased scores, $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$, we see that in general scaling the adjusted mutual information according to Eq. (3.12) leads to smaller values of the $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$ score compared to $I_K(\mathbf{k}_i^*, \mathbf{k}_j^*)$. Especially the diagonal terms, $S(\mathbf{k}_i^*, \mathbf{k}_i^*|\mathbf{w}^{(\mathrm{sym})})$, whose adjusted mutual information scores are $I_K(\mathbf{k}_i^*, \mathbf{k}_i^*) = 1$ by definition (cf. Fig. 3.10), are now weighted by $I_K(\mathbf{w}^{(\mathrm{sym})}, \mathbf{k}_i^{(\mathrm{sym})}) \leq 1$, accounting for the quality of the clustering result with respect to the labels of the ground states of the symmetric case. Consequently, the scaling of the adjusted mutual information score, $I(\mathbf{k}_i^*, \mathbf{k}_j^*)$, via Eq. (3.12) also causes an additional bias to larger values of the $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$ score for clustering results with large respective overlaps between $\mathbf{k}_i^{(sym)}, \mathbf{k}_j^{(\mathrm{sym})}$ and $\mathbf{w}^{(\mathrm{sym})}$ (*i.e.*, commonly labeled ground states of the symmetric case); results with corresponding smaller overlaps of $\mathbf{k}_i^{(sym)}, \mathbf{k}_j^{(\mathrm{sym})}$ and $\mathbf{w}^{(\mathrm{sym})}$

[22] Also for k^*-means clustering we rely on the set of analytically labeled data, $\mathbf{w}^{(\mathrm{sym})}$, of the entire data set, $\mathbf{X}^{(\mathrm{asym})}$, which correspond to the ground state solutions of the symmetric case, $A = 1$, in the evaluation of $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$, given by Eq. (3.12): we respectively compare in $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$ the labels $\mathbf{w}^{(\mathrm{sym})}$ with $\mathbf{k}_i^{(\mathrm{sym})}$ and $\mathbf{k}_j^{(\mathrm{sym})}$, *i.e.*, the fraction of the samples $\mathbf{k}_i^*$ and $\mathbf{k}_j^*$ which respectively corresponds to the known ground state structures of the symmetric Wigner bilayer system.

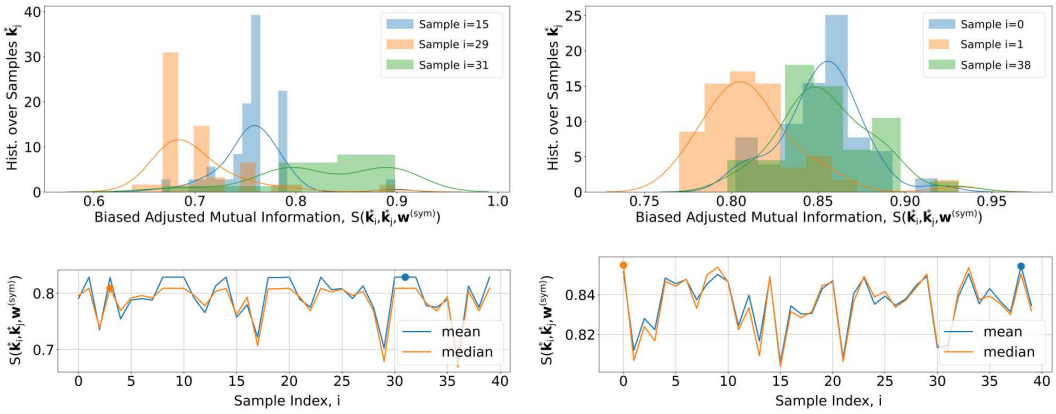


Figure 3.13: Top row and bottom row: same as Fig. 3.11 and Fig. 3.12, respectively, but for the biased adjusted mutual information score, $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$, defined in Eq. (3.12), cf. Fig. A.2 in Appendix A.1.1.

are biased towards smaller values of $S(\mathbf{k}_i^*, \mathbf{k}_j^*|\mathbf{w}^{(\mathrm{sym})})$ (cf. right most bins of sample 0 and sample 38 in the top right panel of Figs. 3.11 and 3.13).

We now assume that "good" clustering results, which are biased towards large values of the $S(\mathbf{k}_i, \mathbf{k}_i|\mathbf{w}^{(\mathrm{sym})})$ score by labeling the symmetric part in the data set as good as possible, occur frequently and perform similarly in terms of the overall quality of the clustering. For such good clusterings also the mean (and the median) of the $S(\mathbf{k}_i, \mathbf{k}_i|\mathbf{w}^{(\mathrm{sym})})$ scores are biased towards larger values, while being biased towards smaller values for qualitatively poor clustering results. We define the mean value, $\bar{k}_i$, of the biased adjusted mutual information score, $S(\mathbf{k}_i, \mathbf{k}_i|\mathbf{w}^{(\mathrm{sym})})$, of the $i, j = 1, \ldots, N_c$ different clustering samples (cf. Figs. 3.10 and A.2), by

$$\bar{k}_i = \frac{1}{N_c} \sum_{j=0}^{N_c-1} S(\mathbf{k}_i, \mathbf{k}_j|\mathbf{w}^{(\mathrm{sym})}). \tag{3.13}$$

With $\bar{k}_i$ we have a reasonably good measure for comparing different clustering results for one given number of clusters, K: we here rely on $\bar{k}_i$ to quantify the quality of a clustering result, $\mathbf{k}_i^*$, of assigning the total number of K clusters correctly, given N_c independent clustering results (cf. Fig. 3.11). We evaluate $\bar{k}_i$ separately for all independent k-means and k^*-means clusterings for several different values of $K = 14$ to $K = 43$: for a given value of K the one sample from the respective $i = 0, \ldots, N_c - 1$ clusterings with the maximum value of $\bar{k}_i$, given by Eq. (3.13), is considered as best clustering results.

We now redraw in Fig. 3.14 the zero temperature ground state phase-diagram of the asymmetric Wigner bilayer system [62–64] (cf. Fig. 3.2) with the phase labeling suggested by the best clustering results of several k-means (left panels) and k^*-means (right panels) clustering procedures, evaluated for different numbers of clusters K; note that the color-coding is arbitrary. By careful inspection of Fig. 3.14 we find that the best K^*=32-means clustering result (central right panel in Fig. 3.14) is, indeed, reasonably

representative for the phase-diagram of the asymmetric Wigner bilayer system as compared to the data from literature: several different phases identified in Refs. [62–64] (such as I through V, II_x, H, P_1, P_3 and S_2) can be resolved very accurately by the best K^*=32-means clustering, while some regions (such as the I_x or V_x phases) are subdivided into much greater detail. We will continue this discussion below and in detail in the following Subsections.

After identifying the sample, $\mathbf{k}_i$, with the largest average biased adjusted mutual information score, $\bar{k}_i$ defined by Eq. (3.13), independently for different numbers of clusters, K ranging from 14 to 43, we now want to compare the best clusterings for different numbers of clusters, K_i and K_j, some of which are depicted in Fig. 3.14. Applying the same strategy as above Fig. 3.15 presents the adjusted mutual information score, $I_K(\mathbf{k}_i, \mathbf{k}_j)$ defined by Eq. (2.84), for the best k-means clusterings compared to the results of all other best clustering results for different numbers of clusters, K_i and K_j; Fig. 3.16 shows the same analysis for k^*-means clusterings.

For both adjusted mutual information scores of k-means and k^*-means clusterings we see that the mean and median of $I_K(\mathbf{k}_i, \mathbf{k}_j)$ have maxima between $K = 31$ and $K = 35$ numbers of clusters, see top panels in Figs. 3.15 and 3.16.

Furthermore, we find that both k-means and k^*-means clustering results nicely correlate when evaluating the adjusted mutual information scores between the k-means and k^*-means clusterings, *i.e.*, $I_K(\mathbf{k}_i, \mathbf{k}_j^*)$ as can be seen in Fig. 3.17. Again we evaluate the mean and median of the adjusted mutual information score of each clustering $\mathbf{k}_i$ with respect to all other clusterings $\mathbf{k}_j^*$ (row-wise) and, analogously, of each clustering $\mathbf{k}_i^*$ with all other clusterings $\mathbf{k}_j$ (column-wise). The maxima of the corresponding mean and the median values are located between a number of $K = 32$ and $K = 36$ clusters in the row-wise and column-wise cases which suggests that a reliable guess of a reasonable number of clusters lies within that range.

In this thesis, our choice for the number of clusters (and thus the final clustering result) is K^*=32, *i.e.*, the best clustering result for a total number of K=32 clusters identified by the k^*-means algorithm (see central right panel of Fig. 3.14), for two reasons:

1. We can uniquely label the data points in $\mathbf{X}^{(\mathrm{asym})}$ corresponding to phase I by identifying all data points which feature a value of x=0 before we employ the k-means clustering procedure to the data set $\mathbf{X}^{(*)}$. It turns out that working with the reduced data set $\mathbf{X}^{(*)}$ instead of $\mathbf{X}^{(\mathrm{asym})}$ removes a large part of the uncertainty of the k^*-means clustering procedure compared to k-means clustering.

2. Overall, there seems to be good agreement between the K^*=32-means clustering results with the different samples from k-means clustering (see Fig. 3.17).

Henceforward, we refer to the best K^*=32-means clustering result (which categorizes the data set of asymmetric Wigner bilayer structures from literature [62–64]) as $\mathbf{k}_{32}^*$-clustering results and to the $c = 1, \ldots, 32$ different clusters (*i.e.*, to the different categories of structural families) as $\mathbf{k}_{32}^{*c}$ families, respectively.

We conclude this rather technical part of Subsection 3.1.4 by replacing in Fig. 3.18 the zero temperature phase-diagram of the asymmetric Wigner bilayer system from Refs. [62–64] (cf. Fig. 3.2) with a phase-diagram based on the $\mathbf{k}_{32}^*$-clustering results.

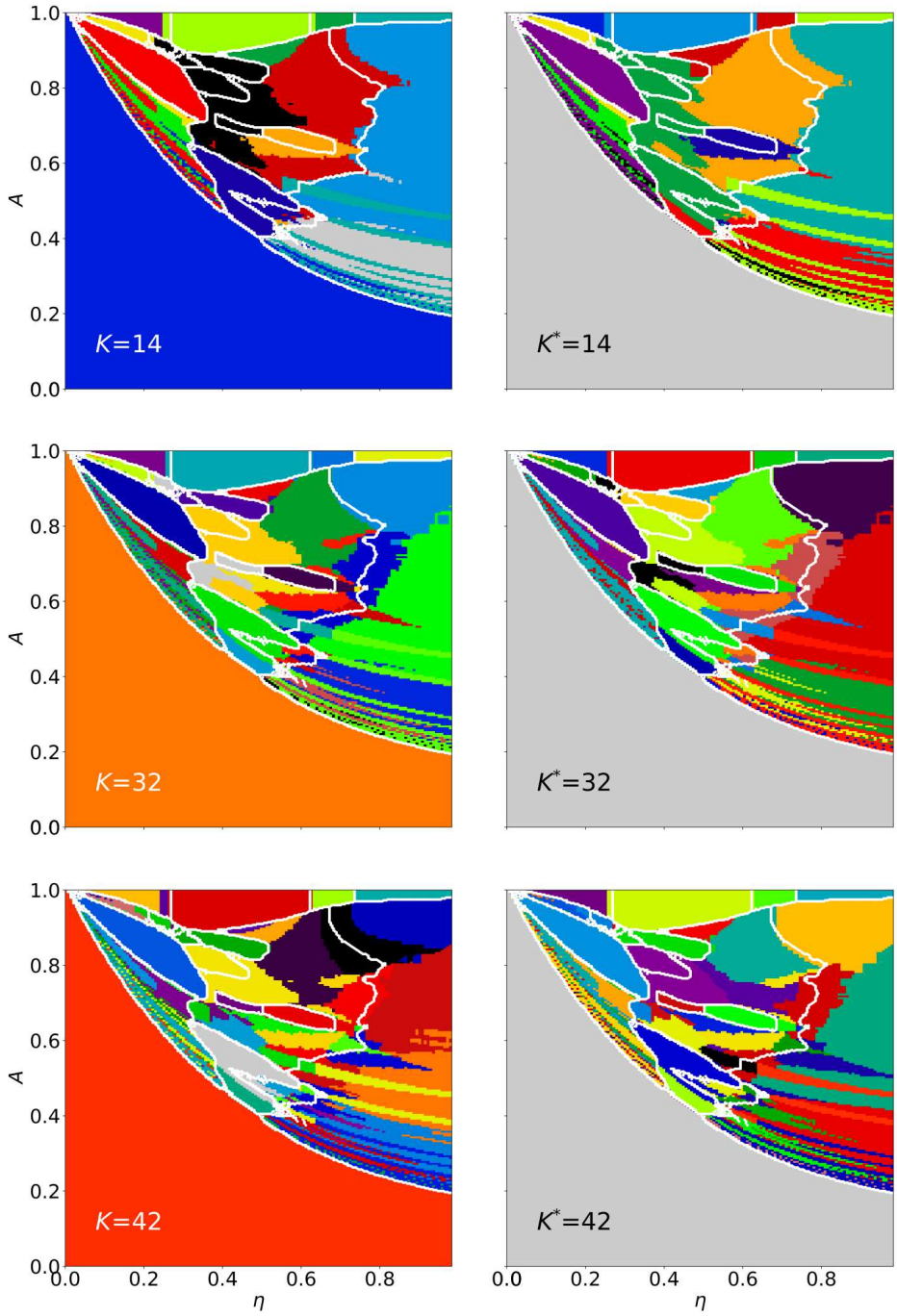


Figure 3.14: Ground state labeling of Ref. [63] structures into 14 (top row), 32 (middle row) and 42 (bottom row) families by k-means (left) and k^*-means (right) clusters (see inset texts). The color-scheme follows no particular order, thick white lines indicate Ref. [63] phase boundaries.

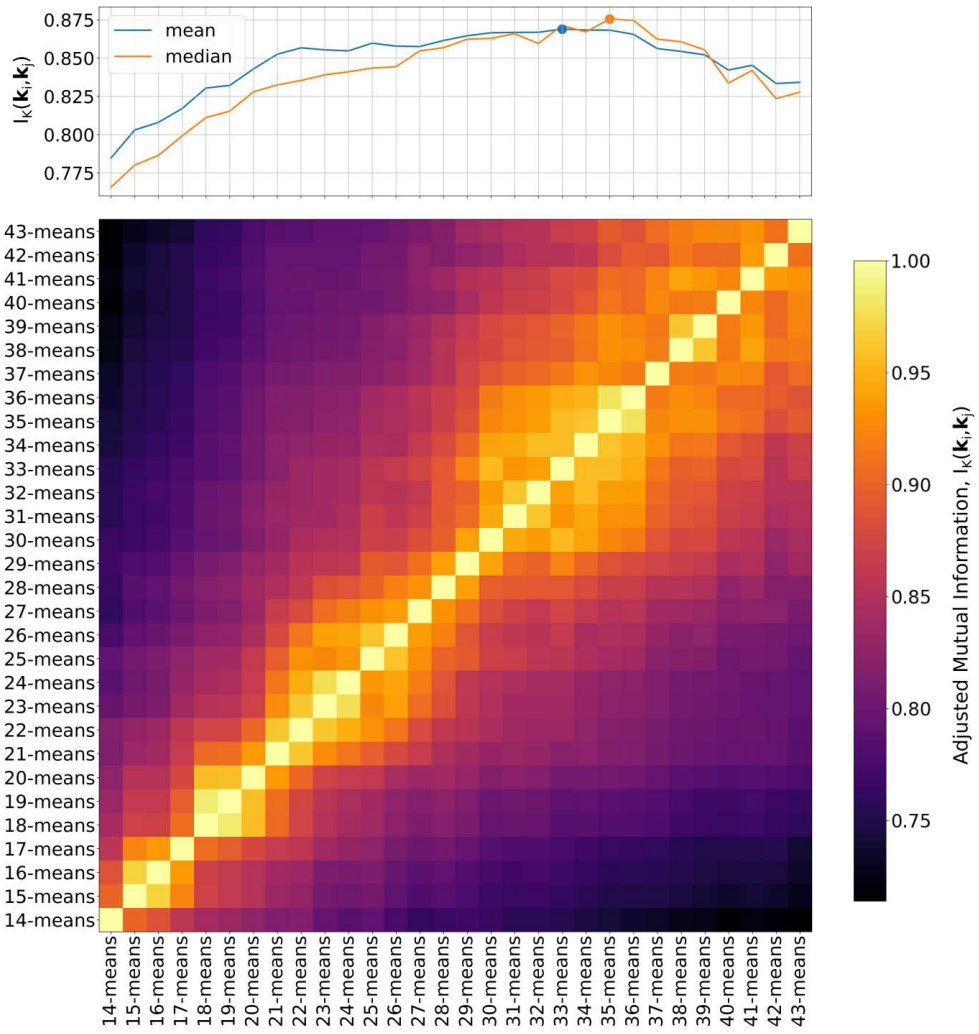


Figure 3.15: The bottom panel shows the adjusted mutual information score, $I_K(\mathbf{k},\mathbf{l})$ defined by Eq. (2.84), for k-means clustering results with different numbers of clusters, K and L ranging from 14 to 43. Every considered k-means clustering result was optimized to maximize Eq. (3.13) over an ensemble of independent runs for the respective number of clusters (details in the text). The top panel shows the (column-wise) mean and the median of the mutual information score of each k-means clustering; maxima are indicated by filled circles.

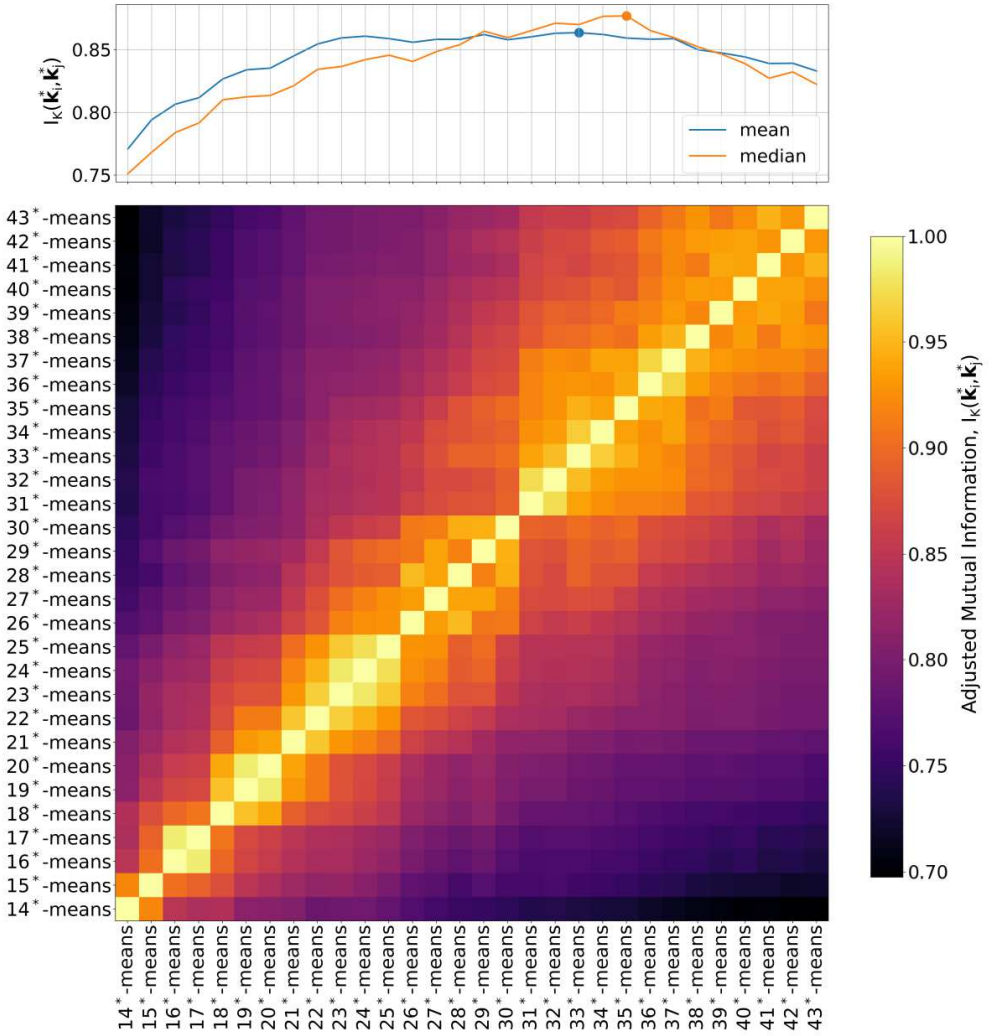


Figure 3.16: Same as Fig. 3.15 but with k^*-means clustering of the data set $\mathbf{X}^{(*)}$, which covers all data points of $\mathbf{X}^{(\mathrm{asym})}$ except for phase I structures with $x = 0$. In the clustering results, $\mathbf{k}_i^*$, all data points of $\mathbf{X}^{(\mathrm{asym})}$ are addressed by cluster labels, but all data points which correspond to phase I structures are manually labeled. The remaining data, $\mathbf{X}^{(*)}$, is then subject to k-means clustering and the corresponding, additional cluster labels are collected in $\mathbf{k}_i^*$.

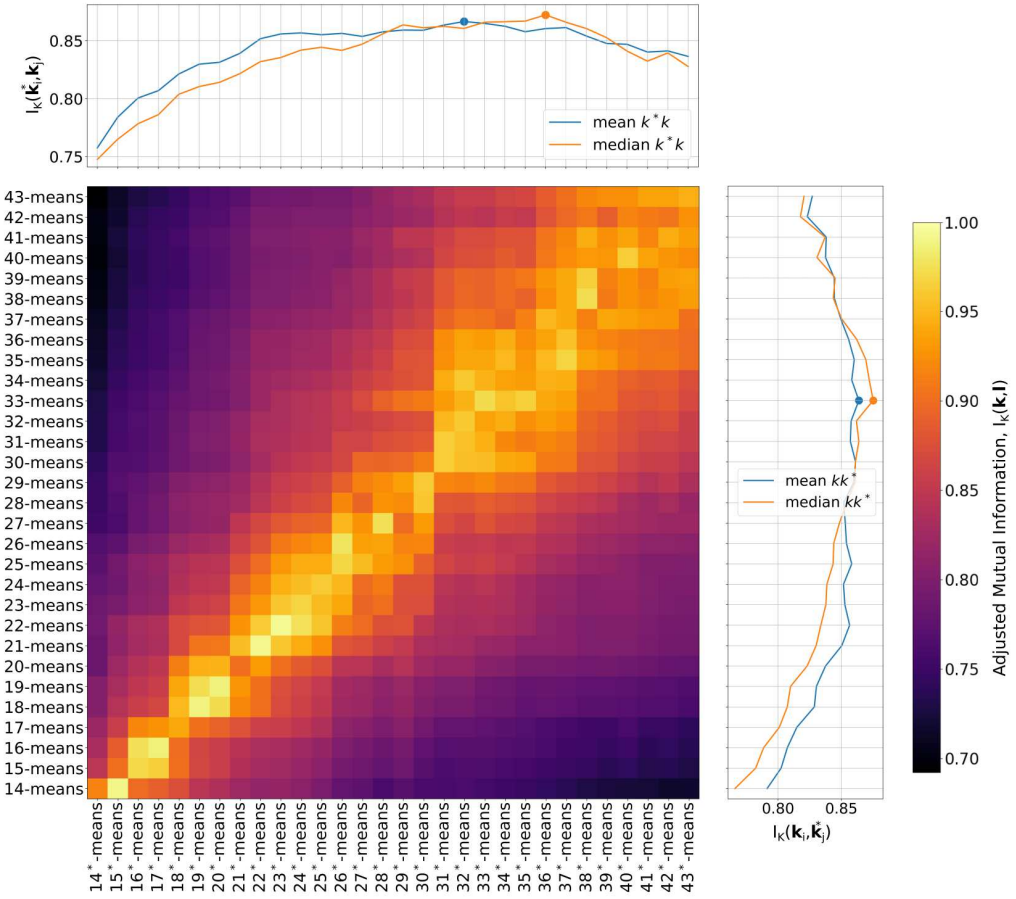


Figure 3.17: Similar to Figs. 3.15 and 3.16 but showing the cross relation of the adjusted mutual information score, $I_K(\mathbf{k},\mathbf{k}^*)$ defined by Eq. (2.84), for the best k-means and k^*-means clustering results; the number of clusters, K and K^*, range from 14 to 43 in both cases. The mean and median values of the adjusted mutual information scores, $I_K(\mathbf{k}_i,\mathbf{k}^*_j)$ and $I_K(\mathbf{k}^*_i,\mathbf{k}_j)$, for each clustering result, $\mathbf{k}_{i=14,\ldots,43}$ and $\mathbf{k}^*_{i=14,\ldots,43}$, with respect to all other clustering results, $\mathbf{k}^*_{j=14,\ldots,43}$ and $\mathbf{k}_{j=14,\ldots,43}$, are depicted in the right panel (row-wise mean and median, kk^*) and the top panel (column-wise mean and median, k^*k), respectively.

Indeed, the $c = 1, \ldots, 32$ families of structures, $\mathbf{k}_{32}^{*c}$, suggested by the $\mathbf{k}_{32}^{*}$-clustering reveal several new structural ground states of the asymmetric Wigner bilayer system. For each of the 32 identified $\mathbf{k}_{32}^{*c}$ families (except for phases I through V, cf. Fig. 3.4, and H, cf. Fig. 3.2) we present in Fig. 3.18 typical examples of bilayer structures which form the ground state of the asymmetric Wigner bilayer system at different system parameters. To better emphasize the respective (η, A)-regions where the different $\mathbf{k}_{32}^{*c}$ families represent the ground state of the system, we split up the presentation of the phase-diagram into four qualitatively different parts, *i.e.*, into panels (a)-(d) in Fig. 3.18:

(a) Panel (a) highlights the region in the phase-diagram of the asymmetric Wigner bilayer system which gives rise to the phases I_x, II, II_x and H at small to medium values of $0 < \eta \lesssim 0.4$ and medium to large values of $0.4 \lesssim A \leq 1$. The ground state structures in this region have in common that they form a hexagonal monolayer if all particles were projected into the same layer. We will discuss the $\mathbf{k}_{32}^{*}$-clustering results of this region in the parameter space in detail in Subsection 3.1.6 (cf. Fig. 3.39).

(b) In panel (b) we collect structural families which feature pentagonal tiles in layer one, *i.e.*, configurations that belong to the broader family of pentagonal structures. Suggested ground states candidates of the asymmetric Wigner bilayer system which fall into this category are P_1, P_2, P_3 and S_2 and the associated range of the system parameters can roughly be given by $0.3 \lesssim \eta \lesssim 0.7$ and $0.4 \lesssim A \lesssim 0.9$. We will discuss the $\mathbf{k}_{32}^{*}$-clustering results corresponding to this region in the parameter space in Subsection 3.1.7 (cf. Fig. 3.56).

(c) Via panel (c) we address $\mathbf{k}_{32}^{*c}$ families which are related to the dodecagonal *type I* family: these structures feature tilings in layer one that are similar to the snub-square structure, S_1, which, in turn, can potentially give rise to ground state configurations of the asymmetric Wigner bilayer system with global twelvefold symmetry [252, 254, 255]. We will discuss these $\mathbf{k}_{32}^{*c}$ families, which form ground states of the asymmetric Wigner bilayer system in the parameter range $0.4 \lesssim \eta \lesssim 0.7$ and $0.4 \lesssim A \lesssim 0.75$, in detail in Subsection 3.1.5 (cf. Fig. 3.23).

(d) In panel (d) we present $\mathbf{k}_{32}^{*}$-clustering results which can be related to the V_x region in the parameter space of the asymmetric Wigner bilayer system, *i.e.*, at large plate separation distances $\eta \gtrsim 0.7$ over a large range of A (cf. Fig. 3.2). We will discuss the $\mathbf{k}_{32}^{*}$-clustering results corresponding to this region in the parameter space in Subsection 3.1.7 (cf. Fig. 3.60).

We refrain here from any further discussion of the phase-diagram depicted in Fig. 3.18 but rather refer to Figs. 3.23, 3.39, 3.56 and 3.60 in Subsections 3.1.5 to 3.1.7 where the respective panels (a)-(d) are discussed in more detail.

In Appendix A.1.2 we present for each $\mathbf{k}_{32}^{*c}$ family the characteristic values and boundaries of the corresponding features, $\mathbf{x} = \{f_1, \ldots, f_{N_f}\}$ defined by Eq. (3.11), in Figs. A.3 to A.6, which explicitly describe the symmetries of the corresponding structural families. Characteristic values and boundaries of the corresponding principal component representation of the structural families identified by the $\mathbf{k}_{32}^{*}$-clustering procedure are

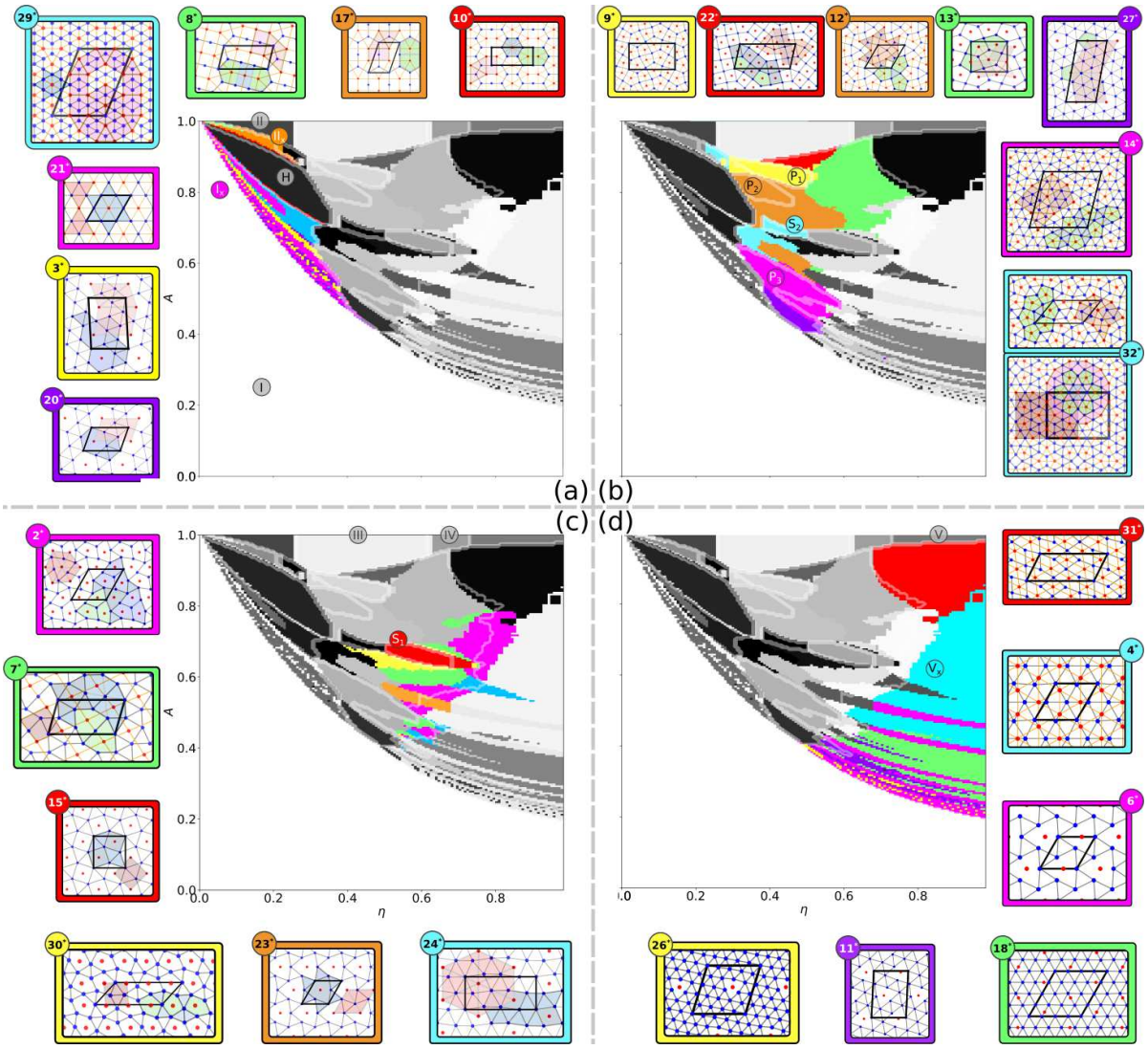


Figure 3.18: Ground state phase-diagram of the asymmetric Wigner bilayer System [62–64] in the (η, A)-plane as identified by the K^*=32-means clustering algorithm; the respective $c = 1, \ldots, 32$ families of structures, $\mathbf{k}_{32}^{*c}$, are color-coded in gray scale in all four panels, (a)-(d), starting at $\mathbf{k}_{32}^{*1}$ (white) and ending at $\mathbf{k}_{32}^{*32}$ (black). We subdivided the presentation of the ground state families of this phase-diagram into four (η, A)-panels, (a)-(d), by symmetry features of the respective occurring ground states (see bullet-points (a)-(d) in the text above). In each panel we highlight the respective parameter regions in bright colors (cyan, yellow, green, orange, red, magenta or purple in no particular order) where certain $\mathbf{k}_{32}^{*c}$ families form the ground state and we present archetypical structures of the respective $\mathbf{k}_{32}^{*c}$ families as insets. In each panel, respectively, the inset structures are labeled by "c^*" to address their association to a certain family $\mathbf{k}_{32}^{*c}$ with $i = 1, \ldots, 32$. The insets are further color-coded in the same way as the ground state regions of the respective $\mathbf{k}_{32}^{*c}$ family in the phase-diagram. Particles in the bottom (top) layer are always colored blue (red) and connections between nearest neighbors in each layer are drawn. Special tiles and features of the different structures are highlighted by colored shapes and the respective unit cell of each structure is emphasized by a thick black frame. The phase-boundaries from literature [62–64] (cf. Fig. 3.2) are indicated by opaque white lines in each panel and, correspondingly, the phases from literature are labeled which are associated to the considered parameter region of the different panels; colored labels indicate, that archetypical structures of the corresponding $\mathbf{k}_{32}^{*c}$ families are presented as insets in the respective panels, gray colored labels I through V and H with indicate, that the corresponding structures are not shown (although the phases have been identified by the $\mathbf{k}_{32}^{*}$-clustering algorithm). Characteristic values and boundaries of the order parameters and of the corresponding principal component representation of all structural families $\mathbf{k}_{32}^{*c}$ are collected in Appendix A.1.2.

presented in Figs. A.7 and A.8 (see Fig. 3.8 for a representation of the principal components in the feature space).

3.1.4.6 New Insights to the Wigner Bilayer System from Unsupervised Learning

A comparision of the identified phases of the asymmetric Wigner bilayer system via methods from literature [62–64] (*i.e.*, with the phase classification scheme given by Table 3.1, cf. Fig. 3.2) and via labeling by unsupervised $\mathbf{k}^*_{32}$-clustering techniques (cf. Fig. 3.14) show very good agreement for several structural families, see Fig. 3.18: phases from the symmetric case, I through V, but also more complex configurations such as II_x, H, S_1, S_2, P_1, P_2 and P_3 can be identified. Phases such as I_x and V_x are, however, subdivided into many categories by the K^*=32-means clustering approach (see Fig. 3.18(a,d), respectively).

As shown in Fig. 3.14 varying the numbers of clusters has a major impact on the final k-means or k^*-means clustering results. Already for only 14 clusters (top row of Fig. 3.14) most of the phases from literature [62–64] can be correctly identified via clustering. Phases I through V are clearly identified with (almost) correct boundaries (first order boundaries are correctly described) and also the honey-comb phase H, phase II_x as well as phase P_3 are identified essentially correctly. Also the phase boundary of the V_x phase is resolved with good accuracy. As illustrated by the top-left and rop-right panels of Fig. 3.14 a total number of K=14 or K^*=14 clusters seems, however, too small to correctly resolve the phases S_1, S_2, P_1 and P_2 all at once. The results from the K=14-means or K^*=14-means clustering algorithms rather suggest to subdivide phases I_x and V_x.

Taking the results from the clustering approach seriously, especially for K^*=32 (see Fig. 3.18), we can indeed identify a rich phase variety in the I_x region which is correlated with the dodecagonal $\Psi^{(4)}_{12}$ "hole" BOOP as already discussed above (cf. blue region within the I_x phase near the respective boundary to H in Fig. 3.3 and the $\mathbf{k}^{*29}_{32}$ family in Fig. 3.18(a) emphasized by cyan color).

Also, the whitish region in the [R, G, B] plot of the $\Psi^{(2)}_{10}$, $\Psi^{(4)}_{5}$ and $\Psi^{(1)}_{6}$ order parameters shown in the bottom left panel of Fig. 3.3 within the P_3 phase can be related to a more complex phase behaviour near phase P_3. The $\mathbf{k}^{*14}_{32}$ and the $\mathbf{k}^{*27}_{32}$ families, but also the S_2 configuration illustrated by the $\mathbf{k}^{*32}_{32}$ family in Fig. 3.18(b) form similar patterns in layer two(distorted rectangles and triangles organizes in a distorted snub-square vertex). However, the decoration of the layer two tiles with particles in layer one becomes increasingly complex (*i.e.*, more particle of layer one are involved per layer two tile) the closer we are to phase I in the parameter space. We will discuss these families of structures in Subsection 3.1.7 in more detail.

Also, the previously [62–64] unclassified structures in the (η, A)-regions in the vicinity of S_1 and S_2 (cf. Fig. 3.2) are classified into several different structural families by the $\mathbf{k}^*_{32}$-clustering algorithm as illustrated by Fig. 3.18(c). The structures depicted in Fig. 3.18(c) have in common that their basic tiles (such as equilateral triangles and squares arranged in a snub-square vertex), with whom their layer one structure is composed, may indicate the existence of a quasicrystalline state with global dodecagonal

symmetry [252, 254, 277]; Subsection 3.1.5 is dedicated to the families of structures illustrated in Fig. 3.18(c).

Even in the V_x region in the phase-diagram some interesting new structural families are identified by the $\mathbf{k}_{32}^*$-clustering procedure, as can be seen in Fig. 3.18(d) and as further discussed in Subsection 3.1.7. Characteristic values and boundaries of the order parameters and of the corresponding principal component representation of all structural families $\mathbf{k}_{32}^{*c}$ depicted in Fig. 3.18, which provide information about the symmetries of these structural families, are collected in Appendix A.1.2.

In general, with this phase-classification tool at hand we can systematically extract information which was previously hidden in the zoo of structures in the database of Refs. [62–64]. Owed to the way the structural data in Ref. [62] was generated (see discussion at the end of Subsection 3.1.3) it was, for instance, rather difficult to faithfully identify energetically competing structures with respect to the ground state at a certain (η, A)-phase point: at a given value of η solutions of the evolutionary algorithm for different numbers of particles per unit cell, N and N', but with the same composition, $x = N_2/N = x' = N_2'/N'$, may end up at (virtually) the same structures with degenerate energies but with different unit cells. Correctly identifying all of the relevant structures in the database from literature [62–64] faithfully by hand can be a tedious task. Via the $\mathbf{k}_{32}^*$-clustering based labeling of the structural database of Refs. [62–64] such degenerate configurations are automatically covered by one structural family. By only considering different structural families we then directly have access to any properties of energetically competing families of structures for any (η, A)-pair.

In Fig. 3.19 we present, for instance, the energy offset of the structures from the second best family (according to the results from the $\mathbf{k}_{32}^*$-clustering) to the corresponding ground state candidate[23] energies of the asymmetric Wigner bilayer system suggested by Refs. [62–64] in the (η, A)-plane. Fig. 3.19 highlights regions in the ground state phase-diagram of the asymmetric Wigner bilayer system where structures belonging to different $\mathbf{k}_{32}^{*c}$ families exhibit very close values of the internal energy per particle (dark areas), a feature which is especially pronounced at phase boundaries (quantified by the black and purple color coding in Fig. 3.19). Orange to yellow areas in Fig. 3.19 indicate a large energetic gap between the ground state and competing structures in the (η, A)-plane.

Fig. 3.20 shows the number of $\mathbf{k}_{32}^{*c}$ families that lie within specified energy intervals, $\Delta E^*/N$, above the ground state candidate energies suggested by Refs. [62–64] in the (η, A)-plane. Both, Figs. 3.19 and 3.20, highlight the energy accuracy required for comparing competing structures for the ground state of potentially very different unit cell sizes and shapes: energy differences of $\Delta E^*/N \approx 10^{-8}$ to $\approx 10^{-6}$ (cf. Figs. 3.19 and 3.20) are still very relevant for typical values of $E^*/N \approx -1$, defined by Eq. (3.9).

As this point we want to stress that, in general, clustering of structural data using principal component analysis (PCA) and k-means clustering (or any other, suited clustering

[23] Although most of the suggested ground state candidates identified by the evolutionary algorithm in Refs. [62–64] are very likely to represent the ground state configurations of the asymmetric Wigner bilayer system at the respective system parameters, there is no rigorous proof that they are, indeed, the ground states. Thus, whenever we write "ground state solutions" in Section 3.1 we really refer to "ground state *candidate* solutions" of the asymmetric Wigner bilayer system.

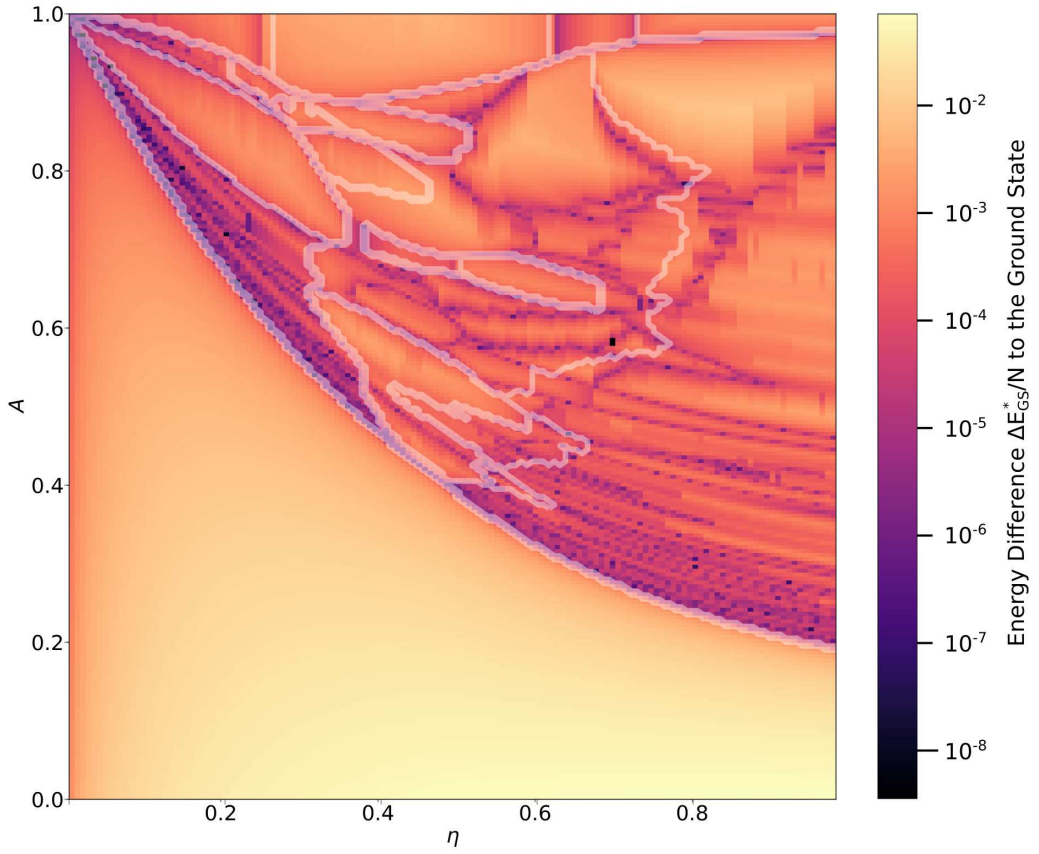


Figure 3.19: Difference, $\Delta E^*_{\mathrm{GS}}/N = (E^*/N - E^*_{\mathrm{GS}}/N)$, between the energy (per particle), E^*/N, of the energetically most favorable structure amongst all non-ground state structural family of the $\mathbf{k}^*_{32}$-clustering results to the suggested ground state energy (per particle), E^*_{GS}/N, of the asymmetric Wigner bilayer system from literature [62–64] at every respective (η, A)-pair. Opaque white lines indicate phase boundaries from literature, cf. Fig. 3.2. In the proximity of the H and I_x phase boundary (within the I_x region) we observe an energetically nearly-degenerate region in the (η, A)-plane with $\Delta E^*_{\mathrm{GS}}/N \approx 10^{-7}$ to 10^{-6} where ground state candidates exhibit large values of the twelvefold symmetric order parameter $\Psi^{(4)}_{12}$ (cf. bottom left panel of Fig. 3.3). This nearly-degenerate region corresponds to the newly identified ground state candidate family $\mathbf{k}^{*29}_{32}$ illustrated by the top left inset structure and cyan emphasized area in the (η, A)-plane of Fig. 3.18(a); this family of structures is further discussed in Subsection 3.1.6.

algorithm or classification algorithm) provides us with an additional feature: PCA is a linear transformation from feature space to latent space and k-means is a mapping of a data point in latent space to a cluster label. Once the clustering algorithm is trained (*i.e.*, once it has converged) it can be used as a classification model[24] and we can ask

[24]Nowadays it is easily possible to train a neural network in a supervised way with the objective of

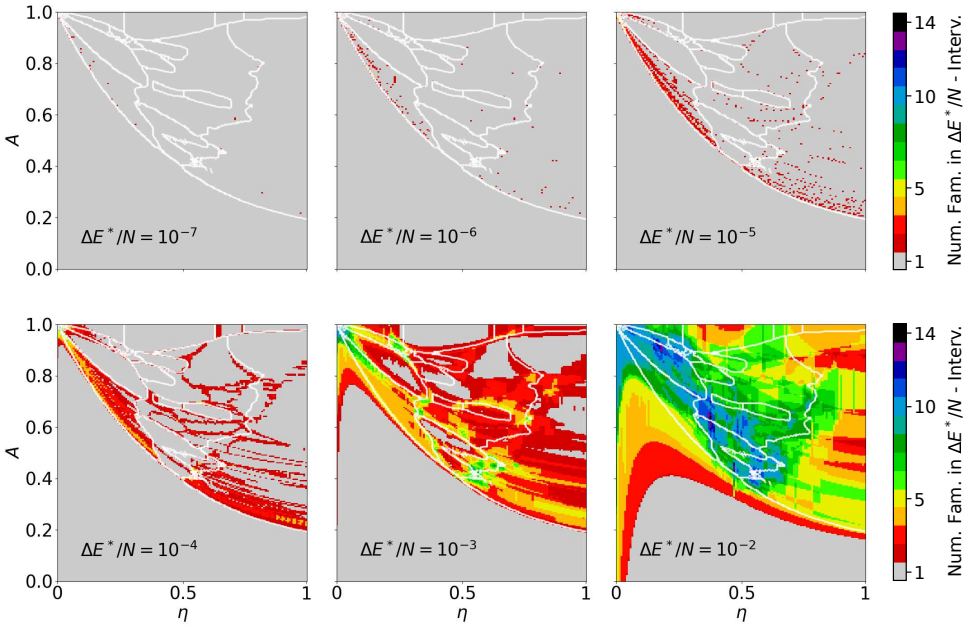


Figure 3.20: Number of $\mathbf{k}_{32}^{*c}$ families that exhibit an energy offset of at most $\Delta E^*/N$ to the ground state candidates of the asymmetric Wigner bilayer system (suggested by Refs. [62–64]) in the (η, A)-plane. $\Delta E^*/N$, ranging from 10^{-7} to 10^{-2}, is chosen differently for each panel (see texts in panels). Regions of gray color emphasize situations where only one structure (*i.e.*, the $\mathbf{k}_{32}^{*c}$ ground state) is present within the respective $\Delta E^*/N$-interval; the colors from red to purple and black (see colorbar) emphasize the level of "$\Delta E^*/N$-degeneracy" at a given pair of the system parameters, *i.e.*, the number of $\mathbf{k}_{32}^{*c}$ families which exhibit an energy difference to the ground state – at a given (η, A)-pair – of at most $\Delta E^*/N$. Phase boundaries from literature [62–64] (cf. Fig. 3.2) are emphasized by white lines.

the following questions for any structure: "what family does it belong to?", "where does it appear in the phase-diagram" and "what are its characteristic features?" (see Appendix A.1.2 for numerical details on the characteristic features of the here employed $\mathbf{k}_{32}^{*}$-clustering classification scheme of the structural data from Refs. [62–64]).

In conclusion of this Subsection we can say that the clustering tools discussed here may be of great help in understanding complex phase-diagrams like the one of the asymmetric Wigner bilayer system. When comparing the different clustering results in Fig. 3.14 it is apparent that especially first order transitions with clear jumps in the order parameters of different phases can be resolved accurately. However, second order phase transitions, such as those between phases II $\rightarrow$ III and III $\rightarrow$ IV (cf.

performing classification tasks [119]. For our purposes, such a task would be to classify structural data into a number of K different categories (identified, for instance, by unsupervised clustering) which would allow as to directly classify a structure from its geometric, structural data, e.g. via coordinates and lattice vectors [278]. The output of the classifier would then be the probability of a structure falling into any of the K clusters or families (when using "softmax" activation in the output layer of the neural network and "categorical cross entropy loss" during training [119]), which may give additional insight when comparing competing structures.

Subsection 3.1.4.3), represent a major challenge for these kinds of algorithms.

Notably, uncertainties in the results of clustering algorithms especially at phase boundaries indicate that the feature vectors (which are abstract representations of a structure) may not provide sufficient information to the clustering algorithm to uniquely resolve the involved different phases – and thereby can teach us about important symmetries of the investigated physical systems. To improve the reliability of clustering algorithms it is therefore important to identify so-far missing features which are necessary to describe a family of structures or to describe the transition between different families of structures (this is usually not an easy task).

On the other hand, if certain data points of a structural data set are frequently labeled qualitative differently by the different results of similar clustering algorithms[25] (or by other classification methods) – *i.e.*, frequent deviations which cannot be related to noise in the data – this can have physical implications: a persistent uncertainty in the clustering results of a certain sub-set of a data set might indicate the proximity of a continuous phase transition between structural families with respect to the system parameters that are related to this specific sub-set of the structural data [279].

[25] By similar clustering algorithms we here mean either differently initialized or slightly differently parametrized clustering algorithms which are conceptually similar (*i.e.*, which are expected to perform the categorization of the data in a conceptually similar way).

3.1.5 Towards Dodecagonal Order: Metastable Super-Structures

At certain values of the plate separation η and the plate-asymmetry A, *i.e.*, in the vicinity of $\eta = 0.55$ and $A = 0.67$, a novel snub-square ground state configuration emerges: the S_1 phase [64], see Fig. 3.1(b,d) and Fig. 3.2. The structure in layer one of S_1 features equilateral triangles and squares arranged in a snub-square tiling [255] (cf. second panel in Fig. 3.21) whose vertices are the charge positions. The tiling of layer of S_1 is a perfect square tiling whose vertices are located at the geometric centers of the square tiles of layer one projected to layer two.

The snub-square tiling is known to be a precursor for dodecagonal quasicrystalline ordering for certain two-dimensional systems (of structures tiled by squares and equilateral triangles) [252, 254, 273, 277]. To investigate the possible emergence of dodecagonal ordering in the asymmetric Wigner bilayer system, we investigate in this Subsection a family of structures which are solely tiled by squares and triangles in layer one and whose vertices in layer two are placed at the projected geometric centers of the squares of layer one (following the above mentioned decoration scheme of S_1 in layer two). We refer to the family of thereby generated structures (or tilings) as the dodecagonal *type I* (DI) family[26] and to the corresponding structures as DI structures; the respective decoration scheme for generating a DI structure from an arbitrary square triangle tiling in layer one is referred to as DI decoration scheme.

In Subsection 3.1.5.1 we first briefly discuss some basic, yet general geometric relations of square-triangle tilings and then provide in Subsection 3.1.5.2 an overview of zero temperature structural ground states of the asymmetric Wigner bilayer system which are structurally related to S_1, *i.e.*, which feature (distorted) square-triangle tilings and traits related to dodecagonal tilings in general. We then present the Stampfli-inflation scheme in Subsection 3.1.5.3 which will allow us to generate ever larger DI structures which gradually, *i.e.*, by increasing the size of the structures, approximate dodecagonal quasicrystalline ordering. We will then introduce an efficient optimization scheme in Subsection 3.1.5.4 to identify DI configurations which minimize the energy (per particle) in the configuration space of DI tilings with a certain unit cell size. After that, we present and discuss in Subsection 3.1.5.5 DI ground state configurations for different DI configuration spaces (each defined by fixed lattice vectors and by a fixed number of particles in the unit cell) and we investigate the mechanical stability of these structures in Subsection 3.1.5.6. Eventually we compare in Subsection 3.1.5.7 the identified DI ground state configurations with the zero temperature ground state structures, suggested by Refs. [62–64].

3.1.5.1 Square-Triangle Tilings: Vertices and Defects

A square-triangle tiling is a two-dimensional tesselation solely composed (*i.e.*, tiled) by two different geometric shapes, *i.e.*, by squares and equilateral triangles, fully covering the plane without overlaps and without gaps.

[26] S_1 is, for instance, a member of the DI family.

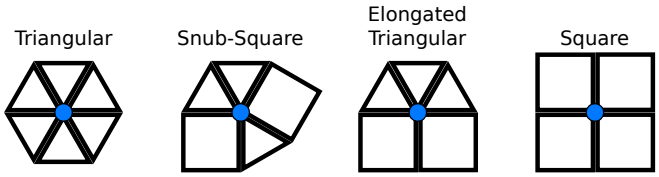


Figure 3.21: Four possible local environments (formed by squares and equilateral triangles) of vertices (emphasized by filled blue circles) of any square-triangle tiling [255]. To obtain $360°$ degrees around a vertex with regular polygons with opening angels of $90°$ (squares) and $60°$ (equilateral triangles) an even number of squares (i.e., zero, two or four) is required. Similarly, the possible number of triangles forming such a vertex can only be six, three or zero. Hence the four depicted vertices are the only geometrically allowed ones. If not combined (and when properly rotated) these four vertices can give rise to (from left to right) a triangular–, a snub-square–, an elongated triangular– and a pure square–tiling [255]. Hence we here label the four vertices according to their associated tiling as *triangular vertex*, *snub-square vertex*, *elongated triangular vertex* and *square vertex*.

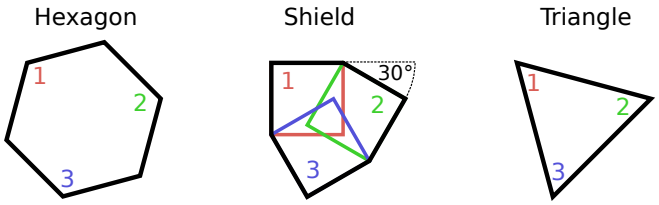


Figure 3.22: Transformation of a hexagon to a shield to a triangle tile (thick black shapes from left to right); fixed vertices (1), (2) and (3) are labeled and color-coded. A perfect shield tile (central panel) is composed by three interlocked squares (see color coded squares) with a relative angle of $30°$ to each other.

In general, every square-triangle tiling can be classified into four local environments which we list and illustrate in Fig. 3.21. We here label these four vertices according to the tiling they are associated to as *triangular vertex*, *snub-square vertex*, *elongated triangular vertex* and *square vertex*: a triangular tiling only features triangular vertices, a snub-square tiling only consists of snub-square vertices (in two possible orientations which are related by $90°$), an elongated triangular tiling is composed by two types of elongated triangular vertices (related by a $180°$ rotation) and a square tiling is solely composed of square vertices [255].

A well known defect in the tesselation of otherwise perfect square-triangle tilings are so called shield tiles [254, 255] (or shields, in shorthand notation), which can be thought of as being composed of three interlocked squares as shown in the central panel of Fig. 3.22. Shield tiles are deformed hexagons whose small opening angles (indicated by the numbered corners in Fig. 3.22) can continuously be varied between $120°$ and $60°$ to form an intermediate tile between a regular hexagon, a perfect shield tile and an equilateral triangle as depicted in Fig. 3.22. Perfect shields can also represent valid tiles in dodecagonal quasicrystals additional to squares and equilateral triangles [253, 280] but we explicitly omit shield tiles in our investigations of the DI family in Subsection 3.1.5.

Other tiles which frequently emerge in dodecagonal quasicrystals are $30°$-rhombi with the same side-lengths as the squares and equilateral triangles of the basic square-

triangle tile-set (not shown here); we do not consider such motives to be part of the tiling in layer one in DI structures in our investigations.

3.1.5.2 Snub-Square (S_1) and Related, Clustered Phases in the Asymmetric Wigner Bilayer System

To gain a deeper understanding of the problem of studying dodecagonal ordering in the asymmetric Wigner bilayer system and to exploit competing structures of the DI family, we present in Fig. 3.23 the structural families identified by the $\mathbf{k}_{32}^{*}$-clustering approach[27] (see Subsection 3.1.4) which are structurally similar to S_1 (or DI), *i.e.*, families of structures featuring (defective and distorted) square-triangle tilings in layer one (see also Fig. 3.18(c)). We highlight the respective regions in the (η, A)–phase diagram where the corresponding $\mathbf{k}_{32}^{*c}$-families represent the ground state candidates of the system. The different structure families are referred to as $\mathbf{k}_{32}^{*c}$, where the index $c = 1, \ldots, 32$ represents the unique (yet arbitrary) family-labels identified by the $\mathbf{k}_{32}^{*}$-clustering algorithm.

Via the structural family $\mathbf{k}_{32}^{*15}$ (emphasized in red in Fig. 3.23) the clustering approach identified the S_1 (snub-square) structure. As already mentioned above, the layer one structure of S_1 is a realization of a square-triangle tiling where two squares and four triangles compose the irreducible unit cell of the configuration. The corresponding structure in layer two also forms a special realization of a square triangle tiling, namely a pure square tiling whose vertices, *i.e.*, whose charge positions, are located at the geometric centers of the squares in layer one. The side length of the squares and equilateral triangles (or equivalently the tiling length of the square-triangle tiling) in layer one is given by a_1 and the tiling length of the square tiling in layer two is $a_2 = a_1 \cdot (1 + \sqrt{3})/2$.

We find that $\mathbf{k}_{32}^{*30}$ (emphasized in yellow in Fig. 3.23), $\mathbf{k}_{32}^{*7}$ (green) and $\mathbf{k}_{32}^{*24}$ (cyan) are distorted square triangle tilings but still resemble S_1: while $\mathbf{k}_{32}^{*30}$ maintains in layer two a (slightly distorted) square tiling, the square-triangle tiling in layer one is distorted such that certain square-triangle pairs may be interpreted as admittedly strongly distorted pentagons (highlighted by green tiles in the $\mathbf{k}_{32}^{*30}$ sub-panel of Fig. 3.23). Structures from the $\mathbf{k}_{32}^{*7}$ family also represent a distorted version of a square-triangle tiling in layer one and the tesselation in layer two is formed by tiles resembling rhombi. Interestingly, the corresponding structure in layer one features an approximately regular yet incomplete dodecagon (cf. blue emphasized area in the $\mathbf{k}_{32}^{*7}$ sub-plot of Fig. 3.23) hosting two distorted pentagons (emphasized by green tiles in the $\mathbf{k}_{32}^{*7}$ sub-plot of Fig. 3.23). The $\mathbf{k}_{32}^{*24}$ family features a distorted square-triangle tiling in layer one and the tiling

[27] The $\mathbf{k}_{32}^{*}$-clustering approach introduced in Subsection 3.1.4 relies on unsupervised k-means clustering (see Subsection 2.4.2) of the database of structural ground state candidate configurations of the asymmetric Wigner bilayer system identified in Refs. [62–64] (or better of the data set of order parameters of the structural database, which we additionally subjected to principal component analysis, see Subsection 2.4.1). Via information-theoretical measures based on the adjusted mutual information score between different clustering results (see Subsection 2.4.5), we could identify a reasonable "guess" for a total number of K^{*}=32 different structural families in the structural database; the asterisk emphasizes that we manually assigned all uniquely defined hexagonal monolayer configurations of the structural database as a family and excluded the corresponding data points from the k-means clustering procedure of the remaining data.

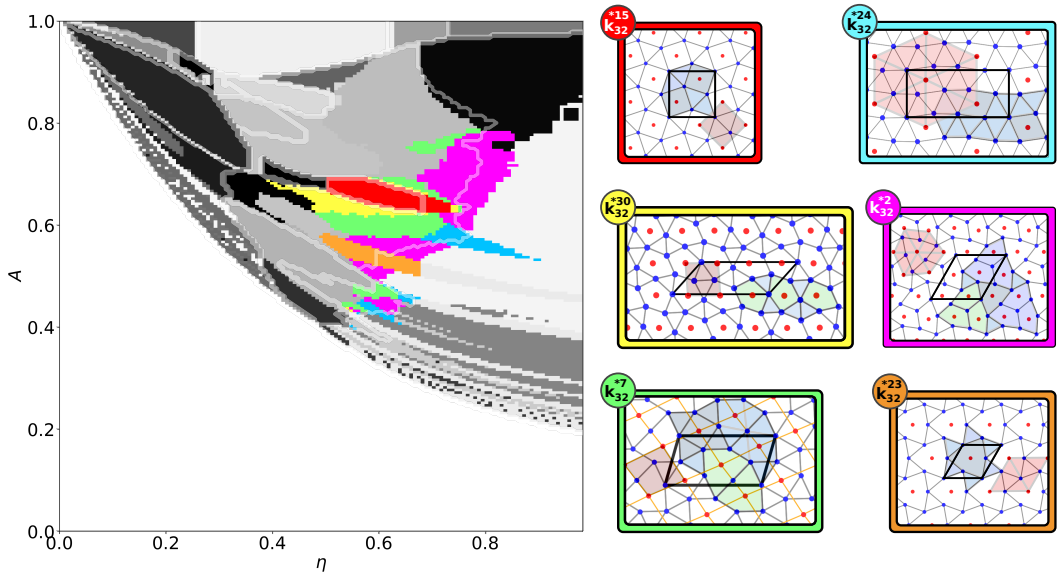


Figure 3.23: Refined zero temperature ground state phase diagram of the asymmetric Wigner bilayer system from literature [62–64] (left) and corresponding archetypical structures which are related to the S_1 configuration (right) based on the $\mathbf{k}_{32}^{*c}$ structural families which were identified by the K^*=32-means clustering (cf. Subsection 3.1.4) subjected to all structural data considered in Refs. [62–64]. The opaque, white lines in the left panel correspond to the phase boundaries from literature [62–64], cf. Fig. 3.2. The associated ground state regions of all $c = 1, \ldots, 32$ $\mathbf{k}_{32}^{*c}$ structural families are emphasized in gray scale in the left panel (each gray scale value – not shown here – represents one family). Representative structures of the $\mathbf{k}_{32}^{*c}$ families, which are related to the S_1 phase (see top left structure of the $\mathbf{k}_{32}^{*15}$ family which is emphasized in red) and, more generally, to the dodecagonal *type I* family, are explicitly depicted in the right part of the plot; special tiles emphasizing the tiling of a structure are highlighted by opaque colors and the respective unit cells are indicated by thick black frames. The different $\mathbf{k}_{32}^{*c}$ families are highlighted by different background colors and the associated region in the phase-diagram (left), where the depicted $\mathbf{k}_{32}^{*c}$ families represent the ground states of the asymmetric Wigner bilayer system, are appropriately color coded. We see that most of the depicted tilings exhibit a (distorted) square triangle tiling in layer one or even feature shield-like tiles. Many tiles of the structures presented here may give rise to dodecagonal motives. For all depicted configurations, the layer two configuration features a very uniform structure composed either of pure squares but also of triangles or (slightly distorted) rhombic tiles. See Appendix A.1.2 for details on characteristic values and boundaries of the order parameters and principal components for the $\mathbf{k}_{32}^{*}$-clustering families of structures.

in layer two (emerging from vertices placed at the geometric centers of the distorted squares of layer one) features rhombi rather than squares.

Two interesting structural families are presented in Fig. 3.23 by $\mathbf{k}_{32}^{*23}$ (emphasized in orange in Fig. 3.23) and $\mathbf{k}_{32}^{*2}$ (magenta): $\mathbf{k}_{32}^{*23}$ is composed of one (close-to-triangle) shield tile (cf. Fig. 3.22) and six adjacent, equilateral triangles (also counting periodic images of the two triangles present in the unit cell). The geometric center of the shield tile is decorated by a charge in layer two which thereby forms a triangular tiling. The layer one tiling of the more complex presented structure of the $\mathbf{k}_{32}^{*2}$ family is composed of one equilateral triangle, one (close-to-triangle) shield tile and three distorted pentagonal tiles arranged almost in a triangle (emphasized with green color). The shield and pentagonal tiles are decorated at their geometric center with a charge in layer two such that a (slightly distorted) triangular tiling emerges in layer two. The different tiles of the $\mathbf{k}_{32}^{*2}$ structure are arranged in an almost self-similar pattern, forming a larger, more complex cluster of tiles resembling a large triangle.

The perfect and distorted square-triangle tilings in layer one of the DI-related ground state structures of the asymmetric Wigner bilayer system can be quite versatile for different system parameters. However, we find that the corresponding structures in layer two appear to be highly uniform: we identify square tiles, triangular tiles and (slightly distorted) rhombic tiles. A mixture of squares and triangles in layer two of ground state candidate configurations of the asymmetric Wigner bilayer system has not yet been reported in literature [62–64].

3.1.5.3 Systematic Stampfli-Inflation

Dodecagonal quasicrystals composed of squares and triangles represent infinitely large structures which feature never repeating patterns of squares and triangles and thereby exhibit global dodecagonal rotational symmetry [252–254, 277]. In an effort to systematically approximate aperiodic dodecagonal quasicrystalline order with DI structures (or DI *approximants*) in the asymmetric Wigner bilayer system, we here introduce a way to increase the complexity of the unit cell of square-triangle tilings using so-called *Stampfli-inflation* [252] (explained in detail below). Stampfli-inflated square-triangle tilings represent a series of finite sized approximants to a dodecagonal quasicrystal: global dodecagonal rotational symmetry is successively approximated via repetitive Stampfli-inflation. Inflating a tiling is a purely geometrical process and in the Subsections to follow we will address the question if large Stampfli-inflated DI structures – more closely approximating a quasicrystal – do (or do not) have energetic advantages compared to smaller, more crystal-like structures in the asymmetric Wigner bilayer system.

During Stampfli-inflation each vertex of a square-triangle tiling is replaced by a regular dodecagon composed of equilateral triangles and squares of side lengths $a_1/(2+\sqrt{3})$; a_1 being the tiling length of the initial square-triangle tiling. In order to avoid overlaps these dodecagons can be assigned to the vertices of the initial square-triangle tiling in two different ways which are related by rotations of 30°, see Fig. 3.24. During Stampfli-inflation this rotation angle is chosen at random with 50% probability [254]. Holes between the dodecagons are filled with appropriate square-triangle motives. Eventually,

the lattice vectors of the structure are rescaled by a factor of $(2+\sqrt{3})$ and thereby a Stampfli-inflated square-triangle tiling with a tiling length of a_1 is generated. Via periodic *stacking* of the unit cell, *i.e.*, by extending the unit cell to its neighboring periodic images along the positive x and y directions, we can additionally quadruple the size of the unit cell. In that way Stampfli-inflation allows us to generate ever larger square-triangle tilings: each inflation step scales the number of particles in the irreducible unit cell by a factor of (approximately) $(2+\sqrt{3})^2$, while stacking increases the number of particles by an additional factor of four (cf. Fig. 3.25).

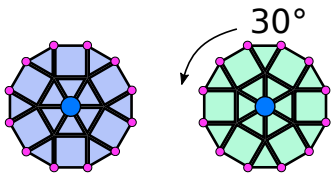


Figure 3.24: Schematic representation of Stampfli-inflation of a single vertex of a square-triangle tiling (indicated by blue filled circles, cf. Fig. 3.21) by a regular dodecagon (twelve outer vertices are highlighted by magenta circles) composed by equilateral triangles and squares. Every vertex of the inflated dodecagon represents (in our case) the position of a charge in layer one in the Stampfli-inflated tiling. There are two valid variants of dodecagons in the inflation procedure (left and right) which are related by a rotation of 30°. During Stampfli-inflation both variants are used at random with 50% probability [254]. In this thesis, we employ the following color-coding: dodecagons with a "horizontal" triangular vertex in the center, *i.e.*, a central hexagon composed of triangles which exhibit one side which is parallel to the horizontal axes) are filled blue (left), those with an additional 30° rotation are filled green.

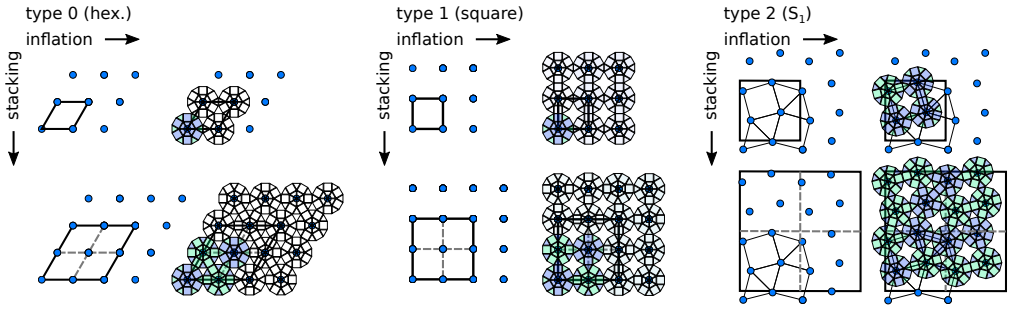


Figure 3.25: Schematic representation of Stampfli-*inflation* and *stacking* of *type 0* (hexagonal lattice, left), of *type 1* (square lattice, middle) and of *type 2* (S_1 or snub-square tiling, right). In each type-specific sub-plot (left to right) the inflation procedure is indicated horizontally (from left to right) and stacking is performed vertically (from top to bottom). The inflated dodecagons have two allowed orientations with a respective angle of 30° to each other, which is indicated by blue and green filling of dodecagons; periodically replicated dodecagons are not filled. Notably, that the tiles in the snub-square type of approximants are additionally rotated by 15° with respect to *type 0* and *type 1*.

By assigning the vertices of Stampfli-inflated square-triangle tilings to particles in layer one of a Wigner bilayer structure and performing DI decoration in layer two we cover DI structures with 6, 22, 82, 306, 1142 and 1224 particles per unit cell[28]. Notably, the number of equilateral triangles and squares of a square-triangle tiling with fixed periodic

[28] The unit cell is scaled in the x and y directions – without the loss of generality [62] – such that the in-plane particle density is constant throughout all configurations considered in this thesis.

boundary conditions (which are determined by the Stampfli-inflation procedure) is also fixed [252, 254]. Consequently, the inflation procedure fixes the number of particles in layer one, N_1, and layer two, N_2, and thus the composition, $x = N_2/N$.

We consider three qualitatively different families of structures which are suitable starting configurations to systematically approximate dodecagonal quasicrystalline order via Stampfli-inflation: the hexagonal lattice (*type 0*), the square lattice (*type 1*) and the snub-square or, equivalently, S_1 structure (*type 2*), see Fig. 3.25. These different families represent three different types of approximants to a dodecagonal quasicrystal with different symmetries, periodic boundary conditions and compositions. To distinguish between different Stampfli-inflated square-triangle tilings of the dodecagonal *type I* family, *i.e.*, DI, we introduce the shorthand notation

$$\mathrm{DI}^{(T)}_{I,S}, \tag{3.14}$$

where the superscript (T) specifies the *type of the approximant* of the initial square-triangle family (left to right in Fig. 3.25), the left index, I, specifies the *inflation order* and the right index, S, the *stacking number* [254]. With this notation, the original lattices depicted in Fig. 3.25 can be identified as $\mathrm{DI}^{(0)}_{0,1}$ (hexagonal lattice), $\mathrm{DI}^{(1)}_{0,1}$ (square lattice) and $\mathrm{DI}^{(2)}_{0,1}$ (snub square structure). Tilings which are inflated once (top right panels of respective types, T, in Fig. 3.25) can be addressed by $\mathrm{DI}^{(T)}_{1,1}$. Zero inflated, once stacked lattices (bottom left panels for respective types, T, in Fig. 3.25) can be written as $\mathrm{DI}^{(T)}_{0,2}$ and once inflated and once stacked lattices (bottom right panels of respective types, T, in Fig. 3.25) are written as $\mathrm{DI}^{(T)}_{1,2}$.

Note that each particular Stampfli-inflation procedure – via choosing T, I and S – gives rise to a respective configuration sub-space of $\mathrm{DI}^{(T)}_{I,S}$ structures which is related to the particular choice of the fixed periodic boundary conditions.

3.1.5.4 Ground State Identification by "Zipper" Annealing and Parallel Tempering

In the Subsection above we have introduced a strategy to generate ever larger square-triangle tilings (or better DI structures) through Stampfli-inflation and stacking. Here, we introduce a way how to efficiently explore the configuration space of DI structures for the respective ground state which minimizes the energy per particle, $E^*/N \to \min$.

In a first step of structure optimization we aim at identifying *the* best DI configuration (featuring a square-triangle tiling in layer one and following the DI decoration scheme in layer two) with a fixed number of particles in the unit cell and with fixed periodic boundaries[29] which is all determined by the specific choice of the inflation procedure, *i.e.*, by the specific choice of T, I and S in $\mathrm{DI}^{(T)}_{I,S}$. In general this is a sampling problem in a discrete yet potentially huge configuration space of equilateral square-triangle tilings in layer one where layer two is, again, always appropriately decorated. To perform the required structure optimization for a given Stampfli-inflated

[29] I.e., with fixed lattice vectors, with a fixed compositions $x = N_2/N$ and, notably, with a fixed number of squares and triangles in the unit cell of layer one.

DI configuration we perform specifically implemented simulated annealing and parallel tempering (or replica exchange Monte-Carlo[30]) methods described below (see also Subsection 2.2.3).

In order to efficiently sample the configuration space of square-triangle tilings (of fixed composition and periodic boundary conditions) we here rely on so called *zipper*-moves [254]. A zipper-move randomly reshuffles a square-triangle tiling according to the following protocol (see Fig. 3.26):

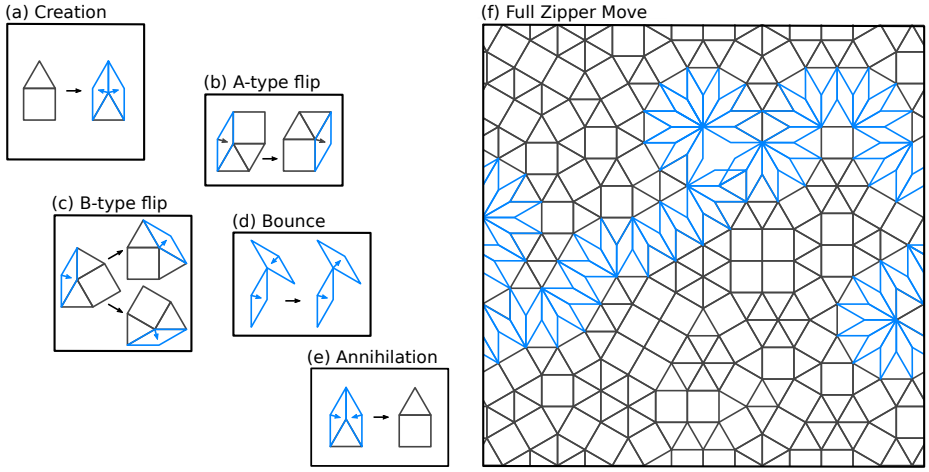


Figure 3.26: Visualization of all basic steps (a-e) between the creation (a) and annihilation (e) of a complete zipper-move (f). This figure is a combination of two recreated figures from Ref. [254], however, in panel (f) we omit all square and triangle tiles which are shuffled by the propagating rhombi (according to panels(a-e)) and we only present the rhombi which represents the entire zipper move in blue color. Panel (f) poses one elementary update move (*i.e.*, a *sweep*) in the Monte Carlo based sampling procedures discussed in the text.

1. Starting from a special type of local environment composed of one equilateral triangle and one square, *i.e.*, a so-called *house* (see Fig. 3.26(a)), two rhombi and one triangle are *created*. Each rhombus is assigned with an outwards pointing arrow labeling its direction of propagation during each step of the zipper move.

2. While the rhombi propagate through (and thereby reshuffle) the square-triangle tiling they can encounter four possible local environments (*i.e.*, the vertices their respective arrows point to) which specify how the rhombi propagate through the tiling via updates of the local environments.

   2.1 If the rhombi encounter one of the three local environments illustrated in Fig. 3.26(b,c,d) updates to these environments are performed which are labeled *A-type flip*, *B-type flip*[31] and *Bounce* and the rhombi propagate through the tiling.

[30] Although simulated annealing is a viable tool, the complexity of larger configurations and the associated dramatic increase in computational costs (especially when evaluating long-ranged Coulomb interactions as accurate as possible) limits its applicability which is why we turned to parallel tempering.

[31] The *B-type flip* can be performed in two possible ways, which is chosen at random in the algorithm.

2.2 If the two rhombi meet again in a local environment which is illustrated by Fig. 3.26(e) they *annihilate* into a house and the zipper-move is complete. Otherwise updates according to step 2.1 are performed until the rhombi eventually annihilate.

A complete zipper-move[32] is schematically depicted in Fig. 3.26(f). Such a zipper-move connects different square-triangle tilings from the same configuration space ergodically, *i.e.*, any square-triangle configuration can be transformed into any other from the same configuration space by a sequence of zipper-moves.

One complete zipper-move of the square-triangle tiling in layer one and the appropriate DI decoration in layer two together represent an elementary update move, *i.e.*, a *sweep*, in the sampling procedure of both, simulated annealing and replica exchange Monte Carlo. Such a zipper-induced *sweep* is evaluated according to the standard Metropolis Hastings [191] acceptance criterion: the sweep is accepted if a uniform random number $r \in [0,1] \leq \min\left(1, \exp\left(-\Delta E^*/T_i^*\right)\right)$ and rejected otherwise; here, $\Delta E^* = E^{*\prime} - E^*$ is the energy difference between the updated sample (or, equivalently, configuration) with an energy of $E^{*\prime}$ and the original sample with an energy of E^*, defined per particle N in Eq. (3.6), and $T_i^* = k_\mathrm{B}T_i/(\sqrt{\rho}e^2)$ is the dimensionless temperature of the current annealing step or the respective replica exchange Monte Carlo sample. Either by cooling down the temperature of a simulated annealing sample slow enough or by applying replica exchange Monte Carlo over a sufficiently long time-span (cf. Subsection 2.2.3) we explore the respective configuration space of $\mathrm{DI}_{I,S}^{(T)}$ structures (determined by the composition and the periodic boundaries) for the best structures which minimize E^*.

3.1.5.5 Self-Similar Super-Structures

In Fig. 3.27 we present results for the structural ground state search executed with simulated annealing for a $N = 82$ particle *type 2* approximant (*i.e.*, a Stampfli-inflated snub-square approximant which we can label $\mathrm{DI}_{1,1}^{(2)}$) which we perform at a plate separation of $\eta = 0.68$ and $A = 0.5$. We start the energy minimization from a random square triangle tiling by applying 100 zipper moves prior to any optimization. The cooling of the sample subject to the simulated annealing procedure happens in 85 equidistant temperature steps starting from $N\beta^* = N/T^* = 0$ and ending at $N\beta^* = N/T^* = 4000$. At each temperature we apply ten zipper induced sweeps to equilibrate the sample.

When analysing the corresponding evolution of the sample energy over the process of annealing (*i.e.*, at the end of each equilibration phase of successive annealing steps) depicted in Fig. 3.27 we see that there appear to be nearly discrete jumps in the sample's energy (especially for low energy configurations at small values of the temperature). This indicates that only very few configurations are competitors to the ground state, a feature which we will discuss in detail below. The final structure is highly ordered such that four regular dodecagons emerge in layer one (two of each kind of the allowed orientations as depicted in Fig. 3.24). Furthermore, when connecting the centers of

[32] The zipper-move got its name from its apparent resemblance to opening and closing a zip fastener [254].

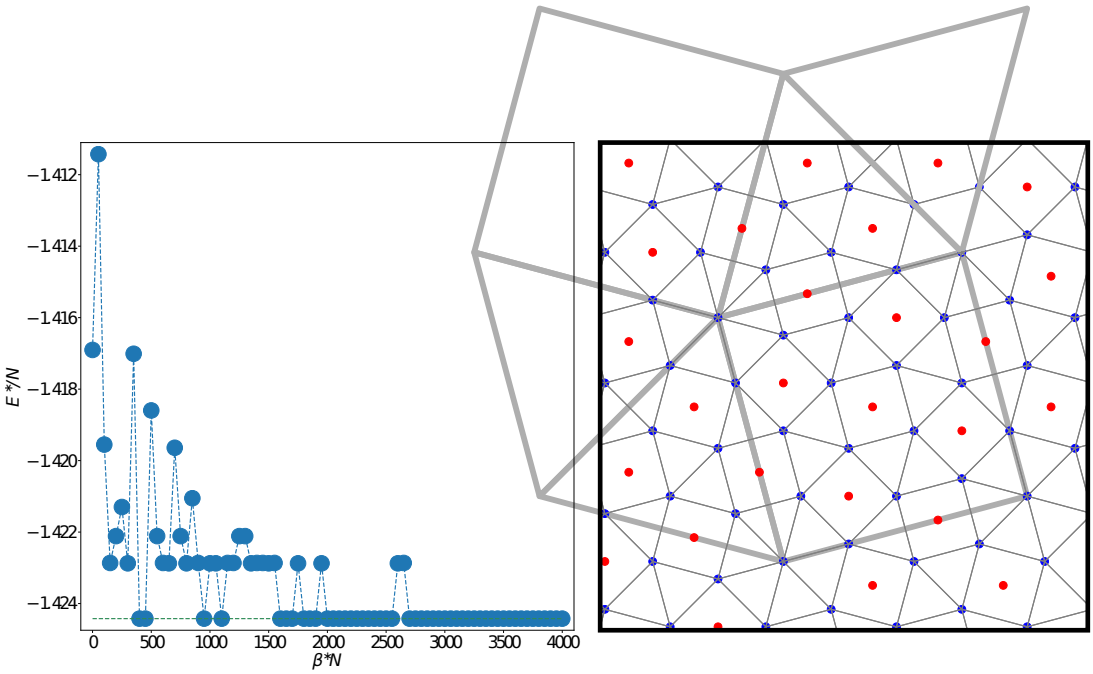


Figure 3.27: Left: Energy per particle, E^*/N, at the end of each equilibration phase of a zipper based simulated annealing procedure in order to identify the ground state for the $\mathrm{DI}^{(2)}_{1,1}$ family of structures (cf. Eq. (3.14)) with $N = 82$ particles in the unit cell as a function of $N\beta^* = N/T^* \in [0, 4000]$ for the system parameters $\eta = 0.59$ and $A = 0.68$; the temperature decreases from left to right in 85 steps. Due to the zipper update moves in layer one particles in layer one (blue dots) always form vertices of equilateral square-triangle tilings and particles in layer two (red dots) are placed, according to the DI decoration scheme, at the projected geometric centers of squares in layer one in each step of the simulation. Right: Unit cell of the highly ordered ground state structure which minimizes E^*/N for the $\mathrm{DI}^{(2)}_{1,1}$ family. Connecting the centers of neighboring dodegacons in layer one (thick gray lines) reveals a super snub-square tiling on a length scale of $d_1 = (2 + \sqrt{3})a_1$; a_1 is the tiling length of layer one. The value of the optimized energy per particle, $E^*_{\mathrm{opt}}/N = -1.424$, has to be compared to the ground state energy from literature [63, 64], $E^*_{\mathrm{GS}}/N = -1.433$ (which corresponds to the S_1 structure), at the same system parameters.

the dodecagons (*i.e.*, the central triangular vertices) a *super*-structure of equilateral squares and triangles emerges on a larger length scale of $d_1 = (2 + \sqrt{3})a_1$: a super snub-square structure; a_1 is the tiling length of layer one and d_1 is the short diameter of the dodecagons. The particles in layer two form either squares (as in S_1) but – necessarily by following the dodecagonal pattern in layer one – also form triangles and large hexagons which are centered (in projection) in the dodecagons of layer one. The pattern in layer two could be described by another, perfect square-triangle tiling on a length scale of $a_2 = a_1 \cdot (1 + \sqrt{3})/2$ if the centers of the dodecagons were also considered as vertices in layer two. However, pulling out the central triangular vertex of each dodecagon of layer one to form a triangular vertex in layer two at the center of the hexagonal tile is energetically less favorable as compared to the structure depicted in Fig. 3.27[33]. The associated optimal energy $E^*_{\mathrm{opt}}/N = -1.424$ (of the $\mathrm{DI}^{(2)}_{1,1}$ ground

[33] During the course of this thesis we also systematically investigated this idea of pulling triangular vertices at the centers of dodecagons in layer one into layer two during Monte Carlo sampling and,

state identified at the system parameters $\eta = 0.59$ and $A = 0.68$) is slightly higher compared to the ground state value from literature [62–64], $E^*_{\mathrm{GS}}/N = -1.433$, which corresponds to the S_1 structure.

For now, we focus on the structural properties of configurations emerging during the zipper-annealing and tempering procedures (see below) for DI approximants of different sizes and symmetries. The effects of subsequent relaxation of the emerging DI ground state structures and a comparison of the relaxed energies of the corresponding configurations with respect to each other and with respect to data from literature [62–64] are discussed in detail below.

Fig. 3.28 depicts the energies and temperatures along a replica exchange Monte Carlo (REMC) run for a $N = 306$ particle $\mathrm{DI}^{(1)}_{2,1}$ approximant (cf. Eq. (3.14)) where an original square lattice (*type 1 approximant*) is inflated twice. After inflation but prior to any optimization the tiling is randomized by 1000 zipper moves. We define a (fixed) range of $N_T = 16$ temperatures covering several magnitudes such that sample diffusion between neighboring temperatures is possible (see top panel in Fig. 3.28). For this setup we perform ≈ 500 tempering steps with 20 zipper induced sweeps in between for sample equilibration at each temperature. Again, the final ground state configuration, now for the $\mathrm{DI}^{(1)}_{2,1}$ family, exhibits a highly regular structure which features several dodecagons in layer one: when connecting triangular vertices which form the centers of dodecagons in layer one we can again identify emerging super-squares and super-triangles on a length scale of d_1. Here, clusters of four super-squares and four super-triangles give rise to yet another super square-triangle structure on a length-scale of $2 \times d_1$, *i.e.*, of twice the diameter of a layer one dodecagon, as emphasized by the cyan colored areas in the ground state inset of Fig. 3.28. The super tiling on the $2 \times d_1$ length scale appears to be a defective super-stacked square-triangle tiling which also features distorted, magenta-colored pentagons as additional super tiles.

An inspection of the entirety of samples generated in the simulated annealing and replica exchange Monte Carlo procedures, respectively depicted in Figs. 3.27 and 3.28, reveals that high-temperature random square-triangle tilings[34] order at low temperatures into self-similar super square-triangle structures on both plates, and on several length scales. We would like to stress that this is not a unique property of the two approximants depicted in Figs. 3.27 and 3.28. We moreover perform analogous optimization procedures for a variety of differently Stampfli-inflated structures – always starting from perfect but randomized DI structures of the corresponding $\mathrm{DI}^{(T)}_{I,S}$ families – and the respective ground state solutions (which could be identified) all show this remarkable emergence of highly ordered self-similar super square-triangle tilings. In Fig. 3.29 we explicitly present self-similar ground state configurations of several $\mathrm{DI}^{(T)}_{I,S}$ families, each evaluated at a fixed composition and with fixed periodic boundaries as determined by the specific choice of the Stampfli-inflation. Note that this analysis is

alternatively, before relaxation. Throughout all investigated configurations this turned out to be energetically less favorable as compared to the respective original DI structures (and even turned out to be unstable).

[34] We will discuss the relevance of random square-triangle tilings below when we investigate the stability of DI configurations with respect to relaxation of all planar particle coordinates and the lattice vectors.

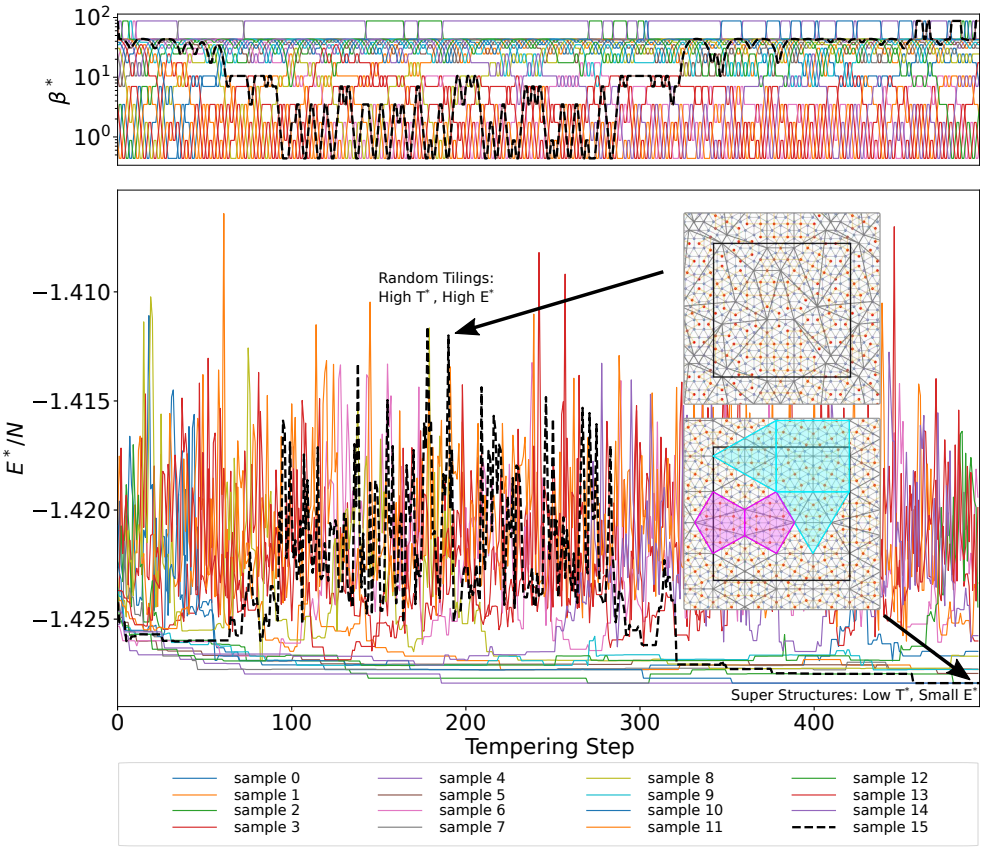


Figure 3.28: Replica exchange Monte-Carlo (REMC) energy minimization of a $N = 306$ vertices $\mathrm{DI}^{(1)}_{2,1}$ approximant (cf. Eq. (3.14)). The top panel shows REMC temperature trajectories, T^*_i, as functions of the tempering steps associated with all $i = 0, \ldots, 15$ samples, used in the sampling procedure, at the end of each equilibration phase between parallel tempering steps: samples are successively swapped between neighboring temperatures in tempering steps and are, respectively, depicted by a unique color throughout the simulation. The bottom panel shows the energy per particle at the end of successive tempering steps for all 16 samples simultaneously, same color coding as top panel. High energy random configurations encountered at high temperatures (top inset) give rise to ground state solutions with self-similar super-tilings at low temperature (see black-dashed trajectory resulting in bottom inset solution). The super tiling (gray lines in top inset and colored areas in bottom inset) is drawn by connecting neighboring centers of regular hexagons in layer one. Note that high energy random tilings are not stable with respect to gradient descent minimization.

performed for an entire range[35] in $\eta \in [0.01, 1.4]/\sqrt{2}$ but still – and practically irrespective of η – all ground state configurations exhibit self-similar behavior in their structural configuration.

In the first row of Fig. 3.29 we present ground state configurations of hexagonal (type $T = 0$) approximants of the $\mathrm{DI}_{I,S}^{(T=0)}$ family which are inflated once (*i.e.*, $\mathrm{DI}_{1,1}^{(0)}$ with $N = 19$, left panel) and inflated and stacked once (*i.e.*, $\mathrm{DI}_{1,2}^{(0)}$ with $N = 76$, middle and right panels). We see that in the left two panels in layer one non-overlapping dodecagons composed of equilateral squares and triangles emerge whose centers form a hexagonal lattice with a lattice constant of d_1 (*i.e.*, a super-triangular tiling). Furthermore, in this case the vertices in layer two describe a rhombitrihexagonal tiling [255], *i.e.*, overlapping, dodecagons (on a hexagonal lattice) with a central hexagonal tile (instead of a triangular vertex) and otherwise composed of squares and triangles with a tiling length of $a_2 = a_1 \cdot (1 + \sqrt{3})/2$. In the top right panel of Fig. 3.29 we rotated (by hand) the central dodecagon by 30° with respect to the top central panel (destroying the rhombitrihexagonal tiling in layer two) which results in a small energy penalty (and also an increased instability against relaxation, as we will see later).

In the second row of Fig. 3.29 we present ground state configurations of different $\mathrm{DI}_{I,S}^{(T=1)}$ approximant families that are based on Stampfli-inflation of a square lattice (type 1), *i.e.*, which are inflated once (*i.e.*, $\mathrm{DI}_{1,1}^{(1)}$ with $N = 22$, left panel), inflated and stacked once (*i.e.*, $\mathrm{DI}_{1,2}^{(1)}$ with $N = 88$, middle) and inflated twice (*i.e.*, $\mathrm{DI}_{2,1}^{(1)}$ with $N = 306$, right). In the left most panel we see that a single dodecagon emerges in the irreducible unit cell in layer one giving rise to super-square tiles (cf. green area) on a square lattice, *i.e.*, a super-square tiling. In contrast to all other tilings depicted in Fig. 3.29 two neighboring squares in layer one (indicated by red area) form two elongated triangular vertices (cf. Fig. 3.21). Such a square-square pair in layer one represents a defect in the otherwise uniform hexagon-square-triangle tiling in layer two (*i.e.*, a tiling composed of hexagons, squares and triangles) of the respective DI structure. Henceforward, we refer to such square-square neighbors in layer one as *atomic defects* and to structures which do not feature atomic defects as *defect free* structures. Stacking of the unit cell (as performed so from the left to the central panel in the second row in Fig. 3.29) enables the parallel tempering minimization to resolve this atomic defect simply by properly (re)orienting neighboring dodecagons in $\mathrm{DI}_{1,2}^{(1)}$ compared to $\mathrm{DI}_{1,1}^{(1)}$: the ground state structure of the $\mathrm{DI}_{1,2}^{(1)}$ family is composed of four dodecagons in the unit cell – their central triangular vertices forming a super-square tiling – whose orientations alternate by 30° (cf. Fig. 3.24) on the corners of the same edges of square tiles, thereby avoiding square-square pairs in layer one. The specific orientation of the dodecagons on the edges of the super-squares gives rise to yet another length-scale of a super square-triangle tiling: a super snub-square tiling emerges in layer two with a tiling length of $2 \times a_2 = (1 + \sqrt{3})a_1$ as indicated by the yellow and orange colored areas in the $\mathrm{DI}_{1,2}^{(1)}$ panel in Fig. 3.29. The right most panel in the second row of Fig. 3.29 shows the ground state configuration of the $\mathrm{DI}_{2,1}^{(1)}$ family, which has already been discussed above (cf. Fig. 3.28 and text): no atomic defects are present here and

[35] I.e., on the same grid $\eta \in [0.01, 1.4]/\sqrt{2}$ in steps of $\Delta\eta = 0.01/\sqrt{2}$ as in Ref. [62].

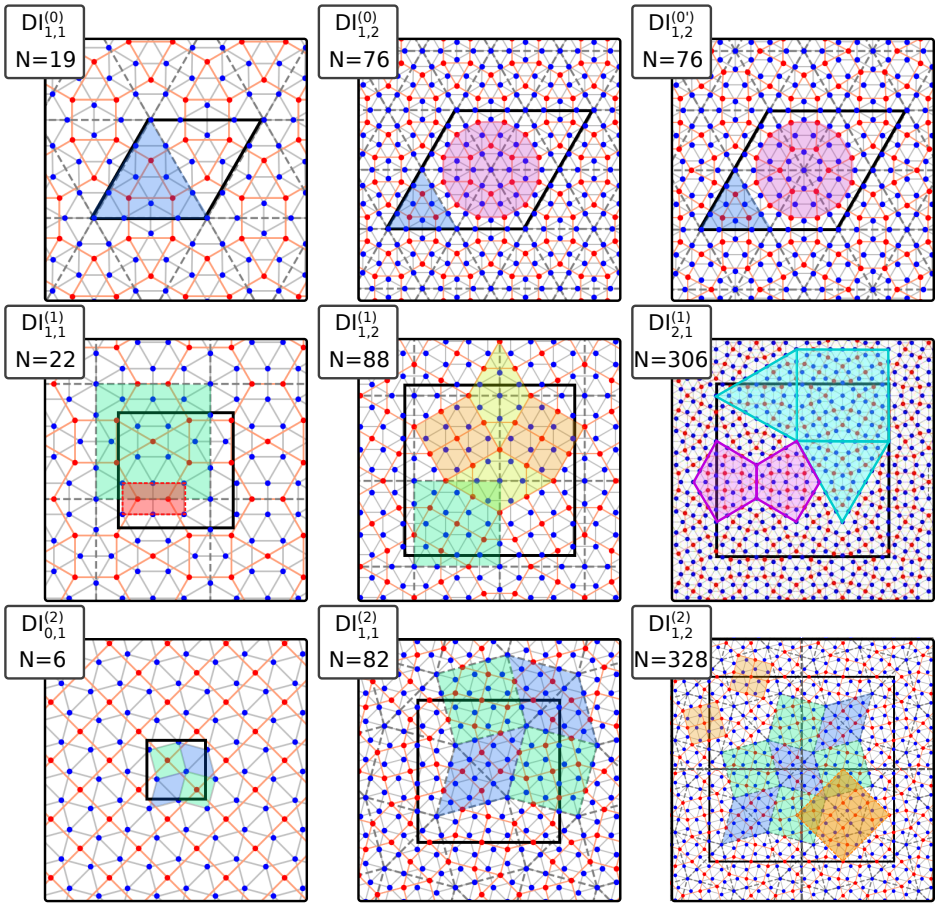


Figure 3.29: Ground-state configurations of several $\mathrm{DI}^{(T)}_{I,S}$ families (see structure labels, cf. Eq. (3.14)) evaluated at their respective periodic boundaries and composition by *zipper* based energy-minimization in a range of $\eta \in [0.01, 1.4]/\sqrt{2}$ [62]) for approximant types $T = 0$ (hexagonal lattice), $T = 1$ (square lattice) and $T = 2$ (snub-square structure) in the upper, middle and bottom row, respectively; particles in layer one (two) are colored blue (red) and the respective unit cells are indicated by thick black frames. The complexity of the unit cells increases from left to right via inflation, I, and stacking, S. All ground states, for the respective $\mathrm{DI}^{(T)}_{I,S}$ families show self-similar behavior (large colored squares, triangles and dodecagons) on several length scales when connecting centers of dodecagons in layer one. Top row: The ground state solutions of the hexagonal approximants show a closest packing of non-overlapping dodecagons in layer one and of overlapping dodecagons in layer two (indicated by magenta areas; *i.e.*, a rhombitrihexagonal tiling [255]). The central dodecagon in the right panel is rotated (by hand) by $30°$ as compared to the middle panel. Large blue-filled triangles indicate the triangular super-tiling. Middle row: The ground state solutions of the different square approximants show dodecagons which are typically arranged on a super-square lattice (green tiles in the left and central panel, cyan square in the right panel). The super tiling of the right most structure is defective (as indicated by the magenta pentagons) and the left most structure features neighboring squares in layer one (red colored squares indicating an elongated triangular vertex, cf. Fig. 3.21). Bottom row: The left most structure (i.e., S_1), is the most stable one in the S_1 region [62–64]. The ground states of the next larger approximants (bottom center and right panels) form self-similar super snub-square tilings when connecting dodecagon-centers in layer one; super tiles are again indicated by colored areas.

super tiles are formed although the $d_1 = (2+\sqrt{3})a_1$ super square-triangle tiling as well as the $2 \times d_1$ "double"-super tilings are defective.

The last row of panels in Fig. 3.29 shows results for the ground state search performed for different snub-square based (type 2) $\mathrm{DI}_{I,S}^{(T=2)}$ families of approximants, specifically for $\mathrm{DI}_{0,1}^{(2)}$ (left panel, $N = 6$), for inflated snub-square approximants $\mathrm{DI}_{1,1}^{(2)}$ (middle panel, $N = 82$) and for inflated and stacked approximants $\mathrm{DI}_{1,2}^{(2)}$ (right panel, $N = 328$). The ground state depicted in the left most panel is equivalent to the S_1 structure itself[36] featuring a snub-square tiling in layer one and a pure square tiling in layer two with respective tiling lengths of a_1 and a_2. The next larger family of approximants, $\mathrm{DI}_{1,1}^{(2)}$, already exhibits $N = 82$ particles in the unit cell and we already discussed the results of the related ground state search via Fig. 3.27 above: similar to the ground state configuration of the related square type of approximants, $\mathrm{DI}_{1,2}^{(S=1)}$ (with $N = 88$ particles in the unit cell), the centers of the four dodecagons in the irreducible unit cell are located on the corners of a super-square tile (green colored areas with an edge length of d_1) with 30° alternating orientations on the corners of the same edges of the super-tiles, thereby avoiding square-square pairs in layer one. Here, however, two different super-square tiles are present in the unit cell which are rotated by ±15° as compared to $\mathrm{DI}_{1,2}^{(1)}$. This additional rotations of the super-square tiles further give rise to four super-triangle tiles (blue areas) in the ground state configuration of the snub-square based $\mathrm{DI}_{1,1}^{(2)}$ family. The entire structure then eventually forms the observed super snub-square tiling when connecting the central triangular vertices of all dodecagons across periodic directions of the square unit cell. In the right most panel of the last row in Fig. 3.29 we also present the results for a ground state search for the $\mathrm{DI}_{1,2}^{(2)}$ family: the additional stacking of $\mathrm{DI}_{1,1}^{(2)}$ has only little effect on the final pattern and a periodically stacked super snub-square phase emerges.

We also investigated larger approximant families, *i.e.*, the snub-square tiling based $\mathrm{DI}_{2,1}^{(2)}$ family with $N = 1142$ particles in the unit cell and the square lattice based $\mathrm{DI}_{2,2}^{(1)}$ family with $N = 1224$ for their ground state configuration with REMC and we present the results in Fig. 3.30. For this family of already considerably complex structures the optimization procedure turned out to be numerically rather demanding such that even after ten days of independent REMC procedures (performed for different values of η on 20 nodes on the *Vienna Scientific Cluster* [281]; each node being equipped with sixteen cores) the optimization did not terminate for any of the simulations. However, we also observe for these families a clear trend to form low energy configuration with self-similar super-tilings. For the (so far) best solution of the snub-square family approximants shown in the left panel of Fig. 3.30 we see the emergence of regular *super-super-dodecagons*, composed of super-square and super-triangle tiles. Furthermore, the centers of the super-super-dodecagons are arranged on super-super-square and super-super-triangle tiles and thereby form a super-super-snub-square tiling (although some super-super-dodecagons are defective and atomic defects are present in the structure).

[36] Notably, the smallest variant of the type 2 approximant families, *i.e.*, $\mathrm{DI}_{0,1}^{(2)}$, is the smallest square triangle tiling presented in this thesis with a non-trivial configuration space composed of two geometrically different configurations, namely of the snub-square tiling and the elongated triangular tiling (cf. Fig. 3.21).

The (so far) best configuration for the square family approximants depicted in the right panel of Fig. 3.30 suggests, that the final pattern exhibits large clusters of (at least) triply stacked super-squares and triangles[37] on a length-scale of $3 \times d_1$.

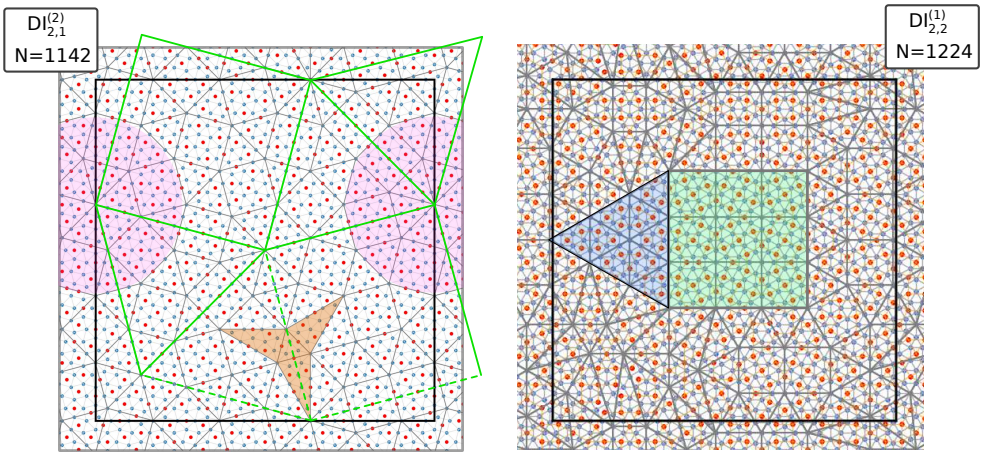


Figure 3.30: The presented tilings are the result of zipper-energy minimization (no relaxation) for approximants with $N = 1142$ (DI$^{(2)}_{2,1}$, left) and $N = 1224$ (DI$^{(1)}_{2,2}$, right) particles in the unit cell (indicated by thick, black, square frames). After ten days of independent parallel tempering simulations on 20 nodes (each equipped with 16 cores), no better results were obtained (both shown configurations are unstable with respect to relaxation). Left: The minimization of the $N = 1142$ approximant is performed directly from the Stampfli-inflated structure without initial randomization (in contrast to the right panel); no energetically more favorable results could be obtained when starting from a random tiling. The magenta area marks a super-super-dodecagon, the orange star marks a defect in the super-super-snub-square structure, which would emerge when connecting all super-super-dodecagonal centers. Right: The blue triangle and green square mark triple-stacked super-triangle and super-square tiles.

For sampling a configuration space of square-triangle tilings (or in our case DI structures) for large, highly ordered, self-similar ground states zipper moves turn out to become increasingly inefficient the larger configuration. Operating on the atomic length scale (*i.e.*, reshuffling the square-triangle tiling in layer one) only very few, very specific zipper moves, such as full rotations of d_1–sized dodecagons, may effectively decrease the energy of highly ordered structures at low temperatures while other, much more probable random zipper moves will result in an energy penalty. In fact, for tilings such as presented in Fig. 3.30 the majority of update moves at low temperatures is rejected since the probability of random, zipper induced sweeps to lower the energy of the sample is quite low; a "to-be-accepted" zipper induced sweep for an already highly ordered structure would have to be related to correctly reorienting or translating entire clusters of super-tiles which are composed of dodecagons. Notably, the symmetry of the self-similar super-structures is very low, resembling the symmetry of the original seed lattice used in the inflation procedure.

[37] In fact, only after observing the large block of super-triangles in the DI$^{(1)}_{2,2}$ approximant we became aware of the *type 0* approximant family: in the code written by Oxborrow *et al.* [254], which we used in this project, Stampfli-inflation and zipper sampling is only implemented for *type 1* and *type 2* approximants and, naturally, we only considered the later types in the beginning of the project.

Ground states of the different $\mathrm{DI}_{I,S}^{(T)}$ families display a tendency to solely form triangular and snub-square vertices in layer one and thereby avoid to form atomic defects (*i.e.*, elongated triangular and square vertices, cf. Fig. 3.21) – wherever this is geometrically possible. The observed triple stacking of dodecagons in super-squares and super-triangles in the right panel of Fig. 3.30 hints at a phase separation of super-square and super-triangular tilings which are here combined to one structure through stacked super-squares and super-triangles with a perfect interface (*i.e.*, without defects). Based on this idea we may circumvent the problem of numerically sampling the entire configuration space for highly ordered configurations: in Appendix A.1.5 we propose an inflation scheme of square-triangle tilings (properly combining Stampfli-inflation and stacking) to avoid square-square neighbors on the atomic length scale. This provides us with a mechanism to generate large self-similar super-structures from scratch – if this is geometrically possible. To this end, further investigations are required and for now – and in what follows – we will focus on the structures presented in Fig. 3.29.

We conclude the discussion of the zipper-minimized ground state structures by emphasizing that, given the periodic boundaries and compositions, the symmetries of the ground state super-structures strongly resemble the symmetry of the initially used type of approximants, *i.e.*, of the hexagonal lattice (type 0), the square lattice (type 1) or the snub-square tiling (type 2). We interpret this result based on the uniformly repulsive and long-ranged Coulomb interaction, that favors essentially identical structures at different length-scales on the two plates. In contrast, the plate-to-plate interaction locks square-triangle arrangements of the two planes into self-similar super-tiles, characteristic of a well-ordered quasicrystal state [45].

3.1.5.6 Relaxation, Noise and Stability

In this Subsection we will investigate the mechanical stability of the $\mathrm{DI}_{I,S}^{(T)}$ ground state structures identified in Fig. 3.29 via relaxation of the corresponding perfect DI configurations – exhibiting a perfect square-triangle tiling in layer one and DI decoration in layer two – as well as of slightly distorted versions thereof. This will allow us to relate the nearly discrete jumps in energy of DI structures during zipper annealing and tempering (cf. Figs. 3.27 and 3.28) with certain, local environments[38] of the DI configurations.

We test if a configuration represents a (local) energy minimum with respect to all (in-plane) particle positions and the lattice vectors via relaxation at a given set of the system parameters, η and A. To be more specific, we numerically minimize $E^*(\mathbf{r}^N; A, \eta)/N$ defined by Eq. (3.6) via gradient-descent minimization[39]: the gradient-descent minimization of $E^*(\mathbf{r}^N; A, \eta)/N$ is performed under the variation of all (in-plane) particle positions in the xy-directions and under the variation of the (in-plane) lattice vectors for a given $\mathrm{DI}_{I,S}^{(T)}$ bilayer structure, while respecting the following constraints: (i) the z-positions of all particles are fixed (fixing thereby the composition

[38] I.e., the already mentioned square-square neighbors (or equivalently *atomic* defects.) in layer one.

[39] Following Ref. [62] we usually employ the limited memory [170] Boyden-Fletcher-Goldfarb-Shanno algorithm [171–174] in its bounded variant [175] (L-BFGS-B) for the gradient-descent based energy minimization of structures of the asymmetric Wigner bilayer system.

x), (ii) the plate separation η is fixed and (iii) the unit cell area is kept constant. We refer to a structure with a vanishing gradient of $E^*(\mathbf{r}^N; A, \eta)/N$ with respect to the (in-plane) particle positions and lattice vectors – under consideration of the constraints (i) to (iii) – as *relaxed configuration*.

A DI structure is called *stable* if the relaxed configuration is not (significantly) different from the original (DI) structure and does not transform into a topologically different configuration (*i.e.*, if neither the square-triangle tiling in layer one nor the DI decoration in layer two are destroyed). We here quantify the similarity between original structures and relaxed structures by comparing all N particle positions of a configuration before, $\tilde{\mathbf{r}}^N = (\tilde{\mathbf{r}}_1, \ldots, \tilde{\mathbf{r}}_N)$, and after relaxation, $\mathbf{r}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. For that purpose, we define the average positional displacement, $\langle \Delta_{xy} \rangle$, between an initial and the corresponding relaxed structure which is given by

$$\langle \Delta_{xy} \rangle = \frac{1}{N} \sum_{i=1}^{N} |\tilde{\mathbf{r}}_i - \mathbf{r}_i|. \tag{3.15}$$

To further investigate the mechanical stability of the class of DI structures, we (potentially) subject all configurations listed in Fig. 3.29 to noise prior to relaxation by adding uniform, random displacements to the in-plane particle positions in the xy-directions, *i.e.*, $\tilde{\mathbf{r}}_i \to \tilde{\mathbf{r}}_i + \xi$, with $\xi = (\delta_x, \delta_y, 0)$ in the interval $-\delta_{xy}/2 \leq \delta_x, \delta_y \leq \delta_{xy}/2$; we refer to δ_{xy} as noise-level. A structure is called mechanically stable *up to a noise level of* δ_{xy} if the average deviation[40] is insignificant, *i.e.*, $\langle \Delta_{xy} \rangle \approx 0$.

Note that before evaluating $\langle \Delta_{xy} \rangle$ in practice we also correct for a possible global shift, $\langle \tilde{\mathbf{r}}^N \rangle \neq \langle \mathbf{r}^N \rangle$ of the mean particle positions before, $\langle \tilde{\mathbf{r}}^N \rangle = \sum_{i=1}^{N} \tilde{\mathbf{r}}_i/N$, and after relaxation, $\langle \mathbf{r}^N \rangle = \sum_{i=1}^{N} \mathbf{r}_i/N$, via $\tilde{\mathbf{r}}^N \to \tilde{\mathbf{r}}^N - \langle \tilde{\mathbf{r}}^N \rangle$ and, analogously, $\mathbf{r}^N \to \mathbf{r}^N - \langle \mathbf{r}^N \rangle$. In that way, we ensure that $\langle \tilde{\mathbf{r}}^N \rangle = \langle \mathbf{r}^N \rangle = (0, 0, xd)$ where $x = N_2/N$ is the composition and d the plate separation distance.

First, we apply the relaxation procedure to all tilings shown in Fig. 3.29 for a range of the relevant[41] system parameter $\eta \in [0, 1]$. We identify all of the self-similar superstructure ground states of the different $\mathrm{DI}^{(T)}_{I,S}$ families – except for $\mathrm{DI}^{(1)}_{1,1}$ depicted in the central left panel of Fig. 3.29 – to be stable in the vicinity of the S_1 phase (cf. noise level $\delta_{xy} = 0$ in Fig. 3.31). We refer to the region in η where a specific DI configuration is stable as respective *stable region* in η (*i.e.*, respective to that configuration). Notably, also in the stable region in η the symmetry of the perfect square tiles in layer one is slightly broken due to relaxation and squares in layer two are transformed into rectangles with an aspect ratio close to unity (see Fig. 3.35 below) which is the reason for the small yet finite values of $\langle \Delta_{xy} \rangle$ throughout the entire η range of all structures presented in Fig. 3.31. The structural properties of the relaxed configurations outside of the respective stable η-regions strongly depend on the value of the composition, x, of the respective configurations and on the precise value of η. Both structures shown

[40] In Appendix A.1.3 we present in an analogues manner the effects of relaxation and noise on the maximum deviation, $\max(\Delta_{xy}) = \max(|\tilde{\mathbf{r}}_i - \mathbf{r}_i|)$.

[41] Note that the value of A does not affect structural properties of a given bilayer structure [62] and consequently does not need to be considered during relaxation. It is, however, important when comparing energies of different structures at a given (η, A)-pair, cf. Eq. (3.9).

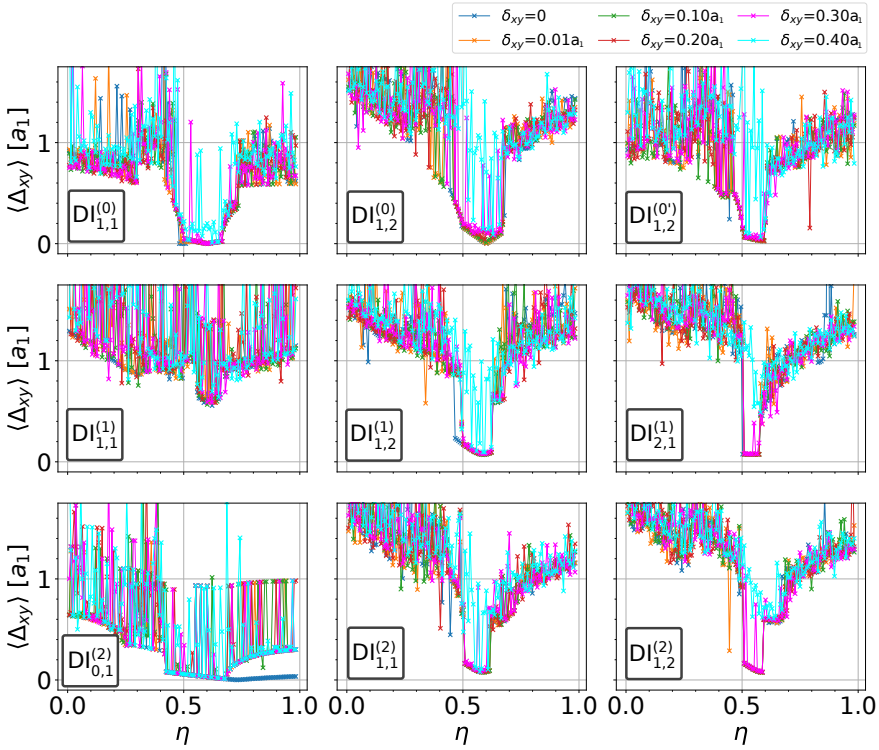


Figure 3.31: The average positional displacement, $\langle \Delta_{xy} \rangle$ defined by Eq. (3.15), in xy-directions between the relaxed and the original ground state configurations of the respective $\mathrm{DI}^{(T)}_{I,S}$ families, depicted in Fig. 3.29, as a function of $\eta \in [0,1]$; the panels here are arranged in the same way as in Fig. 3.29. Prior to relaxation the in-plane positions of all particles of the different structures are subject to uniform, random noise, $\tilde{\mathbf{r}}_i \to \tilde{\mathbf{r}}_\mathbf{i} + \xi$ with a noise-level of $\delta_{xy}/a_1 \in [0, 0.4]$; a_1 is the tiling length in layer one and the results are color coded with respect to the noise-level (see inset). Note that all structures (except for the $\mathrm{DI}^{(1)}_{1,1}$ ground state configuration which features atomic defects in layer one, cf. central left panel) have a broad range of stability in the S_1 related η-region (*i.e.*, near $\eta \approx 0.5$) even for noise levels of $\delta_{xy} = 0.3a_1$.

in Fig. 3.30, *i.e.*, $\mathrm{DI}^{(2)}_{2,1}$ with $N = 1142$ and $\mathrm{DI}^{(1)}_{2,2}$ with $N = 1224$, are unstable in the entire η range, hence we do not show the results of the stability analysis of these two structures here.

Next, we apply different noise-levels between $\delta_{xy} = 0$ and $\delta_{xy} = 0.4\,a_1$ to all configurations presented in Fig. 3.29 prior to relaxation and we collect the respective average positional displacement, $\langle \Delta_{xy} \rangle$, as a function of $\eta \in [0,1]$ in Fig. 3.31; a_1 is the tiling length of the square-triangle tiling in layer one. We see that in the vicinity of the S_1-region (most of) the highly uniform super-structures are mechanically stable up to a noise-level of $\delta_{xy} = 0.3\,a_1$, that is 30% of the tiling length in layer one, over a broad range of η. This remarkable mechanical stability of some DI configurations can be related to the constant composition we assumed during our calculations: if we consider a self-similar super-structure in layer one usually also layer two forms an interlocked super-structure featuring equilateral triangles, squares and hexagons (except for tilings such as $\mathrm{DI}^{(1)}_{1,1}$ which feature atomic defects). As long as the motives of the self-similar DI configurations are present topologically (after being subject to noise) relaxation

guides the structures back to the original self-similar structure (in most cases).

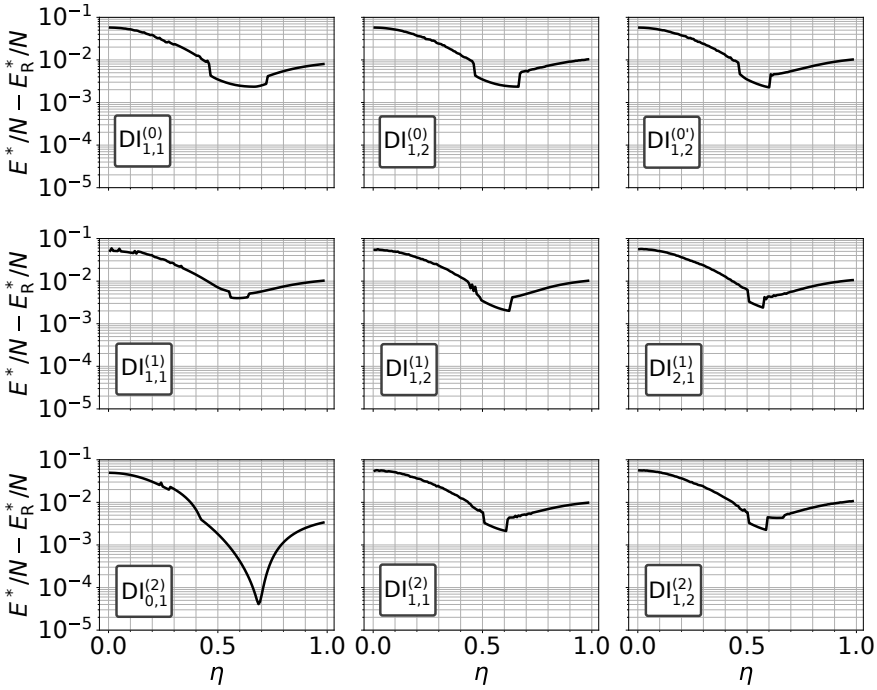


Figure 3.32: The energy difference, $\Delta E_R^*/N = E^*/N - E_R^*/N$, of the energies per particle, E^*/N, of perfect DI tilings, *i.e.*, of the ground states of the respective $\mathrm{DI}^{(T)}_{I,S}$ families, and the corresponding energies after relaxation, E_R^*/N, as a function of η arranged in the same panel ordering as in Figs. 3.29 and 3.31. Note, that the energy, E^*/N, of a perfect DI structure as a function of η is continuous. Consequently, the collapse of the square-triangle tiling in layer one (and/or the collapse of the perfect DI decoration scheme) at the boundaries of the respective stable η-regions is a first order transitions as indicated by the jump of the relaxed energies; this holds for all investigated structures except for $\mathrm{DI}^{(2)}_{0,1} \equiv \mathrm{S}_1$.

To emphasizes the energy scales related to relaxation of different, perfect DI tilings we compare in Fig. 3.32 the energies prior to (*i.e.*, E^*/N) and after relaxation (*i.e.*, E_R^*/N) of the self-similar ground states of the respective $\mathrm{DI}^{(T)}_{I,S}$ families shown in Fig. 3.29 by evaluating $\Delta E_R^*/N = E^*/N - E_R^*/N$ as a function of $\eta \in [0, 1]$. There appears to be a distinct value of $\eta \approx 0.68$ where the perfect S_1 structure (*i.e.*, the smallest possible *type 2* approximant whose results are depicted in the bottom left panel of Fig. 3.32) is particularly stable as indicated by the small change in energy of $\Delta E_R^*/N \approx 4 \times 10^{-4}$. All other approximants analysed in Fig. 3.32 undergo minor structural changes in the respective stable η-regions (also see Fig. 3.31) and, moreover, relaxation is accompanied by rather large changes in energy of the order $\Delta E_R^*/N \gtrsim 2.4 \cdot 10^{-3}$ for the self-similar super-structures. However, and most importantly, the square-triangle tiling in layer one is not destroyed in the stable η-region of defect free super-structures, *i.e.*, for neither of the configurations presented in Fig. 3.29 except for the $\mathrm{DI}^{(1)}_{1,1}$ ground state (cf. central left panel of Figs. 3.29, 3.31 and 3.32): being intrinsically unstable also causes a larger energy difference of $\Delta E_R^*/N \gtrsim 4 \cdot 10^{-3}$. Notably, this also sets the energy scale related to relaxation of the other unstable structures shown in Fig. 3.30, as we will see below.

By carefully analysing unstable configurations – with respect to relaxation – we can identify the main source of instability for all DI structures of the $\mathrm{DI}^{(T)}_{I,S}$ approximant families: the unstable ground state of the $\mathrm{DI}^{(1)}_{1,1}$ family (cf. central left panel of Figs. 3.29, 3.31 and 3.32) and the energetically most favorable, yet unstable configurations of the $\mathrm{DI}^{(2)}_{2,1}$ and $\mathrm{DI}^{(1)}_{2,2}$ families we could identify (cf. Fig. 3.30) all exhibit square-square pairs in layer one that share a common edge. The related DI decorated particles in layer two (placed at the projected geometric centers of layer one squares) are separated by a distance of a_1, *i.e.*, by the tiling length of layer one. This distance is smaller than the usual layer two interparticle separation of $a_2 = a_1 \cdot (1 + \sqrt{3})/2$ of all stable DI configurations which feature either a square tiling (S_1) or hexagon-square-triangle tilings in layer two. Consequently, the corresponding particles in layer two that are separated by a distance of a_1 experience larger repulsive forces (owed to the repulsive Coulomb interaction between the particles) as compared to the other particles which are typically separated by a distance of a_2. In that way atomic defects cause instabilities in the related defective DI structures since these larger forces can usually not be compensated by the other particles in the structure. We empirically find that neighboring square-square pairs in layer one give rise to an energy penalty in (random[42]) DI tilings which scales approximately linear with the number of atomic defects in the unit cell of the related structures, as can be seen in Fig. 3.33. This energy penalty also explains the discrete energy-jumps of the low energy configurations of the simulated annealing procedure depicted in Fig. 3.27: during the annealing procedure the super snub-square structure is encountered several times but is occasionally destroyed by a zipper move leading to a small number (usually one or two) of square-square pairs in the unit cell in layer one thereby causing the discrete jumps in energy observed in Fig. 3.27.

The *type 0* approximants (first line in Figs. 3.29 and 3.31) with their closest packed, overlapping dodecagons in layer two appear to be mechanically stable in a wide η range. In the right most structure shown in Fig. 3.29 the central dodecagon in layer one is rotated by 30° as compared to the structure depicted in the top central panel of Fig. 3.29, breaking thereby the rhombitrihexagonal tiling [255] in layer two. In fact, the structure in layer two changes such that square-square neighbors in layer two emerge (while layer one remains defect free). Comparing the top central and top right panels of Figs. 3.31 and 3.32 suggests that such layer two square-square neighbors may be interpreted as *second order* defects[43] (inspired by the naming *atomic defect* for square-neighbors in layer one) since the respective stable η range is reduced compared to the original, rhombitrihexagonal layer two tiling based configuration: second order defects appear to be an additional source of instability when relaxing the corresponding structure – although not as drastic as atomic defects. Also defects in the super-structure such as the distorted pentagons indicated by magenta areas in the central right panel of Fig. 3.29 appear to affect the range of η-stability compared to the other, fully self-similar super tilings but to this end further work is required.

As a consequence of the instability caused by atomic defects, relaxation of virtually all

[42] I.e., in random square-triangle tilings in layer one which are DI decorated.

[43] This is valid only if also triangles are present in the tiling of layer two since the layer two tiling of the current ground state candidate, *i.e.*, S_1, is only composed of squares.

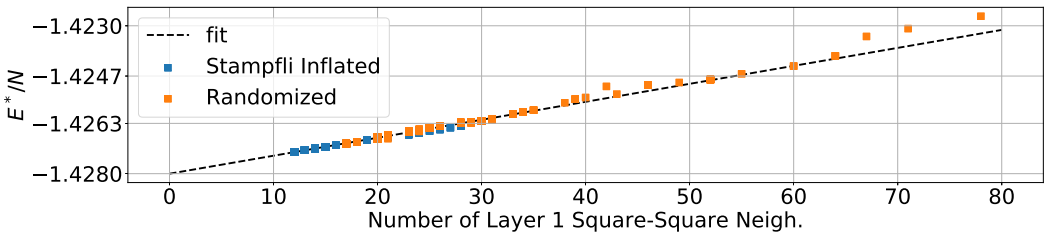


Figure 3.33: Energy per particle E^*/N as a function of square-square neighbors (or, equivalently, atomic defects) in layer one (x-axis) for all configurations encountered in a parallel tempering run of a $N = 1142$ particle approximant ($\mathrm{DI}^{(2)}_{2,1}$, cf. left panel in Fig. 3.30); the data are collected from simulations at $\eta = 0.5869$ and $A = 0.67$ of an initially (i.e., not randomized) Stampfli-inflated $\mathrm{DI}^{(2)}_{2,1}$ structure (blue) and from an initially randomized one (orange). The initially not randomized configuration performs better in energy since it avoids, by construction, from the beginning square-square neighbors in layer one. We see a largely linear scaling of E^*/N with the number of atomic defects present in the structure. This suggests that the final ground state of this approximant, although we could not identify it via simulations, is preferably highly ordered and defect free – if this is geometrically possible. A least means square fit of $E^*/N \approx kx + d$ (black dashed line) indicates a constant energy penalty of $k = \Delta E^*_{\mathrm{ssn}}/N \approx 6{\times}10^{-5}$ for square-square neighbors in layer one and the hypothetical ground state energy of the perfect square-triangle tiling can be evaluated to $d = E^*_0/N \approx -1.428$; the global ground state energy (related to S_1) is given by $E^*_{\mathbf{GS}}/N - 1.4357$ at this (η, A)-state point. However, the predictive power of this fit is rather low since this analysis does not account for relaxation which poses a significant η-dependent contribution to the energy of the order of $\Delta E^*/N \approx 10^{-4} - 10^{-3}$ and is (at least) one order of magnitude larger as the related square-square penalty.

random DI structures (*i.e.*, DI structures based on random square-triangle tilings in layer one) will destroy the perfect (yet random) square-triangle tilings in layer one and, consequently, the symmetry inherent to the DI structures. Only defect free self-similar super-structures are mechanically stable in the vicinity of the S_1-region. Focusing now on the S_1-region in the phase diagram and on related regions which also feature ingredients to form dodecagonal motives (cf. Fig. 3.23) the system rather tends to distort – via relaxation – the perfect square-triangle tiling in layer one of random DI structures to form a layer two tiling which is as uniform as possible (such as a square tiling or a triangular tiling). Hence, what at first sight seems to be a phase separation of random DI structures at high temperatures and highly ordered low-temperature super-structures in Fig. 3.28, depicting the results of a REMC structure optimization procedure, is only an apparent one: with almost certainty random DI structures will not appear at finite values of temperatures in the asymmetric Wigner bilayer system due to their large internal energies and forces associated with atomic defects.

We can therefore draw an important conclusion: we do not expect a random DI tiling would contribute toward entropic stabilization of a highest-symmetry phase quasicrystal [167, 168] at finite temperatures in the quasi-two dimensional Wigner bilayer system governed by pure repulsive, long-ranged Coulomb interaction. In the vicinity of the S_1-region of the phase diagram potential ground state structures appear to exhibit highly uniform tilings in both layers with minimal sets of commensurate tiles (*i.e.*, only featuring squares or squares and triangles). A possible quasicrystalline ground

state of the dodecagonal *type I* family would therefore also need to be highly ordered and highly uniform. The energetically best DI super-structures we identified above are, on the one hand, highly ordered due to their self-similarity features but, on the other hand, the uniformity criterion of the ground states of the asymmetric Wigner bilayer system is violated by the additional hexagonal tile necessary in layer two.

3.1.5.7 Metastable Region of Super-Tilings

Within this project we identified an entire family of self-similar super-structures of the so-called dodecagonal *type I* family and we can provide an explicit series of finite sized quasicrystalline *approximants* in Fig. 3.29 which are mechanically stable in an extended range of η (cf. Fig. 3.31). If the energies per particles of larger approximants is lower compared to smaller ones, quasicrystals – with their self-similarity property – could potentially have an extra energetic advantage relative to approximants. Here we will address the question if the asymmetric Wigner bilayer system can potentially give rise to a zero temperature DI ground state structures of quasicrystalline order within the ensemble of the dodecagonal *type I* family. For that purpose we further need to compare the energies of the related self-similar DI approximants to the ground state energies of the system [62–64] in the relevant (η, A)-proximity of S_1.

Henceforward, we use the shorthand notation $\mathrm{DI}^{(T)}_{I,S}$, which basically addresses the entire configuration space of DI structures for a given triple of the approximant type T, the inflation number I and the stacking number S, only *to address the self-similar ground state structures* of the respective periodic boundary conditions of the $\mathrm{DI}^{(T)}_{I,S}$ configuration space, *i.e.*, to directly address the ground state structures depicted in Fig. 3.29.

We find that the dodecagonal *type I* family does not seem to provide new candidates for global ground states in the asymmetric Wigner bilayer system as can be seen in Fig. 3.34: here we respectively present the offset, $\Delta E^*_{\mathrm{GS}}/N = (E^*_{\mathrm{DI}^{(T)}_{I,S}}/N - E^*_{\mathrm{GS}}/N)$, in the (η, A)-plane of the energies, $E^*_{\mathrm{DI}^{(T)}_{I,S}}/N$, of all relaxed $\mathrm{DI}^{(T)}_{I,S}$ structures shown in Fig. 3.29 to the ground state energies, E^*_{GS}/N, currently suggested by literature [62–64]. Especially in the extended region around S_1 all $\mathrm{DI}^{(T)}_{I,S}$ configurations seem to be energetically less favorable compared to the current ground state energies by a significant amount of $\Delta E^*_{\mathrm{GS}}/N \approx 10^{-3} - 10^{-2}$. The exception is $\mathrm{DI}^{(2)}_{0,1}$ which actually represents the S_1 ($\equiv \mathrm{DI}^{(2)}_{0,1}$) ground state (indicated by white regions[44] emphasizing $\Delta E^*_{\mathrm{GS}}/N = 0$ in the bottom left panel of Fig. 3.34).

Note that outside of the respective stable η-regions, corresponding to each $\mathrm{DI}^{(T)}_{I,S}$ structure[45] (see Figs. 3.31 and 3.32) the system gains energy via relaxation by deforming

[44] Notably, S_1 and H (see Fig. 3.2) exhibit the same composition, $x = N_2/N = 1/3$. Relaxing the S_1 structure in the H region may end up in an H configuration. This behavior is reflected by the offset to the ground state energies of $\Delta E^*_{\mathrm{GS}}/N \approx 0$ of the relaxed $\mathrm{DI}^{(2)}_{0,1}$ configurations in the H region depicted in the bottom left panel of Fig. 3.34.

[45] The η-dependent first order transition of the $\mathrm{DI}^{(T)}_{I,S}$ structures in the vicinity of S_1 is accompanied by jumps in the energies after relaxation, cf. Figs. 3.31 and 3.32, which can be observed in Fig. 3.34.

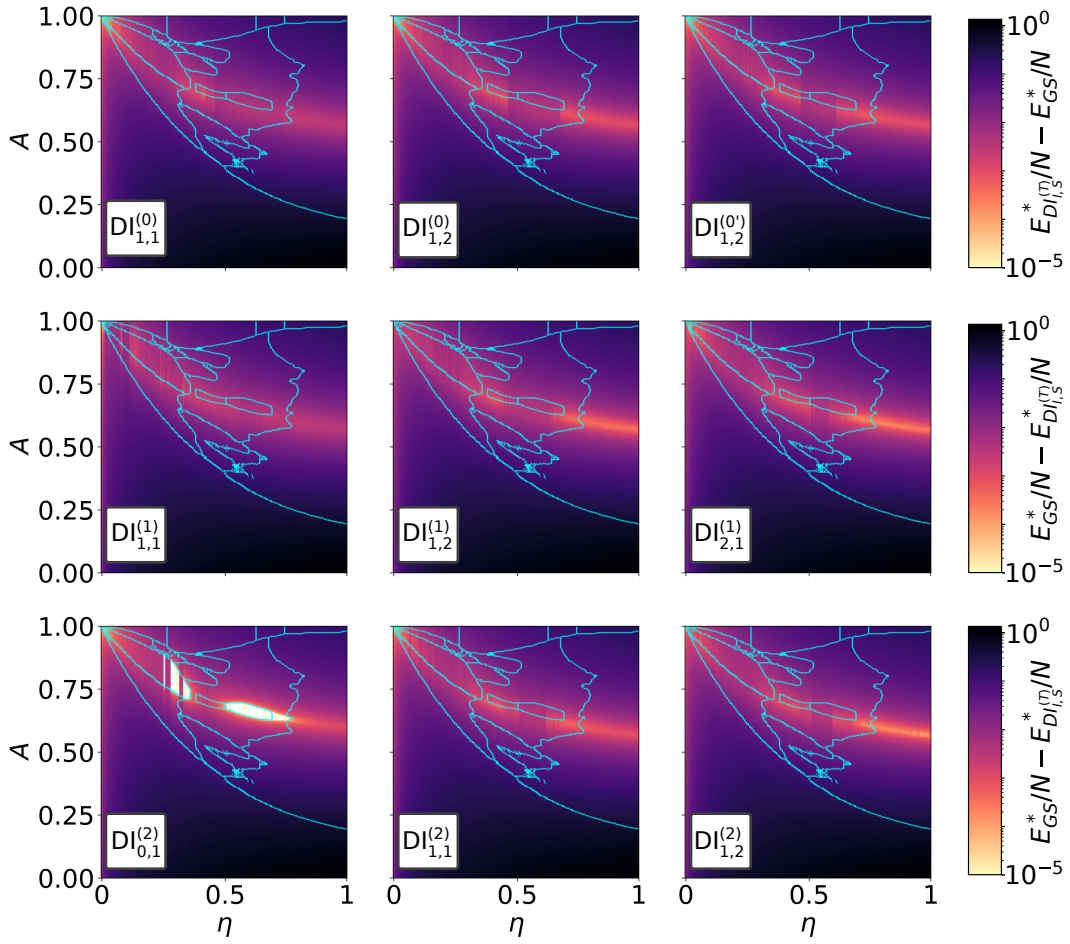


Figure 3.34: Smallest energy offset of all $\mathrm{DI}_{I,S}^{(T)}$ structures presented in Fig. 3.29 (relaxed at the respective values of η and A) to the (suggested) ground state candidates of the asymmetric Wigner bilayer system in the (η, A)-plane. Especially in the S_1 region there seems to be an overall offset of the energies of the relaxed $\mathrm{DI}_{I,S}^{(T)}$ structures to the (suggested) ground state energies from literature of $\Delta E_{\mathrm{GS}}^*/N \approx 10^{-4} - 10^{-3}$ (cf. vertical, shadow-like regions which correspond to the respective stable η-regions of the different structures); no energetically more favorable solutions could be found within the respective $\mathrm{DI}_{I,S}^{(T)}$ families of structures. The cyan lines indicate the phase boundaries from Ref. [62].

(*i.e.*, destroying) the perfect square triangle tilings in layer one. For large values of $\eta \gtrsim 0.7$, especially large relaxed structures such as $\mathrm{DI}_{1,2}^{(2)}$ with $N = 328$ and $N_1 = 224$ (or $\mathrm{DI}_{1,3}^{(1)}$ with $N = 352$ and $N_1 = 240$, not shown here) exhibit energies comparable to the suggested ground state energies from literature [62–64] in a narrow region around $A \approx 0.6$. However, due to relaxation the perfect square-triangle tilings transform into structures of the V_x family (see Fig. 3.2) and now feature hexagonal bilayers at the relevant $\eta \gtrsim 0.7$ and $A \approx 0.6$ combinations. These structures potentially outperform solutions of the evolutionary ground state search [62], the later being restricted to structures with a maximum number of $N \leq 40$ particles per unit cell due to numeri-

cal reasons. In fact, the relaxed $\mathrm{DI}^{(1)}_{1,3}$ structure represents a new global ground state candidate at $\eta \gtrsim 0.7$ and $A \approx 0.6$.

We empirically find that the hexagonal tiles in layer two[46] of $\mathrm{DI}^{(T)}_{I,S}$ structures are responsible for the large energy penalties in comparison to S_1 in the surrounding (η, A)-regions of the S_1 phase (cf. Fig. 3.23). We demonstrate this by directly comparing the energies of the perfect $\mathrm{DI}^{(0)}_{1,1}$ structure with the energies of the perfect S_1 structure in the (η, A)-plane. Notably, $\mathrm{DI}^{(0)}_{1,1}$ features exactly one layer two hexagon in the unit cell. Hence, the energy difference $\Delta E^*_{S_1}/N = (E^*_{\mathrm{DI}^{(0)}_{1,1}}/N - E^*_{S_1}/N)$ of $\mathrm{DI}^{(0)}_{1,1}$ and S_1 as a function of η and A quantifies the energetic cost (or benefit) of maintaining exactly one layer two hexagon compared to the more uniform square tiling in layer two of S_1. The results are present in the left panel of Fig. 3.35.

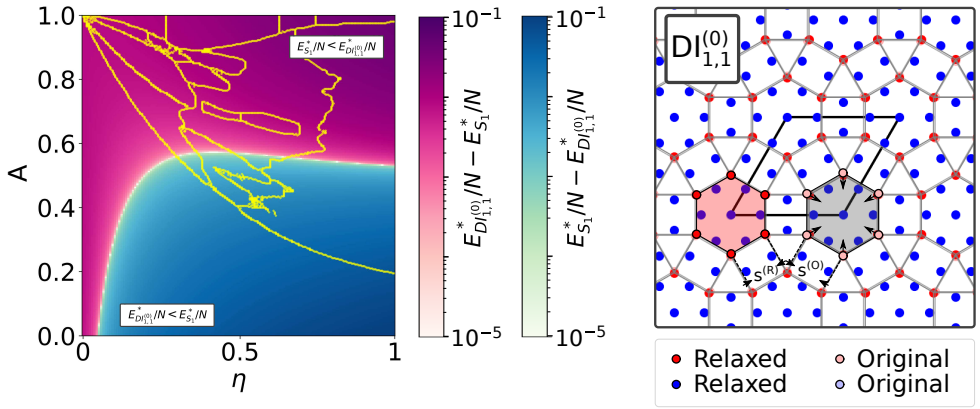


Figure 3.35: Left: Energy differences in a logarithmic scale of the perfect (original, non-relaxed) $\mathrm{DI}^{(0)}_{1,1}$ and the perfect S_1 structure; no other structures are considered. Purple areas indicate regions where $\mathrm{DI}^{(0)}_{1,1}$ is less favorable in terms of E^*/N compared to S_1; in the blue region the situation is reversed and forming layer two hexagons becomes energetically beneficial. The yellow lines indicate the phase boundaries from literature [62–64]. Right: $\mathrm{DI}^{(0)}_{1,1}$ structure relaxed at $\eta = 0.66468$ (solid red and blue circles) on top of perfect, original $\mathrm{DI}^{(0)}_{1,1}$ structure (faint red circles – at the corners of the gray hexagon – and faint blue circles – not visible). While the particle positions in layer one remain virtually the same before and after relaxation (solid blue circles which indicate relaxed vertices completely cover faint blue circles which represent the original vertices), the hexagon in layer two shrinks from the grey emphasized area (with a regular side length of $s^{(O)}$) to the red area (with a regular side length of $s^{(R)}$) by a factor of $(s^{(R)}/s^{(O)})^2 = 0.899$.

We can see that the value of $\Delta E^*_{S_1}/N$ can be positive or negative depending on η and A. However, in the (extended) proximity of the S_1 ground state region the energetic cost of maintaining layer two hexagons is rather large. In the right panel of Fig. 3.35 we compare the perfect $\mathrm{DI}^{(0)}_{1,1}$ structure with another $\mathrm{DI}^{(0)}_{1,1}$ ground state configuration that has been relaxed at $\eta = 0.66468$ (*i.e.*, exactly where $\mathrm{DI}^{(0)}_{1,1}$ is most stable, cf. Figs. 3.31 and 3.32). The vertices of the relaxed $\mathrm{DI}^{(0)}_{1,1}$ configuration exhibits the same topology as the perfect $\mathrm{DI}^{(0)}_{1,1}$ structure and especially the square-triangle tiling in layer one remains

[46] The hexagons in layer two are locked to the squares of the dodecagon in layer one.

virtually unchanged. We can see in Fig. 3.35 that the system reduces the internal energy via relaxation of the $\mathrm{DI}_{1,1}^{(0)}$ configuration (quantified by Fig. 3.32) by slightly compressing the area of the hexagonal tile in layer two which thereby destroys the square-triangle symmetry in layer two: the square tiles in layer two are deformed into rectangular tiles (although with an aspect ratio that is close to unity). We empirically find that the side length of the regular hexagon after relaxation, $s^{(R)}$, is shorter than the tiling length in layer two of the original structure, $s^{(O)} = a_2$, by a relative amount of $s^{(R)}/s^{(O)} = 0.948$ for the situation depicted in Fig. 3.35. The corresponding relative change of the area of the hexagon du to relaxation is $(s^{(R)}/s^{(O)})^2 = 0.899$.

In the (η, A)-proximity of the S_1 region the asymmetric Wigner bilayer system energetically favors structures featuring a layer two tiling which is as uniform as possible, *i.e.*, tilings which rely on a tile-set with as few tiles as possible (such as solely square–, triangular– or rhombic tiles, cf. Fig. 3.23). In contrast, the $\mathrm{DI}_{I,S}^{(T)}$ ground states all feature three types of tiles in layer two, namely an equilateral triangle, a square and a regular hexagon. Especially the energy penalty related to maintaining hexagonal tiles near the S_1 region cannot be compensated through relaxation of the different $\mathrm{DI}_{I,S}^{(T)}$ structures, as we have already shown in Fig. 3.34. As a consequence, the $\mathrm{DI}_{I,S}^{(T)}$ family of dodecagonal structures remains metastable versus competing structures known from literature [62–64] such as S_1 (see also other structures in Fig. 3.23) by a finite amount of energy.

We will now only focus on the family of $\mathrm{DI}_{I,S}^{(T)}$ structures, *i.e.*, on the self-similar ground states of the respective periodic boundary conditions of the different $\mathrm{DI}_{I,S}^{(T)}$ families, and ignore the fact that other structures perform, admittedly, far better in terms of the internal energy per particle throughout the (η, A)-plane. To compare the energies of the different $\mathrm{DI}_{I,S}^{(T)}$ structures we introduce the standard deviation, σ_{DI}, of all related structures by

$$\sigma_{\mathrm{DI}} = \sqrt{\left\langle \left(E^*_{\mathrm{DI}_{I,S}^{(T)}}/N \right)^2 \right\rangle_{[T,I,S]} - \left\langle E^*_{\mathrm{DI}_{I,S}^{(T)}}/N \right\rangle^2_{[T,I,S]}}, \tag{3.16}$$

where we take the average $\langle\cdot\rangle_{[T,I,S]}$ with respect to all available $\mathrm{DI}_{I,S}^{(T)}$ structures, omitting unstable ones such as $\mathrm{DI}_{1,1}^{(1)}$. We evaluate σ_{DI} for the original, perfect $\mathrm{DI}_{I,S}^{(T)}$ structures as well as for the relaxed $\mathrm{DI}_{I,S}^{(T)}$ structures in the (η, A)-plane and we present the results in the left and right panel of Fig. 3.36, respectively. We find that for both situations σ_{DI} extends over several orders of magnitudes, between $\sigma_{\mathrm{DI}} \approx 10^{-5}$ and $\approx 10^{-1}$, depending on the values of η and A.

For the original structures (left panel of Fig. 3.36) we see that the minimal value of $\sigma_{\mathrm{DI}} \approx 10^{-5}$ is obtained over an entire (curve-like) region in the (η, A)-plane emphasizing parameter combinations where all perfect $\mathrm{DI}_{I,S}^{(T)}$ structures are energetically close[47]: apparently, for certain combinations of the system parameters, η and A, where $\sigma_{\mathrm{DI}} \approx$

[47] Notably, the extended region of energetically close $\mathrm{DI}_{I,S}^{(T)}$ structures shown in the left panel of Fig. 3.36 correlates with the results shown in the left panel of Fig. 3.35

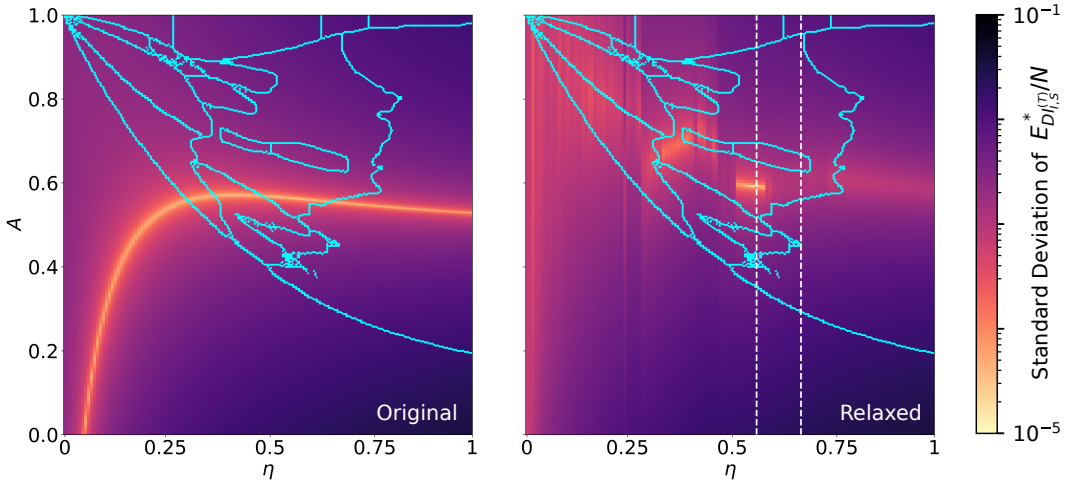


Figure 3.36: Standard deviation σ_{DI}, defined by Eq. (3.16), of all available $\mathrm{DI}_{I,S}^{(T)}$ structures of the original, perfect $\mathrm{DI}_{I,S}^{(T)}$ structures (left) and of the relaxed $\mathrm{DI}_{I,S}^{(T)}$ structures (right) in the (η, A)-plane (except defective structures such as $\mathrm{DI}_{1,1}^{(1)}$, cf. central left panel in Fig. 3.29); the cyan lines indicate the phase boundaries from literature [62–64]. For the original structures we observe a curve of $\sigma_{\mathrm{DI}} \approx 10^{-5}$ (orange/yellow curve, cf. colorbar) which indicates an extended region of combinations of η and A where perfect $\mathrm{DI}_{I,S}^{(T)}$ structures are energetically very similar. For the relaxed case (right) we find a sharp minimum of $\sigma_{\mathrm{DI}} \approx 10^{-5}$ (orange/yellow horizontal line) in the vicinity of $\eta \approx 558614$ and $A \approx 0.585$ indicating a region of energetically similar, mechanically stable $\mathrm{DI}_{I,S}^{(T)}$ tilings. The vertical dashed cuts in the left panel correspond to special values of η for which we present a detailed analysis of the involved structures in Fig. 3.37 and Fig. 3.38, respectively.

10^{-5} (see yellow curve in the left panel of Fig. 3.36) only very little energetic cost is necessary to transform one perfect $\mathrm{DI}_{I,S}^{(T)}$ structure into another one.

The effects of relaxation destroy this feature of the perfect $\mathrm{DI}_{I,S}^{(T)}$ structures (as we interpret $\sigma_{\mathrm{DI}} \approx 10^{-5}$) for most of the (η, A)-plane as we show in the right panel of Fig. 3.36. We observe, however, a narrow yet extended region in the vicinity of $\eta \approx 558614$ and $A \approx 0.585$ which draws our attention (see short, horizontal, yellowish or orange line in the right panel of Fig. 3.36): here, the relaxed $\mathrm{DI}_{I,S}^{(T)}$ structures are (a) mechanically remarkably stable up to a noise level of 30% of the layer one tiling length a_1 (cf. Fig. 3.31) and (b) exhibit a sharp minimum of σ_{DI} and are thus energetically very similar.

To present a clearer picture of the involved energy scales related to the energy offset of the $\mathrm{DI}_{I,S}^{(T)}$ structures to the previously suggested [62–64] ground state candidates of the asymmetric Wigner bilayer system we present in Figs. 3.37 and 3.38 E^*_{GS}/N as a function of A for certain, selected values of η (indicated by vertical cuts in the right panel of Fig. 3.34).

In Fig. 3.37 we present the energy difference between the energies of all $\mathrm{DI}_{I,S}^{(T)}$ structures relaxed at $\eta = 0.66468$ (which is where $\mathrm{DI}_{1,1}^{(0)}$ is most stable, cf. Fig. 3.35) and the

(suggested) ground state energies from literature [62–64]. We find that, indeed, none of the $\mathrm{DI}_{I,S}^{(T)}$ structures (except for S_1, of course) has reasonably small values of the internal energy per particle, E^*/N, to be considered as a serious competitor to the global ground state of the asymmetric Wigner bilayer system in the proximity of the S_1 phase. In the bottom panel of Fig. 3.37 we also present the challenger ranking: we sort all $\mathrm{DI}_{I,S}^{(T)}$ structures, embedded in the ground state candidate database from literature [62–64], in ascending order[48] by E^*/N and we assign integer values to the structures thereby labeling their rank starting with rank 0 for the lowest energy configuration, rank 1 for the energetically second most favorable structure, and so on.

Interestingly, at this particular value of $\eta \approx 0.66468$ (cf. Figs. 3.31 and 3.37) large $\mathrm{DI}_{I,S}^{(T)}$ structures are energetically more favorable than S_1 for $A \lesssim 0.6175$ and they occupy ranks between 2 to 14 for $A \approx 0.6$. However, due to relaxation the majority of $\mathrm{DI}_{I,S}^{(T)}$ configurations deform into structures similar to $\mathbf{k}_{32}^{*30}$ and $\mathbf{k}_{32}^{*7}$ (cf. Fig. 3.23). In fact, only the $\mathrm{DI}_{0,1}^{(2)} \equiv \mathrm{S}_1$ and $\mathrm{DI}_{1,1}^{(0)}$ structures are stable here.

In Fig. 3.38 we present the same analysis as in Fig. 3.37 but for a value of $\eta = 0.558614$ which is located inside of the stable region of all $\mathrm{DI}_{I,S}^{(T)}$ structures shown in Fig. 3.29. Here all $\mathrm{DI}_{I,S}^{(T)}$ structures remain virtually unchanged after relaxation[49], except for the unstable $\mathrm{DI}_{1,1}^{(1)}$ structure. Comparing the energies of the unstable and stable $\mathrm{DI}_{I,S}^{(T)}$ structures depicted in top panels of Figs. 3.37 and 3.38, respectively, emphasizes the energy the system gains through deformation of the perfect self-similar square-triangle tilings into configurations similar to $\mathbf{k}_{32}^{*30}$ and $\mathbf{k}_{32}^{*7}$ (cf. Fig. 3.23). Yet, no serious competitor for the global ground state can be identified in Fig. 3.38.

In Fig. 3.38 we further identify the point $\eta \approx 0.558614$ and $A \approx 0.585$ where the energies of all $\mathrm{DI}_{I,S}^{(T)}$ structures (except for $\mathrm{DI}_{1,1}^{(1)}$) are very similar, *i.e.*, $\sigma_{\mathrm{DI}} \approx 10^{-5}$. This nearly-degeneracy behaviour is also reflected in the bottom panel of Fig. 3.38 which shows the E^*/N sorted ranking of all $\mathrm{DI}_{I,S}^{(T)}$ challenger structures embedded in the literature database [62–64]. We see that all $\mathrm{DI}_{I,S}^{(T)}$ structures occupy ranks between 5 and 15 at $A \approx 0.585$; their ranks are completely reordered as a function of A at this transition point. A ground state transition in the $\mathrm{DI}_{I,S}^{(T)}$ family from S_1 over $\mathrm{DI}_{1,3}^{(1)}$, *i.e.*, a singly inflated, doubly stacked square approximant with $N = 352$ particles per unit cell, to $\mathrm{DI}_{1,1}^{(0)}$ occurs in the vicinity of $A \approx 0.585$ with decreasing A.

Summarizing, in this entire Subsection of investigating the dodecagonal *type I* family we demonstrate the emergence of self-similar super-structures from the $\mathrm{DI}_{I,S}^{(T)}$ family which are mechanically remarkably stable. Although we do not expect to find a new global ground state structure of this family near the S_1 region (or anywhere else in the phase diagram of the asymmetric Wigner bilayer system) we report a region in the vicinity of $\eta \approx 558614$ and $A \approx 0.585$ where $\mathrm{DI}_{I,S}^{(T)}$ structures are metastable. The interesting feature of the energetically similar, mechanically stable DI structures in the proximity of $\eta \approx 0.558614$ and $A \approx 0.585$ suggests that this region may indeed

[48] The entire database [62] was analysed with the K^*=32-clustering approach presented in Subsection 3.1.4 and only the best realization of a family of structures at a given value of η and A is considered in the rank sorting procedure.

[49] Except, of course, for the slightly deformed hexagons as put forward in Fig. 3.35.

give rise to a self-similar, atomic-defect free DI quasicrystal in the ensemble of DI-decorated square-triangle tilings (still neglecting all other possible structures). To this end further investigations are necessary to prove the existence of such a metastable quasicrystalline state in this restricted configuration space of the asymmetric Wigner bilayer system.

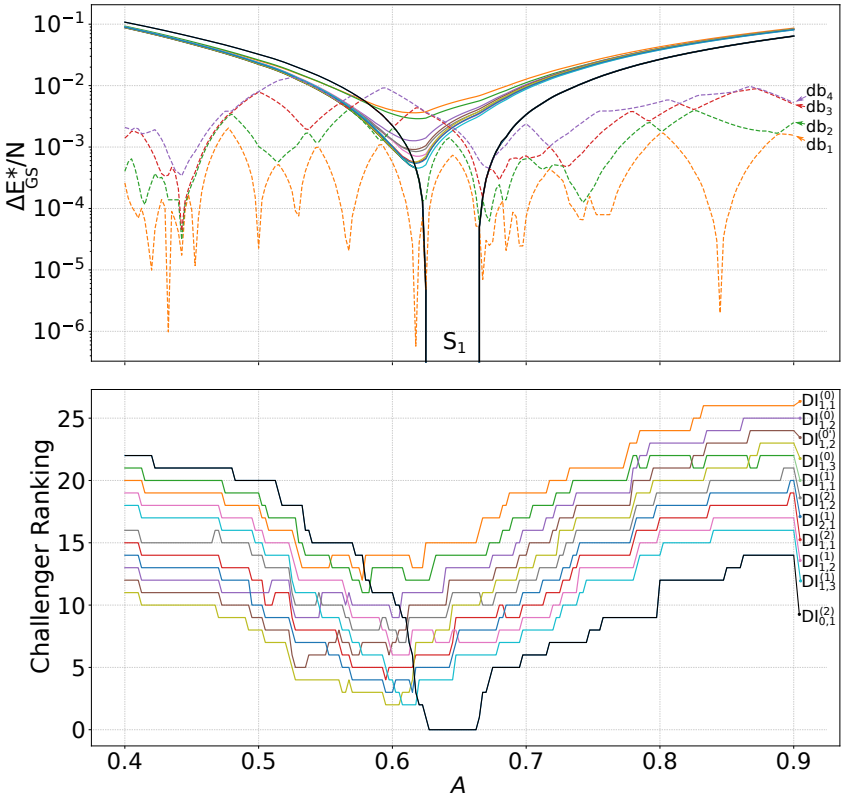


Figure 3.37: Top panel: Energy offset, $\Delta E^*_{\mathrm{GS}}/N$, of relaxed $\mathrm{DI}^{(T)}_{I,S}$ configurations (solid lines, same color-coding in top and bottom panels, labels in bottom panel) to the ground-state energies suggested by Ref. [62] evaluated at $\eta = 0.66468$ (where $\mathrm{DI}^{(0)}_{1,1}$ is most stable, cf. Fig. 3.35) as a function of A. Also the energy offset of the best four challenger families (with respect to the K^*=32-means clustering approach introduced in Subsection 3.1.4) from the literature-database [62], $db_1 - db_4$ (dashed lines) are presented at the respective (η, A)-pairs. Bottom panel: Challenger ranking (i.e., E^*/N-sorting, see text) for all indicated $\mathrm{DI}^{(T)}_{I,S}$ structures embedded in the entire literature-database of structural data from Ref. [62] for relevant (η, A)-pairs.

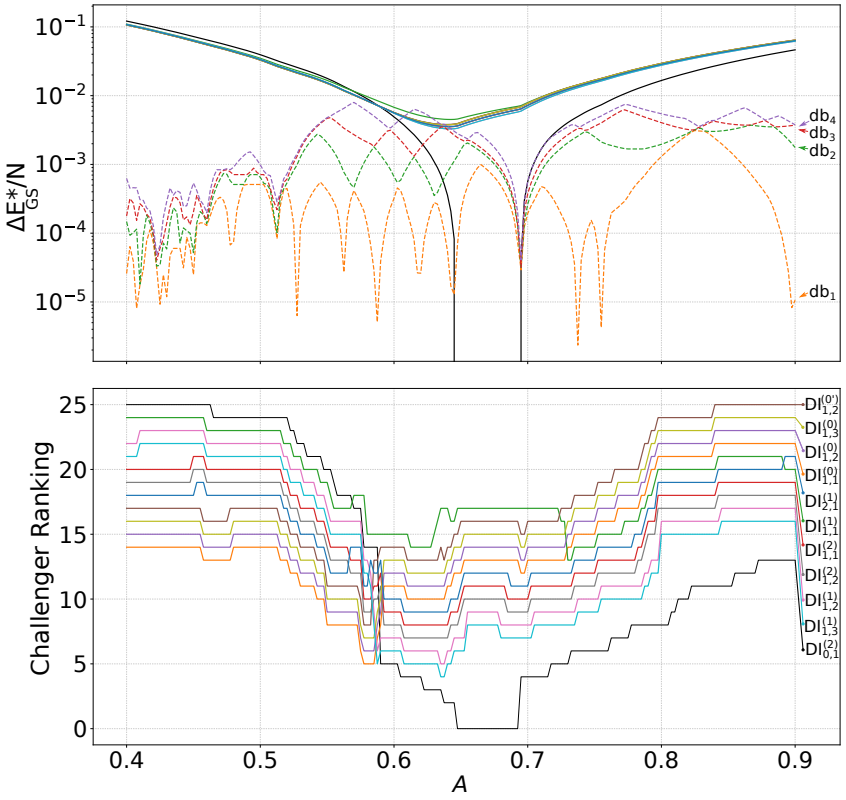


Figure 3.38: Same as Fig. 3.36 but evaluated at $\eta \approx 0.558614$ where all $\mathrm{DI}^{(T)}_{I,S}$ structures are mechanically stable (cf. Fig. 3.31). We see that for $A \approx 0.585$ a ground state transition in the $\mathrm{DI}^{(T)}_{I,S}$ family from S_1 to $\mathrm{DI}^{(1)}_{1,3}$ (double stacked square approximant with $N = 352$) occurs, but the energies of all $\mathrm{DI}^{(T)}_{I,S}$ structures become very similar in the vicinity of this (η, A)-pair as shown in Fig. 3.36 (except for the unstable $\mathrm{DI}^{(1)}_{1,1}$ configuration, which has a finite energy penalty compared to the rest).

3.1.6 Towards a Bilayer-Modulated Hexagonal Wigner Quasicrystal

In the proximity of the phase-boundary between the regions of stability of the trihexagonal- (I_x) and the honeycomb-structure (H) [255] in the parameter space of the asymmetric Wigner bilayer system we report strong signals of the twelve-fold symmetric bond orientational order parameter in layer two, $\Psi_{12}^{(2)}$ (see top right panel of Fig. 3.3). With the help of the $\mathbf{k}_{32}^{*}$-clustering approach[50] introduced in Subsection 3.1.4 we could relate this characteristic feature of the ground state candidates of the system (as suggested by Refs. [62–64]) with a family of structures, *i.e.*, the $\mathbf{k}_{32}^{*29}$ family (depicted as inset in Fig. 3.18(a) and, equivalently, in Fig. 3.39). The $\mathbf{k}_{32}^{*29}$ family is related to both the I_x and the H structures but, moreover, features rectangular tiles in layer two. By mixing the basic motives of the I_x and the H structures (*i.e.*, a large, T_l, and a small, t_s, equilateral triangle) and a corresponding rectangular tile, R, the $\mathbf{k}_{32}^{*29}$ family allows the formation of *pseudo*-dodecagonal clusters – in contrast to the square-triangle based dodecagonal clusters in Subsection 3.1.5. Interestingly, the I_x, the H and the $\mathbf{k}_{32}^{*29}$ families of structures can be considered as *modulations* of the hexagonal monolayer configuration where selected particles are vertically displaced from layer one to layer two, such that tiling like patterns are formed in layer two. In this Subsection we will investigate ground state configurations of the asymmetric Wigner bilayer system related to the $\mathbf{k}_{32}^{*29}$ family.

This part of the thesis is organized as follows: in Subsection 3.1.6.1 we introduce the rectangle-large–triangle-small–triangle (RT_lt_s) family of structures and introduce a related decoration scheme of the rectangular and the two types of triangular tiles. We then discuss general geometric features related to the RT_lt_s family and its relation to modulations of the hexagonal monolayer and to square-triangle tilings in Subsection 3.1.6.2. We continue with investigating ground state configurations of the RT_lt_s families with the help of a specifically implemented replica exchange Monte Carlo (REMC) procedure. In that way we can identify RT_lt_s super-structures which are composed of rectangular and/or triangular super-tiles. In an effort to rationalize the emerge of RT_lt_s super-structures we construct in Subsection 3.1.6.4 an effective tile-Hamiltonian [282, 283] for the RT_lt_s family; based on tile-tile interactions we can explain why the related RT_lt_s super-structures have a clear energetic advantage compared to other, random RT_lt_s configurations. In Subsection 3.1.6.5 we then construct a list of RT_lt_s super-structures from scratch by transforming self-similar DI based square-triangle tilings from above into RT_lt_s super-structure configurations. Some of these RT_lt_s super-structures outperform current ground state configurations of the asymmetric Wigner bilayer system at certain regions in the (η, A)-plane [62–64] as we

[50] The $\mathbf{k}_{32}^{*}$-clustering approach introduced in Subsection 3.1.4 relies on unsupervised k-means clustering (see Subsection 2.4.2) of the database of structural ground state candidate configurations of the asymmetric Wigner bilayer system identified in Refs. [62–64] (or better of the data set of order parameters of the structural database, which we additionally subjected to principal component analysis, see Subsection 2.4.1). Via information-theoretical measures based on the adjusted mutual information score between different clustering results (see Subsection 2.4.5), we could identify a reasonable "guess" for a total number of K^{*}=32 different structural families in the structural database; the asterisk emphasizes that we manually assigned all uniquely defined hexagonal monolayer configurations of the structural database as a family and excluded the corresponding data points from the k-means clustering procedure of the remaining data.

demonstrate in Subsection 3.1.6.6. We investigate inflation rules of the $\mathrm{RT}_l\mathrm{t}_s$ superstructure family in Subsection 3.1.6.7 and present the energy offset of two specific inflated configurations with $N = 1444$ and $N = 1452$ to the previously [62–64] suggested ground state candidates of the asymmetric Wigner bilayer system.

3.1.6.1 The Rectangle-Large–Triangle-Small–Triangle ($\mathrm{RT}_l\mathrm{t}_s$) Family with Pseudo-Dodecagonal Traits

To begin with, we present in Fig. 3.39 all structural families identified by the K^*=32-means clustering procedure put forward in Subsection 3.1.4 (cf. Fig. 3.18(a)) which occur in the vicinity of the (η, A)-regions of the II, II_x, H and the I_x phases (the relevant phases from literature [62–64] are labeled; phases II and H are not explicitly shown in Fig. 3.39, see Fig. 3.2).

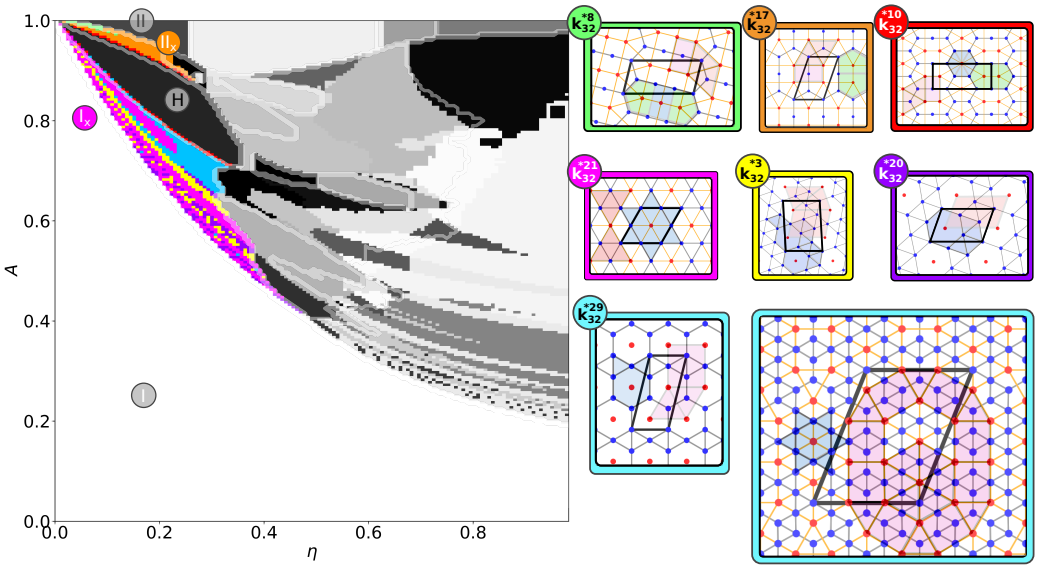


Figure 3.39: Same as Fig. 3.23 but for structures related to the rectangle-large–triangle-small–triangle family (cf. Fig. 3.18(a)). Special features and tiles in the structures shown in the right are, again, highlighted with faint colors and regions in the (η, A)-plane in the left panel where members of the $\mathbf{k}_{32}^{*c}$ structure-families from the right form the ground states are appropriately colored in bright green, orange, red, magenta, yellow, purple and cyan color. The remaining $\mathbf{k}_{32}^{*c}$ families are colored in gray-scale (see also Fig. 3.18). We do not explicitly present the structural families related to H and the II which were reliably identified by the $\mathbf{k}_{32}^{*}$-clustering procedure; they are illustrated in Fig. 3.2. See Appendix A.1.2 for details on characteristic values and boundaries of the order parameters and principal components for the $\mathbf{k}_{32}^{*}$-clustering families of structures.

Via the families $\mathbf{k}_{32}^{*8}$ (green area) and $\mathbf{k}_{32}^{*17}$ (orange area) the K^*=32-means algorithm identified different variants of the II_x structures while the $\mathbf{k}_{32}^{*10}$ family (red area) poses a mixture of the II_x and the H phase. The $\mathbf{k}_{32}^{*15}$ family (magenta area) represents the archetypical I_x structure (*i.e.*, a trihexagonal tiling in layer one) which occupies

the largest fraction of the phase-diagram in the previously identified [62] I_x region: starting from a hexagonal monolayer, *i.e.*, phase I, at $\eta = 0$ and from the symmetric case $A = 1$ the I_x phase (and related structures) extends from regions of small, yet finite to modest values of the plate separation of $0 < \eta \lesssim 0.4$ and asymmetry values of $1 > A \gtrsim 0.4$ between the H and the I regions. We can see in Fig. 3.39 that the K^*=32-means clustering approach suggests to subdivide the I_x region into geometrically more complex sub-families of structures: depending on the actual value of η and A the tiling in layer two can take quite different and rather complex forms, a situation which we illustrate in Fig. 3.39 by presenting selected structures associated to the $\mathbf{k}^*_{32}$-clustering families $\mathbf{k}^{*3}_{32}$, $\mathbf{k}^{*20}_{32}$ and $\mathbf{k}^{*29}_{32}$ (respectively emphasized by yellow, violet and cyan color coding).

In Fig. 3.39 we show two selected configurations of the $\mathbf{k}^{*29}_{32}$ family which represents the ground states at different positions in the (η, A)-plane: the smallest configuration from the $\mathbf{k}^{*29}_{32}$ family with $N = 7$ particles per unit cell forms the ground state, for instance, at $\eta = 0.2262$ and $A = 0.77$ and the largest one with $N = 39$ particles at $\eta = 0.2475$ and $A = 0.735$. Another structural realization of this family is a so-called [255, 276] I_x-*Cairo type of tiling* depicted in Fig. 3.40(a,b) which represents the ground state candidate of the asymmetric Wigner bilayer system in the vicinity of $\eta = 0.106$ and $A = 0.878$ (or also at $\eta \approx 0.226$ and $A = 0.765$ [276] suggesting an entire range of stability).

The layer two tilings of the structures from the $\mathbf{k}^{*29}_{32}$ family feature rectangular tiles and two types of equilateral triangular tiles, whose respective longer, $l = 2\,a_1$ and shorter side lengths, $s = \sqrt{3}\,a_1$, correspond to the two different edge lengths of the rectangular tile (see Fig. 3.40(c); a_1 is the tiling length in layer one). A unique decoration scheme of particles in layer one given for the rectangular and two type of triangular tiles can be given as follows (see Fig. 3.40(c) for an illustration):

- Long edges of triangles and rectangles in layer two are always decorated mid-edge in layer one.
- Small triangles are decorated in the geometric center.
- Short rectangle sides are decorated *off-edge* in the rectangles' interior at a distance of $a_1/2$ to the center of the edge orthogonal to the edge.

Henceforward we will refer to this decoration scheme as rectangle-large–triangle-small–triangle ($RT_l t_s$) decoration scheme and to structures following the $RT_l t_s$ decoration scheme (as for instance the entire $\mathbf{k}^{*29}_{32}$ family) as $RT_l t_s$ structures. Notably, for $RT_l t_s$ structures the two types of triangles always appear in pairs of possible orientations which are related by rotations (or flips) of 90°. As a matter of convenience we here define the *flip-type* $F = 0$ of a structure where small triangles have one side parallel to the horizontal axis[51] and flip-type $F = 1$ if the small triangles are rotated (or flipped) by 90°. Notably, the above introduced decoration scheme of a quasi-2D layer geometry

[51] The I_x-Cairo configuration is related to the type 2 approximant (S_1) structures from Subsection 3.1.5 whose squares in layer one are additionally rotated by ±15° compared to type 0 and type 1. Regarding the flip-type convention (F=0,1) we consider the (horizontal) reference axis for the $RT_l t_s$-decoration scheme for such approximant type 2 based structures to be rotated by 15°.

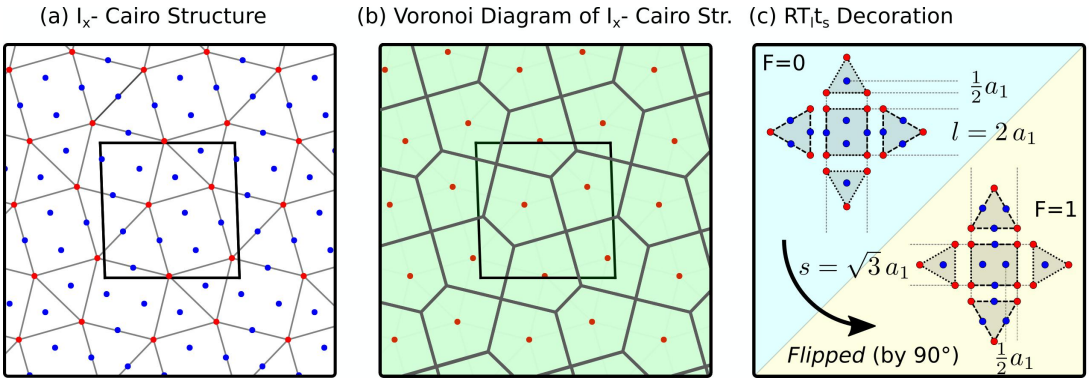


Figure 3.40: I_x-Cairo type of tiling identified as ground state candidate configuration of the asymmetric Wigner bilayer system at $\eta \approx 0.106$ and $A = 0.878$ [276]. Panel (a) shows the I_x-Cairo structure itself (with Delaunay triangulation used to connect the vertices in layer 2, *i.e.*, red particles). The I_x-Cairo configuration is related to S_1 but here a *snub-rectangular* rectangle-large–triangle-small–triangle ($\mathrm{RT}_l\mathrm{t}_s$) based tiling can be identified in layer two rather than the perfect snub-square tiling in layer one of S_1 and here the decoration in layer one is performed according to panel (c). Panel (b) shows the Voronoi diagram [223, 224] (cf. Subsection 2.3.1) of the layer 2 vertices and thereby highlights the four (almost undistorted) pentagonal tiles which form a Cairo type of tiling [255]. The irreducible unit cell in panel (a) and (b) are emphasized by the black square frames, respectively. Panel (c) emphasizes the $\mathrm{RT}_l\mathrm{t}_s$ based decoration scheme of the tiles in layer two: the I_x-Cairo configuration is composed of rectangles and two types of equilateral triangles whose short edges of length $s = \sqrt{3}\,a_1$ (dotted lines) are always off-edge decorated (see text) whose long edges of length $l = 2a_1$ (dashed lines) are mid-edge decorated; a_1 is the layer one tiling length. In that sense, I_x-Cairo perhaps is more similar to S_2 rather than S_1: S_2 can be interpreted as distorted rectangle-triangle tiling in layer two with only off-edge decorated edges as layer one vertices (see Fig. 3.2). The mid-edge and off-edge decoration of $\mathrm{RT}_l\mathrm{t}_s$ tilings (when following the flip-type convention, see text) guarantees that when projecting layer one and two onto the same plate a hexagonal monolayer with a tiling length of a_1 emerges.

has great resemblance with the decoration scheme of three-dimensional Frank-Kasper phases [253, 261].

Structures related to the $\mathbf{k}_{32}^{*29}$ family represent the ground state candidates in the above mentioned (η, A)-region with strong signals of the $\Psi_{12}^{(2)}$ bond orientational order parameter which can be related to the corresponding tiling in layer two: with the two types of triangles and the corresponding rectangular tile it is geometrically possible to construct pseudo-dodecagons. A central triangular vertex formed either by small or by large triangles is surrounded by a ring of six rectangles respectively sharing an edge with the central triangles (appropriately oriented to match the side length of the triangles). Six properly oriented triangles of the other kind (large or small) as those forming the central triangular vertex of the pseudo-dodecagon (small or large) fill the voids between the rectangles by sharing two edges with neighboring rectangles to complete the ring around the central vertex of the pseudo-dodecagon. Indeed, the $N = 39$ particle configuration from the $\mathbf{k}_{32}^{*29}$ family depicted in the bottom right panel of Fig. 3.39 exhibits such pseudo-dodecagonal traits where two pseudo-dodecagons are "glued together" at short rectangle sides (see faint magenta colored tiles in the right most $\mathbf{k}_{32}^{*29}$ family inset structure in Fig. 3.39).

3.1.6.2 Approximate Square-Triangle Tilings with Bilayer Modulations of the Hexagonal Monolayer

In general, we find that the vertices of $\mathrm{RT}_l\mathrm{t}_s$ configurations of both layers one and two form a (possibly slightly distorted) hexagonal monolayer if all particles are projected onto the same plate, a feature which also holds true for II, II_x. The vertices of layer two, which form tiling like patterns for associated families of structures, can thus be seen as vertical displacements of particles from a hexagonal monolayer to the other plate – or as a *bilayer modulation of the hexagonal monolayer*.

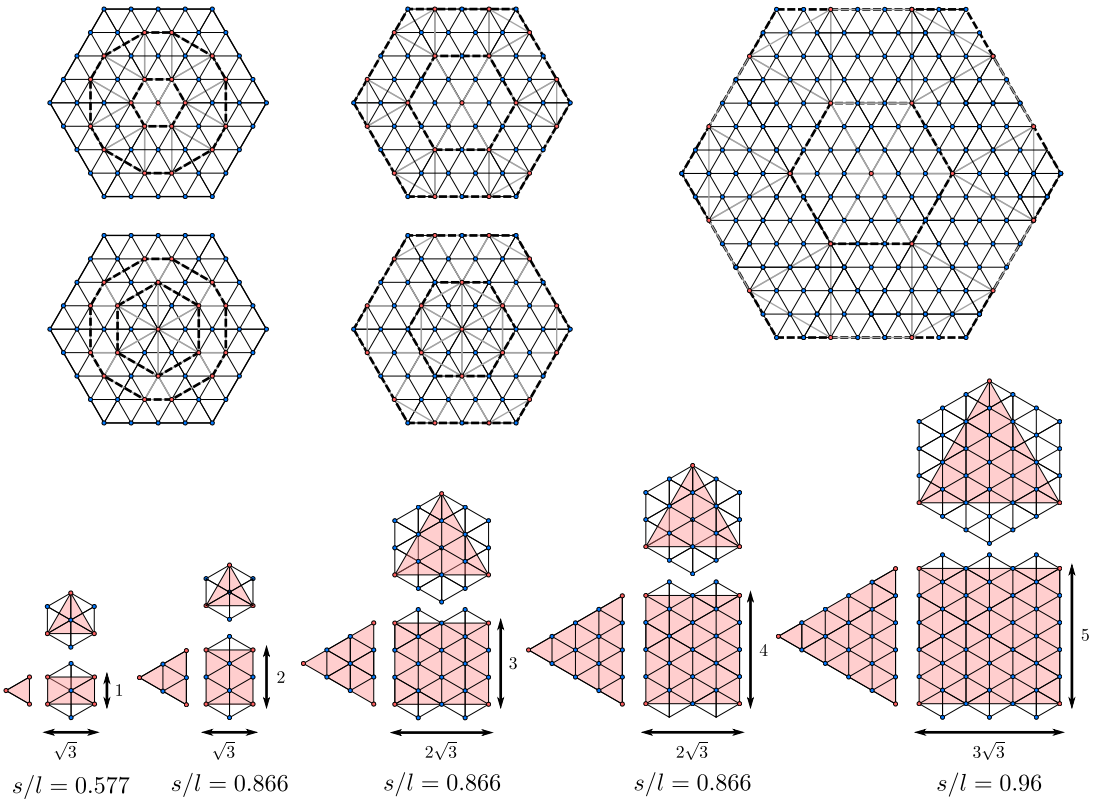


Figure 3.41: Schematic visualization of modulations of a hexagonal monolayer towards hexagonal quasicrystalline structures [264, 274, 284] (only the central pseudo-dodecagons formed by the red vertices are shown). Different, possible choices for the basic $\mathrm{RT}_l\mathrm{t}_s$ tile-sets are shown from left to right with ever larger R, T_l and t_s tiles formed by the red particles (and emphasized in the bottom panels). The aspect ratio of short (s) and long side lengths (l) of the rectangles in the $\mathrm{RT}_l\mathrm{t}_s$ tilings approach unity, $s/l \approx 1$, (i.e., a square-triangle tiling) in the limit of large tiles: $5/(3\sqrt{3}) \approx 0.96$, $7/(4\sqrt{3}) \approx 1.01$, $1000/(577\sqrt{3}) \approx 1.0006$.

In the limit of nearly touching plates, *i.e.*, for $\eta \approx 0$, the Wigner bilayer structure is essentially hexagonal (as in the monolayer case) but *modulated* by patterns of particle displacements from layer one to layer two. Therefore, depending on the value of A, the particles on layer two will form super-structures on the hexagonal lattice with a hexagon-triangle tiling on the denser populated layer. The structure on layer two (*i.e.*, on the sparsely populated layer) follows a maximum uniformity criterion (*i.e.*, particles are distributed *as uniform as possible* on layer two) and will prefer tiling-like

patterns composed of the smallest possible set of tiles (*i.e.*, triangles for H and I_x, rectangles for II, rectangles and triangles for II_x, etc.). For sufficiently large values of A – or equivalently for sufficiently sparse super-lattices in layer two – the ground state structure will always obtain the most uniform triangular super-lattice. However, in the limit of large A – with comparable nearest-neighbor inter-particle distances across both layers – pure triangular super-lattices allow only discrete values for the side lengths of a set of tiles in layer two due to inter-vertex separations on the triangular lattice: $\sqrt{3}\,a_1, 2\,a_1, \ldots,\ a_1$ being the tiling length of the hexagonal lattice. In this limit of large values of A and almost touching plates, smoothly varying the value of A opens a window of possibilities for yet another family of structures, in which super-lattices can *approximate* square-triangle tilings, by providing sets of tiles with intermediate side lengths in between possible pure-triangle super-tilings.

The geometric idea behind the construction of uniform super-lattices on a triangular tiling proceeds by first *approximating* square tiles by a rectangle, R, with an aspect ratio, s/l, close to unity – apparently the first obvious good choice is $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$. The particular choice of R forces a split of the triangle tile into two equilateral triangles, T_l and t_s, whose sides have lengths l and s, respectively, and whose orientations are related by rotations of 90° against each other. Some other geometrically possible choices for the side lengths are $s = a_1$ and $l = \sqrt{3}\,a_1$, $s = 3\sqrt{3}\,a_1$ or $l = 5\,a_1$, $l = 1000\,a_1$ and $s = 577\sqrt{3}\,a_1$ (see Fig. 3.41 for an illustration). Apparently there is a parameter, s/l, (*i.e.*, the aspect ratio of R) which describes how well a square-triangle tiling is approximated by an $\mathrm{RT}_l\mathrm{t}_s$ configuration. Examples of possible values for the aspect ratios, some of which are illustrated in Fig. 3.41, are $s/l = 1/\sqrt{3} \approx 0.577$ for $s = a_1$, $s/l \approx 0.866$ for $s = \sqrt{3}\,a_1$ and $l = 2\,a_1$ and $s/l \approx 1.0006$ for $s = 577\sqrt{3}\,a_1$ and $l = 1000\,a_1$.

In the top panels of Fig. 3.41 we further demonstrate that basic dodecagonal motives – related to dodecagons composed of squares and equilateral triangles – can be transformed into modulated hexagonal configurations, *i.e.*, into $\mathrm{RT}_l\mathrm{t}_s$ tilings, with any size of the rectangular R, large triangular T_l and small triangular t_s tiles; henceforward we refer to this transformation as $\mathrm{RT}_l\mathrm{t}_s$ transformation. Furthermore and more generally, we find that all possible vertices of a square-triangle tiling (*i.e.*, the triangular vertex, the snub-square vertex, the elongated-triangular vertex and the square vertex, cf. Fig. 3.21) can be transformed analogously into modulated hexagonal configurations, again with any size of the R, T_l, t_s tiles. We demonstrate this transformation for $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$ tiles in Fig. 3.42 by providing an explicit list of possible mappings of the square-triangle vertices into two different variants, *i.e.*, in two *flavours* determined by the involved triangles, of $\mathrm{RT}_l\mathrm{t}_s$ vertices which we refer to as *loose* (top row) and *dense* $\mathrm{RT}_l\mathrm{t}_s$ vertices (bottom row):

- a *triangular* vertex (related to the triangular vertex of a square-triangle tiling),
- a *snub-rectangular* vertex (related to the snub-square vertex),
- an *elongated-rectangular* vertex (related to the elongated triangular vertex) and
- a *rectangular* vertex (which comes only in one flavour and is related to the square-vertex).

We also label all possible local edge environments by *edge (1)* through *edge (6)* which connect two separate $\mathrm{RT}_l\mathrm{t}_s$ tiles: *edge (1)* connects two t_s tiles, *edge (2)* connects t_s and R tiles, *edge (3)* connects two R via a short edge, *edge (4)* connects two T_l tiles, *edge (5)* connects T_l and R tiles via a long edge and *edge (6)* connects two R tiles via a long edge.

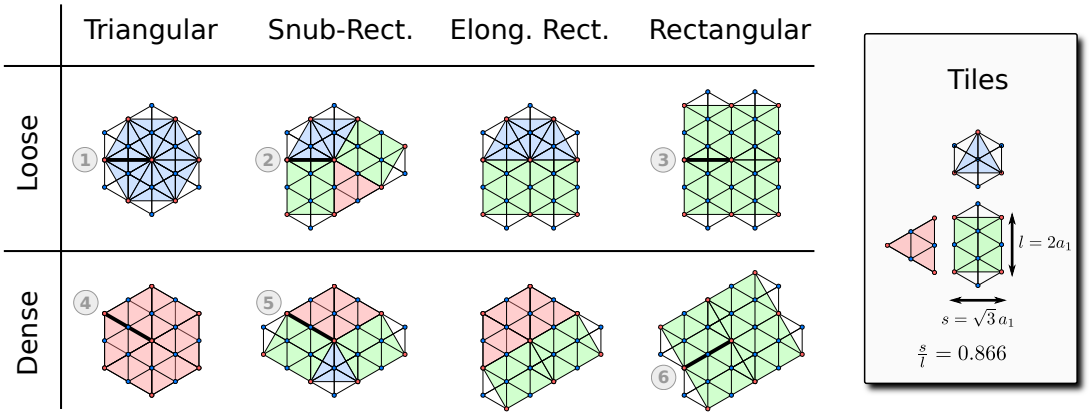


Figure 3.42: Left: schematic visualization of applied transformation rules of the four possible vertices in a square-triangle tiling (*i.e.*, a triangular, a snub-square, a elongated-triangular and a square vertex, cf. Fig. 3.21) to seven different vertices possible for $\mathrm{RT}_l\mathrm{t}_s$ tilings, namely loose (top row) and to dense (bottom row) *triangular* vertices, loose and dense *snub-rectangular* vertices, loose and dense *elongated-rectangular* vertices and a *rectangular* vertex on the specific example of $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$. Different local edge environments of the $\mathrm{RT}_l\mathrm{t}_s$ vertices are labeled by *edge (1)* through *edge (6)* and indicated by thick black lines. Right: Basic $\mathrm{RT}_l\mathrm{t}_s$ tiles formed by red vertices in layer two and decorated by blue vertices in layer one; overlapping or duplicated vertices are deleted, the thin black lines illustrate the triangular tiling when projecting all particles onto one plate.

Obviously, the $\mathrm{RT}_l\mathrm{t}_s$ transformation is valid for tilings without common square-square edges since: all squares are surrounded by triangles and an $\mathrm{RT}_l\mathrm{t}_s$ transformation into two T_l, two t_s and one R tile maintains the position of the geometric center of the associated local environment. The absence of common square-square edges also implies that triangles are either arranged in a triangular or a snub-square vertex and are surrounded either by three other triangles or by three rectangles; also for these motives the geometric centers are symmetric with respect to the $\mathrm{RT}_l\mathrm{t}_s$ transformation. For all of these situations an $\mathrm{RT}_l\mathrm{t}_s$ transformation results in a scaling of the local square-triangle environment. This argument also holds true for transforming four squares of a square vertex into a rectangular vertex. Whenever squares are side by side they cannot always be transformed into parallel rectangles without moving mass centers and a transformation of the unit cell might be necessary for certain structures. A geometrically rigorous proof if any square-triangle tiling can be transformed into an $\mathrm{RT}_l\mathrm{t}_s$ configuration is yet to be given and to this end further investigations are required.

Still, the transformation of all possible local environments of a square-triangle tiling into $\mathrm{RT}_l\mathrm{t}_s$ vertices of two flavours (governed by the two types of triangles) suggests that not only dodecagonal motives can be transformed into $\mathrm{RT}_l\mathrm{t}_s$ structures but there is a one to one mapping of any random square-triangle tiling and $\mathrm{RT}_l\mathrm{t}_s$ tilings. This claim

is supported by an alternative algorithm for performing the RT_lt_s transformation put forward in detail in Appendix A.1.4 where the topology of a square-triangle tiling is never changed when the transformation is executed. We provide a proof by exhaustion for the general validity of RT_lt_s transformations by performing (explicitly for $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$) REMC simulations with RT_lt_s transformed square-triangle tilings – randomized via zipper-moves – with certain periodic boundaries, see below. In other words – and from this point on we assume our claim to be valid and use it as a lemma – any square-triangle tiling can be turned into an RT_lt_s tiling simply by scaling and decorating all tiles appropriately (following the flip-type convention defined in Subsection 3.1.6.1). Thus – purely geometrically speaking – there also exist hexagonal RT_lt_s quasicrystals as super-lattices of the triangular lattice formed by appropriate RT_lt_s transformations of dodecagonal square-triangle quasicrystals, a concept closely related to metallic mean[52] quasicrystals [264, 274].

Notably, also here, the two types of triangles always appear in pairs of possible orientations related by rotations of 90° and we emphasize that deciding on the orientation of one triangle in the RT_lt_s transformation fixes the above introduced flip-type of the entire tiling (*i.e.*, $F = 0$ or $F = 1$). While for structures with real twelve-fold symmetry (such as dodecagonal quasicrystals) 90° flips of local motives do not have an effect on the structure itself the situation can be different for finite-size approximant structures: quasicrystalline approximants always break twelve-fold symmetry in some way and changing the flip-type of such a structure can result in a change of composition, $x = N_2/N$, since the composition of the large and small triangles are, themselves, different.

For $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$ we find for the related I_x and H structures (which are themselves composed only of large or small triangles) a composition of x=1/4 for I_x (cf. dense triangular vertex in Fig. 3.42) and x=1/3 for H (cf. loose triangular vertex in Fig. 3.42). The values of x=1/4 and x=1/3 of the I_x and the H structures give, respectively, the composition per large and small triangle T_l and t_s. The composition of a rectangle tile with $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$ is – equivalent to the composition of large triangles composing I_x– given by $x = 1/4$ (cf. rectangular vertex in Fig. 3.42).

Although the RT_lt_s transformation discussed here and the sampling procedure elaborated in the next Subsection are not limited to any specific values of l and s, we will restrict our discussion from now on the case of $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$ – which is related to the $\mathbf{k}_{32}^{*29}$ family of structures. Henceforward we will use the term rectangle-large–triangle-small–triangle (RT_lt_s) synonymously for $l = 2\,a_1$ and $s = \sqrt{3}\,a_1$ structures.

3.1.6.3 *Zipper*-Inspired Sampling of RT_lt_s Configurations

The one-to-one mapping of square-triangle tilings to RT_lt_s configurations allows us to efficiently sample the configuration space of RT_lt_s structures to find the RT_lt_s configurations that minimize the internal energy of the system at certain values of the system parameters.

[52] The irrational numbers $\phi_k = (k + \sqrt{k^2+4})/2$ form the metallic mean family [284] which give the golden, the silver and the bronze mean for $k = 1, 2, 3$, respectively, and represent the inflation factor for metallic mean quasicrystals [264].

More specifically, the topological equivalence of the tiling in layer two of $\mathrm{RT}_l\mathrm{t}_s$ configurations to square-triangle tilings allows us to use *zipper*-moves to efficiently update the vertices in layer two: by sampling square-triangle tilings and subsequently subject them to the $\mathrm{RT}_l\mathrm{t}_s$ transformation we can perform zipper-induced replica exchange Monte Carlo (REMC) sampling for $\mathrm{RT}_l\mathrm{t}_s$ structures of fixed periodic boundary conditions. Analogously to sampling DI structures (cf. Subsection 3.1.5) also here one zipper move updates the tiling and the $\mathrm{RT}_l\mathrm{t}_s$ decoration[53] defines the rectangle-large–triangle-small–triangle structure and together they represents one sweep in the REMC procedure[54]. Henceforward we refer to the update moves of the vertices in layer two of $\mathrm{RT}_l\mathrm{t}_s$ structures that are based on zipper–moves as pseudo-zipper–moves.

Similar to the zipper based REMC simulations performed for DI structures we present in Figs. 3.43 and 3.44 the results of pseudo-zipper based REMC structure optimizations at $\eta = 0.1061$ and $A = 0.8778$ – where the I_x-Cairo structure represents the ground state candidate [62–64] – for two particular choices of periodic boundary conditions which are inspired by Stampfli-inflation: in Fig. 3.43 we subject a Stampfli-inflated snub-square type of dodecagonal approximant (*i.e.*, a type 2 approximant) to the $\mathrm{RT}_l\mathrm{t}_s$ transformation which results in a $N_2 = 52$ ($N = 209$) particle $\mathrm{RT}_l\mathrm{t}_s$ approximant to a hexagonal quasicrystal [45]. In Fig. 3.43 the same analysis is performed for a Stampfli-inflated and stacked square based approximant (*i.e.*, based on a dodecagonal type 1 approximant) with $N_2 = 60$ ($N = 224$) particles per unit cell. The corresponding best energy from sampling of the $N_2 = 52$ approximant is $E^*_{\mathrm{RTt}}/N = -1.76084604$ and thereby clearly outperforms the current ground state from literature [62–64], *i.e.*, the I_x-Cairo configuration with $E^*_{\mathrm{GS}}/N = -1.76084577$, by $\Delta E^*_{\mathrm{GS}}/N = -2.7 \times 10^{-7}$. The corresponding best energy from sampling the $N_2 = 60$ approximant yields $E^*_{\mathrm{RTt}}/N = -1.76084631$ and outperforms the I_x-Cairo ground state energy by $\Delta E^*_{\mathrm{GS}}/N = -5.4 \times 10^{-7}$.

Since both of the ground state energies of the two $\mathrm{RT}_l\mathrm{t}_s$ REMC sampling procedures outperform the current I_x-Cairo ground state candidate configuration at $\eta = 0.1061$ and $A = 0.8778$ we can at least be fairly sure the global ground state of the Wigner bilayer system at this particular region of system parameters belongs to the $\mathrm{RT}_l\mathrm{t}_s$ family. However, we need to address the question which $\mathrm{RT}_l\mathrm{t}_s$ configuration really represents the ground state of the system – and how we can find it. In practice the pseudo-zipper based $\mathrm{RT}_l\mathrm{t}_s$ transformation is performed according to Appendix A.1.4 which includes relaxation of configurations in each MCMC sampling step (cf. Subsection 2.2.3). This makes the REMC procedure numerically still rather costly such that effectively only small systems – as compared to larger quasicrystalline approximants – can be investigated. In fact, REMC optimization of larger Stampfli based approximants as the ones discussed in Figs. 3.43 and 3.44 are numerically prohibitively costly; we could not reliably identify ground state configurations in ten days of simulation on the *Vienna Scientific Cluster* [281]. To follow up on the idea of identifying global ground states

[53] Here, special care has to be taken to employ the correct flip-types $F = 0$ or $F = 1$ during the decoration procedure to assure a constant number of particles in the system during simulation.

[54] We do not claim that sampling the Wigner bilayer system with these specialized, modified zipper moves will provide us with proper thermodynamics. However, sampling the space of rectangle-large–triangle-small–triangle configurations for ground states is much more efficient using modified zipper moves than relying on Monte Carlo simulations with particles that can move freely in the xy-plane.

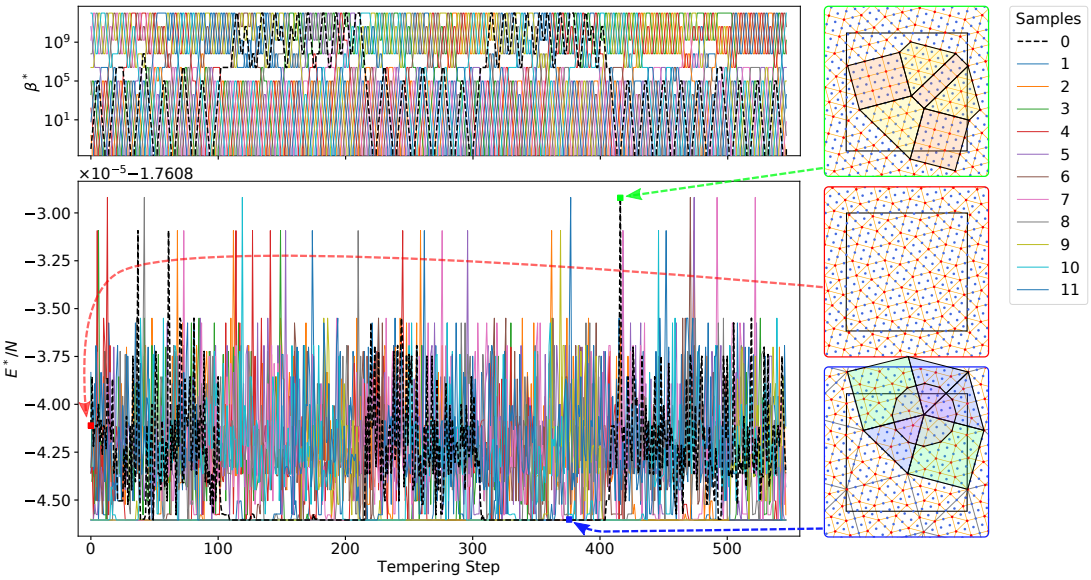


Figure 3.43: Results of REMC simulation (similar as Fig. 3.28) using pseudo-zipper moves to sample layer two of a $N_2 = 52$ ($N = 209$) particle snub-square based $\mathrm{RT}_l\mathrm{t}_s$ approximant at $A = 0.8778$ and $\eta = 0.1061$; the $i = 0, \ldots, 11$ sample trajectories in the $\beta_i^* = 1/T_i^*$ range versus tempering steps are presented in the top left panel and the corresponding energies E_i^*/N of all samples versus tempering steps are presented in the bottom left panel; each sample has an associated color and we highlight the trajectory of *sample 0* via a thick, black-dashed line which transients from medium over high to low values of $E_{i=0}^*/N$. The right three panels (emphasized in green, red and blue) indicate qualitatively different configurations experienced during the REMC-evolution of *sample 0* and the corresponding energies are emphasized with square symbols and indicated by correspondingly colored arrows. At large values of the temperature high energy configurations emerge which may obtain self-similar features (note the triple stacked clusters of squares and triangles forming super-rectangles (orange area) and super-triangle-like tiles (yellow areas) in the green-emphasized structure panel) but the vast majority of high temperature configurations represent random $\mathrm{RT}_l\mathrm{t}_s$ configurations (cf. red-emphasized structure panel). Low energy configurations, again, feature self-similarity and a super snub-square configuration emerges as ground state when connecting the central loose and dense triangular vertices of pseudo-dodecagons (cf. colored tiles in the blue-emphasized structure panel) – the main difference to configurations at larger temperatures (and of larger energies) is the absence of rectangular vertices and dense elongated-rectangular vertices in the ground state. The corresponding best energy from sampling of the super snub-rectangular approximant evaluates to $E_{\mathrm{RTt}}^*/N = -1.76084604$ and clearly outperforms the current ground state from literature, *i.e.*, I_x-Cairo with an energy of $E^*/N = -1.76084577$ by $\Delta E_{\mathrm{GS}}^*/N = -2.7 \times 10^{-7}$. All structures encountered during the pseudo-zipper based REMC sampling depicted here deviate in energy by a maximum of $\Delta E_{\mathrm{RTt}}^*/N \approx 1.5 \times 10^{-5}$. Strictly speaking, the literature database does not cover precisely this value of $\eta = 0.1061$: the ground states from the literature data [62–64] are available for $\eta = 0.15/\sqrt{2}$ and $\eta = 0.16/\sqrt{2}$ and are given, respectively, by I_x-Cairo (at $\eta = 0.15/\sqrt{2}$ with $E_{\mathrm{GS}}^* = -1.76090251$) and by H structures (at $\eta = 0.16/\sqrt{2}$ with $E_{\mathrm{GS}}^*/N = -1.74931247$; the energy of I_x-Cairo here evaluates to $E^*/N = -1.74917621$). At $\eta = 0.15/\sqrt{2}$ we find that the best REMC structure (with a respective value of the internal energy per particle of $E_{\mathrm{RTt}}^*/N = -1.76090248$) is outperformed by the I_x-Cairo structure by $\Delta E_{\mathrm{GS}}^*/N = 3 \times 10^{-8}$. Such close values (and minor changes) of the internal energies of competing structures at the same (and similar) values of the system parameters highlight the delicate numerical sensitivity necessary when searching for ground state configurations in the asymmetric Wigner bilayer system.

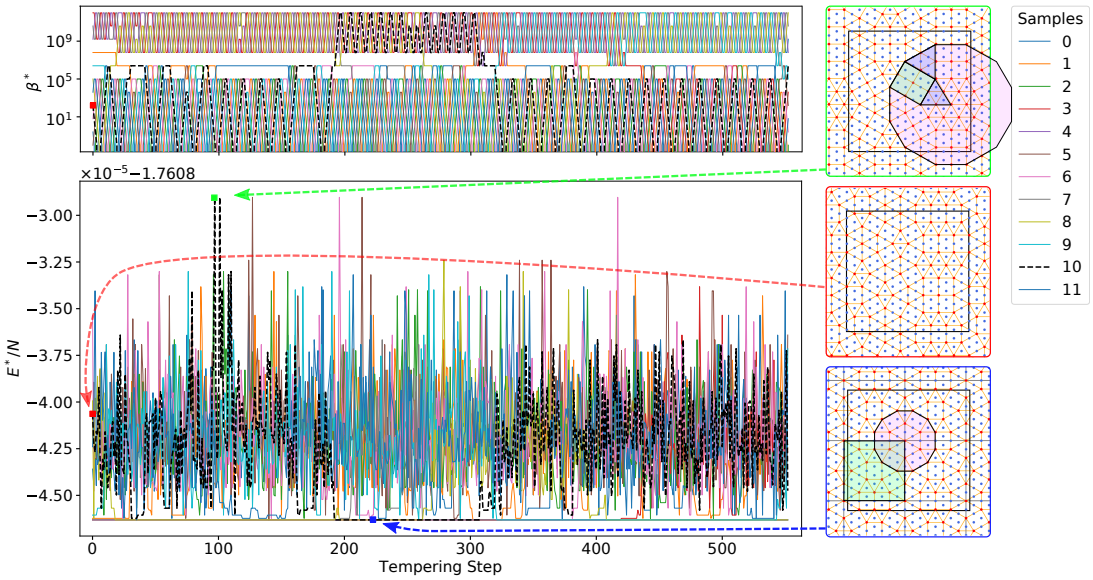


Figure 3.44: Same as Fig. 3.43 but for a $N_2 = 60$ ($N = 224$) particle square based $\mathrm{RT}_l\mathrm{t}_s$ approximant at $A = 0.8778$ and $\eta = 0.1061$. Again, we can see that random $\mathrm{RT}_l\mathrm{t}_s$ tilings at high temperatures order into self-similar (here) super-square configurations at low temperatures which do not feature rectangular and dense elongated-rectangular vertices; in contrast to Fig. 3.43 the super-square tiling only features dense and no loose triangular vertices here. Special tiles and super-tiles are highlighted in color. Note the structure encountered at large values of the temperature featuring a super pseudo-dodecagon (highlighted by the magenta area) in the green-emphasized structure panel: the super pseudo-dodecagon is composed of double stacked super-tiles (which are colored yellow and orange); the vast majority of high temperature configurations represent, again, random $\mathrm{RT}_l\mathrm{t}_s$ configurations (cf. red-empasized structure panel). The corresponding best energy from sampling yields $E^*_{\mathrm{RTt}}/N = -1.76084631$ and outperforms the energy of the currently suggested [62–64] I_x-Cairo ground state candidate configuration by $\Delta E^*_{\mathrm{GS}}/N = -5.4 \times 10^{-7}$. At $\eta = 0.15/\sqrt{2}$ we find $E^*_{\mathrm{RTt}}/N = -1.760903106811$ for the best REMC structure (again relaxed at this value of η) which outperforms the I_x-Cairo structure by $\Delta E^*_{\mathrm{GS}}/N = -6 \times 10^{-7}$.

within the infinitely many $\mathrm{RT}_l\mathrm{t}_s$ structures we first investigate the geometric properties of the respective $\mathrm{RT}_l\mathrm{t}_s$ ground state structures of the two REMC energy minimization procedures. We will further try to analyse the entirety of the encountered structures during REMC for geometric properties such as local environments which govern trends in the ground state structure formation.

Both optimization procedures depicted in Figs. 3.43 and 3.44 lead us to the remarkable conclusion that the ground states form super-triangular and super-rectangular tiles (cf. colored areas in bottom right, blue emphasized inset structures in Figs. 3.43 and 3.44, respectively): when connecting centers of triangular vertices – or, equivalently, of pseudo-dodecagons – we find super-rectangular tiles that have a long and a short side length of $l = 7\,a_1$ and $s = 4\sqrt{3}\,a_1$ (*i.e.*, with an aspect ratio of $s/l \approx 0.99$) and two super equilateral triangular tiles of side lengths $l = 7\,a_1$ and $s = 4\sqrt{3}\,a_1$, respectively. For the energetically most favorable $N_2 = 52$ particle $\mathrm{RT}_l\mathrm{t}_s$ configuration depicted in the bottom right panel of Fig. 3.43 we find that the periodic boundaries allow the formation of a super snub-rectangular tiling. Furthermore, the system in-

deed favors the formation of four pseudo-dodecagons in the unit cell – one with a central loose, three others with central dense triangular vertices – which are exactly arranged in a super snub-rectangular configuration. As depicted in bottom right panel of Fig. 3.44 the ground state of the $N_2 = 60$ particle $\mathrm{RT}_l\mathrm{t}_s$ configuration space forms a super-rectangular tiling – with only one type of pseudo-dodecagons in the unit cell with central dense vertices.

We would like to stress that both of the above discussed ground state structures emerge from randomized $\mathrm{RT}_l\mathrm{t}_s$ configurations. We highlight in Figs. 3.43 and 3.44 the respective sample trajectories which give rise to the respective ground state structures at low temperatures by thick, black-dashed lines. These trajectories show a transient behaviour through the entire range of energies encountered during the respective structure optimization procedure. We can thus also identify the most common and, interestingly, also the energetically least favorable $\mathrm{RT}_l\mathrm{t}_s$ configurations (see green and red emphasized insets in Figs. 3.43 and 3.44) of the respective $\mathrm{RT}_l\mathrm{t}_s$ configuration spaces at this particular (η, A)-pair: the most common structures (see red emphasized sub-panels in Figs. 3.43 and 3.44) exhibit random $\mathrm{RT}_l\mathrm{t}_s$ tilings[55] in layer two at a vanishing cost in energy of $\Delta E^*/N \approx 10^{-6}$ to 10^{-5} as compared to the $\mathrm{RT}_l\mathrm{t}_s$ ground state structures. The energetically least favorable structures (see green emphasized insets in Figs. 3.43 and 3.44) also appear to exhibit large super-tiles. However, these specific "super-tiles" of the energetically least favorable structures can be understood as grains of (phase separated) pure structures (H, I_x and pure rectangles), constraint by the sizes of the respective unit cells and by the corresponding periodic boundary conditions.

We thus conjecture that at this particular (η, A)-pair the system energetically favors mixing the available $\mathrm{RT}_l\mathrm{t}_s$ basic tiles rather than forming clusters (or grains) of pure phases with interfaces between these grains. Furthermore, we observe that in the ground state configurations depicted in the respective bottom right panels of Figs. 3.43 and 3.44 neither dense elongated-rectangular nor rectangular vertices are present; the structures are largely composed of both flavours of triangular and of loose elongated-rectangular vertices.

3.1.6.4 Effective Tile-Hamiltonian

In this Subsection we investigate, if it is possible to construct an effective model, a so-called *tile-Hamiltonian* [282, 283], which is capable of predicting the internal energy of $\mathrm{RT}_l\mathrm{t}_s$ configurations solely based on local *tiling objects* (*i.e.*, local environments such as edges, vertices, cells, etc.). By linearly assigning a particular set of energy coefficients to certain tiling objects present in a configuration a tile-Hamiltonian may (or may not) provide an effective model for both, short and long-ranged interactions of the tiling objects in a system [282, 283]. Thereby, such an effective approach circumvents, in principle, the problem of explicitly evaluating the internal energy of configurations with many positional degrees of freedom – which can quickly become numerically expensive.

[55] For all configurations during the REMC simulations the $\mathrm{RT}_l\mathrm{t}_s$ transformation was applicable.

We explicitly try to identify a tile-Hamiltonian (or, equivalently, an effective model) for the energy per particle of $\mathrm{RT}_l\mathrm{t}_s$ configurations for the system parameters $\eta = 0.1061$ and $A = 0.8778$ for the following three qualitatively different sets of tiling objects: for (a) the number of different vertices, $\mathbf{N}_v(\mathcal{X})$, defined as

$$\mathbf{N}_v(\mathcal{X}) = (N_v, N_{\mathrm{Tr_l}}, N_{\mathrm{Tr_d}}, N_{\mathrm{SN_l}}, N_{\mathrm{SN_d}}, N_{\mathrm{ER_l}}, N_{\mathrm{ER_d}}, N_{\mathrm{R}}), \tag{3.17}$$

(b) the number of different edges, $\mathbf{N}_e(\mathcal{X})$, given by

$$\mathbf{N}_e(\mathcal{X}) = (N_e, N_{\mathrm{e_1}}, N_{\mathrm{e_2}}, N_{\mathrm{e_3}}, N_{\mathrm{e_4}}, N_{\mathrm{e_5}}, N_{\mathrm{e_6}}, N_{\mathrm{e_7}}) \tag{3.18}$$

and (c) the number of vertices and edges combined, $\mathbf{N}_c(\mathcal{X})$, *i.e.*,

$$\mathbf{N}_c(\mathcal{X}) = (N_c, N_{\mathrm{Tr_l}}, \ldots, N_{\mathrm{R}}, N_{\mathrm{e_1}}, \ldots, N_{\mathrm{e_7}}), \tag{3.19}$$

that are present in layer two of the unit cell of a particular realization, $\mathcal{X}$, of an $\mathrm{RT}_l\mathrm{t}_s$ bilayer structure; note that for simplicity we dropped the argument $\mathcal{X}$ for all quantities on the right-hand side of Eqs. (3.17) to (3.19). To be more specific, we collect in $\mathbf{N}_v(\mathcal{X})$ the total number of vertices, N_v, the number of loose and dense triangular vertices, $N_{\mathrm{Tr_l}}$ and $N_{\mathrm{Tr_d}}$, the number of loose and dense snub-rectangular vertices, $N_{\mathrm{SN_l}}$ and $N_{\mathrm{SN_d}}$, the number of loose and dense elongated-rectangular vertices, $N_{\mathrm{ER_l}}$ and $N_{\mathrm{ER_d}}$, and the number rectangular vertices, N_{R}, in layer two of the unit cell of $\mathcal{X}$. In $\mathbf{N}_e(\mathcal{X})$ we collect the total number of edges, N_e, the number N_{e_i} of *edge (1)* through *edge (6)* and the total number of rectangles, N_{e_7}, in layer two of the unit cell of $\mathcal{X}$. In $\mathbf{N}_c(\mathcal{X})$ we collect the total number of vertices and edges, $N_c = N_v + N_e$, and the number of the remaining tiling objects counted by $\mathbf{N}_v$ and $\mathbf{N}_e$ (as given by Eq. (3.19)) in layer two of the unit cell of a particular realization, $\mathcal{X}$, of an $\mathrm{RT}_l\mathrm{t}_s$ structure.

To *predict* the internal energy of an $\mathrm{RT}_l\mathrm{t}_s$ configuration $\mathcal{X}$ we define three different tile-Hamiltonians, *i.e.*, three different linear models $E_v^*(\mathcal{X})$, $E_e^*(\mathcal{X})$ and $E_c^*(\chi)$, by

$$E_v^*(\chi) = \boldsymbol{\epsilon}_v \cdot \mathbf{N}_v(\chi), \tag{3.20}$$
$$E_e^*(\chi) = \boldsymbol{\epsilon}_e \cdot \mathbf{N}_e(\chi) \quad \text{and} \tag{3.21}$$
$$E_c^*(\chi) = \boldsymbol{\epsilon}_c \cdot \mathbf{N}_c(\chi), \tag{3.22}$$

where we introduced the sets of linear coefficients $\boldsymbol{\epsilon}_v = (\epsilon_v, \ldots, \epsilon_R)$, $\boldsymbol{\epsilon}_e = (\epsilon_e, \ldots, \epsilon_{e_7})$ and $\boldsymbol{\epsilon}_c = (\epsilon_c, \ldots, \epsilon_{e_7})$ corresponding energies associated to the three sets of tiling objects $\mathbf{N}_v(\chi)$, $\mathbf{N}_e(\chi)$ and $\mathbf{N}_c(\chi)$ defined by Eqs. (3.17) to (3.19), respectively. We demand from the tile-Hamiltonians $E_{p=v,e,c}^*(\chi)$ given by Eqs. (3.20) to (3.22) to reproduce the "true" Coulomb energies of $\mathrm{RT}_l\mathrm{t}_s$ structures as accurately as possible. As reference data for such "true" energies (to which we also refer to as *target* energies) we here rely on all $i = 1, \ldots, N_{\mathrm{REMC}}$ configurations $\mathcal{X}_i$ and the associated energies $E^*(\mathcal{X}_i)$ obtained during the REMC procedures discussed above (see Figs. 3.43 and 3.44). For the three models defined by Eqs. (3.17) to (3.19) we quantify, respectively, the deviation of the predicted energies, $E_{p=v,e,c}^*(\mathcal{X}_i)$, from the corresponding target energies, $E^*(\mathcal{X}_i)$, of the $\mathrm{RT}_l\mathrm{t}_s$ configurations, $\mathcal{X}_i$, encountered during the REMC procedures via the least-squares error, $\mathcal{L}_{p=v,e,c}$, given by

$$\mathcal{L}_p = \sum_{i=1}^{N_{\mathrm{REMC}}} \left| E_p^*(\mathcal{X}_i) - E^*(\mathcal{X}_i) \right|^2 . \tag{3.23}$$

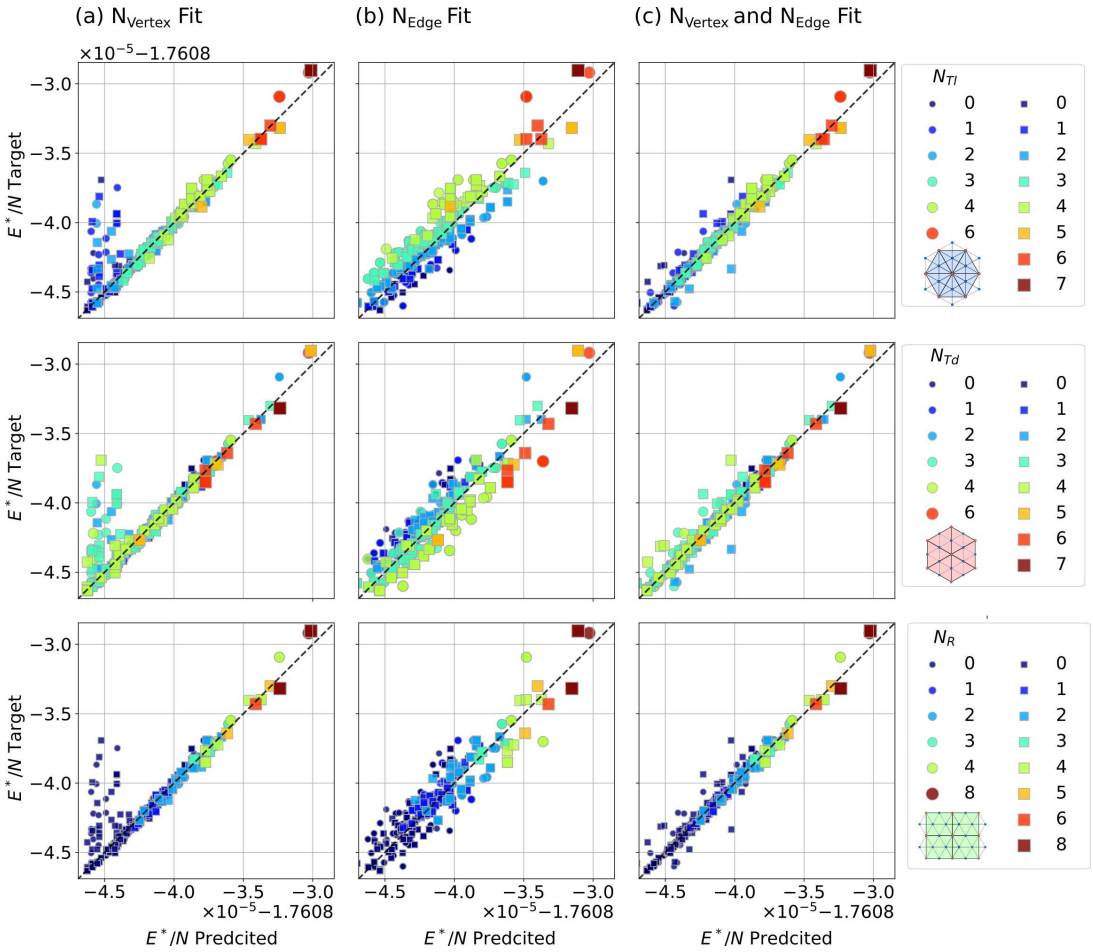


Figure 3.45: Target energies, E^*/N, versus predicted energies, E^*_p/N (via models $E^*_{p=v,e,c}/N$ defined by Eqs. (3.20) to (3.22) in columns (a), (b) and (c), respectively), for samples encountered during pseudo-zipper based REMC sampling of $\mathrm{RT}_l\mathrm{t}_s$ configurations at the system parameters $\eta = 0.1061$ and $A = 0.8778$: each symbol represents a sample from the REMC simulation for a $N_2 = 52$ ($N = 209$) particle $\mathrm{RT}_l\mathrm{t}_s$ transformed snub-square-based approximant (circles, cf. Fig. 3.43) and for a $N_2 = 60$ ($N = 224$) particle $\mathrm{RT}_l\mathrm{t}_s$ transformed square-based approximant (squares, cf. Fig. 3.44). In column (a) we use the number of different vertices, $\mathbf{N}_v$ (cf. Eq. (3.17)), to parametrize the effective linear model, E^*_v/N, given by Eq. (3.20). In column (b) we use the number of edges, $\mathbf{N}_e$ (cf. Eq. (3.18)), to parametrize the effective linear model, E^*_e/N, given by Eq. (3.21). In column (c) we use the number of vertices and edges, $\mathbf{N}_c$ (cf. Eq. (3.19)), to parametrize the effective linear model, E^*_c/N, given by Eq. (3.22). For each column all rows depict the same results of the linear regression with different color-coding: the colors and the sizes of the symbols correspond to the number of *loose triangular* vertices, $N_{\mathrm{Tr_l}}$, (top row), the number of *dense triangular* vertices, $N_{\mathrm{Tr_l}}$,and the number of *rectangular* vertices, N_{R}, of the corresponding $\mathrm{RT}_l\mathrm{t}_s$ configurations as indicated by the respective row-wise legends.

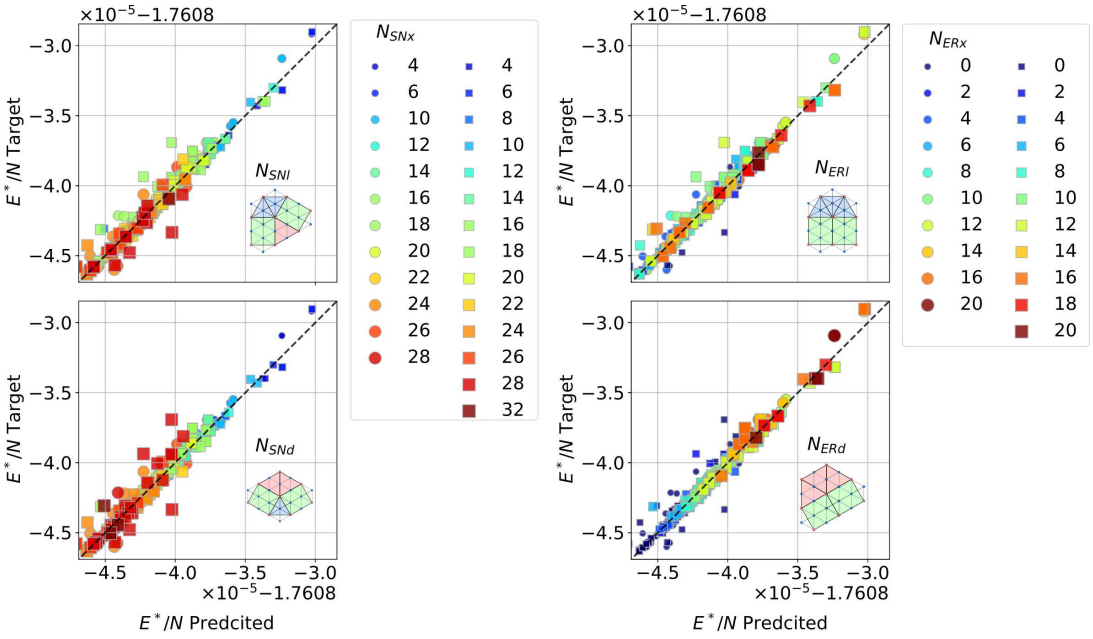


Figure 3.46: Similar to Fig. 3.45 column (c) but with a color-coding based on the number of *snub-rectangular* vertices, $N_{\mathrm{SN_x}}$, (left) and the number of *elongated-rectangular* vertices (right) $N_{\mathrm{ER_x}}$; $x = l, d$ labels either the loose (top) or dense (bottom) flavours of the vertices. Again, each symbol represents a sample from REMC simulations (*i.e.*, the same as discussed in Fig. 3.45) for a $N_2 = 52$ ($N = 209$) particle $\mathrm{RT}_l\mathrm{t}_s$ transformed snub-square approximant (circles) and for a $N_2 = 60$ ($N = 224$) particle $\mathrm{RT}_l\mathrm{t}_s$ transformed square approximant (squares). Left: The symbols in the top and bottom panels are colored and scaled according to the number of loose and dense snub-rectangular vertices $N_{\mathrm{SN_l}}$ and $N_{\mathrm{SN_d}}$, respectively. Right: The symbols in the top and bottom panels are colored and scaled according to the number of loose and dense elongated-rectangular vertices $N_{\mathrm{ER_l}}$ and $N_{\mathrm{ER_d}}$, respectively.

We then fit the linear effective models given by Eqs. (3.17) to (3.19) by minimizing $\mathcal{L}_p$ as a function of the coefficients $\boldsymbol{\epsilon}_p$ via linear regression [285] for $p{=}v, e, c$, respectively.

The target energies E^*/N versus the predicted energies E^*_v/N, E^*_e/N and E^*_c/N (per particle) of all configurations[56] obtained by "sample 0" in the REMC procedure depicted in Fig. 3.43 and by "sample 10" in the REMC procedure depicted in Fig. 3.44 are presented in Figs. 3.45 and 3.46.

When evaluating a tile-Hamiltonian such as defined by Eqs. (3.20) to (3.22) in principle one also has to analyse the tiling objects for possible (linear) dependencies[57] – which are definitely present here. Some dependencies related to summation rules of the tiling objects[58] may be easier to detect. Others, however, are more difficult to identify: for structures with many accumulated rectangular vertices (such as the highest energy

[56] From now on we again drop the argument of the energies.

[57] The occurrence of some of the different tiling objects counted via Eqs. (3.17) to (3.19) may not be independent quantities for $\mathrm{RT}_l\mathrm{t}_s$ structures of a given configuration space: increasing the number of one particular type of vertex, N_i, of a configuration may, for instance, require to reduce the number of another type of vertex, N_j. The corresponding energy coefficients, ϵ_i and ϵ_j may then be correlated and may display unphysical values after minimizing Eq. (3.23) via linear regression.

[58] A certain number of one particular tiling object may imply a certain number of other tiling objects.

configurations depicted in Figs. 3.43 and 3.44) also the number of triangular vertices of both flavours must be large since the number of rectangles and triangles (of each type) is constant in our simulations. Numerous similar, non-obvious constraints are taking place for the here investigated $\mathrm{RT}_l\mathrm{t}_s$ family which require to be analysed carefully in a further study. Thus, we here refrain from explicitly showing the numerical values for the coefficients of Eqs. (3.20) to (3.22) used in the results depicted in Figs. 3.45 and 3.46.

Nevertheless, our analysis shows that the energies of the structural data – based on long-ranged Coulombic energies in a bilayer lattice geometry – can be explained just by counting local tiling objects: we see that the predictions E_v^*/N and E_e^*/N depicted in columns (a) and (b) of Fig. 3.45 show relatively good agreement with the target data E^*/N. However, the fit in (a) does not correctly resolve low energy values and the fit in (b) shows a larger variance over the entire energy range as compared to (a). By combining the tiling objects $\mathbf{N}_v$ and $\mathbf{N}_e$ in the fit depicted in (c) we find that E_c^*/N quite accurately predicts the target energies E^*/N with a lower bias and with higher precision as compared to (a) and (b). Hence, we proved that a reasonable tile-Hamiltonian exists[59].

What is left to do is trying to understand what actually forces the formation of $\mathrm{RT}_l\mathrm{t}_s$ super-tiles in the ground state bilayer configurations at this particular combination of the system parameters. For that purpose we explicitly analyse the occurrence of the involved vertices in the configurations which give rise to the predicted energies presented in Figs. 3.45 and 3.46: we report an almost strictly positive correlation of the number of rectangular vertices, N_R, with E^*/N, a trend which is also present for the number of loose triangular vertices, $N_{\mathrm{Tr_l}}$ (although not as clearly). Low and medium densities of loose triangular vertices can be observed over a large E^*/N range and the number of dense triangular vertices, $N_{\mathrm{Tr_d}}$, is largely independent of E^*/N. In the left panels of Fig. 3.46 we observe a negative correlation of loose and dense snub-rectangular vertices, $\mathrm{SN_l}$ and $\mathrm{SN_d}$, and E^*/N, *i.e.*, the system tries to maximize the number of these vertices in low energy configurations. The right panels of Fig. 3.46 show that the number of loose elongated-rectangular vertices, $N_{\mathrm{ER_l}}$, does not seem to be strongly correlated with E^*/N, *i.e.*, changing $N_{\mathrm{ER_l}}$ is possible without significantly changing E^*/N (omitting possible correlations to other types of vertices here). We see a clear positive correlation of the number of dense elongated-rectangular vertices, $N_{\mathrm{ER_d}}$, and E^*/N indicating an energy penalty for this type of vertices (which is also related to the penalty associated to rectangular vertices). We therefore conclude that rectangular neighbors are allowed in low energy configurations but tend to avoid sharing a short edge, *i.e.*, *edge (5)*.

For the particular choice of the system parameters $\eta = 0.1061$ and $A = 0.8778$ the analysis above allows us to explain the emergence of the $\mathrm{RT}_l\mathrm{t}_s$ super-structures by the system's tendency of avoiding two specific local environments – namely the rectangular and the dense elongated-rectangular vertices – rather than by *maximizing* the number of favorable tiles as in cluster-covering approaches [286–288]. For this (η, A)-combination – and for others which favor $\mathrm{RT}_l\mathrm{t}_s$ super-structures, see below – the

[59] Fixing all necessary, linear independent parameters based on physically meaningful motivations will (most likely) also provide a physically meaningful tile-Hamiltonian.

internal energy of an $\mathrm{RT}_l\mathrm{t}_s$ configuration is minimized by forming pseudo-dodecagonal clusters with central triangular vertices which are encircled by snub-rectangular vertices. The energy is minimized by always explicitly avoiding local environments where four rectangle corners meet at a vertex, or where rectangular tiles share a common short edge in general. neighboring pseudo-dodecagons in low energy configurations can share common snub-rectangular or loose elongated-rectangular vertices and thereby can form triangular and/or rectangular super-tiles.

With our present numerical tools we cannot perform REMC simulations for larger $\mathrm{RT}_l\mathrm{t}_s$ configurations due to prohibitively long computational times. Finding a reliable effective model based on local environments for the $\mathrm{RT}_l\mathrm{t}_s$ family would greatly reduce computational costs – especially because the different kinds of vertices can directly be associated to vertices of different flavours (*i.e.*, loose or dense) of a square-triangle tiling[60] which would make the zipper-based sampling quite efficient. Already at this stage the tile-Hamiltonian defined by Eq. (3.22) is, in principle, suitable to sample different $\mathrm{RT}_l\mathrm{t}_s$ configuration spaces for low energy configurations with reasonable accuracy without involving actual, numerically expensive interactions [282, 283]. Only when comparing qualitatively different structures, especially when they are not of the $\mathrm{RT}_l\mathrm{t}_s$ family, the precise values of the corresponding energies of the configurations become important again.

However, the linear model introduced here would need to be fitted for every particular value of η (when exploiting the symmetry of E^*/N in A, cf. Eq. (3.7)) to be applicable on a wider range of parameters, a task from which we refrain in this thesis. Using other, more flexibly models such as (properly adjusted) artificial neural network potentials [289] which can additionally include system parameters (η, A) as model input parameters could potentially circumvent this problem. Furthermore, if the local environments of the entire asymmetric Wigner bilayer system are exhaustive (at least for ground state configurations) a model sensitive to the possible local environments could potentially comprehend the entire physics of the system (or at least significant portion thereof) solely based on local motives[61].

3.1.6.5 Super-Tilings from $\mathrm{RT}_l\mathrm{t}_s$ Transformation of Self-Similar DI Structures

As we discussed above, sampling of configurations of the rectangle-large–triangle-small–triangle family is numerically quite involved in our present setup. However, we find that the ground state configurations of the $\mathrm{RT}_l\mathrm{t}_s$ family – for their respective periodic boundary conditions – form pseudo-dodecagonal based super-tiles[62] whose emergence we can relate to penalties of specific local environments of the $\mathrm{RT}_l\mathrm{t}_s$ family. This tendency of forming super-structures provides us with important information about the $\mathrm{RT}_l\mathrm{t}_s$ ground state structure formation strategies of the system and allows us to continue our investigations in a different way: nothing prevents us from investigating

[60] The loose or dense vertex flavours are governed by the flip-type, $F = 0, 1$, cf. Fig. 3.40(c).

[61] Probably, also the symmetry of the underlying lattice of a bilayer configuration needs to be taken into account somehow in a faithful model.

[62] In Fig. 3.39 we can see that pseudo-dodecagonal motives (cf. $\mathbf{k}_{32}^{*29}$) have already been present in the structural database in Refs. [62–64] of ground state candidates of the asymmetric Wigner bilayer system but have not been reported yet.

$\mathrm{RT}_l\mathrm{t}_s$ structures solely composed of super-tiles, such as super-triangles (of one or both kinds as identified above in Fig. 3.43) or to form larger structures composed of super-triangles and super-rectangles in very specific setups.

To (semi-)systematically explore possible new candidate structures for the global ground state of the $\mathrm{RT}_l\mathrm{t}_s$ family featuring super-triangular and rectangular tiles – without explicitly performing REMC sampling – we here rely on all self-similar $\mathrm{DI}^{(T)}_{I,S}$ super-square-triangle ground state configurations found in Subsection 3.1.5 (see 3.29). These DI structures form super-triangles and squares by avoiding common edge squares in layer one and further – if triangular vertices in layer one are considered as vertices in both layers – feature square-triangle tilings on both layers. To perform the $\mathrm{RT}_l\mathrm{t}_s$ transformation on these structures we proceed as follows:

- For each $\mathrm{DI}^{(T)}_{I,S}$ structure we can identify two square-triangle tilings, one for each layer, which we assign as layer two vertices of two different, to-be-generated rectangle-large–triangle-small–triangle configurations (to which we refer to as layer one and layer two based structures);
- these are then subject to the $\mathrm{RT}_l\mathrm{t}_s$ transformation in the two possible ways[63], *i.e.*, with flip-type $F = 0$ and with $F = 1$.

We want to emphasize that the super-triangular configuration $\mathrm{DI}^{(T=0)}_{I=1,S=0}$, for which we visualized the $\mathrm{RT}_l\mathrm{t}_s$ transformation scheme in Fig. 3.47, thereby gives rise to four super-triangular structures. We define the layer one based loose and dense super-triangular $\mathrm{RT}_l\mathrm{t}_s$ configurations[64], T_2^L and T_2^D (see bottom middle and bottom right panels of Fig. 3.47) whose super-tiles are, respectively, either all small or all large triangles with a respective side length of $s = 4\sqrt{3}\,a_1$ and $l = 7\,a_1$. The other two configurations are layer two based and feature overlapping loose and dense pseudo-dodecagons in layer two (see top left and top middle panels of Fig. 3.47) with super-triangles of side length $l = 5\,a_1$ and $s = 3\sqrt{3}\,a_1$, respectively. We refer to the latter two structures as loose and dense overlapping super-triangular configurations, TO_2^L and TO_2^D, respectively.

Note that for the dodecagonal square and snub-square DI approximants (types $T = 1$ and $T = 2$) we obtain via the $\mathrm{RT}_l\mathrm{t}_s$ transformation variants of the super snub-rectangular and the super-rectangular configurations discussed above, see Figs. 3.43 and 3.44; variants whose pseudo-dodecagons are oriented in a way such that the structures also lack loose elongated-rectangular vertices and thereby do not exhibit rectangle-rectangle pairs which share a common edge. Since these structures mix loose and dense triangular vertices in the pseudo-dodecagonal based super-tiling as uniform as possible we refer to the structures as *medium density* configurations and label them by M (instead of L and D).

[63] Depending on the symmetry of the initial $\mathrm{DI}^{(T)}_{I,S}$ structure the decoration with $F = 0$ and $F = 1$ may be symmetric as is the case for type 1 (square) and type 2 (snub-square) approximants.

[64] Here we use the terms loose and dense according to the central triangular vertex of the pseudo-dodecagons.

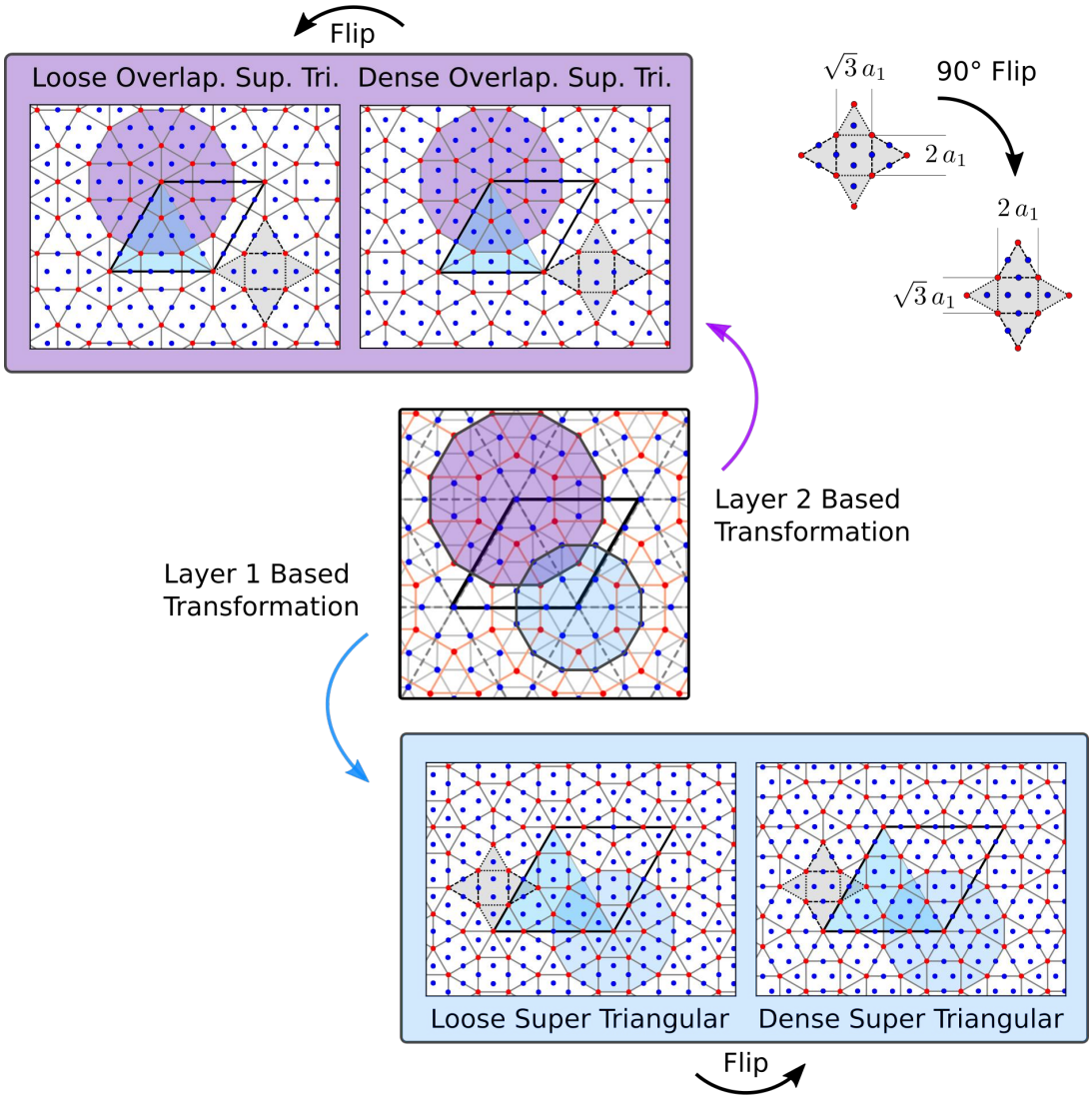


Figure 3.47: Visualization of layer one (bottom panels) and layer two (top panels) based $\mathrm{RT}_l\mathrm{t}_s$ transformations of DI configurations on the example of $\mathrm{DI}^{(0)}_{1,1}$ (see Eq. (3.14) for notation and cf. top left configuration in Fig. 3.47 for the structure in the central panel here). From the $\mathrm{DI}^{(0)}_{1,1}$ structure four $\mathrm{RT}_l\mathrm{t}_s$ structures can be constructed via $\mathrm{RT}_l\mathrm{t}_s$ transformation: Two configurations, *i.e.*, the loose overlapping super-triangular structure (top left, flip-type $F = 1$) and the dense overlapping super-triangular structure (top middle, flip-type $F = 0$), are generated based on the layer two tiling of $\mathrm{DI}^{(0)}_{1,1}$. The two different super-triangular tiles are colored light-blue and the overlapping pseudo-dodecagons are colored light-purple. The other two structures, *i.e.*, the loose super-triangular structure (bottom middle, $F = 0$) and the dense super-triangular structure (bottom right, $F = 1$), are based on layer one of $\mathrm{DI}^{(0)}_{1,1}$. The two different super-triangular tiles (colored light-blue) are larger compared to the layer two based structures; the pseudo-dodecagons (colored light-purple) are of the same kinds but arranged non-overlapping. Notable we can geometrically relate the TO^D_2 structure with adatom super-lattices on (Si,Ge,Sn)-111 diamond-structure surfaces [290].

3.1.6.6 New RT_lt_s Ground State Candidate Super-Structures: The Phase Diagram Refined

In Table 3.3 we list (and label) all qualitatively different RT_lt_s transformed structures which have good chances of being assigned as ground state configuration of the system at certain, selected combinations of η and A and we explicitly present their relaxed energies at $\eta = 0.15/\sqrt{2} \approx 0.1061$ and $A = 0.8778$. At the specified (η, A)-phase point several RT_lt_s super-structures outperform the I_x-Cairo structure, which is the respective ground state candidate suggested by literature [62–64] and we also provide information about the I_x, I_x-Cairo and the H structures. We report a clear energetic advantage of RT_lt_s super-tilings as compared to pure structures which are solely composed of the three basic tiles (*i.e.*, I_x, I_x-Cairo and H) for the particular values of the system parameters of $\eta = 0.106066$ and $A = 0.8778$ of the Wigner bilayer system.

RT_lt_s structure name	Label	N	$E^*/N + 1.76090$	$\Delta E^*_{GS}/N$ to I_x-Cairo
Dense super-rectangular (from REMC Fig. 3.44)	R_2^D	56	-3.11×10^{-6}	-5.9×10^{-7}
Dense super-triangular	T_2^D	49	-3.05×10^{-6}	-5.3×10^{-7}
Medium super-triangular	T_2^M	192	-2.92×10^{-6}	-4×10^{-7}
Super-rectangular (deco. from DI Fig. 3.29)	R_2^M	224	-2.59×10^{-6}	-7.2×10^{-8}
Super-snub-rectangular (deco. from DI Fig. 3.29)	SR_2^M	209	-2.56×10^{-6}	-4.9×10^{-8}
I_x-Cairo	–	15	-2.51×10^{-6}	–
Super-snub-rectangular (from REMC Fig. 3.43)	SR_2^D	209	-2.48×10^{-6}	3.6×10^{-8}
Trihexagonal tiling (only large triangles)	I_x	4	-2.47×10^{-6}	4.3×10^{-8}
Loose super-triangular	T_2^L	48	-2.00×10^{-6}	5.1×10^{-7}
Dense overlapping super-triangular	TO_2^D	27	-1.32×10^{-6}	1.2×10^{-6}
Loose overlapping super-triangular	TO_2^L	25	+3.52×10^{-6}	2.9×10^{-6}
Honeycomb (only small triangles)	H	3	+1.16×10^{-2}	1.16×10^{-2}

Table 3.3: List of RT_lt_s super-structures, corresponding label, number of particles per (irreducible) unit cell, energies per particle at $\eta = 0.15/\sqrt{2} \approx 0.1061$ and $A = 0.8778$ (note the offset of -1.76090 for the given energies, E^*/N) and energy penalty to the current ground state configuration from literature at this combination of η and A, *i.e.*, to the I_x-Cairo structure. The table is sorted by E^*/N in ascending order.

Notably, in the database of ground state configurations of the asymmetric Wigner bilayer system from literature [62–64] only structures of up to $N = 40$ particles per unit cell are considered due to numerical reasons[65]. Hence, neither of the two T_2^D structures with $N = 196$ (which can be reduced to $N = 49$ in the irreducible unit cell), the R_2^D configuration with $N = 224$ (which can be reduced to $N = 56$ in the irreducible unit cell) or SR_2^D ($N = 209$) and SR_2^M (N=209) could have been identified by the evolutionary algorithm in Ref. [62];

We now want to address the question how pseudo-dodecagon based RT_lt_s super-tilings perform at other locations in the (η, A)-plane. For that purpose, all of the RT_lt_s transformed DI structures as well as the results from REMC are now relaxed in an extended (η, A)-range which is relevant for the rectangle-large–triangle-small–triangle family of structures. In Fig. 3.48 we present the corresponding offset, $\Delta E^*_{GS}/N = (E^*_{GS}/N - E^*_{RTt}/N)$, of the lowest energy, E^*_{RTt}/N, of all available relaxed RT_lt_s

[65] The evolutionary algorithm (cf. Subsection 2.2.4) used in Ref. [62] to identify ground state configurations of the asymmetric Wigner bilayer system failed to converge in a reasonable amount of time for structures larger than $N = 40$.

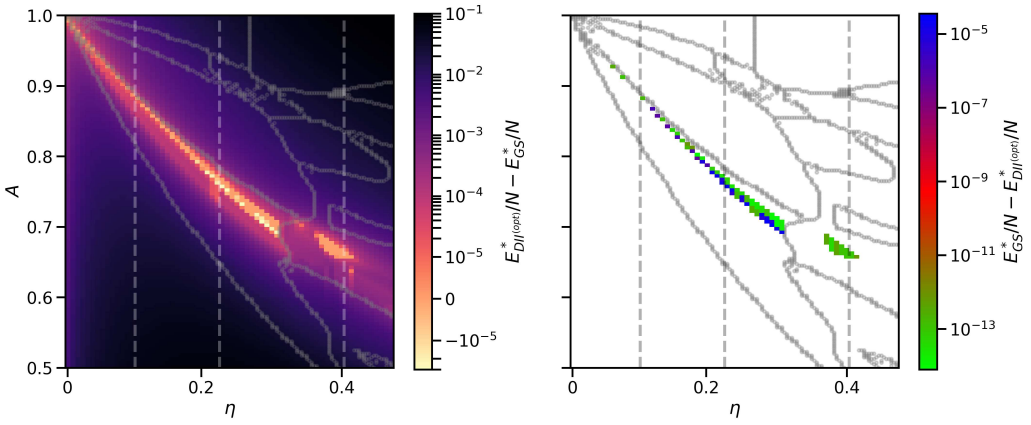


Figure 3.48: Left: Energy offset, $\Delta E^*_{\rm GS}/N$, in log-scale of the (available) energetically most favorable relaxed $\mathrm{RT}_l\mathrm{t}_s$ super-structures to the ground state candidates (as suggest by Refs. [62–64]) of the asymmetric Wigner bilayer system in the (η, A)-plane. Right: Positive values of $-\Delta E^*_{\rm GS}/N$ in log-scale in the (η, A)-plane highlighting regions in phase-space where $\mathrm{RT}_l\mathrm{t}_s$ super-structures are energetically more favorable than ground state candidates from literature [62–64] (see yellow and orange areas in the left panel and colored regions in the right panel): wherever $-\Delta E^*_{\rm GS}/N > 0$ super-structures appear to have an energetic advantage over pure I_x, H or even I_x-Cairo tilings and represents new global ground state candidates of the asymmetric Wigner bilayer system. Solid gray lines indicate phase boundaries from literature [62–64]. Vertical dashed lines correspond to special values of $\eta = 0.106066$, $\eta = 0.226$ and $\eta = 0.403$ which give rise to qualitatively different new ground state candidates of $\mathrm{RT}_l\mathrm{t}_s$ super-structures (at the respective values of A, see text).

super-structures to the currently suggested ground state energy[66], $E^*_{\rm GS}/N$, from literature [62–64] in the (η, A)-plane.

The bright yellow region in the left panel of Fig. 3.48 emphasizes where (some of) the $\mathrm{RT}_l\mathrm{t}_s$ super-structures are either energetically degenerate (*i.e.*, $\Delta E^*_{\rm GS}/N \approx 0$, cf. bright orange area) or even energetically more favorable ($\Delta E^*_{\rm GS}/N \approx -10^{-5}$) than current ground state candidates from literature, labeling thereby new ground state candidate configurations of the asymmetric Wigner bilayer system as compared to Refs. [62–64]. Moreover, this region corresponds to the above mentioned (η, A)-region with strong signals of the $\Psi^{(2)}_{12}$ bond orientational order parameter (cf. top right panel in Fig. 3.3 and Fig. 3.39) which already hosts – in the literature database [62–64] – competing structures of nearly degenerate energies (cf. Fig. 3.19).

To further highlight the improvements made to the ground state phase-diagram by the $\mathrm{RT}_l\mathrm{t}_s$ super-tiling family in the (η, A)-plane we show in the right panel of Fig. 3.48 a heat-map of $-\Delta E^*_{\rm GS}/N$ in the (η, A)-plane in log-scale thereby only emphasizing regions in the parameter space where $E^*_{\rm RTt}/N < E^*_{\rm GS}/N$, *i.e.*, where $\mathrm{RT}_l\mathrm{t}_s$ super-structures are energetically more favorable than literature data. The related energy

[66] All energies are always given for a certain structure at a given (η, A)-pair which we neglect as arguments of the energies for simplicity.

scales range from $-\Delta E^*_{\mathrm{GS}}/N \approx 10^{-13}$ to 10^{-5} which on the one hand indicates that several $\mathrm{RT}_l\mathrm{t}_s$ super-tilings (or at least variants thereof) have already been present in the database [62–64] wherever $-\Delta E^*/N \approx 10^{-13}$ and on the other hand clearly suggests new ground state candidate configurations with $-\Delta E^*/N \approx 10^{-7}$ to 10^{-5}.

In Fig. 3.49 we present a refined phase-diagram for zero temperature ground state candidate configurations of the asymmetric Wigner bilayer system with special focus on the $\mathrm{RT}_l\mathrm{t}_s$ super-structures. We can see, that the region in parameter space where the $\mathrm{RT}_l\mathrm{t}_s$ super-structures represent the ground state candidates of system is rather narrow in A and we want to stress that the resolution in η is imposed by the literature data [62–64].

Special values of $\eta = 0.106066$ and $\eta = 0.226$ are highlighted in Fig. 3.48 via vertical dashed lines where $\mathrm{RT}_l\mathrm{t}_s$ super-structures with qualitatively different super-tiles represent the ground state (at the respective A values), see details below. Furthermore, there appears to be another, distinct region in the phase diagram in the vicinity of $\eta \approx 0.403$ and $A \approx 0.66$ where the relaxed $\mathrm{RT}_l\mathrm{t}_s$ super-tilings are energetically as favorable (or even slightly more favorable) compared to the data from literature [62–64]. Here, relaxation of the $\mathrm{RT}_l\mathrm{t}_s$ transformed super-tilings leads to a deformation of the hexagon-triangle tiling in layer one while the pseudo-dodecagonal rectangle-large–triangle-small–triangle topology in layer two remains intact for some $\mathrm{RT}_l\mathrm{t}_s$ structures. We will discuss this special region – which can be related to the rectangle-large–triangle-small–triangle family via a modified $\mathrm{RT}_l\mathrm{t}_s$ decoration scheme – below.

At plate separation $\eta = 0.106066$ we can identify an A-range featuring new T_2^D and TO_2^D ground states with, respectively, up to 196 particles and 100 per unit-cell. At $\eta = 0.106066$ and $0.8778 \leq A \leq 0.8782$ we find that the T_2^M with $N = 192$, the T_2^D with $N = 196$ and the R_2^D with $N = 224$ configurations improve[67] (at slightly different values of A) the former I_x-Cairo type of ground state candidate by $-\Delta E^*_{\mathrm{GS}}/N \approx \times 10^{-7}$. The smaller version of T_2^D with $N = 49$ (see bottom right panel in Fig. 3.47) is energetically de facto degenerate compared to the $N = 196$ stacked T_2^D configuration, *i.e.*, $\Delta E^*/N \approx 10^{-13}$ depicted in Fig. 3.49. Similarly, also the $N = 224$ particle R_2^D structure shown in Fig. 3.49 can be reduced to $N = 56$ particles per irreducible unit cell at a vanishing cost in energy. The $N = 192$ particle T_2^M structure on the other hand is already represented in Fig. 3.49 with its irreducible unit cell. In the range $0.8782 \leq A \leq 0.88$ (and at $\eta = 0.106066$) the TO_2^D configuration forms the ground state with $N = 100$ particles per unit cell which is a stacked version of the $N = 25$ particle TO_2^D configuration (which both obtain a vanishing energy difference) depicted in the top middle panel in Fig. 3.47. TO_2^D has already been present in the database used to generate the phase diagram [62–64] but has not been reported yet. The newly discovered T_2^M, T_2^D, R_2^D and TO_2^D ground states are framed by I_x for $A \lesssim 0.8778$ and H for $A \gtrsim 0.88$ at $\eta = 0.106066$. All of these newly identified ground state candidate configurations favor dense over loose triangular vertices in the centers of pseudo-dodecagons which themselves form either (exclusively) super-rectangles with

[67] Although an energy penalty as small as $\Delta E^*/N \approx 10^{-7}$ seems like a vanishingly small number, the numerical accuracy of the Ewald summation technique used here, cf. Subsection 2.1.3, is of the order of 10^{-13} with the current set of cutoff parameters.

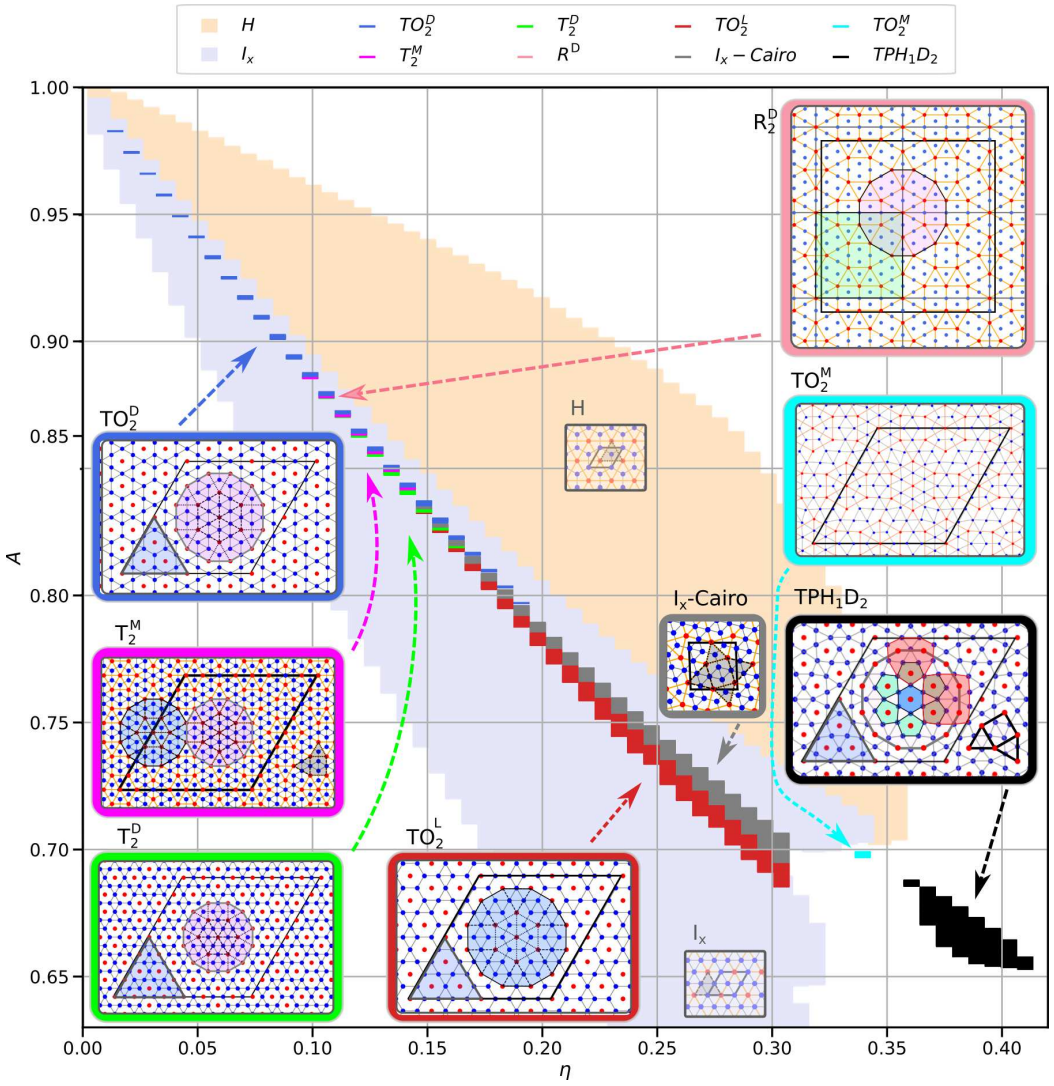


Figure 3.49: Refined phase diagram of zero temperature ground state candidate configurations of the asymmetric Wigner bilayer system including $\mathrm{RT}_l\mathrm{t}_s$ structures which represent newly discovered ground state candidates at selected regions in the (η, A)-plane, see labeled insets (cf. Table 3.1): colored regions indicated by equally colored arrows mark the ground-state regions of several $\mathrm{RT}_l\mathrm{t}_s$ related structures presented as insets which are also correspondingly color-coded. The unit cells of the different structures are indicated by thick black frames and special tiles or motives are highlighted by faint colors: different super-triangles are colored blue, super-rectangles are colored green and dense/loose pseudo-dodecagons are colored purple/blue. The $\mathrm{TPH_1D_2}$ and the TO_2^M structures do not qualify as members of the $\mathrm{RT}_l\mathrm{t}_s$ family but are closely related. The $\mathrm{TPH_1D_2}$ structure features **t**riangles, **p**entagons and **h**exagons in layer one and pseudo-**d**odecagon in layer two, hence the naming. The TO_2^M somehow interpolates between TO_2^D, TO_2^L and the $\mathrm{TPH_1D_2}$ structures by mixing loose and dense overlapping pseudo-dodecagons, thereby necessarily creating small distortions with respect to the $\mathrm{RT}_l\mathrm{t}_s$ family. Notably, for some of the structures which are identified as the ground state candidate configurations smaller versions exist with "unstacked" unit cells – *i.e.*, with $N \to N/4$ particles in the irreducible unit cell, see text. The sizes of the irreducible unit cells which we here identify as ground state candidates are given by $N = 25$ for TO_2^D and $\mathrm{TPH_1D_2}$, $N = 27$ for TO_2^L, $N = 49$ for T_2^D, $N = 56$ for R_2^D, $N = 100$ for TO_2^M and $N = 192$ for T_2^M.

a long side of $l = 7\,a_1$ and a short side of $s = 4\sqrt{3}\,a_1$ in R_2^D or (again exclusively) super-triangles of side length $l = 7\,a_1$ in T_2^D and T_2^M or $s = 3\sqrt{3}\,a_1$ in TO_2^D.

Now we turn to the specific value of $\eta = 0.226$ for the plate separation. For $A \leq 0.751$ we here find the I_x structure to be the ground state candidate (in the I_x region) and for $A \gtrsim 0.765$ we are in the H phase. Between the regions of stability of I_x and H we report for $0.759 < A \leq 0.765$ a I_x-Cairo phase. Further, in the range $0.752 \leq A \leq 0.759$ the stacked loose overlapping super-triangular structure, labeled TO_2^L, with $N = 108$ particles per unit cell represents the ground state candidate configuration; the smaller, "unstacked" TO_2^L variant with $N = 27$ particles in the unit cell – see Fig. 3.47 – is, again, energetically virtually degenerate. The corresponding configuration to TO_2^L with $N = 27$ ($N_1 = 7$ and $N_2 = 20$) seems to be missing in the database [62–64] and we cannot directly compare the results from the evolutionary algorithm for this particular structure[68]. At $\eta = 0.226$ and $A = 0.752$ we find that the TO_2^L structure is energetically more favorable than the direct competitor and former ground state candidate, I_x, by $-\Delta E^*_{\mathrm{GS}}/N = 1.3 \times 10^{-6}$. In comparison, the related TO_2^D structure is here energetically less favorable by a significant amount of $\Delta E^*/N = 3 \times 10^{-4}$ which demonstrates that at this region in the phase diagram loose triangular vertices are clearly favored over dense ones in the centers of the pseudo-dodecagons; the super-triangle tiles of TO_2^L have a side length of $l = 5\,a_1$.

As already mentioned above, there appear to be vanishingly small differences in the energies of the relaxed super-triangular $\mathrm{RT}_l\mathrm{t}_s$ structures and the previously suggested ground state candidates from literature [62–64] in the proximity of $\eta \approx 0.403$ and $A \approx 0.66$. Here, a novel phase emerges with a very special kind of geometry as illustrated in the bottom right inset of Fig. 3.49. Via relaxation of a stacked TO_2^L configuration at the respective, relevant (η, A)-pairs a structure emerges which features (close-to-square) rectangles and (close-to-equilateral) triangles in layer two which are approximately but exclusively mid-edge decorated by layer one particles. This structure represents the ground state candidate in the proximity of $\eta \approx 0.403$ and $A \approx 0.66$. The tiling in layer two can be described by overlapping, pseudo-dodecagons that are placed on a hexagonal lattice which are composed of approximate squares and equilateral triangles. The truly interesting feature of this structure are the tiles in layer one: starting in the center of the pseudo-dodecagon in layer two we find the particles in layer one to be arranged in a hexagon (blue emphasized tile in the bottom right inset of Fig. 3.49) with six adjacent triangles, a motive which is similar to the I_x phase. However, the triangles are then framed by six (only slightly distorted) pentagons with layer two particles placed in their geometric centers. This inner ring of six pentagons in layer one is surrounded by an outer ring of twelve distorted layer one pentagons[69], again with layer two particles placed in their geometric centers which thereby form the outer vertices of the pseudo-dodecagon in layer two. Henceforward, we refer to this structure as $\mathrm{TPH_1D_2}$ due to the **t**riangular, **p**entagonal, **h**exagonal tiles in layer one and the pseudo-**d**odecagonal arrangement of particles in layer two.

[68] Consequently, we see the relatively large improvements in energy of $-\Delta E^*/N \approx 10^{-5}$ of these structures compared to the previously suggested ground state values from literature in Fig. 3.48 in the proximity of $\eta \approx 0.226$ and $A \approx 0.752$.

[69] Groups of three or four neighboring pentagonal tiles of these two rings of pentagons form super-tiles reassembling distorted triangles (or rather shield-like tiles) or rectangles, respectively, see red indicated motives in the bottom right inset of Fig. 3.49.

We note that the TPH_1D_2 structure with $N = 100$ particles in the unit cell can be reduced to $N = 25$ particles in the irreducible unit cell which has already been identified by the evolutionary algorithm in Ref. [62] but has not been reported. We also stress the close connection of the decoration scheme of the layer two tile-set of (close-to-square) rectangles – with an aspect ratio of $s/l \approx 0.99$ – and triangles by particles in layer one with the RT_lt_s decoration scheme: starting from a perfect square-triangle tiling in layer two the decoration scheme here is a variant of the RT_lt_s decoration scheme introduced in Subsection 3.1.6.1 with the modification that particles in layer one are only placed at the projected mid-edge positions of all layer two tiles (with subsequent relaxation of the entire structure). A triangular vertex in layer two thereby generates hexagonal tiles in layer one and snub-square vertices[70] in layer two automatically give rise to pentagonal motives in layer one.

Another type of newly discovered RT_lt_s based ground state candidate structure at $\eta \approx 0.339$ and $A \approx 0.698$ is given by the medium density overlapping super triangular structure which we label TO_2^M (with $N = 100$ particles in the irreducible unit cell, not listed in Table 3.3). The TO_2^M structure slightly violates the RT_lt_s decoration scheme and somehow interpolates between TO_2^D, TO_2^L and the TPH_1D_2 structures by mixing loose and dense triangular centers in two distorted overlapping pseudo-dodecagons on an approximate super-triangular tiling. Thereby the TO_2^M structure necessarily creates deviations and small distortions with respect to RT_lt_s structures (cf. Fig. 3.49).

Summarizing, the loose TO_2^L, the dense TO_2^D and the dense T_2^D and R_2^D structures with $N = 25$, $N = 27$, $N = 49$ and $N = 56$ (or with $N = 100$, $N = 108$, $N = 196$ and $N = 224$ in their stacked versions), respectively, as well as the TPH_1D_2 configuration with $N = 25$ ($N = 100$ when stacked) represent newly reported ground state candidates of the asymmetric Wigner bilayer system for special (η, A)-regions. Furthermore, also the medium density TO_2^M and T_2^M configurations with $N = 100$ and $N = 192$ in the respective irreducible unit cell, represent new ground state candidates, see Fig. 3.49:

- TO_2^L and TPH_1D_2 have already been present in the literature database [62–64] but have not been reported.
- TO_2^D has apparently been missed in Ref. [62] (although the evolutionary algorithm could have identified it) and we report a new structural ground state candidate of the asymmetric Wigner bilayer system which improves previously suggested solutions by $-\Delta E^*_{GS}/N \approx 10^{-5}$ in parts of the phase diagram.
- T_2^M, T_2^D, R_2^D and TO_2^M are completely new structures as compared to the ones in the literature database [62–64] and improve previously suggested solutions by $-\Delta E^*_{GS}/N \approx 10^{-7}$ to 10^{-6} in parts of the phase diagram.

All of the RT_lt_s super-structures are based on RT_lt_s modulations of the hexagonal monolayer – or exhibit slight modifications thereof, cf. TPH_1D_2 and TO_2^M– and are either twofold (for R_2^D and TO_2^M) or sixfold rotationally symmetric. They feature super-rectangular and/or two types of differently sized super-triangular tilings when connecting the centers of their layer two pseudo-dodecagons. The emergence of these complex super-structure ground state candidates is truly remarkable for a system whose

[70] Snub-square vertices are composed of three triangles and two squares, *i.e.*, five tiles.

physics is governed solely by long ranged Coulomb interaction between charged particles which are confined to two oppositely charged plates.

3.1.6.7 Inflated $\mathrm{RT}_l\mathrm{t}_s$ Super-Structures

The basic tile-set of the $\mathrm{RT}_l\mathrm{t}_s$ decoration scheme highlights the simplest possible structures that can be generated with these three tiles, namely the H phase, the I_x phase and a configuration of pure rectangles. From the high energy results of the REMC simulations (cf. green emphasized structure insets in Figs. 3.43 and 3.44) we find, that phase boundaries of grains of such pure structures – restricted by the respective periodic boundary conditions and depending on the values of η and A – are energetically unfavorable; the basic tiles rather mix and thereby form super-tiles such that the ground states of the $\mathrm{RT}_l\mathrm{t}_s$ family tend to form super-structures. Which kind of super-tiles are realized in the end is dictated by the kind of triangular vertices that are favored by the system at specific values of η and A as central motives of the pseudo-dodecagons in layer two. The newly discovered global $\mathrm{RT}_l\mathrm{t}_s$ super-tiling ground state candidates (see Fig. 3.49) all exhibit such super-tiles but, respectively, only one kind of these super-tiles is usually present in these configurations: either small or large super-triangles or super-rectangular tiles. In this Subsection we investigate larger, more complex super-structures of the $\mathrm{RT}_l\mathrm{t}_s$ family which uniformly mix dense and loose central vertices of non-overlapping pseudo-dodecagons. We then compare these structures against the previously suggested [62–64] ground state candidates of the asymmetric Wigner bilayer system for different values of the system parameters.

In Table 3.4 we see that the SR_2^D structure obtained from REMC is energetically slightly less favorable than I_x-Cairo while SR_2^M (which is based on a super-snub-square DI structure) is energetically more favorable. SR_2^D and SR_2^M are related by exactly one rotation of a pseudo-dodecagon in the unit cell: SR_2^D only has one loose and three dense triangular vertices while SR_2^M has two of each kind in the irreducible unit cell and thereby eliminates not only all rectangular but also all snub-rectangular vertices in the tiling in layer two. Here we follow up on this idea of completely avoiding rectangular neighbors in layer two which share a common edge since avoiding atomic defects in Subsection 3.1.5 provides us a unique construction scheme for super-tilings based on pseudo-dodecagons in layer two: in Appendix A.1.5 we present design rules based on square-triangle tilings how to construct large Stampfli-inflated self-similar super-structures from scratch, simply by avoiding neighboring squares at the smallest length-scale. The method itself is straightforward but somehow lengthy and feels out of place here, which is why we moved it to the Appendix. The key ingredient is not only to inflate a square-triangle lattice but also to perform the stacking correctly, *i.e.*, we do not simply periodically stack the unit cell but stack the tiles instead. In that way Stampfli-inflated dodecagons are inscribed – notably with a very particular sequence of orientations – to stacked super-tiles, six to a stacked triangle and nine to a stacked square.

Here we used a square-triangle structure depicted in Fig. A.16 and subjected it to $\mathrm{RT}_l\mathrm{t}_s$ transformations, once with $F = 0$ and once with $F = 1$, which results in two possible, rectangle-neighbor-free quasicrystalline approximants with $N = 1444$ and $N = 1452$

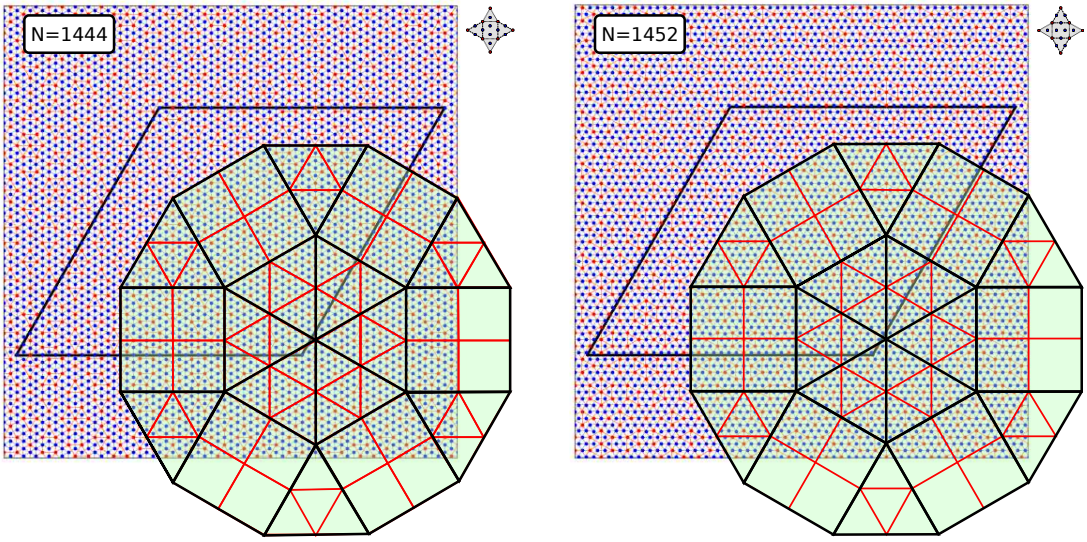


Figure 3.50: The left and right panel show super-stacked $\mathrm{RT}_l\mathrm{t}_s$ configurations with $N = 1444$ (flip-type $F = 0$) and $N = 1452$ (flip-type $F = 1$) particles per irreducible unit cell, respectively. Both structures are based on the defect-free inflated square-triangle structure shown in Fig. A.16 which we subjected to $\mathrm{RT}_l\mathrm{t}_s$ transformations. The initial square-triangle structures both exhibit a hexagonal lattice (black, rhombic unit cells) with overlapping dodecagons inscribed into the unit cells in a first inflation step (green colored dodecagon composed of black framed rectangles with an aspect ratio of $s/l = 8\sqrt{3}/14 \approx 0.99$ and two types of triangles). An additional stacking of the inflated rectangles and triangles (indicated by thick red lines) mark the vertices which are decorated with non-overlapping, pseudo-dodecagons of the $\mathrm{RT}_l\mathrm{t}_s$ family (see thin red lines).

particles per unit-cell (see Fig. 3.50). We label the $N = 1444$ particle approximant by $\mathrm{IpD}^{\mathrm{L}}$ (for hexagonal–pseudo-dodecagonal) and the $N = 1452$ approximant by $\mathrm{IpD}^{\mathrm{D}}$.

After relaxing the $\mathrm{IpD}^{\mathrm{L}}$ and $\mathrm{IpD}^{\mathrm{D}}$ approximants in the (η, A)-range relevant to the $\mathrm{RT}_l\mathrm{t}_s$ family (at $0 \leq \eta \lesssim 0.4$ and $0.6 \lesssim A \leq 1$) we can compare the energies of the two structures, *i.e.*, E^*_{1444}/N and E^*_{1452}/N, with the previously suggested ground state energies from literature [62–64], E^*_{GS}/N. Similar to Fig. 3.48 we present in Fig. 3.51 the energy offset $\Delta E^*_{\mathrm{GS}}/N = (E^*_{\mathrm{IpD}'}/N - E^*_{\mathrm{GS}}/N)$ in the (η, A)-plane for a very fine grid in A, where $E^*_{\mathrm{IpD}'}/N = \min(E^*_{1444}/N, E^*_{1452}/N)$ for a given pair of η and A. Stunningly, we find that the $\mathrm{IpD}^{\mathrm{L}}$ structure is energetically more favorable than previously suggested [62–64] ground state candidate configurations by $\Delta E^*_{\mathrm{GS}} \approx 10^{-7}$ to 10^{-6} in selected, very narrow η and A regions ($\mathrm{IpD}^{\mathrm{D}}$ is energetically always slightly more unfavorable than $\mathrm{IpD}^{\mathrm{L}}$). The right panel of Fig. 3.51 emphasizes the system parameters where the $\mathrm{IpD}^{\mathrm{L}}$ approximant's energy is below the previously suggested ground state energies from literature [62–64] and the improvements to these previously suggested ground state energies are color coded.

To improve the readability of the energy offset, $\Delta E^*_{\mathrm{GS}}/N$, depicted in Fig. 3.51 in the (η, A)-plane – and to emphasize the delicate, narrow parameter range where $\mathrm{IpD}^{\mathrm{L}}$ is energetically more favorable than the current ground state candidates from literature –

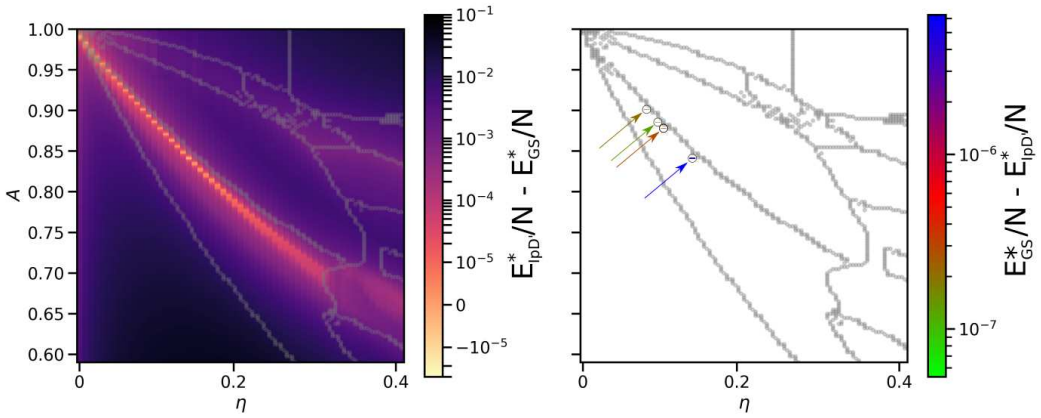


Figure 3.51: Same as Fig. 3.48 but only for the $N = 1444$ and $N = 1452$ particle $\mathrm{IpD^L}$ and $\mathrm{IpD^D}$ approximants depicted in Fig. 3.50.

we present $\Delta E^*_{\mathrm{GS}}/N$ in Fig. 3.52 separately for each available value[71] of η as a function of A. We can see that for selected values of $0.084 \leq \eta \leq 0.141$ very narrow[72] minima of $\Delta E^*_{\mathrm{GS}}/N$ emerge as a function of A.

However, we also need to compare the newly discovered $\mathrm{IpD^L}$ configurations against the rest of the $\mathrm{RT}_l\mathrm{t}_s$ super-tiling family, since the T_2^D, TO_2^D, R_2^D and T_2^M (cf. Fig. 3.49) are serious competitors for the structural ground state in the (η, A)-region where the $\mathrm{IpD^L}$ structure is energetically more favorable than the suggested ground state candidates from literature. We thus investigate the energy offset $\Delta E^*_{\mathrm{GS}'}/N = (E^*_{1444}/N - E^*_{\mathrm{GS}'}/N)$ of the $\mathrm{IpD^L}$ structure to the suggested extended ground state energies $E^*_{\mathrm{GS}'}/N$ which now also includes the $N = 25$ ($N = 100$), $N = 49$ ($N = 196$), $N = 56$ ($N = 224$) and $N = 192$ particle super-tilings TO_2^D, T_2^D, R_2^D and T_2^M. Evaluating the energy offsets of $\mathrm{IpD^L}$ to all of these structures on the same grid as in Fig. 3.52 reveals, that the smaller configurations always win by approximately $\Delta E^*/N \approx 10^{-7}$ to 10^{-6} (which still is remarkably close). Since the minima of $\Delta E^*_{\mathrm{GS}}/N$ in Fig. 3.52 are so narrow we explicitly search for the optimal values of $A_{(\mathrm{opt})}$ which minimize $\Delta E^*_{\mathrm{GS}'}(A;\eta)/N$ for all relevant (fixed) values of η by functional optimization using the *SciPy-optimize-minimize* framework [176]. The optimized value[73], $A_{(\mathrm{opt})}$, as a function of η is presented in the left panel of Fig. 3.53 and we present the associated minimal energy offset $\Delta E^*_{\mathrm{GS}'}(A_{(\mathrm{opt})};\eta)/N$ in the respective right panel. For the matter of completeness,

[71] The literature database structures [62] were generated on a grid in $\eta \times \sqrt{2} = [0., 2.]$ in steps of $\Delta\eta = 0.01/\sqrt{2}$. Extending the database to a finer grid in η was out of scope of this thesis.

[72] The ground state phase diagrams presented in Refs. [62–64] are evaluated on a grid in A with a step size of $\Delta A = 0.01$ which would not allow us to identify the sharp minima of $\Delta E^*_{\mathrm{GS}}/N$ in Fig. 3.52.

[73] Notably, we did not consider the results obtained by REMC (R_2^D and SR_2^D) in the minimization of the extended energy offset $\Delta E^*_{\mathrm{GS}'}/N$ of $\mathrm{IpD^L}$ and $\mathrm{IpD^D}$ presented in Fig. 3.53, but since neither of the two structures seems to be energetically more favorable than the rest of the $\mathrm{RT}_l\mathrm{t}_s$ super-structures (especially T_2^D, T_2^M and TO_2^L, cf. Fig. 3.49) this is also not necessary.

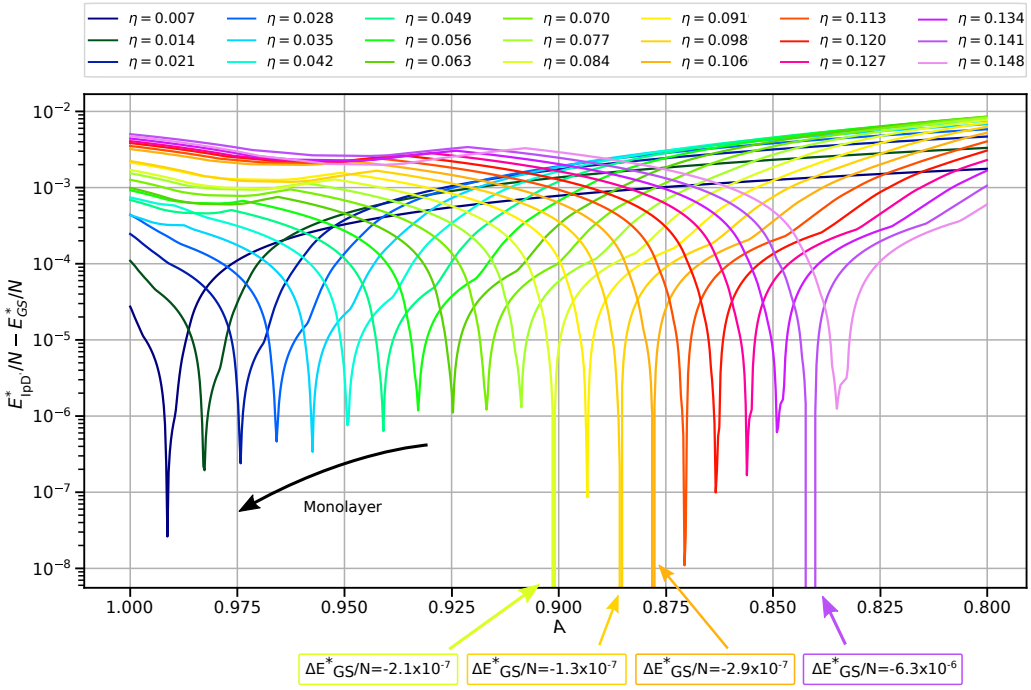


Figure 3.52: The energy offset $\Delta E^*_{\mathrm{GS}}/N = (E^*_{1444}/N - E^*_{\mathrm{GS}}/N)$ of the N =1444 approximant, $\mathrm{IpD}^{\mathrm{L}}$, to the suggested ground state candidates from literature [62–64] for every available value of the plate separation η (color-coded according to the legend at the top) as a function of A. The (η, A)-combinations where $\mathrm{IpD}^{\mathrm{L}}$ is energetically more favorable than the ground state candidates from literature are indicated and the related energy offsets are given (see colored arrows and boxes at the bottom). Note that the small values of $\Delta E^*_{\mathrm{GS}}/N$ for $\eta \leq 0.035$ corresponds to the condensation of the bilayer system into a hexagonal monolayer configuration.

we also minimize $\Delta E^*_{\mathrm{GS}'}/N$ for the $N = 1452$ particle $\mathrm{IpD}^{\mathrm{D}}$ approximant (which is energetically less favorable than $\mathrm{IpD}^{\mathrm{L}}$) and present the results in Fig. 3.53.

We can see that for the $N = 1444$ structure ($\mathrm{IpD}^{\mathrm{L}}$) the minimal value of $\Delta E^*_{\mathrm{GS}'}/N$ is obtained near $\eta \approx 0$, which corresponds to a condensation of layer two and layer one into a hexagonal monolayer. The smallest value of the offset $\Delta E^*_{\mathrm{GS}'}/N$ apart from $\eta \approx 0$ occurs at $\eta_{(\mathrm{opt})} = 0.15/\sqrt{2} \approx 0.1060066$ and $A \approx 0.87784427$, where a local minimum with $\Delta E^*_{\mathrm{GS}'}(A_{(\mathrm{opt})}; \eta_{(\mathrm{opt})})/N \approx 10^{-7}$ can be identified. Thus, $\mathrm{IpD}^{\mathrm{L}}$ is energetically less favorable compared to the smaller, competing $\mathrm{RT}_l\mathrm{t}_s$ super-structures which, respectively, feature only one type of super-tile in the unit cell, see Fig. 3.49.

Given the size of the $N = 1444$ approximant the energy differences of $\Delta E^*_{\mathrm{GS}'}/N \approx 10^{-7}$ appears to be vanishingly small and numerical accuracy – although the here employed Ewald summation method is highly reliably and in principle exact [92] – might become an issue when comparing the energies of structures of such vastly different complexities of the unit cells. At this point we would like to mention that we used the same Ewald cutoff parameters[74] as were also used in Ref. [62] for all structure evaluations in the

[74] Specifically, we use a real space cutoff of $r_c = 15$, a reciprocal space cutoff of $k_c = 10$ and an Ewald parameter of $\alpha = 0.4$ which are defined as unit-less quantities by fixing the volume (or better the

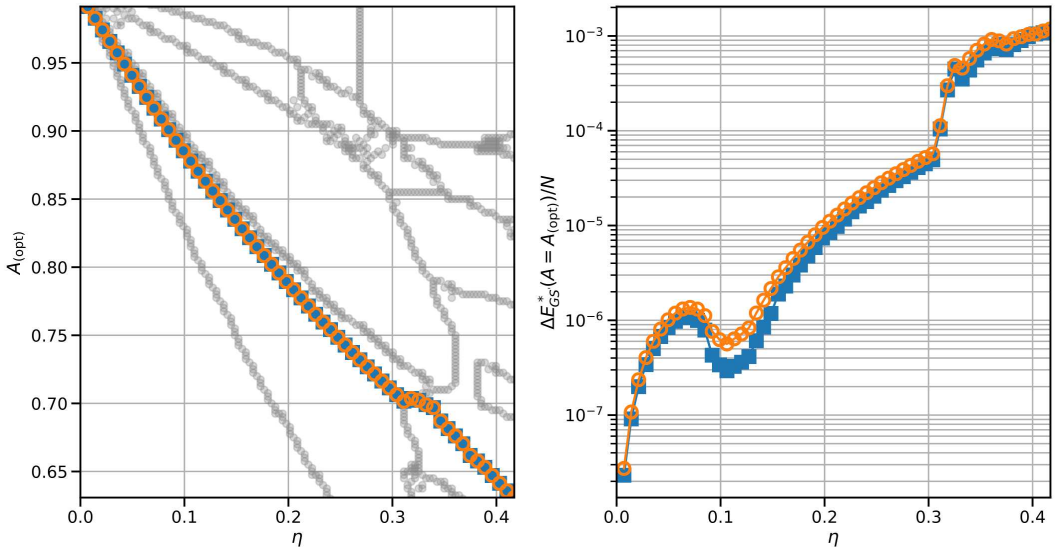


Figure 3.53: The optimal value of $A_{(\mathrm{opt})}$ (left) minimizing $\Delta E^*_{\mathrm{GS}'}(A;\eta)/N$ for different values of η and the corresponding minimal value of $\Delta E^*_{\mathrm{GS}'}(A;\eta)/N$ (right) for the $N = 1444$ (blue) and $N = 1452$ (orange) $\mathrm{IpD}^{\mathrm{L}}$ and $\mathrm{IpD}^{\mathrm{D}}$ approximants (cf. Fig. 3.50). $A_{(\mathrm{opt})}$ is evaluated separately for $\mathrm{IpD}^{\mathrm{L}}$ and $\mathrm{IpD}^{\mathrm{D}}$ by minimizing $\Delta E^*_{GS'}/N(A;\eta))$ with respect to A at fixed values of η.

manuscript. Notably, also the spacing in η is relatively coarse for our purposes of investigating $\mathrm{RT}_l\mathrm{t}_s$ ground state configurations (*i.e.*, the step size in η as used by Ref. [62] is $\Delta\eta = 0.01/\sqrt{2}$). We find that small changes in η can have a crucial impact on the ground state structure formation of the asymmetric Wigner bilayer system. However, refining the η-grid to $\Delta\eta = 0.001/\sqrt{2}$ did not significantly improve the energy offset of $\mathrm{IpD}^{\mathrm{L}}$ and $\mathrm{IpD}^{\mathrm{D}}$ to the other $\mathrm{RT}_l\mathrm{t}_s$ super-structures[75].

Indeed, the minimal value of the energy offset $\Delta E^*_{\mathrm{GS}'}/N$ which we identified in Fig. 3.53 for the $N = 1444$ particle approximant $\eta^\times \approx 0.1060066$ and $A^\times \approx 0.87784427$ is a promising point in the phase diagram to investigate the possible emergence of quasicrystalline order in the asymmetric Wigner bilayer system. In the remaining part of this Subsection we will perform our analysis solely for this point in the phase-diagram to which we refer to as $\mathrm{IpD}^\times = (\eta^\times, A^\times)$.

In an effort to gain a better understanding of the energy landscape at the $\mathrm{IpD}^\times = (\eta^\times, A^\times)$ phase-space point we present in Figs. 3.54 and 3.55 the (extended) energetic offset of all ~ 401 corresponding, previously identified structures in the database of Refs. [62–64] and of all $\mathrm{RT}_l\mathrm{t}_s$ super-tiling structures introduced above (see Figs. 3.43 and 3.49) to the current ground state candidate R_2^D (at $\mathrm{IpD}^\times$). Detailed information about several relevant structures is listed in Table 3.4.

In Fig. 3.54 we present the (extended) energy offset $\Delta E^*_{\mathrm{GS}'}/N$ for all competing structures sorted by the internal energy E^*/N of each configuration in ascending order. In

area) per particle to unity, see also Subsection 2.1.4.

[75] Notably, there is no data for the ground state energies from literature [62–64] for other (relevant) η-values than listed in Fig. 3.52. Performing an evolutionary ground state search for the additional values of η is out of the scope of this thesis.

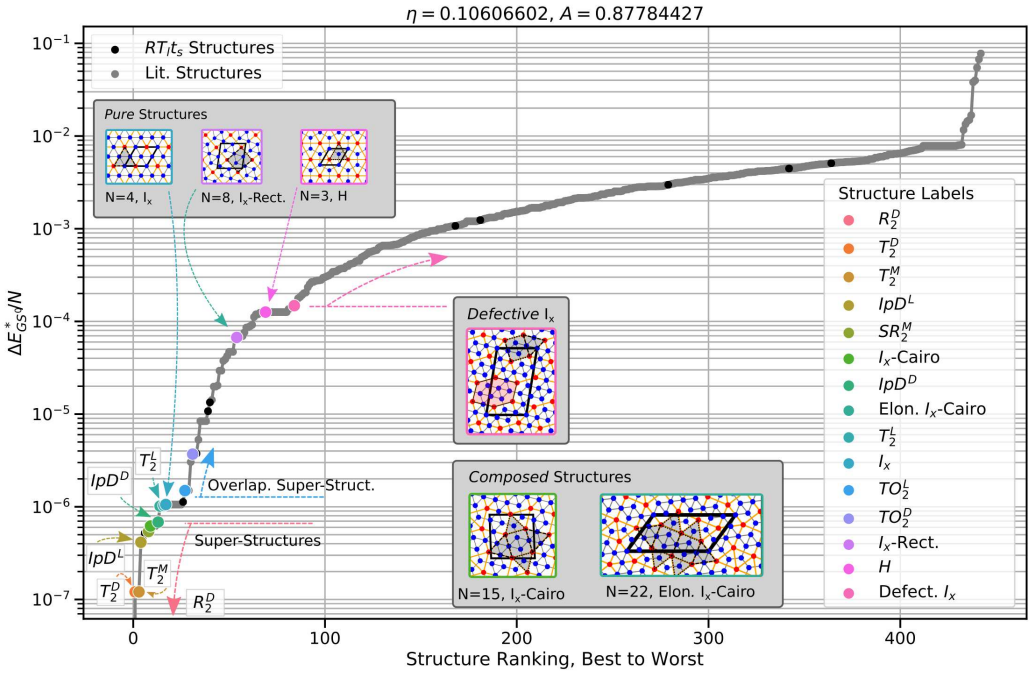


Figure 3.54: (Extended) energy offset $E^*_{\mathrm{GS}'}/N$ of all structures from the database (gray, labeled "Lit. Structures") and all $\mathrm{RT}_l\mathrm{t}_s$ super-structures (black) to the newly discovered R_2^D ground state candidate at $\eta = 0.15/\sqrt{2}$ and $A \approx 0.8778447$ sorted by ascending energy values E^*/N (from "best" to "worst") along the horizontal axis (labeled "structure ranking"). This (η, A)-point corresponds to the smallest encountered value of the extended energy offset of the $N = 1444$ particle $\mathrm{IpD}^{\mathrm{L}}$ structure at (significant) finite values of η (cf. Fig. 3.53). Special structures are indicated via color-coding of the symbols and are partly depicted as insets (see Figs. 3.49 and 3.50 for the larger $\mathrm{RT}_l\mathrm{t}_s$ configurations, numerical details are listed in Table 3.4). The unit cells of the structures are indicated by thick black frames and special tiles and motives related to the $\mathrm{RT}_l\mathrm{t}_s$ decoration scheme or to deviations thereof are emphasized.

Fig. 3.55 we present the (extended) energy offset $\Delta E^*_{\mathrm{GS}'}/N$ for all competing structures as a function of the composition $x = N/N_2$ of the respective structures. The pure structures solely composed of the three basic $\mathrm{RT}_l\mathrm{t}_s$ tiles, *i.e.*, I_x, I_x-rect. and H, are presented as insets in Fig. 3.54 and also selected structures such as I_x-Cairo (and an elongated version thereof which we label elongated I_x-Cairo) and an archetypical defective structure are indicated in this figure. Other visual representations of relevant $\mathrm{RT}_l\mathrm{t}_s$ structures are collected in Figs. 3.49 and 3.50.

The newly suggested ground state candidate at $\mathrm{IpD}^{\times}$ is represented by the R_2^D structure. We can see that the direct competitors to the R_2^D structure are represented by the T_2^D and T_2^M structure (which perform slightly worse than T_2^D by $\Delta E^*/N \approx 10^{-11}$). The next competitor to the R_2^D structure with an energy penalty of $\Delta E^*_{\mathrm{GS}'}/N \approx 10^{-7}$ is indeed the $N = 1444$ particle $\mathrm{RT}_l\mathrm{t}_s$ super-stacked-structure which we labeled $\mathrm{IpD}^{\mathrm{L}}$. Also the I_x-Cairo tiling, the $N = 1452$ particle $\mathrm{IpD}^{\mathrm{D}}$ approximant, and other layer one based $\mathrm{RT}_l\mathrm{t}_s$ transformed structures exhibit only minute energy penalties of $\Delta E^*_{\mathrm{GS}'}/N < 10^{-6}$ (see Table 3.4 for details). The above introduced elongated I_x-Cairo

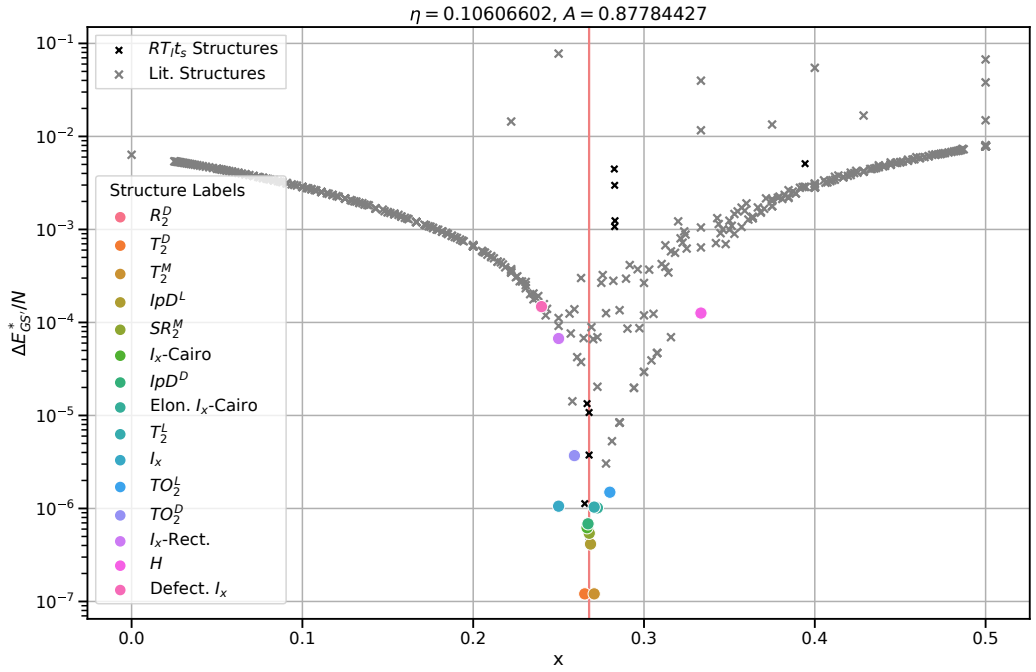


Figure 3.55: Similar to Fig. 3.54 but here $\Delta E^*_{\mathrm{GS}'}/N$ is shown as a function of the composition $x = N_2/N$. Black and gray crosses mark unlabeled $\mathrm{RT}_l\mathrm{t}_s$ structures and solutions from the evolutionary algorithm from literature [62–64] (labeled "Lit. Structures") for $\eta = 0.15/\sqrt{2}$ and $A \approx 0.87784427$, respectively. Colored circles mark the same structures that are also emphasized in Fig. 3.54. The vertical red line indicates the composition of the R_2^D ground state candidate.

structure, the pure I_x structure and layer two based $\mathrm{RT}_l\mathrm{t}_s$ transformed structures exhibit slightly larger energy penalties of $\Delta E^*_{\mathrm{GS}'}/N \approx 10^{-6}$. A configuration purely composed of $\mathrm{RT}_l\mathrm{t}_s$ rectangles (here labeled I_x-rect.) as well as the honeycomb phase H already display a larger energy penalty of $\Delta E^*_{\mathrm{GS}'}/N \approx 10^{-4}$ but are still composed of the basic rectangle and small triangle tiles of the $\mathrm{RT}_l\mathrm{t}_s$ family. For energetically less favorable structures than the honeycomb phase defects start to emerge where the symmetry of the $\mathrm{RT}_l\mathrm{t}_s$ family is lost at a cost of $\Delta E^*_{\mathrm{GS}'}/N > 10^{-4}$.

From Fig. 3.55 we can learn that at the $\mathrm{IpD}^\times = (\eta^\times, A^\times)$ phase-space point the system appears to prefer a very particular composition of $x = N_2/N = 15/56 \approx 0.268$ which is realized by R_2^D. The three pure structures, I_x, I_x-rect. and H, mark in Fig. 3.55 the numerical boundaries of the allowed compositions for structures of the $\mathrm{RT}_l\mathrm{t}_s$ family, that is $x = 1/4$ for I_x and I_x-rect. and $x = 1/3$ for H. Compositions of mixtures of the $\mathrm{RT}_l\mathrm{t}_s$ tiles must lie between these boundaries and the sharp minimum of the (extended) energy offset around $x \approx 0.268$ can be related to the compositions of the $\mathrm{RT}_l\mathrm{t}_s$ tiles which compose a pseudo-dodecagon in layer two: one such pseudo-dodecagon consists of six large triangles and six rectangles, both with a composition of $x = 1/4$, and six small triangles with $x = 1/3$, *i.e.*, 18 tiles in total. The compositions of the R_2^D, T_2^D and T_2^M structures but also of IpD^L and IpD^D are almost perfectly framed by the composition of such a "pseudo-dodecagonal mixture", $x_{\mathrm{mix.}} = (1/4 + 1/4 + 1/3)/3 =$

Rank	Structure Labels	N	N1	N2	x	E^*/N	$E^*_{\mathrm{GS}'}/N$
0	R_2^D	56	41	15	0.268	-1.76089838	–
1	T_2^D	49	36	13	0.265	-1.76089826	1.2×10^{-7}
3	T_2^M	192	140	52	0.271	-1.76089826	1.2×10^{-7}
4	$\mathrm{IpD^L}$	1444	1056	388	0.269	-1.76089797	4.1×10^{-7}
8	SR_2^M	209	153	56	0.268	-1.76089784	5.4×10^{-7}
9	I_x-Cairo	15	11	4	$0.2\dot{6}$	-1.76089776	6.2×10^{-7}
13	$\mathrm{IpD^D}$	1452	1064	388	0.267	-1.76089770	6.8×10^{-7}
14	Elongated I_x-Cairo	22	16	6	0.273	-1.76089737	1.02×10^{-6}
16	T_2^L	48	35	13	0.271	-1.76089735	1.03×10^{-6}
17	I_x	4	3	1	0.25	-1.76089733	1.06×10^{-6}
27	TO_2^L	25	18	7	0.28	-1.76089689	1.5×10^{-6}
31	TO_2^D	27	20	7	0.259	-1.76089468	3.7×10^{-6}
54	I_x-Rect.	8	6	2	0.25	-1.76083119	6.7×10^{-5}
69	H	3	2	1	$0.\dot{3}$	-1.76077227	1.26×10^{-4}
84	Defective I_x	25	19	6	0.24	-1.76075042	1.48×10^{-4}

Table 3.4: Numerical details of labeled structures in Figs. 3.54 and 3.55 for the $\mathrm{IpD}^\times = (\eta^\times, A^\times)$ phase-space point with $\eta^\times = 0.15/\sqrt{2}$ and $A^\times \approx 0.87784427$; structures T_2^D and T_2^M have a relative energy offset of 2.28×10^{-11} per particle. The columns $N = N_1 + N_2$, N_1 and N_2 list the total number of particles, the number of particles in layer one and the number of particles in layer two in the unit cell of the respective structures, $x = N_2/N$ is the corresponding composition.

$0.2\dot{7}$, of the pure structures. Also the compositions of the other structures which are serious competitors to the ground state are very close to $x_{\mathrm{mix.}}$, see Table 3.4 for details. However, some other configurations are close to the "optimal" composition but obtain energies that are significantly larger compared to the configurations R_2^D, T_2^D and T_2^M, or even $\mathrm{IpD^L}$ and $\mathrm{IpD^D}$; such large energy penalties are related to unfavorable local environments of the structures (for the $\mathrm{IpD}^\times$ phase-space point, in particular). We would like to stress that other possible super-structures similar to $\mathrm{IpD^L}$ or $\mathrm{IpD^D}$ which potentially only host dense triangular vertices would also be reasonable candidates for the structural ground state at the $\mathrm{IpD}^\times$ phase-space point.

In general, the $\mathrm{RT}_l\mathrm{t}_s$ structures discussed in this Subsection may all be interpreted as realizations of an ensemble of random $\mathrm{RT}_l\mathrm{t}_s$ tilings with minute differences in energy of the order $\Delta E^*_{\mathrm{GS}'} \sim 10^{-7}$ to $\lesssim 10^{-4}$. Defects in $\mathrm{RT}_l\mathrm{t}_s$ configurations are penalized with $\Delta E^*_{GS'}/N > 10^{-4}$ and are energetically much less favorable especially compared to $\mathrm{RT}_l\mathrm{t}_s$ super-tilings. Furthermore, we show that there is also an ensemble of $\mathrm{RT}_l\mathrm{t}_s$ super-tilings composed of super-rectangular and super-triangular tiles which obtain a vanishingly small energy offset of $\Delta E^*_{\mathrm{GS}'} \sim 10^{-7}$ to 10^{-6}. This suggests that at small finite temperatures of the order of $1/\beta^* \sim 10^{-7}$ to 10^{-6} the most probable state the system will attain will be a random $\mathrm{RT}_l\mathrm{t}_s$ super-tiling configuration. However, such a "random-tiling" scenario [167, 168] which stabilizes the highest-symmetry phase quasicrystal via entropic contributions of arbitrary random $\mathrm{RT}_l\mathrm{t}_s$ (super-)tilings does not seem physically plausible: the considerable size of the super-tiles would presumably make such entropic contributions to the finite-temperature stability marginal.

3.1.7 Miscellaneous

In this Subsection we discuss miscellaneous topics encountered during our investigations of the asymmetric Wigner bilayer system. We present the current state of our research related to pentagonal and V_x-based ground state candidate solutions of the asymmetric Wigner bilayer system and we provide relations between some of the ground state candidate structures to three-dimensional phases from intermetallics and to adatom based surface structures.

Pentagonal Structures: Parallels to Intermetallics

There is a whole family of zero temperature ground state structures of the asymmetric Wigner bilayer system which is dominated by (distorted) pentagons, filling as much space as possible in the densely-populated layer one (cf. green highlighted area in Fig. 3.1(d)). Amongst this wider range of structural families featuring pentagonal motives in layer one are, for instance, the P_1, P_2, P_3 and S_2 configurations known from literature [62–64] (cf. Fig. 3.2). Depending on the actual value of η and A the particles of layer two arrange themselves in some sort of uniform configuration similar to triangular, square or square-triangle tilings but even layer one structures that are intermediate between triangular and square tilings are possible. Since (regular) pentagons cannot fill the two-dimensional plane without gaps or overlaps, the energetically most favorable structures which feature pentagonal tiles either complement these motifs with triangles or deform pentagons to the point that they are more similar to hexagons (or a square plus two more triangles).

In Fig. 3.56 we present the structural ground state families identified by the $\mathbf{k}_{32}^*$-clustering procedure[76] put forward in Subsection 3.1.4 which qualify as members of the wider range of (layer one) pentagonal structures (see also Fig. 3.18(b)). The situation here seems to be even more complicated and versatile as compared to the DI and RT_lt_s structures discussed in the previous Subsections 3.1.5 and 3.1.6 and certainly requires a comprehensive study on its own. Hence, we refrain here from discussing every structural family identified by the $\mathbf{k}_{32}^*$-clustering algorithm in detail, but will emphasize instead special features of the respective configurations whenever it is necessary or helpful for the discussion in this Subsection.

We first investigate a structure inspired by Pt_7Zn_{12} [291], a three-dimensional structure well-known from intermetallics. Pt_7Zn_{12} can be related to a bilayer configuration whose vertices in layer one form as many pentagons as possible such that particles in layer two, which are placed at the projected geometric centers of the pentagons of layer one, form

[76] The $\mathbf{k}_{32}^*$-clustering approach introduced in Subsection 3.1.4 relies on unsupervised k-means clustering (see Subsection 2.4.2) of the database of structural ground state candidate configurations of the asymmetric Wigner bilayer system identified in Refs. [62–64] (or better of the data set of order parameters of the structural database, which we additionally subjected to principal component analysis, see Subsection 2.4.1). Via information-theoretical measures based on the adjusted mutual information score between different clustering results (see Subsection 2.4.5), we could identify a reasonable "guess" for a total number of K^*=32 different structural families in the structural database; the asterisk emphasizes that we manually assigned all uniquely defined hexagonal monolayer configurations of the structural database as a family and excluded the corresponding data points from the k-means clustering procedure of the remaining data.

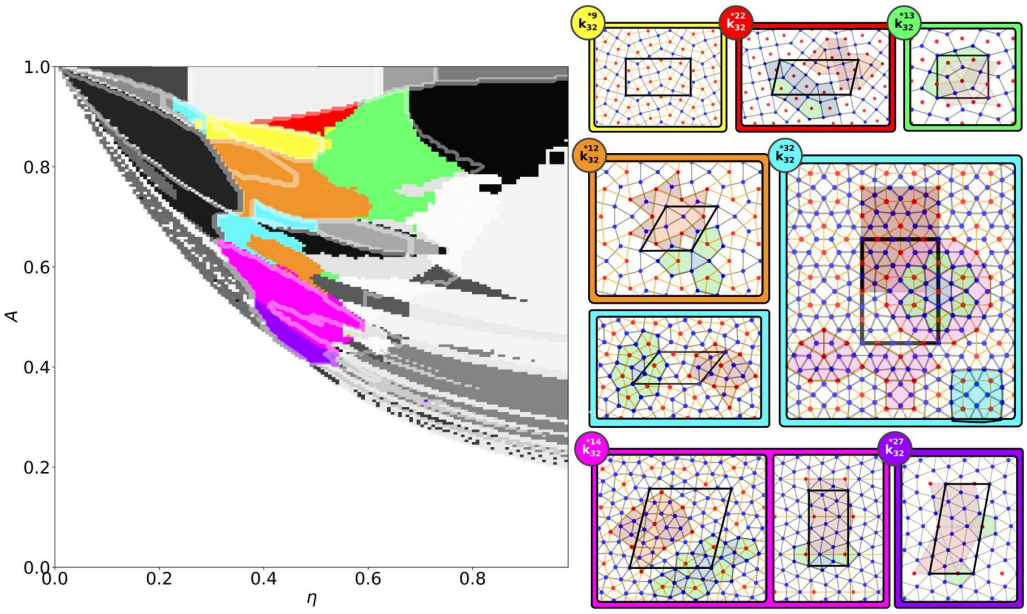


Figure 3.56: Same as Figs. 3.23 and 3.39 but for structural families identified by the $\mathbf{k}^*_{32}$-clustering procedure (cf. Subsection 3.1.4) which feature pentagonal motives in layer one. In the left panel the phase boundaries from literature [62–64] are emphasized by faint white lines, the gray scales indicate the regions of stability of all $c = 1, \ldots, 32$ structural families, $\mathbf{k}^{*c}_{32}$. The particular $\mathbf{k}^{*c}_{32}$-families related to pentagonal structures are highlighted with bright colors and certain representative configurations of the families that represent the ground state candidates of the system at certain combinations of the system parameters are depicted in the right (cf. $\mathbf{k}^{*c}_{32}$-labels); the structures are color-coded according to the color-scheme in the (η, A)-panel. Faintly colored areas in the right panels emphasize special motives and tiles of the structures and thick black frames indicate the respective unit cells. See text for a discussion of (some of) the highlighted structures. See Appendix A.1.2 for details on characteristic values and boundaries of the order parameters and principal components for the $\mathbf{k}^*_{32}$-clustering families of structures.

an approximate triangular lattice there (cf. inset in the left panel of Fig. 3.57). In the left panel of Fig. 3.57 we present the energy offset ΔE^*_{GS} between the Pt_7Zn_{12}-inspired bilayer configuration and the suggested ground state solutions from literature [62–64] in the (η, A)-plane. Remarkably, such a configuration – after relaxation in the relevant (η, A)-region – directly represents a ground state candidate solution of the system in the proximity of $\eta \approx 0.396$ and $A \approx 0.775$ where $\Delta E^*_{\mathrm{GS}} \approx 0$.

In the right panel of Fig. 3.57 we investigate a variant of the S_2 structure which is inspired by yet another structure well-known from intermetallics, namely the *BRuSc*-structure [272] (see inset in the right panel of Fig. 3.57). In the right panel of Fig. 3.57 we present the energy offset of the *BRuSc*-inspired bilayer configuration (after relaxation) to the suggested ground state energies from literature [62–64] in the (η, A)-plane. The smallest energy offset is indeed obtained in the vicinity of the S_2 phase at smaller values of A (see bright yellow area in the right panel of Fig. 3.57) but only evaluates to $\Delta E^*_{\mathrm{GS}} \approx 10^{-4}$. The *BRuSc*-inspired structure is thus no competitor to the ground states of the asymmetric Wigner bilayer system.

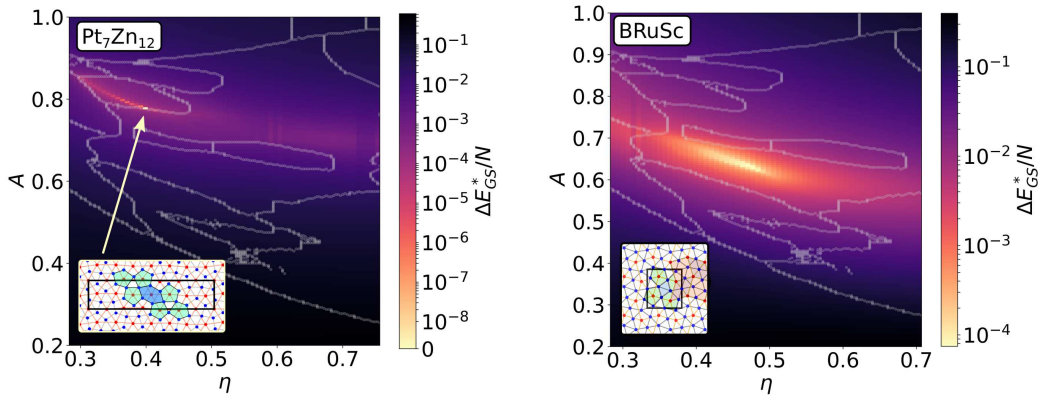


Figure 3.57: Energy offset ΔE^*_{GS} of two bilayer configurations inspired by the three-dimensional Pt_7Zn_{12}-structure (left, see inset) and the *BRuSc*-structure (right, see inset), both well-known from intermetallics, to the ground state candidate solutions of the asymmetric Wigner bilayer system suggested by Refs. [62–64] in the (η, A)-plane; the Pt_7Zn_{12}- and *BRuSc*-inspired bilayer configurations were relaxed at the respective η and A values. The Pt_7Zn_{12}-inspired structure indeed represents a ground state candidate of the system near $\eta \approx 0.396$ and $A \approx 0.775$ (indicated by arrow) while the *BRuSc*-inspired structure is no serious competition at any values of η and A (see color-bar). The phase boundaries from Refs. [62–64] are emphasized by gray lines.

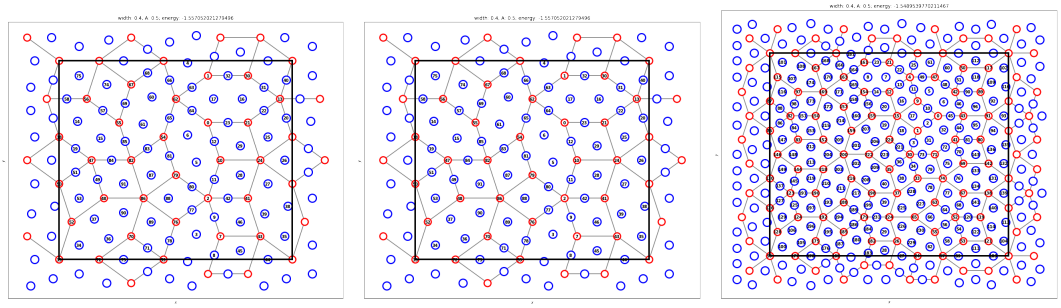


Figure 3.58: Bilayer configurations which represent approximants to a decagonal quasicrystal in layer two related to the S_2 phase and the *BRuSc*-inspired structure (left to right: T_{29}, T_{47}, T_{76}). The structures are unstable although some pentagonal motives may survive relaxation (relaxed structures not shown here). Particle indices in the structures are indicated by numbers.

Both, the S_2 and the *BRuSc*-inspired structure have a remarkable geometry: local pentagonal symmetry in layer one coexists with motives that could potentially give rise to global *decagonal* symmetry, *i.e.*, a decagonal quasicrystal in layer two. We also started investigating decagonal quasicrystalline approximants related to S_2 and *BRuSc* – featuring their own inflation rules [272] – through structures which we labeled T_{29}, T_{47}, T_{76} (cf. Fig. 3.58) and we compared them against current ground state candidates of the system known from literature [62–64]. However, none of these three configurations are stable, although some local pentagonal motives in layer two – so-called stars or flowers – may survive a relaxation procedure (relaxed structures are not shown

here). Thus, neither of the T_{29}, T_{47}, T_{76}-structures represent serious competitors to existing ground state structures of the system. To this end, further investigations are required: to stabilize these structures other decoration schemes of the tiles in layer two by particles in layer one seem necessary.

Pentagonal Holes in the Hexagonal Monolayer: a Relation to $\mathrm{RT}_l\mathrm{t}_s$

As can be seen in Fig. 3.56, the $\mathbf{k}_{32}^{*}$-clustering algorithm identified S_2 as part of the $\mathbf{k}_{32}^{*32}$ family whose smallest (*i.e.*, $N = 12$) and whose largest structure (*i.e.*, N=40) are depicted in Fig. 3.56 and emphasized with cyan color. Also the $\mathrm{TPH_1D_2}$ structure discussed in Subsection 3.1.6 and depicted in Fig. 3.49 classifies as a member of this family. Interestingly, we may also interpret the $\mathbf{k}_{32}^{*32}$ family to feature (distorted) rectangles and triangles in layer two and, furthermore, we can relate $\mathbf{k}_{32}^{*32}$ to the rectangle-large–triangle-small–triangle ($\mathrm{RT}_l\mathrm{t}_s$) decoration scheme discussed above in Subsection 3.1.6:

S_2, for instance, can be considered as being composed of two distorted, $\mathrm{RT}_l\mathrm{t}_s$-decorated rectangles (related by a $\approx 60°$ rotation) and four distorted triangles which are now purely center decorated. Thereby, the splitting into the two types of triangles related to the $\mathrm{RT}_l\mathrm{t}_s$ structures, *i.e.*, a large T_l and a small t_s triangle, is avoided; all triangles are basically of one kind (*i.e.*, t_s triangles), but are slightly distorted.

Furthermore, the $\mathrm{TPH_1D_2}$ structure (see Subsection 3.1.6) and the N=40 realization of the $\mathbf{k}_{32}^{*32}$ family (see Fig. 3.56) are composed of (close-to-square) rectangles and (close-to-equilateral) triangles in layer two which are purely mid-edge decorated (*i.e.*, T_l triangles). Via relaxation in the respective stable (η, A)-region in the proximity of $\eta \approx 0.403$ and $A \approx 0.66$ (cf. Subsection 3.1.6) these structures obtain small deviations of the purely mid-edge decorated particle positions compared to perfect $\mathrm{RT}_l\mathrm{t}_s$ structures. In that way, the aspect ratio of every rectangle approaches unity, $s/l \approx 0.99$, for both the $\mathrm{TPH_1D_2}$ and the N=40 particle $\mathbf{k}_{32}^{*32}$ configurations such that almost perfect squares are realized.

We would like to stress that the $\mathrm{TPH_1D_2}$ structure exhibits remarkable geometry in which triangular, pentagonal and hexagonal tiles in layer one coexists with (approximate) dodecagons in layer two which are composed of approximate squares and equilateral triangles. Notably, the pentagonal motives enter the situation through snub-rectangular vertices in layer two (cf. Fig. 3.42): a snub-rectangular vertex is formed by two squares and three equilateral triangles (*i.e.*, five tiles) which are arranged as rectangle-triangle-triangle-rectangle-triangle. A pure mid-edge decoration of the edges of the neighboring tiles forming this vertex (as well as a pure center decoration of the corresponding tiles) leads automatically to a distorted pentagonal motive in layer one [62]. Also this class of structures requires further investigation.

Through Fig. 3.56 we see that when we decrease the value of A the $\mathbf{k}_{32}^{*32}$ family (cyan) transforms into the $\mathbf{k}_{32}^{*14}$ family (magenta) and eventually into the $\mathbf{k}_{32}^{*27}$ family (purple) which are related to phase P_3. For $\mathbf{k}_{32}^{*14}$ the structure formation in layer two is quite similar to S_2: we identify distorted rectangles and triangles as can be seen by the emphasized layer two motives of the left $\mathbf{k}_{32}^{*14}$ structure inset shown in Fig. 3.56 (which

is, in fact, a prototype configuration of phase P_3). The corresponding structure in layer one resembles a defective hexagonal lattice with pentagonal *holes* around layer two vertices, a feature which is also valid for the $\mathbf{k}_{32}^{*27}$ family[77] (highlighted by green areas in the respective inset of Fig. 3.56).

Since reducing A lowers the particle density in layer two and (consequently) increases the particle density in layer one, the transition from S_2 to P_3 is not completely surprising. However, there are notable parallels of the $\mathbf{k}_{32}^{*32}$, the $\mathbf{k}_{32}^{*14}$ and the $\mathbf{k}_{32}^{*27}$ families to the RT_lt_s modulation of the hexagonal monolayer discussed in Subsection 3.1.6.2 (cf. Fig. 3.41): a decreasing value of A is accompanied by sparser population in layer two whose vertices form ever larger (distorted) triangular, rectangular or even rhombic-like tiles (cf. $\mathbf{k}_{32}^{*14}$ and $\mathbf{k}_{32}^{*27}$ from left to right in Fig. 3.56). In contrast to the RT_lt_s modulations of the hexagonal monolayer via vertical displacements of particles from one to the other layer, here, vertices in layer two give rise to pentagonal *holes* in layer one and thereby locally break the hexagonal symmetry. In the limit of very sparse population of layer two there are many discrete possibilities to form tiles which may even give rise to (approximate) local pentagonal symmetry in layer two[78]. Closely studying the decoration scheme of the $\mathbf{k}_{32}^{*14}$ and $\mathbf{k}_{32}^{*27}$ families could help to stabilize the structures depicted in Fig. 3.58. Thus, an interesting, open issue is to investigate the possible emergence of global decagonal symmetry, *i.e.*, decagonal quasicrystals in the asymmetric Wigner bilayer system based on a modulated hexagonal lattice in the proximity of the I–P_3 phase boundary.

More Parallels to Intermetallic Phases: *Yb*-Based Structures, Penta-Graphene and Shield-Tilings in the Low Coupling Regime of the V_x-region

The $\mathbf{k}_{32}^{*12}$ family of configurations shown in Fig. 3.56 (emphasized in orange) can be related to the P_2 phase (which we use synonymously for $\mathbf{k}_{32}^{*12}$ from now on). Taking a closer look at the lattice in layer two reveals an emerging shield tile (cf. Fig. 3.22) with six adjacent equilateral triangles. Each vertex in layer two is surrounded with a pentagonal hole in layer one.

Interestingly, the P_2 structure can be found in three-dimensional *YbInPd*-based systems [280]: inspired by the so-called *HP9* structure [280] we define the $HP9_{InPd}^{Yb}$ bilayer configuration where *Yb* atoms of the *HP9* structure are used to form the lattice in layer two and *In* and *Pd*-atoms form the pentagonal holes in layer one[79]. In the left panel of Fig. 3.59 we present the energy offset of the $HP9_{InPd}^{Yb}$ configuration to the current ground state candidates of the asymmetric Wigner bilayer system suggested by Refs [62–64]. Indeed, in the parameter range where P_2 is stable this structure represents a ground state candidate of the asymmetric Wigner bilayer system, *i.e.*, $\Delta E_{GS}^{*}/N = 0$, as indicated by the bright yellow area in the left panel of Fig. 3.59.

[77] Different structures of the $\mathbf{k}_{32}^{*14}$ and the $\mathbf{k}_{32}^{*27}$ families should therefore be characterized by strong signals in $\Psi_5^{(2,4)}$ and $\Psi_6^{(1)}$ as can indeed be seen by the bright, whitish/green signal in bottom left panel in Fig. 3.3 in the P_3 region, see also Fig. A.6.

[78] The distorted rhombi featured in the right configuration of the $\mathbf{k}_{32}^{*14}$ family in Fig. 3.56 exhibit a small opening angle of $\alpha_{\tilde{P}} \approx 76°$ which is close to perfect pentagonal symmetry of $\alpha_P = 360°/5 = 72°$.

[79] Without loss of generality we assume layer two to sit above layer one, hence the super– and sub-script notation of $HP9_{InPd}^{Yb}$.

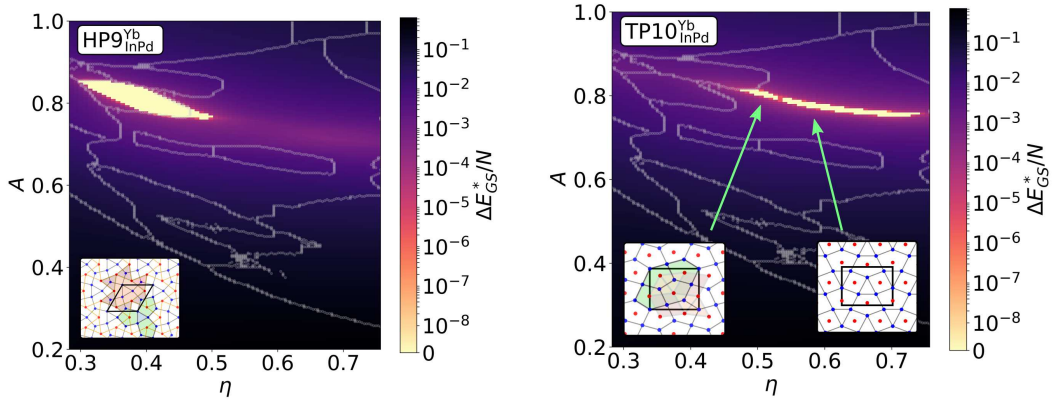


Figure 3.59: Same as Fig. 3.57 but for the S_2P_1 structure (left inset) relaxed at all relevant (η, A)-pairs before evaluating the respective energy offset to the ground state candidates of the asymmetric Wigner bilayer system suggested by Refs. [62–64]. The arrow corresponding to the S_2P_1 structure marks the S_2P_1 ground state region. Note the relation to the Cairo tiling type of Voronoi diagram of the I_x-Cairo structure shown in Fig. 3.40(b). The right inset shows the S_2P_1 structure after relaxation at larger values of η (as compared to the respective ground state region) giving rise to a herringbone pattern in layer one; the respective arrow indicates the associated ground-state region of the depicted structure in the phase diagram.

Another quite interesting structural family is given by $\mathbf{k}_{32}^{*13}$ (cf. green emphasized top right panel in Fig. 3.56). Here, the tiling in layer two is similar to S_2 but with a distorted rectangle-triangle tiling that is even more distorted[80] as for S_2. The interesting part about this structure, which we label S_2P_1, is the tiling in layer one which is purely composed of distorted pentagons such that a (sheared) Cairo type of tiling [255] emerges in layer one which reassembles the structure of penta-graphene [292]. This type of tiling can be generated by placing particles in layer one at the geometric center of both the distorted triangles and rectangles in layer two of the S_2 configuration. Another physical realization of S_2P_1 can again be found in *YbInPd*-based systems through the so-called TP10 structure [280]: again, Yb atoms represent the layer two vertices of a corresponding bilayer structure and In and Pd atoms form the respective lattice in layer one. We label the respective bilayer configuration $TP10^{Yb}_{Pd}$ and present in the right panel of Fig. 3.59 the energy offset of the S_2P_1 (or equivalently $TP10^{Yb}_{Pd}$) structure to the ground state candidates of the asymmetric Wigner bilayer system suggested by Refs [62–64]. We see that this structure represents the ground state candidate of the asymmetric Wigner bilayer system in the proximity of $\eta \approx 0.5$ and $A \approx 0.81$. For larger values in η the structure deforms – via relaxation – such that layer one exhibits a herringbone pattern [255] (see right insets in the right panel of Fig. 3.59) which represents the ground state candidate structure in an adjacent region to the S_2P_1 phase in the phase diagram.

To fully complete the analysis of the K^*=32-means clustering put forward in Sub-

[80] In fact, layer two of the $\mathbf{k}_{32}^{*13}$ family more closely resembles a distorted triangular lattice.

section 3.1.4 we conclude this Subsection by showing in Fig. 3.60 the associated $\mathbf{k}^*_{32}$-clustering families of the zero temperature ground state structures of the asymmetric Wigner bilayer system suggested by Refs. [62–64] in the V_x region (related to hexagonal bilayer structures) of the parameter space (see also Fig. 3.18(d)). Similar to Figs. 3.23, 3.39 and 3.56 we can see, that several new structural details can be differentiated as compared to earlier investigations [62–64] which we briefly discuss below.

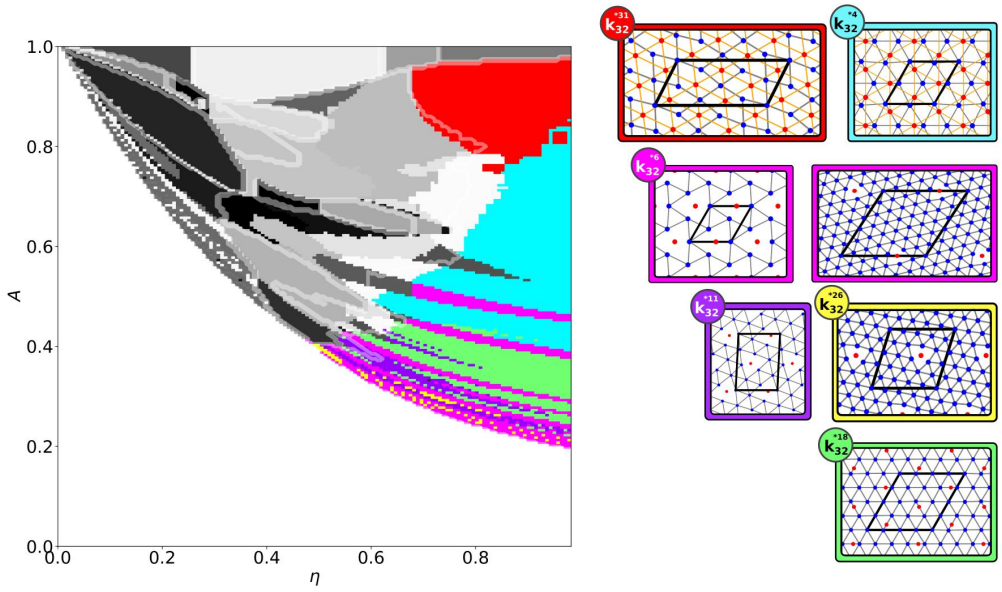


Figure 3.60: Same as Figs. 3.23, 3.39 and 3.56 but for structural families related to the V_x phases identified by $\mathbf{k}^*_{32}$-clustering (cf. Subsection 3.1.4) based on all structures considered in Refs. [62–64]. See text for a discussion of (some of) the highlighted structures. See Appendix A.1.2 for details on characteristic values and boundaries of the order parameters and principal components for the $\mathbf{k}^*_{32}$-clustering families of structures.

The structural family $\mathbf{k}^{*31}_{32}$ (emphasized in red in Fig. 3.60) exhibits a herringbone pattern in layer one and an approximate hexagonal lattice in layer two. The $\mathbf{k}^{*4}_{32}$ (emphasized in cyan in Fig. 3.60) and the $\mathbf{k}^{*18}_{32}$ families (green) show hexagonal bilayers with different particle densities in layer two. Interestingly, we can identify through the structural families $\mathbf{k}^{*6}_{32}$ (magenta), $\mathbf{k}^{*11}_{32}$ (purple) and $\mathbf{k}^{*26}_{32}$ (yellow) different types of shield-tilings in layer one: vertices in layer two force the hexagonal layer one to open up shield-like holes whose densities are determined by the tiling length of the hexagonal layer two.

In Fig. 3.60 we present two archetypical configurations of the $\mathbf{k}^{*6}_{32}$, one with N=4 and another one with N=40, which represent the ground state candidates of the system at different values of system parameters. Both configurations display close-to-triangle shield-like tiles in layer one and a hexagonal lattice in layer two (with layer two vertices centered at the projected geometric centers of the shield-like tiles of layer one). However, the density of the shield-like tiles for the N=4 structure is much larger compared to the N=40 configuration. The two structures certainly exhibit similar features but may also be considered as different types of structures, unless polymorphs of this family

with different densities of the shield-like tiles can be associated to the values of the composition, x, similarly as previously suggested for the V_x structures [62–64]. Moreover, as can be seen from Figs. 3.23 and 3.60 the $\mathbf{k}_{32}^{*23}$ family and the N=4 particle variant of the $\mathbf{k}_{32}^{*6}$ family topologically describe the same structure, only the angles of the shield-like tiles vary as a function of η (from close-to-square for $\eta \approx 0.45$ and close-to-triangle for $\eta \approx 1$). This clearly shows, that the predictions of the clustering approach discussed in Subsection 3.1.4 need to be considered with caution (cf. also family $\mathbf{k}_{32}^{*32}$ depicted in Fig. 3.56 which comprises the S_2 and the TPH_1D_2 structures). Nevertheless, the clustering approach introduced in this thesis allowed us to efficiently identify particularly interesting regions in the parameter space of the asymmetric Wigner bilayer system and, in general, represents a valuable tool for studying complex phase diagrams.

Notably, also the N=4 shield-tiling structure of the $\mathbf{k}_{32}^{*6}$ and $\mathbf{k}_{32}^{*23}$ families has a physical realization in the above mentioned HP9 structure [280]: by inverting the assignment of the Yb atoms from layer two (cf. left panel of Fig. 3.59) to layer one and considering all positions of In atoms located in the geometric center of the shield-like arranged Yb atoms as particle locations in layer two (hence, the labeling $HP9^{In*}_{Yb}$) we can construct the $N = 4$ shield-tiling of the $\mathbf{k}_{32}^{*6}$ and the $\mathbf{k}_{32}^{*23}$ families. Eventually, we present in Fig. 3.61 the respective energy offset of this configuration (after relaxation) to the ground state candidates of the asymmetric Wigner bilayer system suggested by Refs. [62–64] in the (η, A)-plane. We can identify an extended parameter range of $0.5 \lesssim \eta \lesssim 0.72$ and $A \approx 0.6$ where this kind of shield-tiling forms a ground state candidate of the asymmetric Wigner bilayer system.

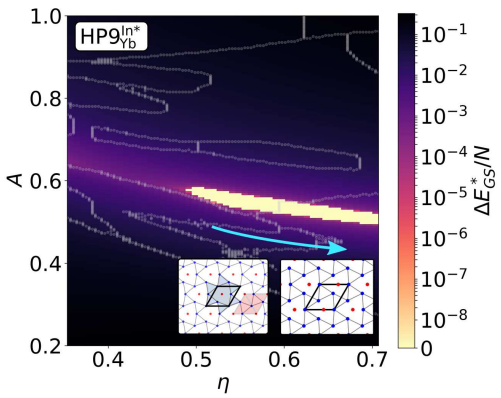


Figure 3.61: Same as Fig. 3.57 but for the N=4 shield-tiling of the $\mathbf{k}_{32}^{*6}$ family (see insets). The bright yellow area marks the associated region in parameter space where this configuration represents the ground state candidate of the asymmetric Wigner bilayer system and the cyan arrow indicates the successive (continuous) deformation (via relaxation) with increasing η of an initially almost perfect shield tile (left inset, cf. Fig. 3.22) to a nearly triangle shaped shield tile variant (see right inset).

3.1.8 Conclusions and Outlook

In Ref. [62] an evolutionary algorithm was employed to systematically identify ground state configurations of the Wigner bilayer system and has proven to be a highly reliable and very general framework for identifying trends in the structure formation process of this physical system. The analysis in Ref. [62] provides a solid foundation for further, in-depth studies of structure formation processes of very specific families of structures, *i.e.*, for specialized problems which need to be treated with great precaution. In this thesis, we investigated the asymmetric Wigner bilayer system for possible zero temperature quasicrystalline ground state solutions related to the snub-square (S_1) structure and the trihexagonal phase (I_x), a fruitful yet delicate endeavor. In general, specialized problems usually require tailored approaches which are capable of efficiently exploring the problem specific, yet usually still considerably large configuration space. We therefore employ specifically implemented simulated annealing and replica exchange Monte Carlo methods which rely on specialized update moves related to the geometry of the investigated sub-families of structures.

Revisiting the Complex Phase Diagram with Unsupervised Learning

We propose a new scheme based on k-means clustering [239–242] of order parameters of structural data of the asymmetric Wigner bilayer system [62–64] in order to classify and group families of related structures. Indeed the different phases in Refs. [62–64] could be resolved by this very simple unsupervised machine learning approach which, moreover, provides us with a deeper insights into the structural-landscape of ground states of the asymmetric Wigner bilayer system. Especially in the proximity of the S_1, pentagonal, I_x and the V_x ground state regions, but also along phase boundaries in general we could identify an even richer phase-behaviour as previously reported [62–64]. This clustering tool of structural data further allowed us to identify sweet spots in the phase-diagram which may give rise to structures with dodecagonal or decagonal symmetries.

Towards Dodecagonal Order: Metastable Super-Structures

We systematically investigate the snub-square (S_1) region of the phase diagram for possible new square-triangle based ground states which successively approximate a dodecagonal quasicrystal. For that purpose we repetitively employ Stampfli-inflation [252] to square-triangle tilings and perform specifically implemented simulated annealing and replica exchange Monte Carlo simulations based on so-called zipper-moves [254] to explore the configuration space of inflated square-triangle tilings for optimal solutions to the structural ground state. By applying this procedure to different configuration spaces of quasicrystalline approximants with different (unit cell) complexities a cascade of ground states emerges – one ground state for every configuration space determined by the specific choice of the inflation – which all exhibit self-similar super-square-triangle-tilings on multiple length scales.

However, this family of structures seems to be energetically less favorable compared to S_1 – or other global ground state candidates of the asymmetric Wigner bilayer system in general. Especially the large hexagonal tiles – in addition to an otherwise perfect square-triangle tiling – emerging in layer two of the related self-similar super-structures have proven to cause energy penalties as compared to the uniform layer two square tiling of the S_1 structure.

We report a metastable phase of nearly-degenerate self-similar super-structures.

Towards a Bilayer-Modulated Hexagonal Wigner Quasicrystal

We report a remarkable ordering of rectangle-large–triangle-small–triangle (RT_lt_s) configurations into super-structures which represent global ground state candidates of the asymmetric Wigner bilayer system. Between the regions of stability of the trihexagonal phase (I_x) and the honeycomb phase (H) we observe emerging pseudo-dodecagonal motives in layer two – composed of RT_lt_s tiles and arranged to form different super-triangular and super-rectangular tiles – which can be interpreted as vertical modulations of the hexagonal monolayer in layer one. The modulated hexagonal monolayer structures have direct geometric connections to metallic mean quasicrystals [274] but also to dodecagonal, square-triangle based quasicrystals [252, 254]. The discrete possibilities of the side lengths, l and s, of the rectangular and the large and small triangular tiles are dictated by the hexagonal lattice but the ratio l/s is a free parameter for these kind of structures and in the limit of $l \approx s$ dodecagonal order can be approximated by hexagonal RT_lt_s structures. In this thesis, we focused on $l = 2a_1$ and $s = \sqrt{3}a_1$, where here a_1 is the tiling length of the hexagonal (equilateral triangular) lattice.

Inspired by the close connection of RT_lt_s structures and square-triangle tilings we implemented a replica exchange Monte Carlo structure optimization algorithm which employs modified zipper-moves to reshuffle the RT_lt_s tiling instead of a square-triangle tiling. We observe, that ground state realizations of the RT_lt_s family in the asymmetric Wigner bilayer system have a tendency of avoiding specific local environments, namely four rectangle corners meeting at a vertex, or rectangular tiles sharing a common edge in general. We conjecture that the emergence of super-tiles and large pseudo-dodecagonal cluster motifs among low energy RT_lt_s super-structures is a consequence of the system's tendency to avoid these types of local environments, rather than to maximize the number of favorable motifs as in cluster-covering approaches [287]. If the particle decorations of the super-tiles were matching at the edges such that adjoining tiles (RT_l or Rt_s) did not lead to conflicts by edge-sharing, vanishingly small energy differences of $\Delta E^*_{GS'} \sim 10^{-7}$ to 10^{-6} between arbitrary random RT_lt_s tilings would contribute toward entropic stabilization of the highest-symmetry phase quasicrystal [167, 168] at elevated (small) values of the temperature. The super-tile decorations can also be used to define an "inflation rule" and consequently prove that a perfect quasicrystalline state is possible, although we point out that the geometric feasibility of such an inflation procedure does not warrant that the process minimizes the energy. Instead, the so called "energetic scenario" explains the emergence of quasicrystals in physical systems as a consequence of so-called *matching rules* [293] or even appropriate cluster-motif maximization [286], but neither of these scenarios seems to be in action in Wigner bilayer

system. Similarly, the "random-tiling" scenario does not seem physically plausible due to the size of the super-tiles, that would presumably make such entropic contribution to the finite-temperature stability marginal.

In this thesis, we solely observe RT_lt_s crystalline ground state candidates which exhibit exclusively super-triangular or super-rectangular motives, either formed by overlapping or by non-overlapping pseudo-dodecagons. We could identify newly discovered global ground state candidate RT_lt_s super-structures with 25, 49, 56, 100 and 192 particles per unit cell in the here investigated Wigner bilayer system. However, "true" ground states of this family are increasingly hard to identify via simulations due to the existence of energetically almost degenerate competing structures of vastly different complexity – which, moreover, cannot easily be transformed into each other.

General Remarks on the Super-Structure Formation

The emergence of quasicrystals or even of finite sized quasicrystalline approximants especially in soft-matter systems [43, 50, 273, 294, 295], is often justified by inter-particle potentials which exhibit distinct features, either in real or in reciprocal space. For instance, a suitably parametrized pair-potential [44, 49, 264, 273, 296] may favor particle arrangements of very specific nearest neighbor and second nearest neighbor separations – usually two characteristic length-scales are involved. Together with three body interactions between particles such systems may trigger the formation of a quasicrystal [42, 43, 49, 297].

In contrast, the physics of the here investigated asymmetric Wigner bilayer system is fully determined by featureless long-ranged, purely repulsive Coulomb interaction between the particles which are confined between two oppositely charged plates. Hence, we find it truly remarkable that this system can give rise to such complex metastable DI or ground state RT_lt_s super-structure arrangements which we observed for very specific values of the system parameters.

The ground state assembly of the asymmetric Wigner bilayer system is geometrically guided by the specific choice of the plate separation distance and the charge ratio of the two plates. The purely repulsive Coulomb potential of the point-charges leads to lateral, tiling-like particle arrangements which are distributed as uniform as possible on the two oppositely charged, co-planar layers. Thereby, different, interlocked, tiling like patterns on both layers are formed with minimal tile-sets in general, a process which is apparently able to trigger the formation of the here observed complex super-structures for certain combinations of the system parameters.

Parallels to Intermetallics and Atomistic Structures

A very interesting physical analogy to the RT_lt_s super-structures is given by adatom superlattices on (Si,Ge,Sn)-111 diamond-structure surfaces [290], which realize exactly the kind of super-structures on the hexagonal lattice as we observe in this thesis. We hypothesize, that structures of the RT_lt_s family may be able to explain the $BaTiO_3$ *on Pt(111)* surface, which obtain dodecagonal quasicrystalline order on a perfectly

hexagonal surface [298]. However, especially the absence of a second layer makes this analogy with adatom based surface structures not obvious but rather interesting: a physically motivated hypothesis of this connection might be that the asymmetric Wigner bilayer system might represent an effective, geometrically inspired model for special kinds of vertical phonon excitations of the adatoms on (111) surfaces.

In the pentagonal regions of the phase diagram of the asymmetric Wigner bilayer system the structure formation is less clear compared to the here investigated dodecagonal *type I* and rectangle-large–triangle-small–triangle families, but we identified routes towards understanding the structure formation of configurations featuring distorted pentagonal tiles. We report a remarkable relation of pentagonal ground state of the asymmetric Wigner bilayer system with three-dimensional structures from intermetallics, such as *BRuSc* [272] and the Pt_7Zn_{12} structures [291], which can be mapped to bilayer structures. Furthermore, we report analogies of ground state configurations of the asymmetric Wigner bilayer system with *Yb*-based structures [280] and even with penta-graphene [292].

All in all it is quite remarkable, how many parallels can be drawn from ground state configurations of the asymmetric Wigner bilayer system to three-dimensional atomistic structures known from intermetallics.

3.2 Supramolecular Ordering of Complex Molecules under Electrochemical Conditions

The contents of this Section is largely based on Ref. [86], a work entitled *"Reliable Computational Prediction of the Supramolecular Ordering of Complex Molecules under Electrochemical Conditions"*, which emerged during the course of this thesis and is modified – wherever necessary – to align with the scope of this document.

3.2.1 Introduction

Supramolecular chemistry deals with intermolecular interactions and structure formation beyond individual molecules, and as such lies at the basis of many nano- and mesoscopic structures found in biology. In recent decades, impressive progress in the experimental branches of this field have resulted in at least two Nobel Prizes in chemistry. By contrast, the theoretical understanding and especially the *in silico* prediction of supramolecular ordering has lagged behind somewhat. This is easily understood if one considers the sheer size of the systems under study, requiring in many cases consideration of a solid substrate, a sufficiently large number of molecular building blocks or tectons, and a condensed matter medium (*i.e.*, a solvent or electrolyte solution). The interaction of these three components, each with their intrinsic properties, and with optional extrinsic steering (*e.g.*, by light, heat, electric field), will determine the observed supramolecular structures and govern the transitions between them [21, 22].

In this thesis, we propose a theoretical framework to predict supramolecular ordering of complex molecules at an electrochemical solid–liquid interface. The calculations were inspired by a recent experimental work [83] in which particularly clear-cut transitions between supramolecular structures were observed as a function of the applied electric field at a metal-electrolyte interface. The target molecules whose supramolecular ordering is considered constitute an organic salt that consists of a large, disc-shaped polyaromatic cation (PQP^+) and a much smaller, inorganic anion (perchlorate, ClO_4^-) [81, 84, 85].

The concept of choice to investigate these scenarios would rely (i) on a faithful description of the properties of the system (notably a reliable evaluation of its energy) via *ab initio* simulations and (ii) in a subsequent step the identification of the optimized (ordered) arrangement of the molecules on the substrate by minimizing this energy via efficient and reliable numerical tools; this optimization has to be performed in a high dimensional search space, spanning all possible geometries of the unit cell and all possible coordinates and orientations of the molecules within that cell. Both these approaches, considered separately from each other, are conceptually highly complex and from the numerical point of view very expensive, which precludes the application of this combined concept even for a single set of external parameters (such as temperature, density, and external field); it is thus obvious that systematic investigations of the self-assembly scenarios of such systems are definitely out of reach.

In an effort to overcome these limitations we propose the following strategy: in a first step we map the *ab initio* based energies onto the energy of a related classical model (or classical force field), where the atomistic units of the molecules are featured as spherical, charged units with Lennard-Jones type interactions and where the electrolyte is treated as a homogeneous, dielectric medium; the interaction between the atomic entities and the metallic surface is modeled by a classical, perfectly conductive, Lennard-Jones like wall potential. The as yet open parameters of the resulting force field (energy- and length scales, charges, etc.) are fixed by matching the *ab initio* energies of the system with the related energies of this force field: this is achieved by considering archetypical configurations of the system's building blocks (molecules and surface) and by systematically varying characteristic distances between these units over a representative range. These *ab initio* energies were then fitted along these "trajectories" by the parameters of the classical force field: the energy- and length-scales of the involved interatomic Lennard-Jones or Mie potentials as well as the atom-wall interaction parameters.

It turns out that this force field is indeed able to reproduce the *ab initio* based energies along these "trajectories" faithfully and with high accuracy. Even though the emerging force field is still quite complex (as it features both short-range as well as long-range Coulomb interactions and involves mirror charges) it is now amenable to the aforementioned optimization techniques which thus brings systematic investigations of the self-assembly scenarios of these molecules under the variations of external parameters within reach (see Fig. 3.62 for a schematic visualization of our approach).

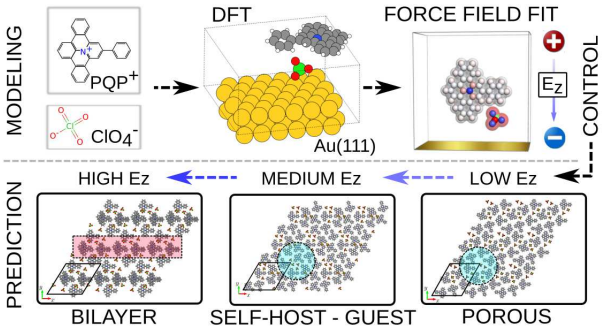


Figure 3.62: Schematic visualization of our theoretical approach [86] to investigate supramolecular self-assembly scenarios at metal surfaces on the example of the PQP-ClO_4 system on Au(111) [83].

As a benchmark test for our approach we have considered the above mentioned system, studied in recent experimental investigations: the cation is PQP^+ (9-phenylbenzo[1,2] quinolizino[3, 4, 5, 6-fed] phenanthridinylium, a disk-shaped polyaromatic molecule), while the anion is perchlorate, ClO_4^-; the self-assembly of these ions on a Au(111) surface under the influence of an external electric field was studied. The high accuracy with which the ensuing energies calculated from the force field reproduce the *ab initio* simulation data make us confident about the applicability of the force field for the subsequent optimization step: using an optimization technique which is based on ideas of evolutionary algorithms we have then identified the self-assembly scenarios of the ions on the Au surface, for a given set of external parameters (temperature, density, and external field). These first results provide evidence that our approach is quite

promising. This concept is furthermore completely flexible as it can easily be extended to other organic molecules of similar (or even higher) complexity. The computational cost of this optimization step is still substantial. Therefore, a detailed, quantitative and, in particular, systematic investigation of the self-assembly scenarios of the PQP^+ and the ClO_4^- ions on the Au surface for a broad range of external parameters is out of the scope of this thesis. Instead we demonstrate in this Section for selected sets of parameters that our approach is indeed able to reproduce several of the experimentally identified self-assembly scenarios.

In this context it has to be emphasized that such a type of optimization problem is highly non-trivial since the huge number of possible local minima in the potential energy surface (embedded in a high-dimensional parameter space) increases exponentially with the number of particles (and their degrees-of-freedom) of the system [96]; thus exhaustive search strategies hit the computational limits or even exceed the capacities of present day supercomputers. Yet another complication in structure prediction algorithms is caused by the fact that different polymorphs of a system can be kinetically trapped and a vast number of other minima, having values of the internal energy comparable to the global minimum may also play an important role in structure formation processes [89, 96].

At this point we owe an explanation to the reader why we have chosen the possibly unconventional approach. Of course, it is obvious that an optimization of the molecular configurations on the basis of full *ab initio* calculations is from the computational point of view by far out of reach. However, one can argue that suitable force fields (available in literature) or machine-learning potentials such as high-dimensional neural network potentials [289, 299–304], kernel-based ML methods [305] (such as Gaussian approximation potentials [247, 306–309]) or more specialized, effective potentials for selected molecular motives used by, for instance, the SAMPLE approach [54, 89, 90] might represent a more conventional strategies to tackle this problem (note that the field of machine-learning potentials is rapidly growing and the above list is far from comprehensive). With such machine-learning potentials the computational cost to faithfully evaluate force fields related to atomistic or molecular systems can be reduced by orders of magnitude compared to *ab-initio* calculations while maintaining the high precision as well as the high flexibility of *ab-initio* based models) [302]; even the formation and breaking of chemical bonds between atoms can be described with these machine-learning based approaches. Hence, such arguments represent fully legitimate objections against our approach.

The problem we are addressing in this part of the thesis is however a non-standard problem and thus requires to be treated with a custom force field: the justification for our strategy is that we wanted to endow the atomic units of the molecules with "real" physical properties (such as "size" or "charge"), which will help us to perform the second step in our structural search that we have envisaged (and that we are currently working on): as the computational costs of our approach are still considerably, large scale investigations are still prohibitively expensive. In an effort to overcome these limitations we plan to proceed to even more coarse-grained models which grasp, nevertheless, the essential features of our complex molecules. On the basis of such models we would then be able to identify with rather low computational costs first

trends in structural identification processes. Investigations along this direction are topic of future works.

Finally we point out that we are well aware of the limitations and deficiencies of our present model. Features such as the response of the metallic electronic distribution of the gold surface due to the presence of an external bias, variable electrostatic properties of the molecular species (allowing thus for polarization effects), or a space dependent permittivity cannot be included in our concept. However, at this point it is fair to say that, to the best of our knowledge, none of the aforementioned alternative approaches (such as the use of conventional force fields or machine learning frameworks) are able to take all these effects faithfully into account, either.

This part of the thesis is organized as follows: In Subsection 3.2.2 we describe the essential features of the experimental setup, introduce an *ab initio* and a classical representation thereof and discuss the mapping procedure between those different instances. In Subsection 3.2.3 we put forward the adaptation and parametrization of the memetic optimization procedure based on ideas of evolutionary algorithms (see Subsection 2.2.4) in order to identify ordered ground state configurations of complex molecules under electrochemical conditions and in Subsection 3.2.4 we present selected numerical results which demonstrate a semi-quantitative agreement with the experimentally observed self-assembly scenarios of PQP^+ and ClO_4^- ions on an Au(111)-electrolyte–interface under the influence of an external electrostatic field. We conclude our findings in Subsection 3.2.5.

3.2.2 The System and its Representations

3.2.2.1 System

Both the DFT calculations and the related force field are based on a framework that mimics the essential features of the experimental setup, put forward (and discussed) in [83]; this framework is schematically depicted in Fig. 3.63: PQP^+ and ClO_4^- ions are immersed into an electrolyte (aqueous 0.1M perchloric acid). From below, the system is confined by a Au(111) surface, which in the experiment serves as the solid substrate for adsorption. An electric field, E_z, can be applied between a reference electrode located within the electrolyte and the Au surface. The PQP^+ and the ClO_4^- ions are first treated via DFT based *ab initio* calculations (see Subsection 3.2.2.2). The calculated energies are then used to fix the force fields of classical particles (notably their sizes, energy parameters, and charges) which represent the atomic entities of the respective ions; the interaction between the atomic entities and the Au(111) substrate is described by means of a classical wall-particle force field (see Subsection 3.2.2.3). Throughout the electrolyte molecules are not considered explicitly. The electrolyte is rather assumed to be a homogeneous effective medium with a permittivity of water, *i.e.*, $\epsilon_r = 78.36$, at $T = 25°C$ [310–312], corresponding to the temperature at which the experiments by Cui *et al.* [83] were carried out and assuming that the low concentration of perchloric acid does not change the value of ϵ_r substantially [313–316]. Hence, in this Section we use "electrolyte" as synonym for "solvent" unless explicit use is required.

We emphasize at this point that in the experiment, an exact specification of the electric field strength is not possible: as detailed in the supplementary information of Ref. [83], the authors of the related experimental investigations have estimated rather the degree of charge compensation on the Au surface by the adsorbed PQP^+ ions as a function of their changing coverage, which does not allow to estimate the electric field directly. This fact limits the degree of quantitative comparison between experiment and theory.

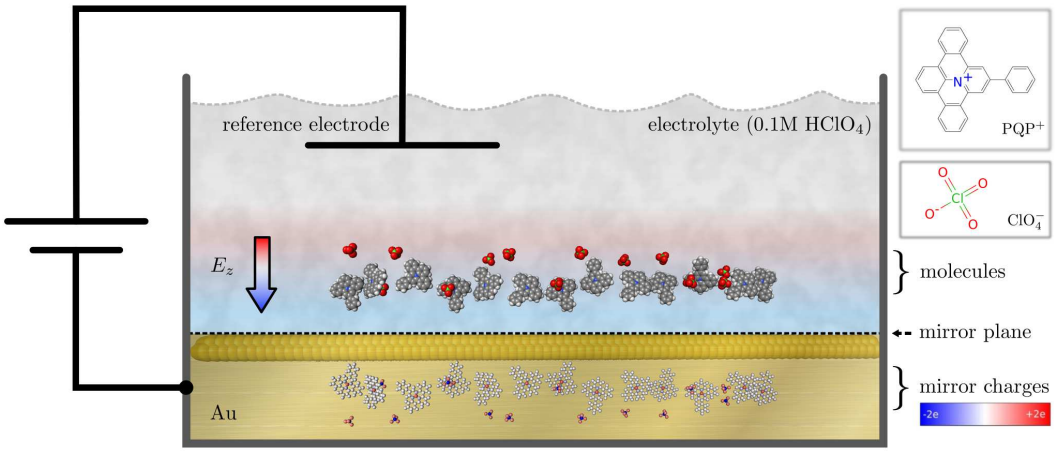


Figure 3.63: Schematic visualization of the experimental setup to control the pattern formation of PQP^+ (and ClO_4^-) molecules (structure formulas given in top right insets) close to a Au(111)-surface: two Au-layers are explicitly shown, the golden, shiny area represents the conductive Au-bulk, the black dashed line marks the surface of the electronic density which we interpret as mirror-plane. The ions are immersed into an electrolyte (gray, shaded region), which is considered as an effective, homogeneous medium. In the region close to the Au surface (red to blue shaded areas) a homogeneous, electrostatic field E_z (bold, colored arrow), oriented in the z-direction, features the electrostatic potential drop between the Au-surface and the reference-electrode inside the electrolyte. The colors of the atoms in the electrolyte correspond to their type, while the color of the mirror-atoms (located in the Au-bulk) specify their partial charges, quantified by the colorbar (see bottom right) in units of the electron charge, e.

3.2.2.2 *Ab initio* Simulations

The density functional theory calculations[81] were performed with the software package GPAW [317, 318] and the structures handled by the atomic simulation environment [319]. The electronic density and the Kohn-Sham orbitals were represented within the projector augmented wave method [320], where the smooth parts were represented on real space grids with grid spacing of 0.2 Å for the orbitals and 0.1 Å for the electron density. The exchange-correlation energy is approximated as proposed by Perdew,

[81] The calculations and contents of Subsection 3.2.2.2 is largely based on Ref. [86] and were primarily conducted by our collaborators O. Brügner, S. Sharma and M. Walter.

Burke and Ernzerhof (PBE) [321] and weak interactions missing in the PBE functional are described as proposed by Tkatchenko and Scheffler (TS09) [322]. The TS09 approximation assumes that long range dispersive contributions are absent in the PBE functional, such that these can be applied as a correction. The total energy which is written as

$$E = E_{\mathrm{PBE}} + w_S E_{\mathrm{vdW}} \tag{3.24}$$

where E_{PBE} is the PBE energy and E_{vdW} is the TS09 correction. A weight factor w_S was introduced that will allow to incorporate electrolyte effects into the dispersive contributions as discussed below. For interactions in vacuum $w_S = 1$. The presence of the aqueous environment on the electronic and nuclear degrees of freedom included in E_{PBE} is modeled by a continuum solvent model [323].

Molecular interactions are studied on simulation grids with Dirichlet (zero) boundary conditions. Neumann (periodic) boundary conditions were applied in x- and y-directions in the surface plane for simulations involving the gold surface, while zero boundary conditions were applied in the perpendicular z-direction. The simulation grid was chosen such that at least 4 Å of space around the position of each atom in the non-periodic directions was ensured. The Au(111) gold substrate was modeled by two layers of 54 atoms, each using the experimental lattice constant of fcc gold of $a = 4.08$ Å. These settings result in a rectangular unit cell of 26.0×15.0 Å^2. The Brillouin zone was sampled by 3×3 Monkhorst-Pack [324] distributed k-points in the periodic directions.

Potentials are scanned by fixing all gold atoms and a central atom of PQP^+ (the nitrogen atom) and/or of ClO_4^- (the chlorine atom) to given positions while all other atoms were allowed to relax without any symmetry constraints until all forces were below 0.05 eV/Å.

The interaction of two perchlorate anions in dependence of their distance is shown in Fig. 3.64 for different approximations for the total energy in Eq. (3.24). As expected, the potentials follow the screened electrostatic repulsion $[\varepsilon_\infty(0) R_{\mathrm{Cl-Cl}}]^{-1}$ for large distances$R_{\mathrm{Cl-Cl}}$, where $\varepsilon_\infty(0) = \varepsilon_r$ is the static relative permittivity of water. There is a slight attractive part in the potential around $R_{\mathrm{Cl-Cl}} \simeq 5.2$ Å already in the PBE potential which leads to a very shallow local minimum. The main reason for this minimum is the decrease in the effective surface ΔA when the solute cavities (the solvent excluded regions) begin to overlap. This decreases the energetic cost to form the surface due to the effective surface tension $\gamma = 18.4$ dyn cm^{-1} (γ also contains attractive contributions and is therefore much lower than the experimental surface tension of water [323]). Subtracting $\gamma \Delta A$ nearly removes all of the minimum as demonstrated in Fig. 3.64. Including the full dispersion contribution $[\varepsilon_\infty(\omega_{\mathrm{opt}}) = 1]$ this local minimum substantially deepens and becomes the total minimum of the potential. An attractive contribution to the potential is not to be expected for the interaction of two anions and needs further discussion. An overestimation of dispersion interactions is suspected if these are treated as in vacuum and no screening through the electrolyte is considered.

The aqueous environment influences the van der Waals (vdW) interactions as these are of Coulombic origin [325]. In order to derive an approximate expression for the

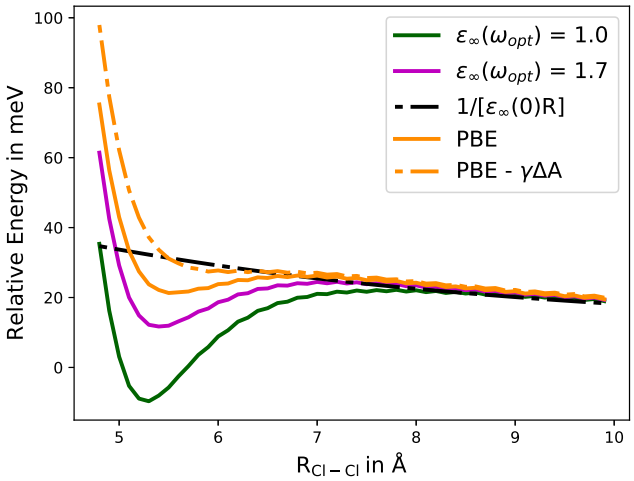


Figure 3.64: The relative energy of two ClO_4^- anions as a function of the distance between their chlorine atoms, $R_{\mathrm{Cl-Cl}}$, where the separated anions define the energy reference; $\varepsilon(\omega_{\mathrm{opt}}) = 1$ with full van der Waals (vdW) corrections and $\varepsilon(\omega_{\mathrm{opt}}) = 1.7$ with scaled vdW corrections. The dash-dotted line shows the PBE energy where the energy contribution of effective surface tension $\gamma\Delta A$ is subtracted (see text).

screened vdW interaction of two molecules A and B at distance R inside the electrolyte, the C_6 coefficient defining the vdW energy C_6/R^6 is expressed by the Casimir-Polder integral [322, 326]

$$C_6 = \frac{3}{8\pi^2\varepsilon_0}\int_0^\infty \alpha^*_A(i\xi)\alpha^*_B(i\xi)\,\phi(i\xi)\,d\xi \tag{3.25}$$

where $\alpha^*_{A,B}$ are the polarizabilities of the interacting molecules and ϕ is determined by propagation of the electric field through the embedding medium [326] with $\phi = 1$ in vacuum. Both $\alpha^*_{A,B}$ and ϕ are modified relative to vacuum in solution. In the simplest model [326] $\phi(i\xi) = \varepsilon_\infty^{-2}(i\xi)$ may be written with the frequency dependent relative permittivity of the electrolyte ε_∞. The effects of the electrolyte on the polarizabilities $\alpha^*_{A,B}$ should, at least partly, already be included in the TS09 description through the effective atomic polarizabilities derived from the self-consistent electron density calculated within the electrolyte. What is left is the effect of the permittivity entering through the function $\phi(i\xi)$ in Eq. (3.25). It is assumed that the main contribution of $\phi(i\xi)$ is at the resonance frequencies of α^*_{AB}, which are in the optical region for usual molecules. It is further assumed that $\varepsilon_\infty(\omega_{\mathrm{opt}})$ is approximately constant in this frequency region, such that $\phi = [\varepsilon_\infty(\omega_{\mathrm{opt}})]^{-2}$ may be pulled out of the integral. This factor scales the C_6 coefficient and therefore the vdW contribution. In other words, the weight $w_s = [\varepsilon_\infty(\omega_{\mathrm{opt}})]^{-2}$ is applied in Eq. (3.25) with the experimental permittivity of water in the optical region of $\epsilon_\infty(\omega_{\mathrm{opt}}) = 1.7$, see Ref. [327]. This approach reduces considerably the depth of the suspiciously deep minimum as seen in Fig. 3.64 such that only a shallow local minimum remains similar to the PBE potential. The reduction obtained is quite strong in respect of the small contributions of Axilrod-Teller-Muto interactions commonly assumed [328, 329]. The quantitative connection between the screening of dispersive interactions in polarizable media and the many-

body effects neglected in TS09 [330] are not immediately clear and is certainly worth further investigation. In what follows the same scaling for all the vdW contributions of the DFT potentials is used.

3.2.2.3 Force Field Model

In this Subsection we describe how we cast our setup into force fields where the atoms in the molecular constituents are described as spherical particles, each of them carrying a charge. The mapping is guided by the energies obtained via the *ab initio* simulations detailed above.

The Au(111) surface is modeled as a flat and perfectly conducting surface involving mirror charges as detailed below. However, we note, that the position of the corresponding surface in the DFT calculations does not coincide with the position of the atoms. Before proceeding the following comment is in order: in this mapping procedure the distance of a point-charge to a metallic surface is unambiguously defined through the electrons leaking out of the potential defined by the nuclei [331, 332]. This feature can explicitly be seen in jellium models [333], but emerges also in implicit calculations [334] where electrons spill out of the surface of metal clusters [335]. From the latter study we estimate an effective spill out of the surface of 0.5 Å, a value that agrees qualitatively with estimates from the jellium models, extrapolated to large structures [332]. This value will be used in the following for our problem.

3.2.2.4 The Atomistic Model

In our atomistic model the molecules are represented as rigid entities composed of atomistic constituents. The molecules are immersed into a microscopic electrolyte, which is treated as a continuous medium of given permittivity. From below the system is confined by a conducting Au(111)-surface (which is assumed to extend in the x- and y-directions), an external field (with respect to the electrolyte) can be applied in z-direction, *i.e.*, perpendicular to the surface (or wall). Fig. 3.63 schematically depicts all details of this atomistic model for the PQP^+ ClO_4^- system, confined by the Au-surface.

In order to specify the different entities of the system and their force fields we use the following notation:

(i) Each of a total number of N molecules is uniquely labeled by capital Latin indices I: for each of these units this index is assigned to its center-of-mass (COM) position vector, $\mathbf{R}_I$, to a vector $\mathbf{P}_I$, specifying its orientation within the lab-frame in terms of the angle-axis framework [336, 337] (see Appendix A.2.1.2 for more details), and to the set of coordinates, $\mathbf{r}^{N_I}$, of the respective N_I atomistic constituents of the molecule in its COM-frame (to which we also refer as its *blueprint*). The set of COM-positions and orientation-vectors of all N molecules

are denoted by $\mathbf{R}^N$ and $\mathbf{P}^N$. The set of all $n = \sum_{I=1}^{N} N_I$ atom positions in the lab-frame is given by $\mathbf{r}^n$, and the position of each atom in the lab-frame is uniquely defined by a vector $\mathbf{r}_i$, labeled with Latin indices ($i = 1, \ldots, n$).

(ii) Between all atoms we consider long-range Coulombic interactions (index 'C'),

$$U^{(\mathrm{C})}(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q_i q_j}{r_{ij}} \qquad i \neq j \tag{3.26}$$

with the inter-atomic distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and charges q_i and q_j of the units i and j; the dielectric constant ϵ_0 and the relative permittivity ϵ_r specify the implicit electrolyte. Further, we introduce short-range force fields (index 'S') for which we have considered two options: first, a Lennard-Jones potential (index 'LJ'), *i.e.*,

$$U^{(\mathrm{LJ})}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]; \tag{3.27}$$

for the energy- and length-parameters, ϵ_{ij} and σ_{ij}, we have opted for the standard Lorentz-Berthelot mixing rules [338], *i.e.*, $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$, respectively.

Alternatively, we have also considered for the short-range interactions the Mie potential [339] (index 'Mie'), which can be considered as a generalization of the LJ interaction; its functional form is given by

$$U^{(\mathrm{Mie})}(r_{ij}) = C_{ij}\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{\gamma_{ij}^{(R)}} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{\gamma_{ij}^{(A)}}\right]; \tag{3.28}$$

and allows for a variation of the exponents of the repulsive and attractive contributions to the potential, $\gamma_{ij}^{(R)}$ and $\gamma_{ij}^{(A)}$, respectively. ϵ_{ij} and σ_{ij} are again parameters for the energy- and the length-scales. The C_{ij} are defined as functions of the exponents [339]:

$$C_{ij} = \left(\frac{\gamma_{ij}^{(R)}}{\gamma_{ij}^{(R)} - \gamma_{ij}^{(A)}}\right)\left(\frac{\gamma_{ij}^{(R)}}{\gamma_{ij}^{(A)}}\right)^{\left(\frac{\gamma_{ij}^{(A)}}{\gamma_{ij}^{(R)} - \gamma_{ij}^{(A)}}\right)}; \tag{3.29}$$

for the exponents we apply arithmetic mixing laws, *i.e.*, $\gamma_{ij}^{(R)} = \frac{1}{2}(\gamma_i^{(R)} + \gamma_j^{(R)})$ and $\gamma_{ij}^{(A)} = \frac{1}{2}(\gamma_i^{(A)} + \gamma_j^{(A)})$.

(iii) We assume the Au-surface to be perfectly conductive, consequently we need to explicitly consider mirror-charges in our model; when further assuming $z = 0$ as the plane of reflection, the Coulombic interaction becomes

$$U'^{(\mathrm{C})}(r_{ij}) = U^{(\mathrm{C})}(r_{ij}) + U^{(\mathrm{C})}(r_{ij'}) + U^{(\mathrm{C})}(r_{i'j}) + U^{(\mathrm{C})}(r_{i'j'}) \tag{3.30}$$

with the mirror charges $q_{i'} = -q_i$ and their positions $\mathbf{r}_{i'} = (x_{i'}, y_{i'}, z_{i'}) = (x_i, y_i, -z_i)$.

(iv) We describe the solid–liquid interface in terms of a slab-geometry with a lower confining wall, *i.e.*, we assume periodicity in the x- and y-directions, but a finite extent, d, of the geometry in the z-direction which is chosen such that no restriction in the orientation of any molecule occurs, thus $d \approx 1.2-2$ nm, given their size and the slab-width. We define the (orthorhombic) lattice vectors, $\mathbf{a}_1 = (a_{11}, 0, 0)$, $\mathbf{a}_2 = (a_{21}, a_{22}, 0)$, and $\mathbf{a}_3 = (0, 0, d)$, which, without the loss of generality, define the volume of the unit cell, $V = a_{11}\, a_{22}\, d$, and which we collect within the matrix $\mathbb{B} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. Together with the molecular basis, given by $\mathbf{R}^N$, $\mathbf{P}^N$ and all N (rigid) molecular blueprints, $\mathbf{r}^{N_I}$, we now define the supramolecular lattice[82]

$$\mathcal{X} = \mathcal{X}(\mathbf{R}^N, \mathbf{P}^N, \mathbb{B}) = \{\mathbf{R}^N, \mathbf{P}^N, \mathbb{B}\}, \tag{3.31}$$

which gives rise to all atomic coordinates in the lab-frame, $\mathbf{r}^n$, *i.e.*, the molecular crystal structure of the system (see Appendix A.2.1.2).

(v) The force field between the atomic entities and the Au-surface is described via an LJ-type wall potential [340],

$$U^{(\text{wall})}(z_i) = 2\pi\epsilon_{\text{w}i}\left[\frac{2}{5}\left(\frac{\sigma_{\text{w}i}}{z_i}\right)^{10} - \left(\frac{\sigma_{\text{w}i}}{z_i}\right)^{4} - \frac{\sqrt{2}\sigma_{\text{w}i}^3}{3(z_i + (0.61/\sqrt{2})\sigma_{\text{w}i})^3}\right]; \tag{3.32}$$

in the above relation, z_i is the height of atom i above the surface, $\sigma_{\text{w}i}$ and $\epsilon_{\text{w}i}$ are the length- and energy-parameters of the interactions of each atom i with the wall, respectively.

(vi) Finally, we express the electrostatic interfacial potential between the electrode and the Au-surface by an external, homogeneous electrostatic field, E_z (*i.e.*, oriented perpendicular to the surface): we account for this potential via $U^{(\text{field})}(z_i) = z_i\, q_i\, E_z$ [341].

Thus and eventually the total potential energy of our model is given by the expression

$$U(\mathbf{r}^n, \mathbb{B}; E_z) = \sum_{i\neq j}^{n\ *}\left[U'^{(\text{C})}(r_{ij}) + U^{(\text{S})}(r_{ij})\right] + \sum_{i=1}^{n}\left[U^{(\text{wall})}(z_i) + U^{(\text{field})}(z_i)\right], \tag{3.33}$$

with 'S' standing for 'LJ' or 'Mie'; we recall that $\mathbf{r}^n$ is the set of all n atomic positions $\mathbf{r}_i$ in a lattice with slab-geometry (defined by the unit cell $\mathbb{B}$). If not present (and not explicitly addressed) the electric field will be dropped in the argument list of Eq. (3.33), that is $U(\mathbf{r}^n, \mathbb{B}; E_z = 0) \equiv U(\mathbf{r}^n, \mathbb{B})$. The notation '$\sum^*$' indicates that summation is only carried out over atoms, labeled with Latin indices i and j, which belong to different molecules I and J (with $I \neq J$); molecules being labeled with capital indices. The

[82] Although, strictly speaking, we here refer to a supramolecular crystal structure but we adopted the naming convention from Ref. [82].

energy given in Eq. (3.33) and the corresponding force fields are efficiently evaluated using the software-package *LAMMPS* [93].

To evaluate the long-range Coulomb term, $\sum_{i\neq j}^{n} {}^{*}U'^{(\mathrm{C})}(r_{ij})$ in the given slab-geometry we use numerically reliable and efficient slab-corrected 3D Ewald-summation techniques [92, 148, 149] (see also Section 2.1). The other terms in Eq. (3.33) are evaluated via direct lattice summation techniques.

3.2.2.5 Parametrizing the Classical Model *via Ab-Initio* Calculations

In our approach, the blueprint of each molecule $\mathbf{r}^{N_I}$ is obtained from electronic structure calculations based on density functional theory (DFT), using dispersion corrected *ab initio* structure optimization [321, 322], as described in Subsection 3.2.2.2. The partial charges of the atoms, q_i, are parametrized via a Bader analysis [342] and are collected in Tables A.2 and A.3 in Appendix A.2.1.3. These charges are directly transferred to the atomic entities. We repeat that throughout the electrolyte molecules have not been considered explicitly: instead, we treat within the force field the electrolyte as an effective, homogeneous medium, introducing the electric permittivity of water ϵ_r.

In order to fix the remaining model parameters that specify the interactions in Eq. (3.33) we search for each atomistic entity (labeled i) the set of atomistic model parameters (specified below) which reproduces via Eq. (3.33) the *ab initio* energies as good as possible. On one side we consider either the length- and the energy parameters of the LJ-potential (denoted by $\mathcal{L} = \{\sigma_i, \epsilon_i\}$) or the length- and the energy parameters together with the exponents of the Mie-potentials (denoted by $\mathcal{M} = \{\sigma_i, \epsilon_i, \gamma_i^{(R)}, \gamma_i^{(A)}\}$), as well as the wall parameters, $\mathcal{W} = \{\sigma_{\mathrm{w}i}, \epsilon_{\mathrm{w}i}\}$. To fix these parameters we proceed as follows:

(i) We first perform *ab initio* structure optimization for different, characteristic molecular configurations, specified below. Here, molecules are either positioned next to each other (without considering the wall) or above the Au-surface: in the former case we fix the positions of two selected atoms belonging to different molecules, the atoms being separated by r_{ij}; in the latter case we keep the height, z_k, of one selected atom above the surface constant. Relaxation of all other degrees of freedom leads in the *ab initio* simulations to spatially and orientationally optimized molecular structures; they are denoted by $\mathbf{d}^n(r_{ij})$ and $\mathbf{d}^n(z_k)$, respectively, with corresponding energies $U_{\mathrm{DFT}}\left(\mathbf{d}^n(r_{ij}), \mathbb{B}\right)$ and $U_{\mathrm{DFT}}^{(\mathrm{wall})}\left(\mathbf{d}^n(z_k), \mathbb{B}\right)$; they are, themselves, functions of the inter-atomic distance, r_{ij}, and the atom-wall separation, z_k, of the selected atoms.

(ii) For every optimized *ab initio* structure, $\mathbf{d}^n(r_{ij})$ and $\mathbf{d}^n(z_k)$, obtained in this manner we define a corresponding molecular configuration $\mathbf{r}^n(r_{ij})$ and $\mathbf{r}^n(z_k)$, which is based on the above introduced atomistic model via the rigid molecular blueprints $\mathbf{r}^{N_I}$ (with the index I running now over all N molecules present in the respective DFT structure). To this end we synchronize the COM-positions of each molecule I in the *ab initio* simulation with the corresponding COM-positions $\mathbf{R}_I$ of its classical counterparts and align their orientation $\mathbf{P}_I$ accordingly.

(iii) Finally we evaluate the corresponding energies with the help of the force field via Eq. (3.33) at zero electric field, *i.e.*, $U_{\mathcal{L}/\mathcal{M}}\left(\mathbf{r}^n(r_{ij}),\mathbb{B}\right)$ and $U^{(\text{wall})}_{\mathcal{L}/\mathcal{M},\mathcal{W}}\left(\mathbf{r}^n(z_k),\mathbb{B}\right)$. We search for the best set of parameters $\mathcal{L}$ (or $\mathcal{M}$) and $\mathcal{W}$ via simultaneously minimizing

$$\mathcal{F}_{\mathcal{L}/M} = \sum_{\{r_{ij}\}} \left| U_{\text{DFT}}\left(\mathbf{d}^n(r_{ij}),\mathbb{B}\right) - U_{\mathcal{L}/\mathcal{M}}\left(\mathbf{r}^n(r_{ij}),\mathbb{B}\right)\right|^2 \tag{3.34a}$$

$$\mathcal{F}^{(\text{wall})}_{\mathcal{L}/M,\mathcal{W}} = \sum_{\{z_k\}} \left| U^{(\text{wall})}_{\text{DFT}}\left(\mathbf{d}^n(z_k),\mathbb{B}\right) - U^{(\text{wall})}_{\mathcal{L}/\mathcal{M},\mathcal{W}}\left(\mathbf{r}^n(z_k),\mathbb{B}\right)\right|^2 . \tag{3.34b}$$

Of course, in the model the same unit cell, $\mathbb{B}$, and the same number of particles, n, as in the respective *ab initio* simulations have to be used. Note that in Eq. (3.34a) the wall-term included in Eq. (3.33) is obsolete since the surface atoms are not considered.

These fits are based on five particularly chosen, archetypical configurations, to be discussed in the following. In the panels of Fig. 3.65 we display schematic sketches of these configurations of the PQP^+ and ClO_4^- molecules; these panels show the corresponding energy curves obtained from the force field, with parameters based on a fitting procedure to the *ab initio* energy profiles.

(a) *Tail-to-tail* configuration (see inset of in panel (a) in Fig. 3.65): We have considered a series of *ab initio* structure optimizations at constant, but successively increasing nitrogen-nitrogen distances, r_{NN}, in the x-direction (while keeping y_{NN} and z_{NN} constant) of an anti-parallel oriented pair of PQP^+ molecules; both cations are vertically decorated with a ClO_4^- molecule. The aromatic parts of the PQP^+ molecules lie flat in the x- and y-directions such that their tails face each other.

(b) *Face-to-face* configuration (see inset in panel (b) in Fig. 3.65): In this case we consider anti-parallel oriented, but vertically stacked PQP^+ molecules (both being horizontally decorated by ClO_4^- molecules) under the variation of the nitrogen-nitrogen distance, r_{NN}, in z-direction (while now keeping x_{NN} and y_{NN} constant). Again, the aromatic parts of the PQP^+ molecules lie flat in x- and y-directions; however, and in contrast to case (a) these units face each other.

(c) *ClO_4^-–ClO_4^-* configuration (see inset in panel (c) in Fig. 3.65): Here two ClO_4^- molecules are considered, varying the chlorine-chlorine x-distance, r_{ClCl}, while keeping y_{ClCl} and z_{ClCl} constant.

(d) *Face-to-wall topped* configuration (see inset in panel (d) in Fig. 3.65): In this case a single PQP^+ molecule, lying flat and parallel to the (x,y)-plane, is located above two layers of Au and is vertically decorated by a ClO_4^- molecule. The cell geometry is assumed to be periodic in the x- and y-directions and finite along the z-axis; in an effort to scan along the z-direction, we have performed a series of *ab initio* based structure optimizations for selected fixed values of z_{N}, *i.e.*,

the z-position of the nitrogen in PQP^+ above the Au-surface. The LJ 10-4-3 potential [340] has been used between the Au(111) surface and the molecules.

(e) *Face-to-wall beside* configuration (see inset in panel (e) in Fig. 3.65): In contrast to case (d), the PQP^+ cation is now horizontally decorated by the ClO_4^- anion such that both molecules are adsorbed on the Au-surface. Again, the LJ 10-4-3 potential [340] defined in Eq. (3.32) has been used between the Au(111) surface and the molecules.

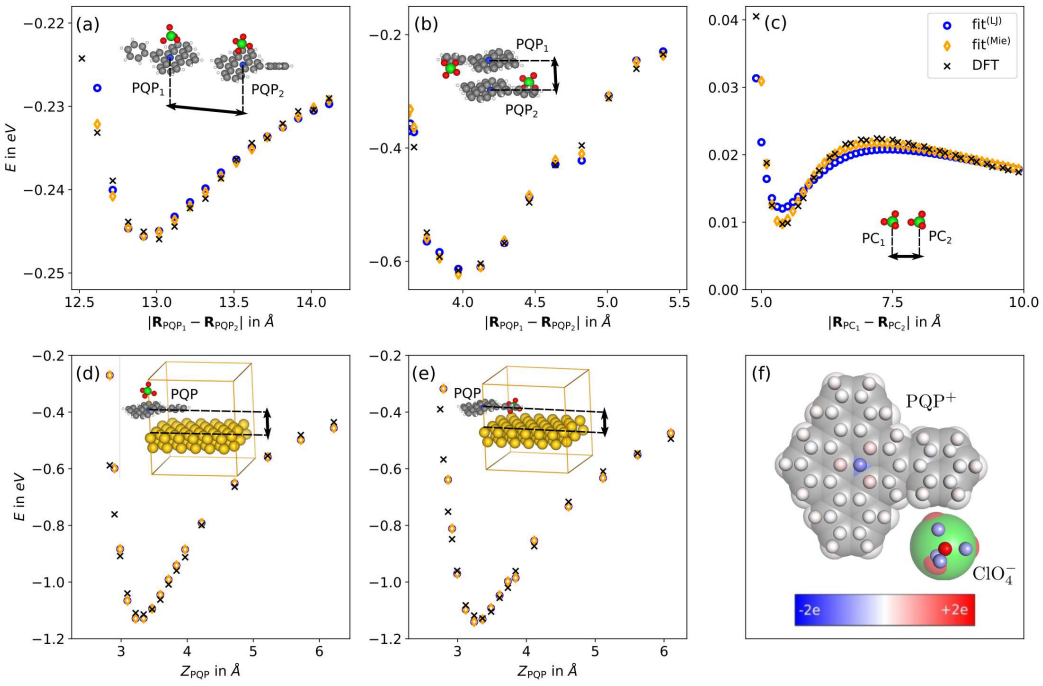


Figure 3.65: Energies as obtained in *ab initio* simulations (black crosses) and fitted data, using the force field (involving an LJ interactions – open blue circles – or Mie interactions – open orange diamonds), see Subsection 3.2.2.4; also shown are – with labels (a) to (e) – five schematic sketches of the five archetypical configurations of the molecules (along with their relative displacements, schematically indicated via the arrows as the distances vary along the abscissa); the related energy curves are used to fit the parameters of the force field, as outlined in the text; the labels correspond to the itemization (a) to (e) used in Subsection 3.2.2.5. Panel (f): PQP^+ and the ClO_4^- molecules, drawn to scale and using the Mie force field for the short-range interactions: atomic entities are shown as transparent spheres with their diameters fixed by their respective optimized σ_i-values and their Bader charges (see color code).

In practice we first optimize $\mathcal{F}_{\mathcal{L}/\mathcal{M}}$, given in Eq. (3.34a), involving thereby *all* interatomic force field parameters; their values are listed in Table 3.5 for the LJ and the Mie models. These parameters are then kept fixed and are used in the subsequent calculations to optimize the wall force field parameters via optimizing $\mathcal{F}^{(\text{wall})}_{\mathcal{L}/\mathcal{M},\mathcal{W}}$, specified in Eq. (3.34b); the emerging parameters are listed in Table 3.6. In panel (f) of Fig. 3.65 we present a visualization of the molecules PQP^+ and ClO_4^-, using these optimized parameters and providing information about the charge of the atomic entities via the color code.

	σ_{H}	σ_{C}	σ_{N}	σ_{O}	σ_{Cl}	ϵ_{H}	ϵ_{C}	ϵ_{N}	ϵ_{O}	ϵ_{Cl}
$\mathcal{L}$	2.243	3.658	3.743	2.865	5.953	3.052	1.204	3.311	7.396	0.172
$\mathcal{M}$	2.236	3.703	3.328	2.428	4.956	3.999	0.946	2.021	11.481	5.289
	$\gamma_{\mathrm{H}}^{(R)}$	$\gamma_{\mathrm{C}}^{(R)}$	$\gamma_{\mathrm{N}}^{(R)}$	$\gamma_{\mathrm{O}}^{(R)}$	$\gamma_{\mathrm{Cl}}^{(R)}$	$\gamma_{\mathrm{H}}^{(A)}$	$\gamma_{\mathrm{C}}^{(A)}$	$\gamma_{\mathrm{N}}^{(A)}$	$\gamma_{\mathrm{O}}^{(A)}$	$\gamma_{\mathrm{Cl}}^{(A)}$
$\mathcal{L}$	12	12	12	12	12	6	6	6	6	6
$\mathcal{M}$	6.263	7.136	8.659	8.743	15.455	7.500	12.299	13.854	17.193	4.684

Table 3.5: Numerical results for the optimized LJ and Mie parameters, $\mathcal{L} = \{\sigma_i, \epsilon_i\}$ and $\mathcal{M} = \{\sigma_i, \epsilon_i, \gamma_i^{(R)}, \gamma_i^{(A)}\}$, for each element i, for the results depicted in the Fig. 3.65 (σ_i in Å and ϵ_i in meV). Reference values from the literature are listed in Table A.1.

	$\sigma_{\mathrm{w[H,C,N]}}$	$\sigma_{\mathrm{w[O,Cl]}}$	$\epsilon_{\mathrm{w[H,C,N]}}$	$\epsilon_{\mathrm{w[O,Cl]}}$
$\mathcal{W}^{(\mathrm{LJ})}$	3.197	3.625	3.741	15.781
$\mathcal{W}^{(\mathrm{Mie})}$	3.208	3.630	3.698	20.167

Table 3.6: Top row: Numerical results for LJ-length and well-depth parameters, $\sigma_{\mathrm{w}i}$ in Å and $\epsilon_{\mathrm{w}i}$ in meV, between the wall and each element $i = [\mathrm{H}, \mathrm{C}, \mathrm{N}]$ and $j = [\mathrm{O}, \mathrm{Cl}]$, grouped by the molecules they belong to (PQP^+ and ClO_4^-), for intermolecular short-range LJ parameters listed in Table 3.5. Bottom row: corresponding $\sigma_{\mathrm{w}i}$ and $\epsilon_{\mathrm{w}i}$ parameters for intermolecular short-range Mie-parameters also listed in Table 3.5.

3.2.3 Identifications of Self-Assembly Scenarios

With the classical force field for the PQP^+ and ClO_4^- molecules introduced in Subsection 3.2.2.4 at hand we are now ready to identify the ordered ground-state configurations of these molecules as they self-assemble on the Au-surface – immersed into an electrolyte and exposed to an electric field. While we leave a more comprehensive and systematic investigation of these self-assembly scenarios to future studies, we focus in this thesis on the technical details of our approach and on a few selected sets of external parameters (i.e, the electric field strength and the particle density).

Our overall objective is to find for our system the global minimum of the total free energy, F, at $T = 0$ K as a function of the positions and orientations of all molecules per unit cell for a given value of cell volume and E_z; at $T = 0$ this task reduces to the minimization of the internal energy U.

Since the experimental observations [83, 85] provide evidence of a structural organization of the molecules into supramolecular lattices, the center-of-mass coordinates of the molecules, $\mathbf{R}^N$, and their orientations, $\mathbf{P}^N$, as well as the parameters defining the unit cell, $\mathbb{B}$, (see Subsection 3.2.2.4 and Fig. 3.63 for details), are the variables which have to be optimized for the search of ground-state configurations: we minimize $U(\mathbf{r}^n, \mathbb{B}; E_z)$, defined in Eq. (3.33), with respect to $\mathbf{R}^N$, $\mathbf{P}^N$, and $\mathbb{B}$, keeping the number of molecules N, the unit cell volume V (with fixed slab width d), and the electrostatic field strength E_z constant.

The energy minimum has to be found in a huge-dimensional parameter space, spanning the positions and orientations of the molecules and by the parameters specifying the unit cell. To be more specific, the dimensionality is set by the number of parameters

to be optimized, which read 64, 76, and 88 for five, six, and seven molecules per unit cell, respectively.

For this purpose we use a memetic evolutionary algorithm [139] which combines the search strategies of evolutionary algorithms (EA) [63, 104–109] and local, steepest gradient descent procedures (LO) [169, 176] following Subsection 2.2.4: initially, a total (fixed) number of $i = 1, \ldots, N_{\mathrm{EA}}$ different lattice-configurations, $\mathcal{X}_i = (\mathbf{R}_i^N, \mathbf{P}_i^N, \mathbb{B}_i)$ as defined in Eq. (3.31), is generated, mostly at random[83]. This *population* is exposed to concepts of natural (or, rather, artificial) selection.

At every iteration step of the EA a new configuration $\mathcal{X}_{i\oplus j}$, *i.e.*, an *offspring*, is created via a *crossover operation* between two selected *parent* configurations, $\mathcal{X}_i$ and $\mathcal{X}_j$, from existing configurations of the most recent population (the index $i \oplus j$ emphasizes the respective crossover operation). This new offspring configuration, $\mathcal{X}_{i\oplus j}$, is then subjected to *mutation operations* of the different variational attributes of $\mathcal{X}_{i\oplus j}$ (always under the restrictions of fixed system parameters N, V, E_z and d and within predefined numerical boundaries).

Subsequently, the (potentially mutated) offspring configuration is subjected to an LO, an operation which represents by far the most time consuming task in our algorithm and is thus performed in parallel[84] using the "*mpi4py*" framework [214, 215, 219]. The relaxed configurations $\mathcal{X}_{i\oplus j}$ are gathered[85] by the algorithm which then decides – via a criterion primarily based on the respective internal energy of the configurations – whether the new relaxed molecular arrangements are *accepted* as new members of the population or if they are *rejected*.

In this acceptance or rejection stage of the algorithm, we employ the *energy-nichening operation* (see Subsection 2.2.4), *i.e.*, we discard offspring configurations which are too close in energy to any configuration of the most recent population. However, this procedure alone cannot cope with "degenerate" configurations, *i.e.*, if structurally distinct configurations have essentially the same energy values (within the specified nichening tolerance). In our approach we allow configurations to enter the population only if their structures differ significantly from those of the competing, degenerate configurations. In order to quantify the structural difference between configurations we associate a feature vector, $\mathbf{x}_i$ (*i.e.*, a set of order parameters), which collects a set of order parameters pertaining to configuration $\mathcal{X}_i$ (see Appendix A.2.2.2 for details). The degree of similarity between two configurations, $\mathcal{X}_i$ and $\mathcal{X}_j$, is then evaluated by taking the Euclidean distance between the corresponding feature vectors, *i.e.*, $\Delta_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$; similar configurations will have a small distance, while unlike configurations will have a large distance. If Δ_{ij} is above a certain threshold value, the offspring configuration, $\mathcal{X}_{i\oplus j}$, will not be discarded by the energy-nichening operation.

[83] Among those configurations we have also intentionally included as "educated guesses" molecular configurations, inspired by the experimental self-assembly scenarios identified in Ref. [83]. However, it should be emphasized that a tentative configuration is only available for the PQP^+ ions as the experiment does not provide any information about the locations of the perchlorate ions.

[84] For an optimal load-balance we additionally spawn a master-thread on one of the MPI-processes to asynchronously distribute optimization tasks of offspring configurations among all idle MPI-processes, basically following the "work-stealing" philosophy.

[85] The gathering of relaxed configurations is performed asynchronously by the master-thread.

Summarizing, the complexity of the problem at hand forces us to use all the above listed advanced optimization tools, including a basin hopping memetic approach combining the heuristic nature of evolutionary strategies with deterministic local gradient descent algorithms. The gradient descent method[86] deterministically evaluates every local minimum of the basin with high accuracy (which is additionally sped up by the "basin dropping" procedure, see Subsection 2.2.4) while the evolutionary search gradually adapts its population towards the energetically most favorable solution. In that way the algorithm explores the search space for the global optimum, *i.e.*, for the best molecular arrangement, $\mathcal{X}_{\text{opt}}$, which minimizes the energy $U(\mathbf{r}^n_{\text{opt}}, \mathbb{B}_{\text{opt}}; E_z)$ for a given set of system parameters.

For a more detailed description of the memetic evolutionary algorithm (*i.e.*, EA+LO) we refer to Subsection 2.2.4.

From a computational point of view we want to stress that a large portion of efforts was dedicated to implementing a faithful binding of the software package *LAMMPS* [93] to our evolutionary algorithm. *LAMMPS* provides a large number of different force fields – partly implemented to be executable on GPUs – many of which can be combined, *i.e.*, hybridized, to simulate a physical system. We are dealing with a non-standard problem requiring a tailored force-field, which in our case is realized by a combination of Coulomb and Mie interactions in a rigid body setup of the molecules, using mirror charges which have to be "dragged along" with their associated atoms, including a wall potential and an electric field, all that in an orthorhombic slab geometry (only accessible via triclinic cells in *LAMMPS*). To the best of our knowledge *LAMMPS* cannot handle such a setup directly in a molecular dynamics simulation and this combination of force fields and geometric constraints does not allow GPU support. However, our approach is completely flexible and other interactions may simply be included, opening the way for both directions of further investigation: (i) to reduce the complexity of the system even more by using coarse grained potentials as further addressed at the end of this Section and (ii) going to more detailed descriptions of the physics of the system; the latter strategy may involve using neural network potentials [289] (which do have a *LAMMPS* binding [343]) or similar machine learning approaches to gain accuracy and flexibility.

To round up this Subsection it should be noted that a variety of techniques has been used in literature for related optimization problems; among those are: Monte Carlo or molecular dynamics-based techniques such as simulated annealing [97, 98], basin-hopping [99–101], minima hopping [102, 103], and evolutionary approaches such as genetic algorithms [63, 104–115]. The decision on the method of choice relies on the specific problem: for instance, Hofmann *et al.* used the SAMPLE technique (see Refs. [54, 89, 90]), relying on a discretization of the search space into limited, archetypical, intermolecular motives and elaborate data fitting of emerging force fields to describe intermolecular interactions of supramolecular monolayers[87]. To the best of our knowledge, this promising approach has neither been applied to supramolecular systems beyond monolayer configurations, so far, nor has it been used in combination with charged

[86] Energy- and gradient-evaluations of this Section rely on calculations using the open-source software package *LAMMPS* [93].

[87] Also commensurability measures of the supramolecular lattice and the substrate are considered.

molecules, with systems featuring multiple components or with an external control parameter, such as an electric field. In general, the fact that the number of archetypical inter-molecular motives grows rapidly with the increasing size of the molecules bears the risk of hitting very soon the limits of computational feasibility. However, suitable adaptations of this strategy and/or a combination with evolutionary search strategies or with reinforcement learning [123, 124] – which has, for instance, very successfully been applied to protein folding problems [126] in a similar way as AlphaGo [125] was able to master the infamous board game – might represent a viable route to circumvent the aforementioned limitations. Thus, future investigations of such intricate problems as the complex monolayer to bilayer transition, as addressed in this thesis, might come within reach.

3.2.4 Results for the PQP-ClO_4 System

3.2.4.1 General Remarks and System Parameters

In the following we present selected results for self-assembly scenarios of PQP^+ and ClO_4^- molecules on an Au(111)-electrolyte–interface under the influence of an external electrostatic field, as obtained via the algorithm presented in the preceding Subsections. Our choice of parameters is guided by the experimentally observed molecular configurations [83]. We demonstrate that our proposed strategy is indeed able to reproduce on a semi-quantitative level the experimentally observed self-assembly scenarios [83]. As a consequence of the still sizable costs of the numerical calculations a more detailed investigations (where we systematically vary the system parameters) and a quantitative comparison of our results with the related experimental findings [83] is out of scope of this thesis and requires further investigations.

To be more specific we have used the following values for the (external) system parameters:

- an indication for the number of molecules per unit cell is provided by the experiment [83]: we have considered unit cells containing ten, twelve, and 14 pairs of PQP^+ and ClO_4^- molecules. These numbers in molecules include, of course, also the related mirror molecules and correspond to 630, 636, and 742 atomic entities per unit cell, respectively (which interact via short-range and long-range potentials, which are subject to particle wall interaction and which are sensitive to an external electrostatic field);
- also the actual values of the surface area A is motivated by estimates taken from experiment [83]: we have varied A within the range of 6.5 nm^2 to 12.25 nm^2, assuming a step size of typically 0.5 nm^2; systems will be characterized by the surface density of the PQP^+ molecules, defined as $\sigma_{PQP} = N_{PQP}/A$, N_{PQP} being the number of PQP molecules per unit cell;
- the range of the experimentally realized values for the electrostatic field strength E_z is, however, difficult to estimate since the major drop in voltage occurs near the negatively charged Au-surface and the nearby layers of cations [341], which is not directly accessible in experiment. Therefore we have covered – at least

in this first study on this topic – several orders of magnitude in the value for E_z within a range that extends (on a logarithmic grid) from $E_z = -1$ V/nm to $E_z = -10^{-3}$ V/nm; in addition, we have also performed calculations at zero electrostatic field.

It should be mentioned that we have used in all these calculations the Mie potential within the classical model, since the related LJ model is not able to fit the *ab initio* data with a comparable and sufficient accuracy (see also discussion in Subsection 3.2.2.4).

We have covered in total approximately 176 combinations of these parameters (that is the unit cell volume V with a constant slab width d, the number of molecules N, and the electrostatic field strength E_z); for each of these we performed independent evolutionary searches with a population size of typically $N_{\mathrm{EA}} = 40$ configurations (which are successively updated during the evaluation of the algorithm).

In order to offer the reader an insight into the computational complexity of the EA+LO based energy minimization of this Section we outline via a few characteristic numbers the computational limitations: the bottleneck of the identifications of self-assembly scenarios are (i) the huge number of calls of energy-evaluations in the optimization steps. As an example, we have at least 10^4 calls of the energy kernel per generation. For each state point we need at least 10^4 generations, which leads to an absolute minimum of 10^8 calls of the energy kernel for *one (!)* set of system parameters. (ii) The optimization of the energy is performed in a high dimensional search space (as specified above), spanning ~ 60 to 90 degrees of freedom, depending on the number of molecules. It is a particular strength of our optimization algorithm (as detailed in the following) to identify in an efficient and reliable manner minima in such high dimensional search spaces. More details about the numerical costs of our calculations can be found in Appendix A.2.3.1.

3.2.4.2 Lateral Particle Arrangements

In this Subsection we discuss the lateral self-assembly scenarios of the PQP^+ and of the ClO_4^- molecules. Selected results for our numerical investigations are presented in Fig. 3.66 and in – on a more quantitative level – in Table 3.7. The actual values have been chosen in an effort to reproduce – at least on a qualitative level – the results obtained in the experimental investigations. Indeed, the sequence of the obtained ordered ground state configurations (shown in panels (a) to (c)) clearly indicates the transition from a stratified bilayer configuration (identified at a rather strong electrostatic field strength of $E_z = -0.3$ V/nm), over an auto-host–guest mono-layer structure (obtained by reducing the field down to $E_z = -0.1$ V/nm), and eventually to an open-porous configuration (identified at $E_z = -0.01$); similar observations have been reported in the related experimental study [83] as we illustrate in Fig. 3.67.

From the results of our investigations (which are shown only selectively) we learn that an electrostatic field strength of $E_z = -0.3$ V/nm always leads to bilayer configurations, similar to the one shown in panel (a) of Fig. 3.66. This stratified bilayer configuration represents the energetically most favorable one as we vary at fixed E_z the volume of the unit cell and the number of molecules within the respective ranges,

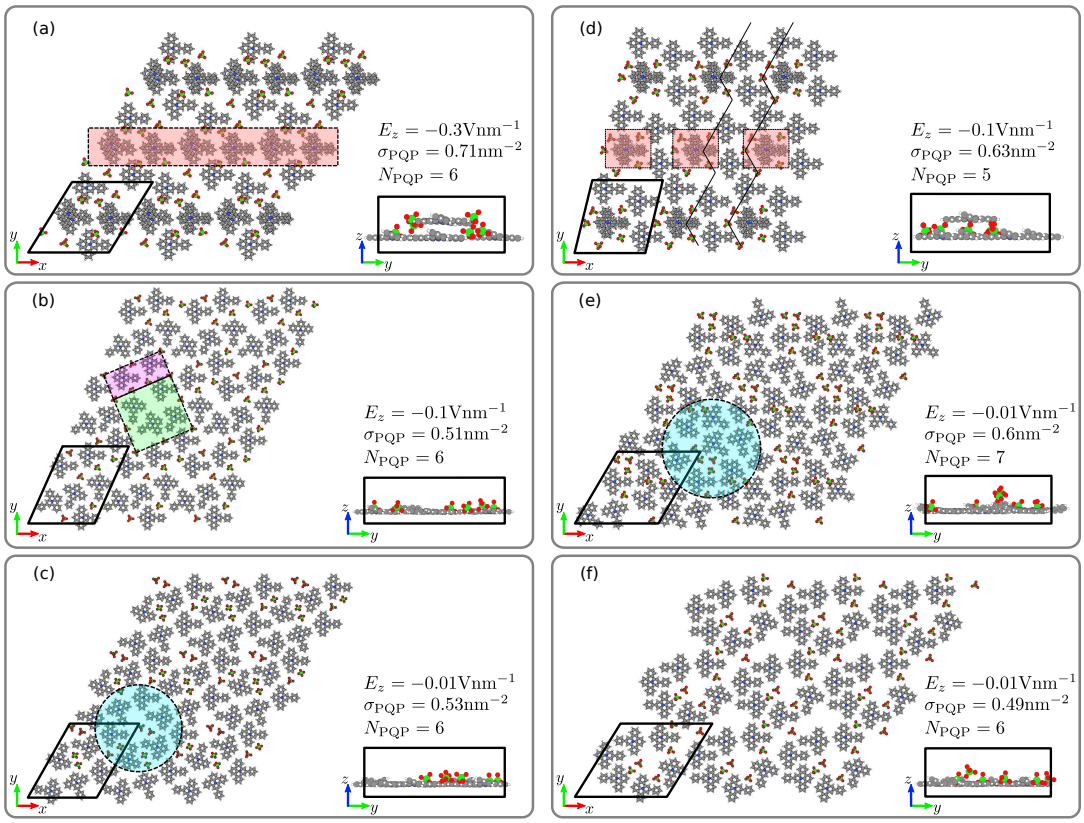


Figure 3.66: Results for the ground state configurations of PQP^+ and ClO_4^- molecules, adsorbed on a Au(111) surface under the influence of an external electrostatic field E_z, as they are obtained via the numerical procedure, as specified in Subsection 3.2.3 and Subsection 2.2.4; calculations are based on the classical model for the molecules, involving the Mie potential (for details see Subsection 3.2.2.4). In the main panels configurations are shown in a periodically extended view as projections onto the (x, y)-plane and in the respective insets as projections onto the (y, z)-plane; in the main panels the respective unit cells are highlighted by thick black lines. Results are shown for different values of the number of PQP^+ molecules, the surface density σ_{PQP} and the electrostatic field E_z: see labels in the different panels and Table 3.7 for details. The red shaded areas, framed with dashed lines, in panel (a) and (d) emphasize PQP^+ molecules which sit on top of other cations, starting to form a bilayer structure. The dashed, shaded, magenta rectangular and green square areas in panel (b) represent tilings formed by perchlorate molecules within the dense PQP^+ monolayer configuration. The dashed, shaded, cyan circles in panel (c) and (e) emphasize, quantitatively, the porous and auto-host–guest motives identified in experiment (see Fig. 4A and Fig. 4C in Ref. [83], respectively, cf. Fig. 3.67).

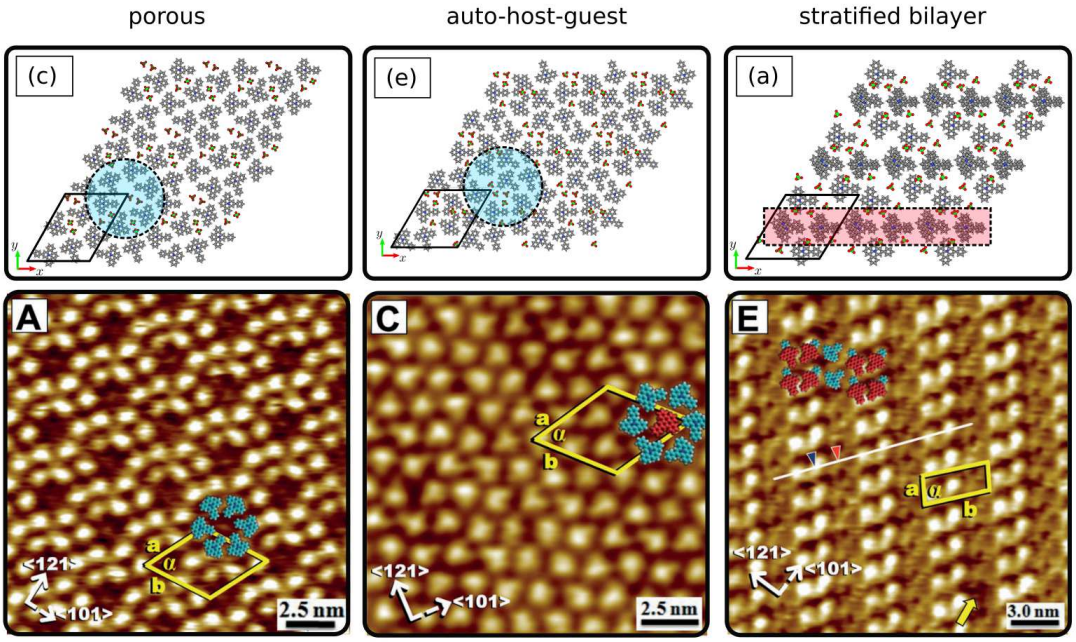


Figure 3.67: Comparison of our numerical findings (top row) as presented in panels (c), (e) and (a) of Fig. 3.66 with, respectively, scanning tunneling microscope (STM) images of experimental results (bottom row) presented in panels (A), (C) and (E) of Fig. 2 in Ref. [83] (copyright Wiley-VCH GmbH, reproduced with permission). The labeling in the top-row panels corresponds to those in Fig. 3.66, the labels in the bottom row to the ones in Fig. 2 in Ref. [83]; the respective electrostatic field strength at which the supramolecular configurations shown in panels (c), (e) and (a) have been identified are listed in Table 3.7. Black (yellow) frames in the panels of the top (bottom) row indicate the respective unit cells of the supramolecular lattices. The cyan shaded areas in the panels (c) and (e) indicate the qualitative porous and auto-host–guest motives formed by the PQP^+ molecules identified by our two-stage approach (cf. tentative molecular configurations shown in panels (A) and (C), respectively). The shaded red area in panel (a) indicates a bilayer arrangement of PQP^+ molecules which is in qualitative agreement with the stratified bilayer configurations found in the experiment (cf. red colored PQP^+ molecules in panel (E)).

specified in the preceding Subsection; the numerical data of the related internal energy are compiled in Table 3.7.

As we proceed to $E_z = -0.1$ V/nm we observe self-assembly scenarios as the ones depicted in panels (b) and (d) of Fig. 3.66, which correspond to auto-host–guest configurations observed in experiment [83]; for the data presented in these panels two different values for N_{PQP} (and hence for σ_{PQP}) have been considered: the mono-layer configuration shown in panel (b) has a slightly lower value for the internal energy (per molecule) than the rhombohedral bilayer configuration shown in panel (d). However, as can be seen from Table 3.7 the energy differences are very tiny: differences of the order of 10^{-4}eV correspond to values where we hit the numerical accuracy of the *ab initio* based energy values.

Eventually, we arrive at the so-called open-porous structures, observed in experiment [83]: the ground state configurations depicted in panels (c), (e), and (f) of Fig. 3.66 are evaluated at the same electrostatic field strength of $E_z = -0.01$ V/nm,

assuming different values for N_{PQP} and σ_{PQP}; the open-porous pattern emerging in panels (c) is the most favorable one in terms of energy per molecule (see Table 3.7 for the numerical details). There are, however, several serious competing structures with minute energy differences at this value of the electric field strength: another open-porous structure, depicted in panel (f), with an energy penalty of less than 8.1 meV per PQP^+ molecule compared to case (c) and a considerably denser configuration, depicted in panel (e), with an internal energy value worse by only 11.3 meV compared to case (c) and by 3.2 meV compared to case (f).

From the numerical point of view the following comments are in order: for a fixed state point, the energy differences of competing structures attain values which hit the limits of the accuracy of the *ab initio* based simulations, which can be estimated to be of the order of 0.1 eV to 0.01 eV per molecule for dispersive interactions [322, 344–346]. These values set the limits of our numerical accuracy. For completeness we note that for the results for the energies obtained via the classical force field (which are based on *LAMMPS* calculations) we estimate that our results are numerically reliable down to $\sim 10^{-6}$ eV per atom; within the range of such minute energy differences no competing structures have been found in our investigations – although the physical relevance of such tiny energy differences given the considerably larger uncertainties in the DFT data raises valid concerns. In general we observe that the energy differences for the energetically optimal ground state configurations become smaller as the electrostatic field tends towards zero. Even though the optimization algorithm (as outlined in Subsection 2.2.4) has turned out to be very efficient and reliable, we observe (in particular for smaller values of the external field) that new configurations are included in the population of the best individuals even after a large number of optimization steps.

	E_z [$\mathrm{Vnm^{-1}}$]	U/N_{PQP} [eV]	N_{PQP}	A [$\mathrm{nm^2}$]	N_{PQP}/A [$\mathrm{nm^{-2}}$]
(a)	-0.30	-1.5804	6	8.5	0.705882
(b)	-0.10	-1.7276	6	11.75	0.510638
(d)	-0.10	-1.7274	5	8.0	0.625000
(c)	-0.01	-1.6445	6	11.25	0.533333
(f)	-0.01	-1.6364	6	12.25	0.489796
(e)	-0.01	-1.6332	7	11.75	0.595745

Table 3.7: Results of evolutionary ground-state search for different electric field strengths, E_z, for different unit cell areas, A and number of PQP^+ molecules, N_{PQP}, each line represents a evolutionary search. The respective structures are presented in Fig. 3.66.

3.2.4.3 Vertical Particle Arrangements

In Fig. 3.68 we present in separate panels the height distributions of PQP^+ and ClO_4^- as functions of the electrostatic field, E_z (which is binned for the six different values of E_z that were investigated); along the vertical axis we count (for a given value of E_z) the occurrence of the respective molecules in bins of one Å, and normalize by the total number of all considered configurations identified by the evolutionary

algorithm, which are located within an interval of at most $1 k_B T$ (or 43 meV) above the configuration with the best energy, which are of the same order of magnitude as the values presented in Table 3.7 for different electric field strengths, E_z. Note in this context that the distance of the first layer of PQP^+ molecules can be directly estimated by the vertical equilibrium position of carbon atoms, $z_C^{(eq)} = 3.166\text{Å}$, obtained by minimizing Eq. (3.32) for a single carbon atom with $\sigma_{wC}^{(Mie)} = 3.208\text{Å}$, taken from Table 3.6.

For the high values of the field strength (*i.e.*, for $E_z = -10$ V/nm and -1 V/nm) the PQP^+ ions are preferentially adsorbed onto the gold surface as a closely packed monolayer (see left panel of Fig. 3.68), while the perchlorate anions are strongly dissociated and assemble as far from the gold surface as possible (corresponding in our investigation to the numerical value of the slab height, which we fixed to 12 Å) – see right panel in Fig. 3.68. This situation represents an extreme case in the sense that neither the Coulomb nor the short-range Mie interactions between ions and anions can compensate for the strong negative surface potential.

Decreasing now the magnitude of the electrostatic field to the more moderate value of $E_z \sim -0.3$ V/nm reveals the emergence of a bilayer structure, formed by the PQP^+ molecules, with pronounced peaks located at $z_{PQP^+}^{(1)} \sim 3\text{Å}$ and $z_{PQP^+}^{(2)} \sim 7\text{Å}$, with relative weights of 72% and 21%, respectively. Such stratified bilayer configurations (as depicted in panel (a) of Fig. 3.66) are in competition with structures similar to ones shown in panel (d) of Fig. 3.66. Note that in parallel a far more complex height distribution of the perchlorate molecules sets in as soon as the now moderate electrostatic field allows them to proceed towards the interior of the slab: now, more than half of the ClO_4^- anions are located "in between" the PQP^+ "layers", trying on one side to compensate the charges of one or several PQP^+ "partners" in the slab region and "filling spatial holes" wherever they can, on the other side. A large portion of perchlorate ions is even allowed to adsorb on the surface at a distance of $z_{ClO_4^-}^{(1)} \sim 4.074\text{Å}$; note that these COM positions above the interface are larger than the minimal height of the PQP^+ cations due to two reasons: if one face of the oxygen tetrahedron is oriented towards the interface (*i.e.*, parallel to the gold surface), the COM of the ClO_4^- ion is increased by a value of $z_{Cl} - z_O \approx 0.492\text{Å}$ with respect to the oxygen atoms. These atoms themself have an equilibrium distance to the surface of $z_O^{(eq)} \approx 3.582\text{Å}$(evaluated by minimizing Eq. (3.32) for a single oxygen atom with $\sigma_{wO}^{(Mie)} = 3.630\text{Å}$, cf. Table 3.6), summing up to the presented minimal COM distance of the adsorbed ClO_4^- molecules from the interface. Note that the height distribution of the ClO_4^- ions is now rather broad (see Fig. 3.69), which is definitely owed to their relatively smaller size and their considerably higher mobility, as compared to their cationic counterparts (see also discussion below); these features make a conclusive interpretation of the roles of the ClO_4^- ions in the structure formation of the entire system rather difficult. Our interpretation is that the perchlorate ions are – due to their small spatial extent and their high mobility – able to compensate for local charge mismatches and to act as spatial spacers between the cations.

Decreasing further the magnitude of the electrostatic field down to $E_z \sim -0.1$ V/nm and $E_z \sim -0.01$ V/nm provides unambiguous evidence that the formation of the PQP^+

ions into bilayer structures become energetically more and more unfavorable, as the upper peak in the height distribution of the cations vanishes gradually. Concomitantly, an increasing number of perchlorate molecules approach the gold surface and are predominantly located there; possibly they act as a space filler on the surface itself while at the same time the small values of the electrostatic field keep the PQP^+ molecules near to the surface. In this context it should be noted that decreasing the electrostatic field is equivalent to decreasing the surface potential; thus and in combination with the adsorbed perchlorate molecules a transition from an auto-host–guest to a porous structure is plausible.

Eventually, at zero electric field the system exclusively gains energy from intramolecular interactions and adsorbtion on the gold surface. Since the perchlorate molecules are rather spherical in their shape they can efficiently adsorb onto the gold surface (in an orientation explained above), while the PQP^+ molecules are able to efficiently stack, especially without a guiding electrostatic field.

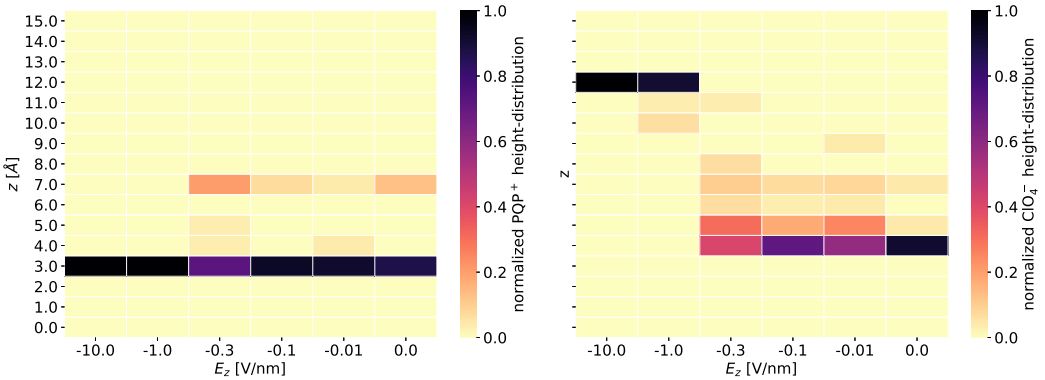


Figure 3.68: Height-distribution of PQP^+ (left) and ClO_4^- (right) molecules as functions of the considered values of the electrostatic field, E_z, normalized by the number of respective molecules (see colour code at the right hand side of the panels: a value of one means, that all respective molecules in all considered configurations are counted in one specific bin), see text for the energy considered in this analysis. Along the vertical axes the binning is performed in steps of one Å: $z = 0$ marks the position of the gold surface, the slab width amounts to 12Å.

3.2.4.4 The Role of the Perchlorate Anions

We come back to the above mentioned volatility of the perchlorate ions: in Fig. 3.69 we present results from yet another evolutionary analysis: we now fix the positions and orientations of PQP^+ molecules as well as the extent and the shape of the unit cell of some optimized configuration (as, for instance, depicted in panel (a) of Fig. 3.66) and vary only the degrees of freedom of the perchlorate anions. Fig. 3.69 shows – for fixed cell geometry and fixed PQP^+ positions and orientations – four structurally different perchlorate arrangements whose energy ranges within an interval of 38 meV (per PQP^+–ClO_4^- pair): the fact that we obtain completely different configurations of the perchlorates (with essentially comparable energies) undoubtedly indicates the high mobility of the ClO_4^- ions. Changes in the structure, as one proceeds from left

to right, are highlighted by respective circles (specifying the position of the "moving" perchlorate ion) and related arrows. The ClO_4^- molecules exhibit a remarkable freedom in their rotation without (or only marginally) changing the energy of a configuration; this fact has rendered the minimization of the energy very difficult. However, it should also be noted that even translations can be performed without a substantial change in energy.

We emphasize that the analysis of these different structures was achieved by using a so-called t-SNE [244] analysis on the leading five principal components of a principal component analysis [238] of order parameters of all configurations identified by the evolutionary algorithm; for more detailed information on this rather technical issue we refer to Subsections 2.4.1 and 2.4.4 and Appendix A.2.3.2.

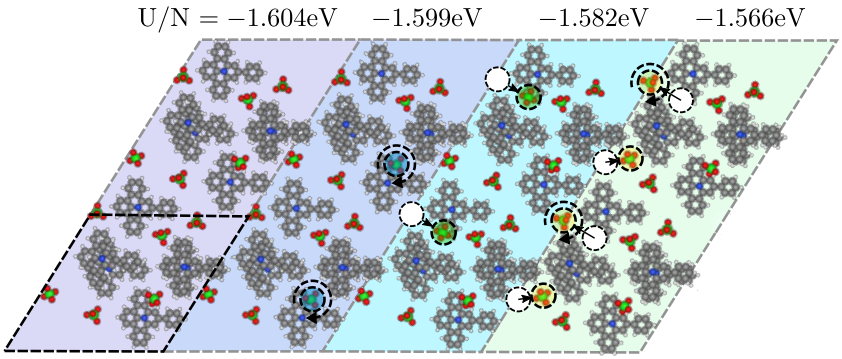


Figure 3.69: Four structurally different configurations of perchlorate ions (framed by gray, dashed lines and differently shaded areas) at an optimized, fixed cell geometry (indicated in the bottom left corner by the black, dashed line) and optimized, fixed positions and orientations of the PQP^+ ions. The energies of these four configurations range within an interval of 38 meV (per PQP^+–ClO_4^- pair), as obtained in an evolutionary search procedure for the energy minimizing configuration under the variation of the degrees of freedom of the ClO_4^- ions only, starting from the configuration depicted in Fig. 3.66(a). Changes in the structure as one proceeds from left to right are highlighted by respective circles (specifying the position of the "moving" and "rotating" perchlorate ion) and arrows.

3.2.5 Conclusions and Outlook

The prediction of supramolecular ordering of complex molecules at a metal–electrolyte interface using DFT based *ab initio* calculations is in view of the expected gigantic computational costs, and despite the availability of peta-scale computers, still an elusive enterprise.

In Section 3.2 we have proposed a two-stage alternative approach: (i) DFT-based *ab initio* simulations provide reference data for the energies introduced in a classical model for the molecules involved, where each of their atomic entities are represented by a classical, spherical particle (with respective size, energy parameters, and charges). We modeled the interaction between the atomic entities and the metallic surface by a classical, perfectly conductive, Lennard-Jones like wall potential; the electrolyte was treated as a homogeneous, dielectric medium. The inter-particle and particle-wall

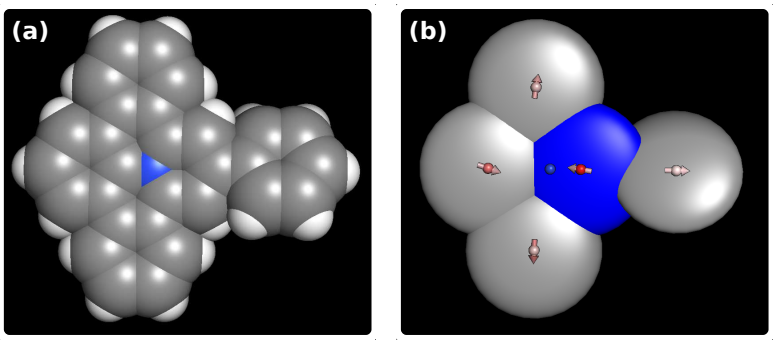


Figure 3.70: Panel (a): atomistic model of the PQP^+ molecule as used in this thesis (white: hydrogen atoms, grey: carbon atoms, and blue: nitrogen atom), see also Fig. A.20 in the Appendix A.2.1.3; panel (b): related coarse-grained model in a hierarchy of ever simpler models, using, *e.g.*, Gay-Berne potentials [347] to account for the van der Waals interaction of all atoms in the specific rings and a multi-pole expansion to second order (monopole as coloured points, dipole moments as small arrows) for the electrostatic interaction.

parameters were obtained via the following procedure: considering archetypical configurations (involving pairs of ions and/or ions located close to the surface) DFT energies were fitted by the related energy values of the classical model. (ii) The second step identifies the ordered ground state configurations of the molecules by minimizing the total energy of the now classical system. This optimization is based on evolutionary algorithms, which are known to operate efficiently and reliably even in high dimensional search spaces and for rugged energy surfaces.

Our new two-stage strategy overcomes the hitherto prohibitive computational cost of modeling the full system, while reproducing the key observations of a well-documented experimental system consisting of disc-shaped PQP^+ cations and ClO_4^- anions: as a function of increasing electric field at the metal–electrolyte interface, the molecular building blocks are seen to self-organize into an open porous structure, an auto-host–guest pattern and a stratified bilayer. Future work will focus on verifying the extent of predictive power of our method towards molecular self-assembly under electrochemical conditions, and on strategies to further streamline and reduce the computational cost of our approach, without sacrificing the reliability of the predicted results.

In view of the high computational costs and the conceptual challenges encountered in our investigations we have pondered the question if the complexity of the current model (which – as a classical model – is comprehensive in the sense that it contains all atomistic features) could possibly be further reduced, avoiding thereby conceptual and computational bottlenecks. The idea behind this strategy is to develop – starting from the present model – a hierarchy of ever simpler models where, for instance, larger sub-units of the molecule (such as aromatic rings) are replaced by disk-shaped units carrying higher electrostatic moments, as schematically visualized in Fig. 3.70. Such a model might provide a first, semi-quantitative prediction of the self-assembly of the PQP^+ and of the ClO_4^- ions at considerably reduced costs and might help to pre-screen possibly promising portions of the huge parameter space for subsequent investigations of the full model. Efforts in this direction are currently pursued.

4 Conclusions and Outlook

In this thesis, we investigated the ground state self-assembly of different physical systems composed of particles or molecules featuring long-range Coulomb interactions which are confined to quasi–two-dimensional bilayer geometries or to slab geometries near solid–liquid-interfaces under electrochemical conditions. Our studies were conducted with the help of specifically implemented numerical tools based on simulated annealing, replica exchange Monte Carlo, and evolutionary algorithms.

In this Chapter, we first present in Section 4.1 general conclusions of our findings of Sections 3.1 and 3.2. We then discuss in Section 4.2 possible future directions of investigations related to the numerical studies of the asymmetric Wigner bilayer system and of the supramolecular ordering of complex molecules at solid–liquid interfaces under electrochemical conditions. Eventually, we present in Section 4.2 a brief outline of two projects which emerged during the course of this thesis (and which are related to the topics of this document), namely the self-assembly of ionic liquid crystals into smectic phases and reinforcement learning of self-navigation strategies for active microswimmers.

4.1 General Conclusions

Towards Quasicrystalline Order in the Asymmetric Wigner Bilayer System

As identified by Antlanger *et al.* [62–64] the asymmetric Wigner bilayer system features an incredibly rich plethora of ordered crystalline ground state configurations – forming tiling like patterns – in a parameter space spanned by the plate separation distance and the ratio of the uniform charge densities of the two plates. In Section 3.1 we studied the possible emergence of dodecagonal and hexagonal quasicrystalline order in the asymmetric Wigner bilayer system.

In an effort to gain a deeper understanding of the ground state phase behavior of the asymmetric Wigner bilayer system we employed in Subsection 3.1.4 unsupervised clustering algorithms (based on principal component analysis [238] and k-means clustering [239–242]) to categorize all available ground state candidate configurations of the system known from literature [62–64] into families of structures. We demonstrated, that our clustering approach is indeed able to relate the different ground state candidates to the different phases identified in literature with high accuracy. Moreover, with the clustering approach, we found new structural families of ground state candidates of the asymmetric Wigner bilayer system which have not been reported previously and we were able to identify promising system parameters at which quasicrystalline order may be favored in the asymmetric Wigner bilayer system.

We further investigated the regions of stability of the snub-square (S_1) and the trihexagonal (I_x) structures in the parameter space of the asymmetric Wigner bilayer system in Subsections 3.1.5 and 3.1.6 for quasicrystalline ground state configurations with, respectively, global dodecagonal and hexagonal (*pseudo*-dodecagonal) symmetry. We used specifically implemented numerical tools based on simulated annealing and replica exchange Monte Carlo, utilizing so-called *zipper*-update moves [254] to reshuffle the respective tilings to efficiently conduct the structural ground state searches for the different structural families at the corresponding relevant system parameters.

We realized in Subsection 3.1.5 that the S_1 structure is a representative of an entire family of square-triangle structures in which the local motives are almost exactly squares and equilateral triangles, but global ordering may plausibly lead to an ordered state without translational symmetry, *i.e.*, a dodecagonal quasicrystal. (Note: generically, the local motives in Wigner bilayer structures do not have such a "discrete" regular character.) We discovered the S_1 structure generalizes to a special subset of square-triangle tilings, such that square-square assembly is energetically disfavored. The remaining, energetically more favorable local square-triangle-based motives still allow the formation of qualitatively different square-triangle tilings which may feature dodecagonal clusters and super-clusters, important precursors of aperiodic quasicrystalline ordering. Strikingly, when special dimensions of the unit cell were chosen such that self-similar super-tilings are geometrically possible, the latter indeed represent the ground state for that particular choice of the periodic boundary conditions. However, this family of self-similar super-structures, which we labeled the dodecagonal *type I* (DI) family, remains metastable with respect to the S_1 structure and other, qualitatively different ground state candidates of the asymmetric Wigner bilayer system known from literature [62–64].

Similarly, in Subsection 3.1.6 we identified a series of new ground state candidate superstructures featuring rectangles, R, and two different types of corresponding equilateral triangles, T_l and t_s, as basic tiles, *i.e.*, the RT_lt_s family of structures. For certain system parameters, these basic RT_lt_s tiles are found to be arranged in pseudo-dodecagonal clusters (*i.e.*, clusters composed of RT_lt_s tiles rather than of squares and equilateral triangles) which form super-structures in low energy configurations of the RT_lt_s family. Complex super-triangular and super-rectangular RT_lt_s structures with 25, 49, 56, 100 and 192 particles per irreducible unit cell are seen to be energetically more favorable than ground state candidates suggested in Refs. [62–64] (with at most 40 particles per unit cell) for certain combinations of the system parameters.

For both the DI and the RT_lt_s families of structures we conjecture that the emergence of super-tiles and large (pseudo-)dodecagonal cluster motifs among the low energy DI and RT_lt_s super-structures is a consequence of the system's tendency to avoid certain types of local environments of the respective basic tile-sets, rather than to *maximize* the number of favorable motifs as in cluster-covering approaches [287].

We emphasize, that the emergence of the here identified, considerably complex metastable DI and ground state candidate RT_lt_s super-structures is the more remarkable as the physics of the asymmetric Wigner bilayer system is governed by repulsive long-ranged Coulomb interactions between the charged particles. The ground state self-assembly of the system is guided by the delicate interplay between the commensura-

bility of the attained configurations in both layers and charge neutralization of the long-range Coulomb interaction [64]. In comparison, the emergence of quasicrystals is often linked to several (usually two) pronounced length scales in the interparticle potentials of a system in the presence of three-body interactions [42, 43, 49].

For a more detailed conclusion of Section 3.1 we refer to Subsection 3.1.8.

Supramolecular Ordering of Complex Molecules under Electrochemical Conditions

In recent experiments conducted by Cui *et al.* [83] the spontaneous and reversible transition between two- and three-dimensional self-assembly scenarios of a supramolecular system ($PQPClO_4$) at a solid–liquid interface under electrochemical conditions [Au(111) in 0.1 M $HClO_4$] has been identified and thoroughly discussed. By a simple variation of the interfacial potential it was possible to selectively organize the disc-shaped PQP^+ cations (a polyaromatic molecule entitled 9-phenylbenzo[1,2] quinolizino[3, 4, 5, 6-fed] phenanthridinylium) in an open porous pattern, to fill these pores to form an auto-host-guest structure or to stack the building blocks in a stratified bilayer. In Section 3.2 our complementary theoretical investigations were dedicated to rationalize these experimental observations.

We proposed a computationally lean, two-stage approach to computationally predict the supramolecular ordering of complex, charged molecules at a solid–liquid interface under electrochemical conditions.

In the first stage, we used *ab initio* simulations which provided reference data to fit the parameters of a conceptually much simpler and computationally less expensive model of the molecules: classical, spherical particles represent the respective atomic entities, a soft but perfectly conductive flat wall potential represents the metallic surface and the electrolyte is treated as a homogeneous, dielectric medium. To obtain the inter-particle and particle-wall parameters we considered archetypical configurations with *ab initio* accuracy (involving pairs of ions and/or ions located close to the surface) and fitted the energies evaluated via density functional theory with our classical model.

In the second stage, we then aimed at identifying ordered ground state configurations of the molecules by minimizing the internal energy of the corresponding classical system. This was achieved by employing specifically implemented optimization techniques based on memetic evolutionary algorithms, which are known to operate efficiently and reliably even in high-dimensional search spaces and for rugged energy surfaces.

Our two-stage strategy overcomes the hitherto prohibitively large computational cost of describing the full supramolecular system by means of purely *ab initio* based models. We were able to reproduce the key observations of the well-documented experimental system mentioned above consisting of disc-shaped PQP^+ cations and ClO_4^- anions [83]: as a function of increasing the electric field at the metal–electrolyte interface, the molecular building blocks are seen to self-organize into an open porous structure, an auto-host–guest pattern and a stratified bilayer.

Future work will focus on verifying the extent of predictive power of our method towards molecular self-assembly under electrochemical conditions. Furthermore, we will pursuit

strategies to further streamline and reduce the computational cost of our approach, without sacrificing the reliability of the predicted results.

A more detailed conclusion of Section 3.2 can be found in Subsection 3.2.5.

4.2 Outlook and Future Work

The Asymmetric Wigner Bilayer System

Typically, identifying ground state candidates for the asymmetric Wigner bilayer system is a challenging task. On the one hand the configuration space of possible ground state candidates of the system is tremendously large and on the other hand the relative energy differences of competing structures in the sevenths or eights digit are very relevant for ground state considerations of the system.

Investigating the emergence of quasicrystalline order in the asymmetric Wigner bilayer system turned out to be a highly complex endeavor. On the one hand, we were able to determine that structures related to the DI family (except for the S_1 structure) do not represent competitors for the global ground states of the system at any values of the system parameters. On the other hand, we were not able to conclusively answer whether the RT_lt_s super-structure ground states candidates identified in Subsection 3.1.6 represent all related RT_lt_s based ground state candidates of the system in the corresponding range in the parameter space, or if even more complex configurations need to be considered. For instance, we identified RT_lt_s super-structures of up to 1444 and 1452 particles per unit cell, which are energetically only slightly less favorable compared to the less complex RT_lt_s ground state candidate super-structures with 49 or 56 particles per unit cell. It is left to future work to resolve this issue in more detail.

We realized the RT_lt_s family of structures can be interpreted as a *bilayer-modulation* of a hexagonal monolayer structure: by vertically displacing selected particles (from the hexagonal monolayer to the other plate) a tiling composed of rectangular, R, and two different types of equilateral triangular tiles, T_l and t_s, can be formed with a long to short side-length ratio l/s that is governed by the discrete interparticle distances of the hexagonal monolayer. We thus conjecture, that the RT_lt_s family of structures – giving rise to large pseudo-dodecagonal clusters and super-structures in the asymmetric Wigner bilayer system – might be able to explain the $BaTiO_3$ on Pt(111) surface, which obtains dodecagonal quasicrystalline order on a perfectly hexagonal surface [298]. Further investigations are necessary to study this analogy more carefully.

As we briefly discussed in Subsection 3.1.7 there are promising regions in the parameter space of the asymmetric Wigner bilayer system, where local pentagonal motives are favored by the system. In turn, these parameter space regions might give rise to ordered ground state configurations with global tenfold symmetry, *i.e.*, decagonal quasicrystals. An in-depth study of the related ground state self-assembly strategies in the asymmetric Wigner bilayer system is topic of future work.

In general, numerical tools based on Mont Carlo sampling or on memetic evolutionary algorithms [62–64] are legitimate approaches to explore the configuration space of bilayer structures for ground states of the asymmetric Wigner bilayer system for the entire parameter space of the system. For selected values of the system parameters the relevant configuration space to draw ground state candidate structures from may be geometrically highly restricted (as is the case for the DI and RT_lt_s families). Very general optimization strategies can then be rather inefficient and need to be fine-tuned for the particular problem at hand. In Subsection 3.1.6 we demonstrated, that it is possible to construct an effective model specifically for RT_lt_s bilayer structures, which is capable of predicting the internal energy of the related structures with high accuracy and precision purely based on counting the occurrence of special local motives in the configurations. A promising route to proceed with the search for ground states of the asymmetric Wigner bilayer system could be to (i) identify important local motives of the relevant bilayer structures at selected values of the system parameters and (ii) to construct a (more) general effective model of the system, possibly relying on machine learning potentials. (iii) Structural energy minimization procedures based on identifying the energetically most favorable combination of such local motives (at given values of the system parameters) may then potentially go well beyond the capabilities of existing numerical structure-optimization tools.

As a final remark, we want to emphasize that it is quite remarkable, how many parallels can be drawn between ground state configurations of the asymmetric Wigner bilayer system real-world atomistic structures known from literature (cf. Subsections 3.1.7 and 3.1.8). In general, modern data-driven machine-learning algorithms [209] might be of great help to systematically investigate scientific literature for more similarities between the ground state configurations of the asymmetric Wigner bilayer systems and real-world atomistic structures.

Supramolecular Ordering of Complex Molecules under Electrochemical Conditions

As an obvious next step of our studies on the computational predictions of supramolecular ordering of complex molecules under electrochemical conditions discussed in Section 3.2 we aim at performing a more quantitative analysis of the ground-state self-assembly of the investigated $PQPClO_4$ system. In an effort to significantly reduce the computational costs of our approach, we will employ a further coarse-grained model of the supramolecular system, which maintains the essential features of the complex molecules: larger sub-units of the PQP^+ molecules (such as the aromatic rings) are approximated by disk-shaped units, or by oblate ellipsoids, carrying higher electrostatic moments. With this, we hope to perform more efficiently global ground state searches for a much finer grid of the system parameters as performed in Section 3.2. First numerical experiments in this direction have already been conducted.

In the course of the above-mentioned improvements to the quantitative analysis of the studies performed in Section 3.2, we also plan to extend our investigations to other, related supramolecular systems that have already been investigated experimentally by our collaborators in Refs. [82, 85]: the supramolecular ordering of polyaromatic cations

$PQPC_6^+$ (*i.e.*, PQP^+ molecules with an additional linear C_6H_{13} tail) with organic- (benzene sulfonate, BS^-, and anthraquinone disulfonate, $AQDSA^{2-}$) and inorganic-anions (perchlorate, ClO_4^-) at the Au(111)/octanoic acid interface has previously been examined. With adequate (minor) adaptations our approach can directly be applied to these other supramolecular systems. In contrast to the $PQPClO_4$ system investigated in Section 3.2 the $(PQPC_6)_2AQDSA$ and the $PQPC_6BS$ systems have experimentally largely been studied without modifications of the surface potential. Furthermore, especially the size of the $AQDSA^{2-}$ anions is comparable to the size of the $PQPC_6^+$ cations which makes these units better visible in the experimentally obtained scanning tunneling microscope images, as opposed to the highly mobile ClO_4^- units. Consequently, a comparison between the experiment and theory should be easier and may allow us to exploit the predictive power of our approach to investigate the surface self-assembly scenarios of these molecules at different values of the applied electric field.

Conceptually different yet very promising approaches to tackle global energy minimization problems of supramolecular lattices at solid interfaces are based on machine-learning and optimal design theory, such as *Surface Adsorbate polyMorph Prediction with Little Effort* (*SAMPLE*) [90] and *Bayesian Optimization Structure Search* (*BOSS*) [120, 121]. Quantitative investigations of such intricate problems as the complex monolayer to bilayer transition of supramolecular lattices at solid–liquid interfaces under electrochemical conditions, addressed in this thesis, might come within reach by employing suitably adapted variants of these (or similar) methods to our problem.

Self-assembly of Ionic Liquid Crystals into Smectic Phases

During the course of this thesis, a related project together with R. Wanzenböck in the scope of a diploma thesis emerged concerning the ground state self-assembly capacities of ionic liquid crystals [213], *i.e.*, of anisotropic, charged particles of ellipsoidal shape in three-dimensional lattice geometries.

Such ionic liquid crystals have attracted a steadily increasing interest during recent years both in academic and in industrial research. With their anisotropic shape and their ability to carry charges, they combine properties of charged particles and liquid crystals which are, for instance, reflected in their complex self-assembly capacities, making them technologically highly interesting [217, 218]. Ionic liquid crystals often form columnar or smectic phases and may consequently display high conductivity into selective spatial directions. Thus, they are promising candidates for anisotropic electrolytes in batteries [348–350] or for electrolyte constituents in solar cells [351, 352].

With this first study [213] on ordered ground-state structures of such complex particles – where we employed and extended tools based on memetic evolutionary algorithms (cf. Subsection 2.2.4) – we gained a better understanding of the ground state self-assembly of these anisotropic, charged units. To this end, we tested and recorded the impact of several model parameters on the self-assembly scenarios of these particles in R. Wanzenböck's diploma thesis; further investigations are envisaged.

Reinforcement Learning of Self-Navigation Strategies for Active Microswimmers

Another very promising project concerning self-navigation strategies in active matter systems [353] was initiated together with A. Zöttl and was elaborated together with M. Hübl in the scope of a project and diploma thesis [354].

We considered a simple model of a biological (or artificial) microswimmer [355], consisting of three concentrically aligned beads linked by two muscles in a hydrodynamic environment. With the help of reinforcement-learning [123, 124], specifically, by employing the *NeuroEvolution of Augmented Topologies* (*NEAT*) genetic algorithm [356, 357], we identified efficient swimming gaits, *i.e.*, periodic, non-reciprocal contraction and expansion patterns of the two muscles: a neural-network-based sense-response machinery evolved via NEAT allowed the microswimmer to self-navigate in a complex environment. A corresponding publication of our findings is currently under peer review [207].

As a next step, we plan to investigate self-navigation strategies or collective phenomenon of more complex microswimmer models compared to the one-dimensional active microswimmer described in Ref. [207]. In general, this topic is fascinating and challenging since it combines many aspects related to physics, chemistry, biology, computer science and artificial intelligence.

Acknowledgments

First of all, I would like to express my profound gratitude to my advisor Prof. Gerhard Kahl for giving me the opportunity to work under his supervision over the past few years. I am thankful for his trust and the guidance that I received during my Ph.D., and for the encouragement regarding my extracurricular experiments with liquid matter.

I am grateful for the financial support during my Ph.D., initially via projects of Gerhard Kahl and later via a DOC-fellowship of the Austrian Academy of Sciences. Also, I would like to acknowledge the granted computation time at the Vienna Scientific Cluster and at the supercomputing center in Jülich.

Furthermore, I am thankful for the many fruitful collaborations that emerged during my Ph.D., which allowed me to work with and learn from experts in different fields. I would like to especially thank Marek Mihalkovič, with whom I had the privilege to intensively collaborate either in Vienna, Bratislava or Modra. The numerous chats about physics, computer science or the thermodynamics of global economics always inspired me and I am grateful for his support in many ways. Also, I would like to thank Michael Walter, who was kind enough to host me for a week in Freiburg, Stijn F. L. Mertens for many interesting discussions and Emmanuel Trizac and Martial Mazars for hosting me in Paris on several occasions.

Necessary for a healthy scientific environment are not only external collaborations but also the general scientific setup in the own lab. I truly enjoyed being part of the soft matter theory group at the TU Wien for the past few years. I want to thank all the members of the group for the interesting seminars, the intense yet amusing discussions during lunch and coffee breaks and especially for taking care of *the* lemon-tree whenever I could not make it to the office.

Of course, I am thankful for my family, my loved ones and my friends, who kept me going during this intense time. I particularly want to mention Anela Lolić. I am grateful for her support both on a scientific and on a personal level and for her patience with me. She has always had an open ear for my thoughts and ideas and encouraged me to face new challenges in my own way. My brothers, Andreas and Thomas Hartl, being ten and eleven years older than me, took a great part in my education when I was little and have become great sources of life-experiences when I grew older. Already in kindergarten, Andreas challenged me with math problems while Thomas introduced me to computers, for which I am honestly thankful.

Eventually, I would like to deeply thank my father, Sigmund Hartl, whom I miss from the bottom of my heart, and my mother, Martha Hartl, who is the bravest and most warm-hearted person I know. Without their support, without their trust and without their love I would have neither been able to achieve a goal like this nor to become the person that I am today.

A Supplementary Theory and Information

A.1 The Asymmetric Wigner Bilayer System

A.1.1 Unsupervised Clustering of Order Parameters

In a first attempt to automatically cluster the structural data of the asymmetric Wigner bilayer system from Refs. [62–64] we applied the DBSCAN algorithm [243] described in Subsection 2.4.3 on the data set of feature vectors $\mathbf{X}^{(\mathrm{asym})} = (\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\mathrm{asym}}})$, with $\mathbf{x}_j \in \mathbb{R}^{N_f}$ defined by Eq. (3.11), of all configurations in the literature database of the asymmetric Wigner bilayer structures [62–64]. The DBSCAN algorithm is, in principle, able to group data points, $\mathbf{x}_j$, into clusters of data points with similar feature vectors, solely relying on local density measures of the data points of the data set $\mathbf{X}^{(\mathrm{asym})}$ in the feature space (*i.e.*, on an ϵ-neighborhood of the data points, cf. Subsection 2.4.3) spanned by the $N_f = 30$ features f_i introduced via Eq. (3.11). A minimum number of N_m data points needs to be specified which the DBSCAN algorithm recognizes as a cluster. In contrast to k-means clustering [239–242] discussed in the main text Subsection 3.1.4.5 the number of clusters does not need to be specified in advance for the DBSCAN algorithm.

In Fig. A.1 we present the zero temperature ground state phase diagram of the asymmetric Wigner bilayer system [62–64] in the (η, A)-plane with structure labels identified via DBSCAN of the nine dimensional data points of the latent space representation $\mathbf{L}^{(\mathrm{asym})}$ of the data set $\mathbf{X}^{(\mathrm{asym})}$ (corresponding to the leading nine principal component directions [238] of $\mathbf{X}^{(\mathrm{asym})}$, cf. Subsection 3.1.4.4). In total, 229 families of structures (for the DBSCAN parameters $\epsilon = 0.5$ and $N_m = 10$) have been identified by the DBSCAN algorithm, by far too many to present in a reasonable way in the scope of this thesis. DBSCAN can be used as an additional tool for phase identification purposes and can be put into relation with the results from k-means clustering (cf. Figs. 3.14 and 3.18), but it is obvious that DBSCAN is not the method of choice for our purposes.

Although the DBSCAN algorithm is able to resolve many phases correctly also many so-called "noise points", *i.e.*, data points which cannot be associated to a cluster by the algorithm – arise from this analysis, since DBSCAN is very sensitive to noise in the data, especially if the feature space dimension is large; such a sensitivity to noise also implies that data points near a second order transition, which is related to smoothly varying order parameters or features of the respective data points, may (falsely) be considered as noise points by the DBSCAN algorithm. DBSCAN is more reliable when studying lower-dimensional data (e.g. a data set with two or three features) which can be structured by local density measures in their respective frame of reference. A combination of t-SNE analysis (cf. Subsection 2.4.4) and DBSCAN, similar to the approach in the Supplementary Information of Ref. [86], could be a promising route to tackle the phase classification scheme of the asymmetric Wigner bilayer system.

In this thesis, the method of choice to identify families of ground states of the asymmetric Wigner bilayer system is given by k-means clustering in combination with the concept of adjusted mutual information [245, 246] (see Subsection 3.1.4.5). In Fig. A.2 we represent the biased adjusted mutual information score, $S(\mathbf{k}_i, \mathbf{k}_j|\mathbf{L}^{(\mathrm{sym})})$ defined by Eq. (3.12), for a total number of 40 independent K^*=14-means clustering results (left)

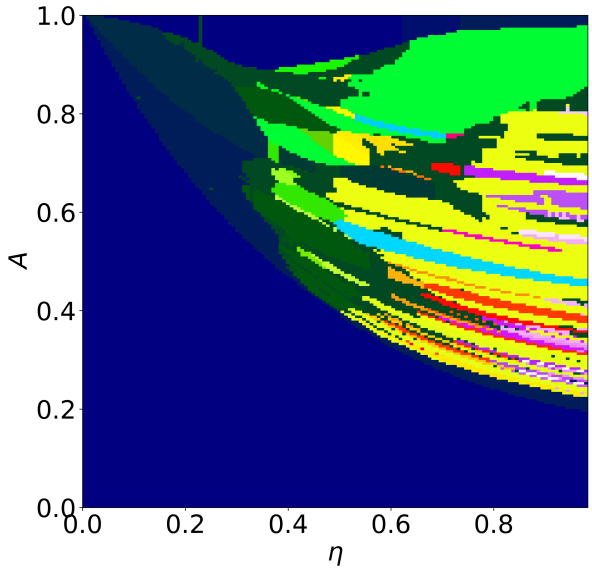


Figure A.1: Zero temperature ground state phase diagram of the asymmetric Wigner bilayer system in the (η, A)-plane [62–64] labeled (via color coding) by the results of a DBSCAN [243] analysis of the nine dimensional data points of the latent space data set $\mathbf{L}^{(\mathrm{asym})}$ (corresponding to the nine principal component directions [238], cf. Subsection 3.1.4.4 of the data set $\mathbf{X}^{(\mathrm{asym})}$) of the database of asymmetric Wigner bilayer structures from literature [62–64]. Several phases from literature can be resolved (cf. Fig. 3.2) and new phases can be identified (such as the $\mathrm{TPH_1D_2}$ region shown in Fig. 3.49, the $\mathrm{S_2P_1}$ region depicted in Fig. 3.59 or the shield region highlighted in Fig. 3.61) but there also appear to be many "noise points" (not shown here) within the DBSCAN classification. A more careful analysis of all the emerging clusters is required but is out of the scope of this thesis.

and K^*=32-means clustering results (right); this figure needs to be compared with the adjusted mutual information scores, $I_K(\mathbf{k}_i, \mathbf{k}_j)$ defined in Eq. (2.84), of the same clustering results shown in Fig. 3.10.

A.1.2 Ground State Symmetries of Clustering Results: Order Parameters and Principal Components

In Subsection 3.1.4 we performed k-kmeans clustering [239–242] of the data set of order parameters, $\mathbf{X}^{(\mathrm{asym})} = (\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\mathrm{asym}}})$, which quantify the spatial symmetries of asymmetric Wigner bilayer structures from a database of $j = 1, \ldots, N_{\mathrm{asym}}$ different configurations, $\mathcal{X}_j$ (cf. Eq. (3.5)), from literature [62–64].

Each feature vector $\mathbf{x}_j = \{f_1(\mathcal{X}_j), \ldots, f_{N_f}(\mathcal{X}_j)\}$, defined by Eq. (3.11), represents a total number of $i = 1, \ldots, N_f$ features (or values of order parameters), $f_i(\mathcal{X}_j)$, of a particular structure, $\mathcal{X}_j$, from the database. To be more specific, each feature vector $\mathbf{x}_j$ represents a set of values of different realizations of bond orientational order parameters $\Psi^{(1,2,3,4)}_{[3,4,5,6,8,10,12]}(\mathcal{X}_j)$, defined in Subsection 3.1.4.1 based on Eq. (2.62), of the composition

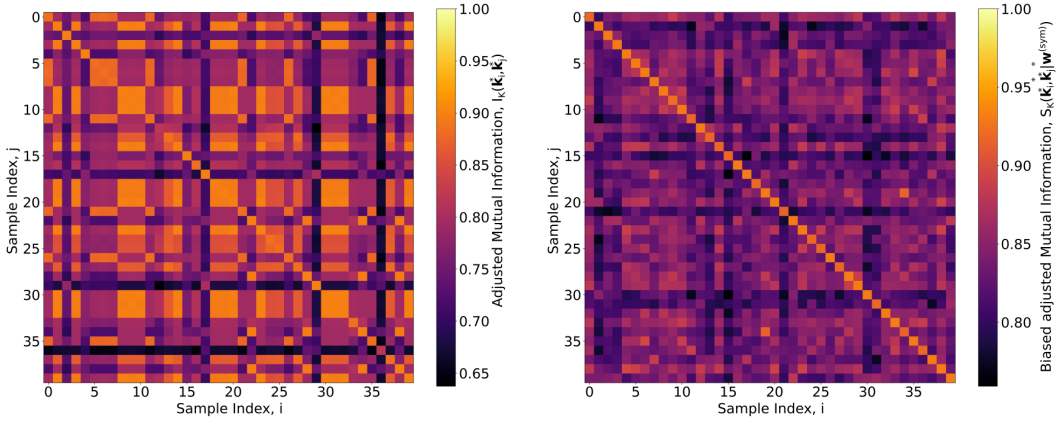


Figure A.2: Same as Fig. 3.10 but showing the biased adjusted mutual information score, $S(\mathbf{k}_i, \mathbf{k}_j|L^{(\mathrm{sym})})$, defined by Eq. (3.12).

$x(\mathcal{X}_j) = N_2(\mathcal{X}_j)/N(\mathcal{X}_j)$, defined by Eq. (3.8), and of the intralayer nearest neighbor ratio order parameter $g_r(\mathcal{X}_j)$, defined by Eq. (3.10), of a particular configuration $\mathcal{X}_j$.

Via principal component analysis we projected the data set, $\mathbf{X}^{(\mathrm{asym})}$, defined in the $N_f = 30$-dimensional feature space onto an $N_l^* = 9$-dimensional latent space. Each data point $\mathbf{l}_j \in \mathbb{R}^{N_l^*}$ from the latent space representation, $\mathbf{L}^{(\mathrm{asym})} = (\mathbf{l}_1, \ldots, \mathbf{l}_{N_{\mathrm{asym}}})$, of the data thereby corresponds to a data point, $\mathbf{x}_j \in \mathbb{R}^{N_f}$, in the data set $\mathbf{X}^{(\mathrm{asym})}$ and hence to a structure, $\mathcal{X}_j$, in the database [62–64]. We then performed k-means clustering on the latent space representation, $\mathbf{L}^{(\mathrm{asym})}$, of the structural database [62–64] in order to assign each structure from the database to one of a total number of K clusters (see Subsection 3.1.4.5 for details).

In that way, we were able to identify a K^*=32-means clustering of the structural database from literature [62–64] (see Fig. 3.18, for details on the "asterisk" notation see Subsection 3.1.4.5). Thus, we could label each of the $j = 1, \ldots, N_{\mathrm{asym}}$ structures in the database with an identifier, $c_j \in 1, \ldots, 32$, and thereby algorithmically group the $N_{\mathrm{asym}} \approx 64000$ different structures into 32 families of structures, which we refer to as the $\mathbf{k}_{32}^{*c=1}, \ldots, \mathbf{k}_{32}^{*c=32}$ structural families, respectively.

Here we present in Figs. A.3 to A.6 characteristic values (mean, median, boundaries and extreme values) of the order parameters related to each of the $\mathbf{k}_{32}^{*c=1}, \ldots, \mathbf{k}_{32}^{*c=32}$ families of structures of asymmetric Wigner bilayer configurations; in Figs. A.7 and A.8 we present characteristic values of the corresponding principal components of the $\mathbf{k}_{32}^{*c=1}, \ldots, \mathbf{k}_{32}^{*c=32}$ families.

In this way, we can relate each $\mathbf{k}_{32}^{*c}$ family of structures with characteristic symmetries.

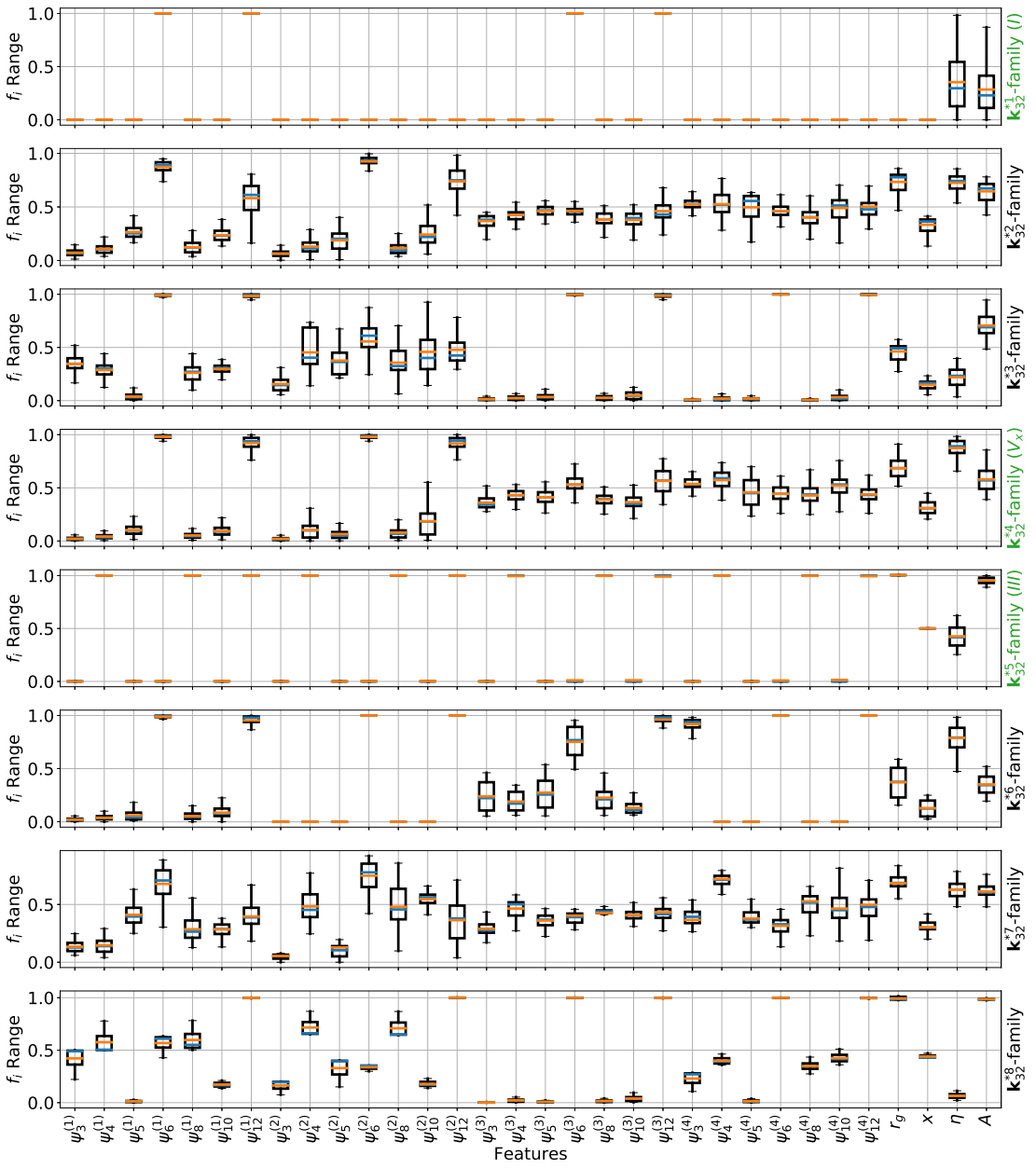


Figure A.3: Characteristic values of the order parameters, $\mathbf{x} = \{f_1, \ldots, f_{N_f}\}$ defined by Eq. (3.11), and of the system parameters, η and A, (all listed along the horizontal axis) of the ground state configurations of the asymmetric Wigner bilayer system [62–64] associated with the $\mathbf{k}_{32}^{*c}$ families of structures (see text, Subsection 3.1.4.5 and Fig. 3.18) for $c = 1$ to $c = 8$ in eight corresponding "boxplot"-panels from top to bottom. The labels of the $\mathbf{k}_{32}^{*c}$ families are indicated on the right-hand-side of each panel, respectively; whenever a $\mathbf{k}_{32}^{*c}$ family coincides with a phase from literature [62–64] (cf. Table 3.1 and Fig. 3.2) the corresponding $\mathbf{k}_{32}^{*c}$ label is emphasized in green and extended with the associated literature phase label. The median (horizontal orange lines), the mean (horizontal blue lines), the first and third quartiles (black boxes) and the whiskers (black vertical lines) of the different order parameters and of the related system parameters of all ground state configurations associated with the $\mathbf{k}_{32}^{*c}$ families of the asymmetric Wigner bilayer structures from the literature database [62–64] are presented; the first and third quartiles describe the characteristic range of the order parameters, *i.e.*, the coordinate values of the corresponding feature space directions where 25% of all data points (*i.e.*, the feature space representations of the corresponding ground state structures) are located at smaller and at higher coordinate values, respectively. The whiskers emphasize the minimum and the maximum values of the corresponding features.

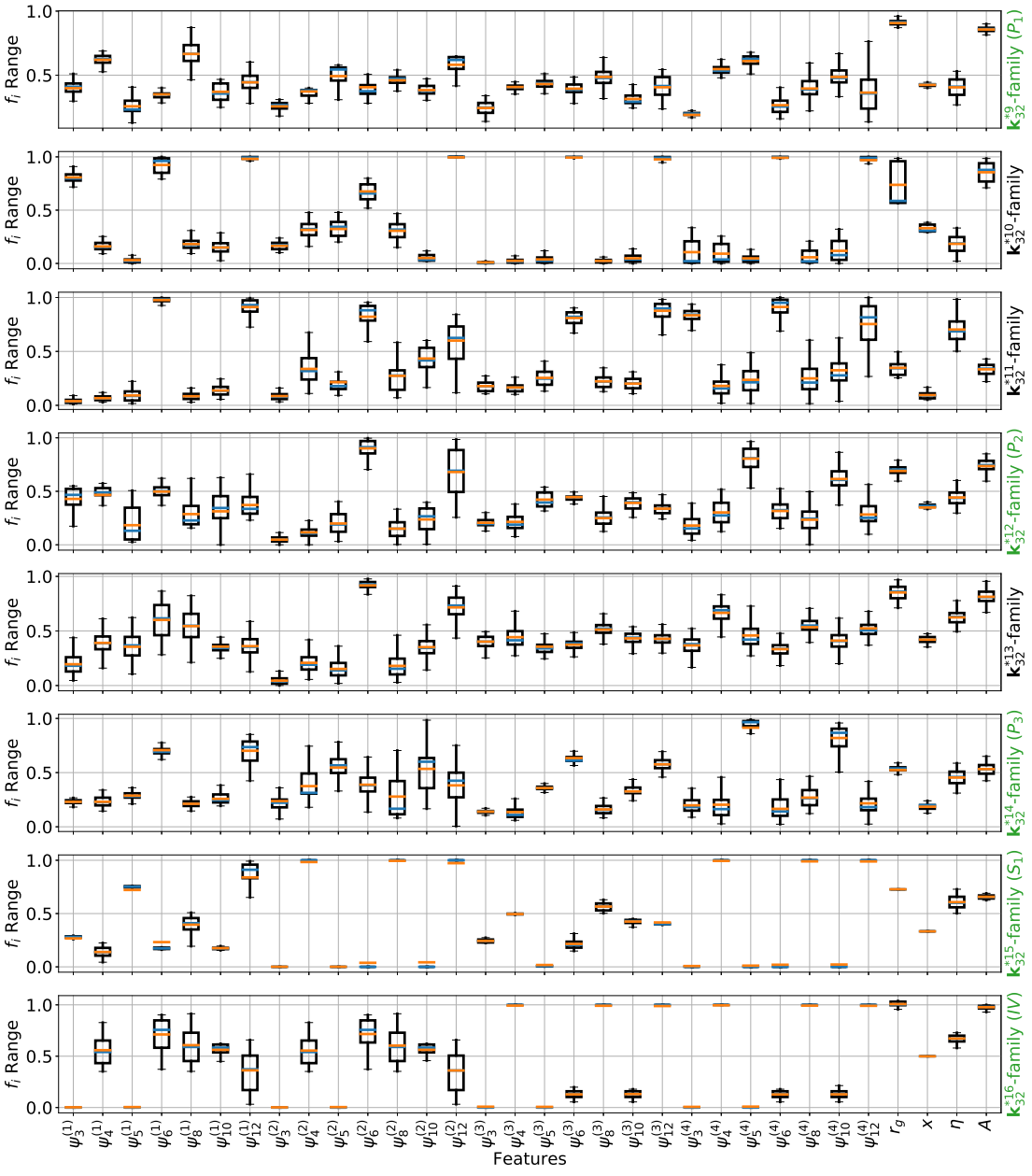


Figure A.4: Same as Fig. A.3 but for $\mathbf{k}_{32}^{*c}$ families with $c = 9$ to $c = 16$ from top to bottom.

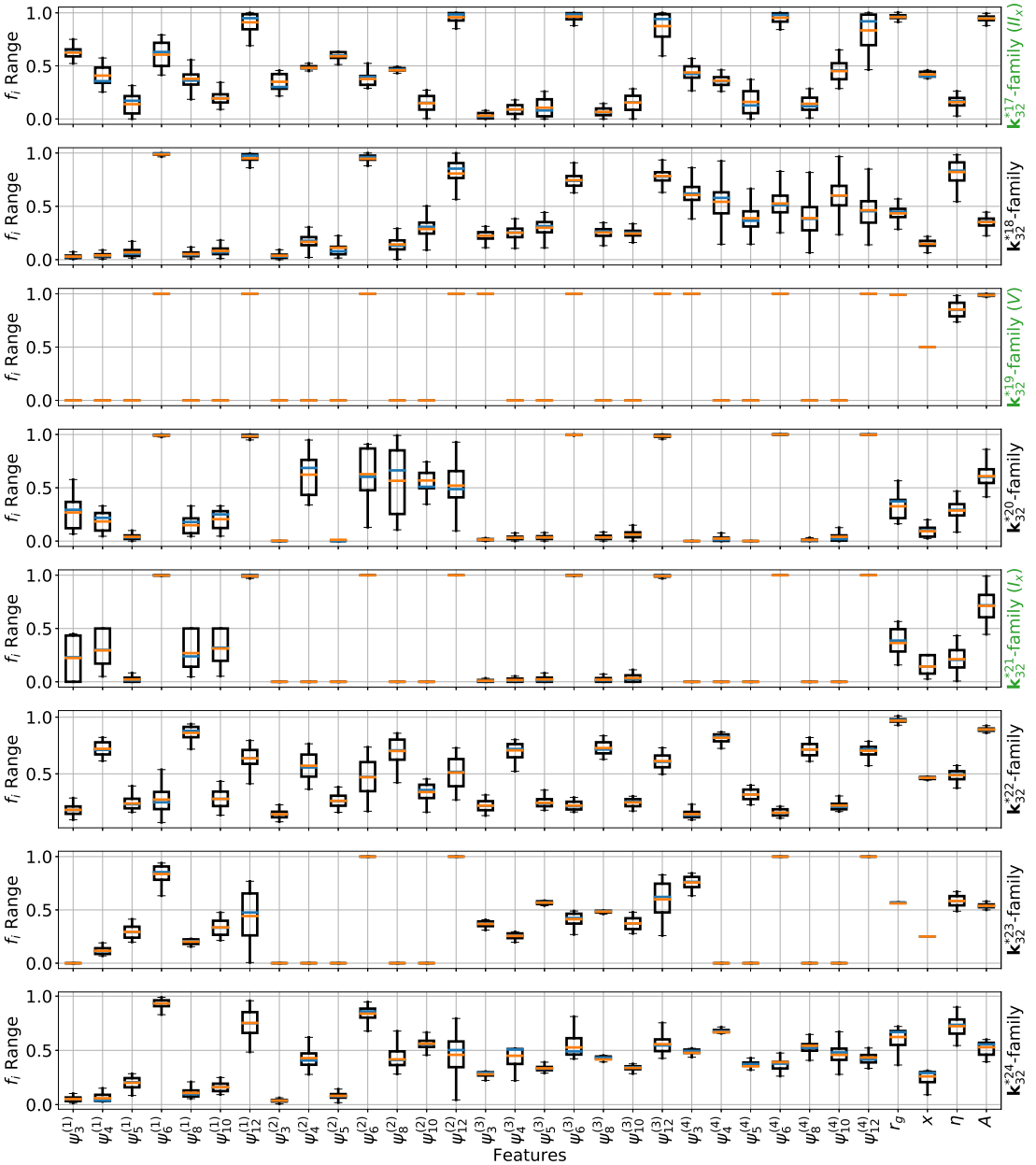


Figure A.5: Same as Fig. A.3 but for $\mathbf{k}_{32}^{*c}$ families with $c = 17$ to $c = 24$ from top to bottom.

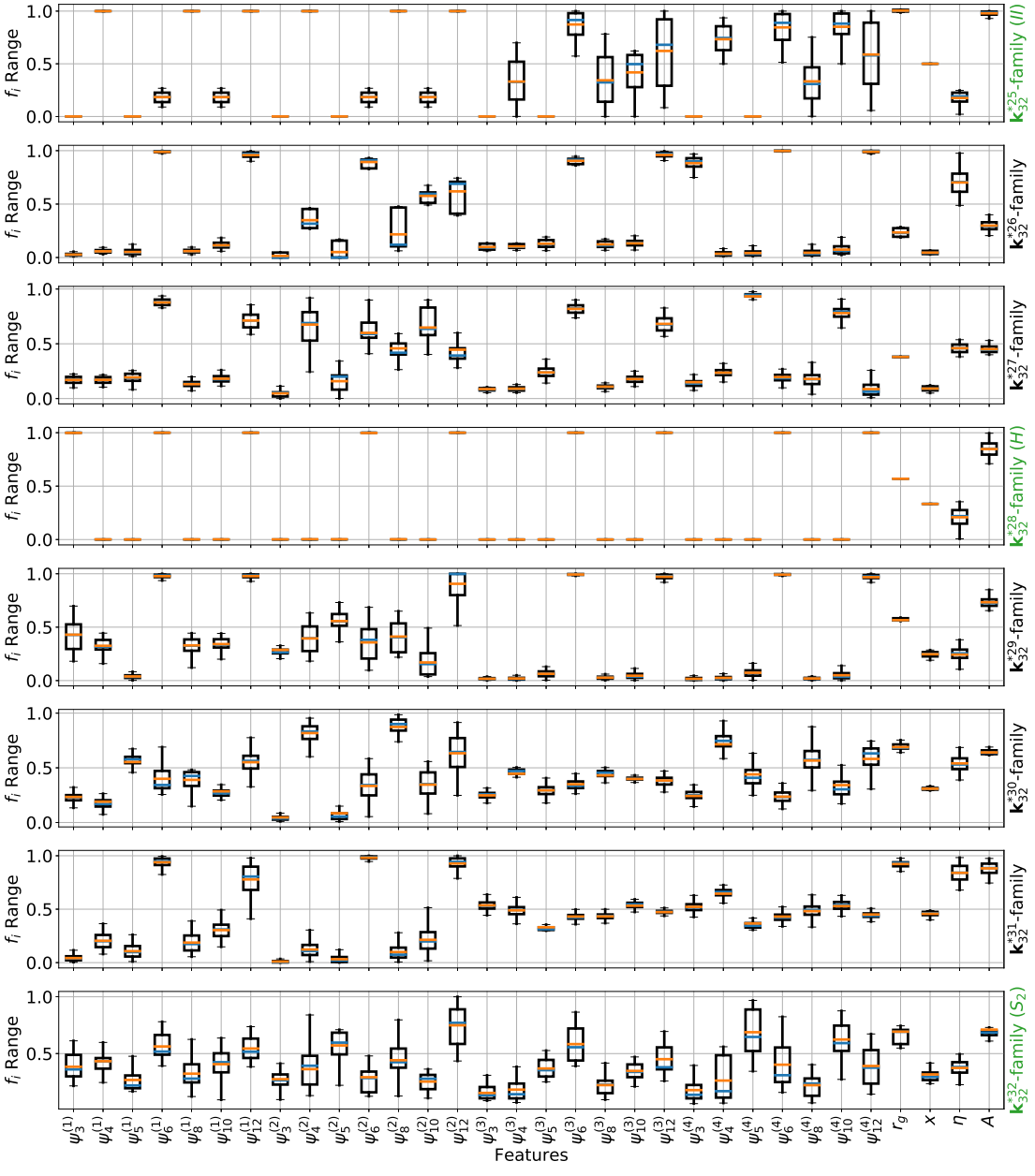


Figure A.6: Same as Fig. A.3 but for $\mathbf{k}_{32}^{*c}$ families with $c = 25$ to $c = 32$ from top to bottom.

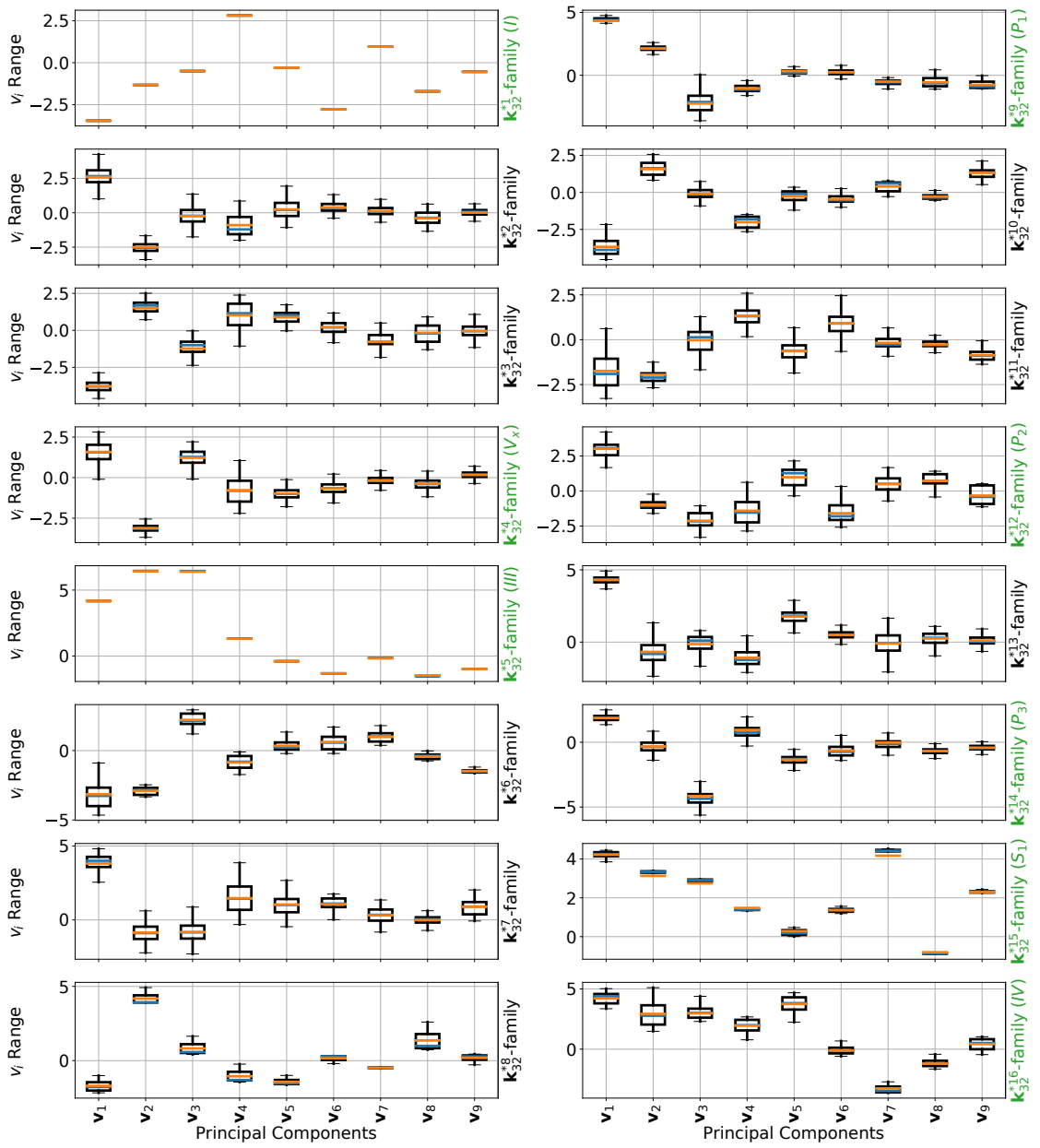


Figure A.7: Similar to Fig. A.3, but showing the characteristic values of the leading nine principal components $\mathbf{v}_1, \ldots, \mathbf{v}_9$ – instead of all 30 features – associated to the ground state configurations of the asymmetric Wigner bilayer system [62–64] which are related to the $\mathbf{k}_{32}^{*c}$ families of structures for $c = 1$ to $c = 16$. The nine principal components – representing the frame of reference of the latent space representation of the data, $\mathbf{L}^{(\mathrm{asym})}$ – are visualized in Fig. 3.8 to emphasize the directions they represent in the feature space (and thereby quantify the relation between order parameters and a particular principal component direction). The characteristic values for η and A for the different $\mathbf{k}_{32}^{*c}$ families are presented in Figs. A.3 and A.4.

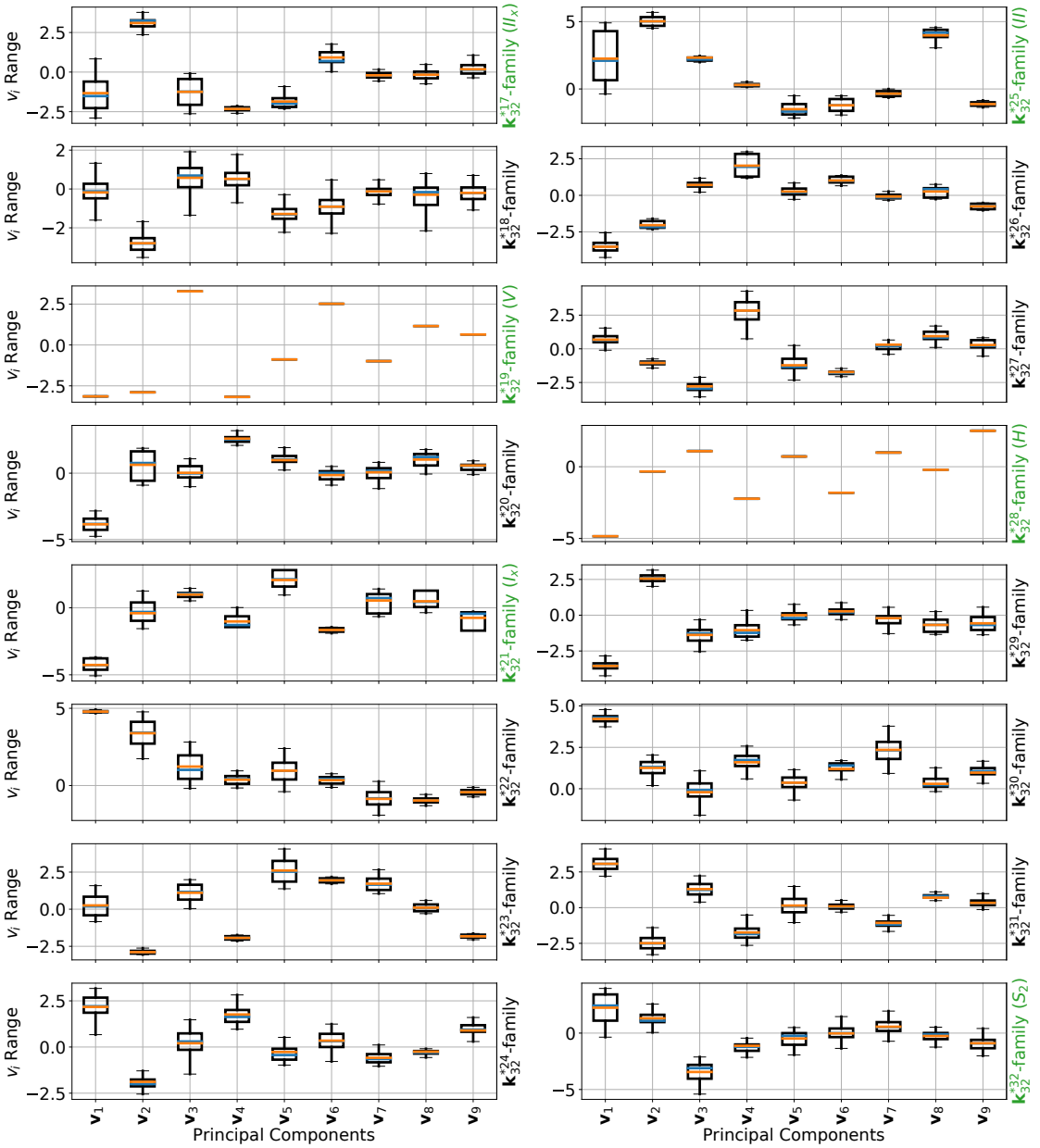


Figure A.8: Same as Fig. A.7 but for the $\mathbf{k}_{32}^{*c}$ families with $c = 17$ to $c = 32$. The characteristic values for η and A for the different $\mathbf{k}_{32}^{*c}$ families are presented via Figs. A.5 and A.6.

A.1.3 Mechanical Stability of Super Square-Triangle Tilings

Similar to the average particle position deviation $\langle \Delta_{xy} \rangle$, defined by Eq. (3.15), we define here the maximum deviation, $\max(\Delta_{xy})$, of a relaxed structure of the asymmetric Wigner bilayer system to the initial structure as

$$\max(\Delta_{xy}) = \max(|\tilde{\mathbf{r}}_i - \mathbf{r}_i|) \text{ for } i = 1, \cdots, N, \tag{A.1}$$

where the $\tilde{\mathbf{r}}_i$ represent the $i = 1, \ldots, N$ particle positions within the unit cell of the initial structure and $\mathbf{r}_i$ the particle positions of the corresponding relaxed structure. Similar to the presentation of $\langle \Delta_{xy} \rangle$ in Fig. 3.31 we here illustrate in Fig. A.9 the results of $\max(\Delta_{xy})$ related to the dodecagonal *type I* family of structures depicted in Fig. 3.29.

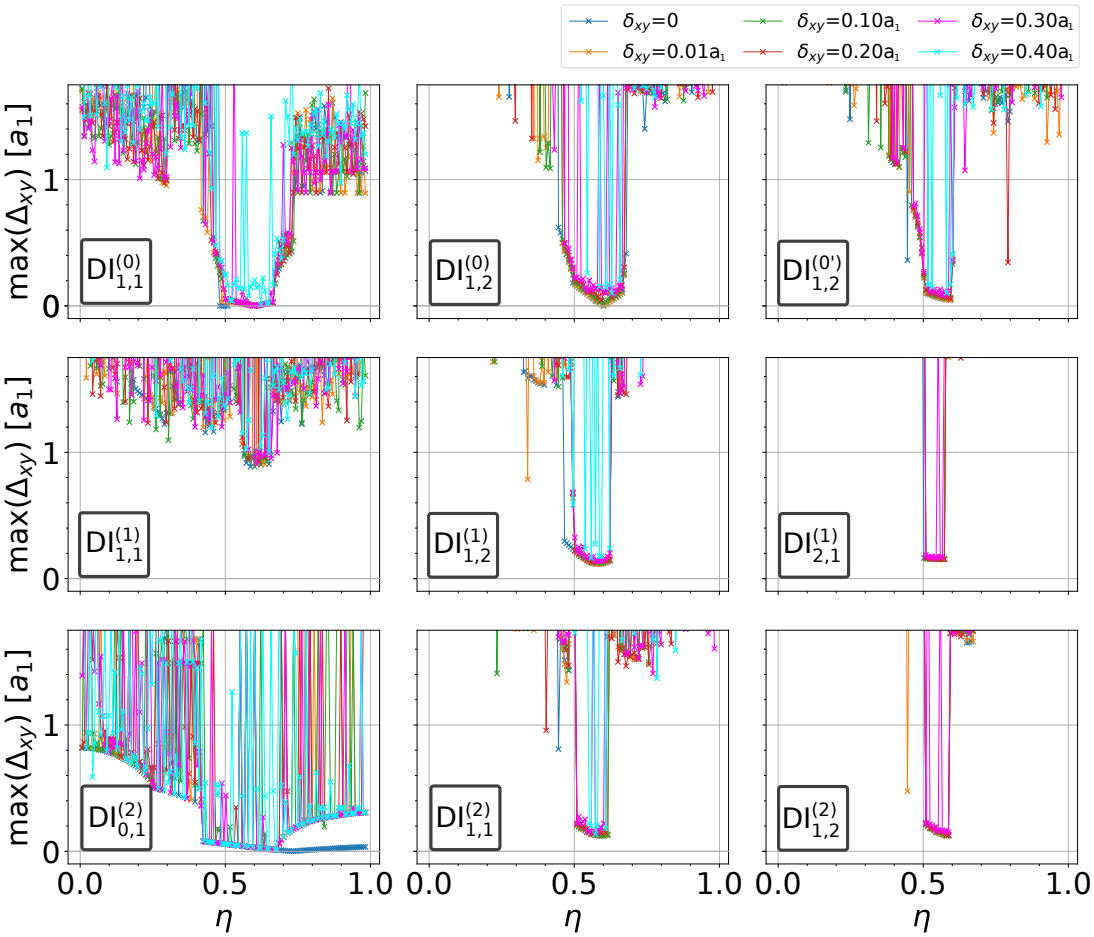


Figure A.9: Same as Fig. 3.31 but showing the maximum value, $\max(\Delta_{xy})$ defined by Eq. (A.1), of the deviation of a relaxed asymmetric Wigner bilayer structure to the initial structure.

A.1.4 Rectangle-Large–Triangle-Small–Triangle decoration

The rectangle-large–triangle-small–triangle ($\mathrm{RT}_l\mathrm{t}_s$) decoration scheme introduced in Subsection 3.1.6 can practically be employed on perfect square triangle tilings as follows (see Fig. A.10):

- Start from an arbitrary square triangle tiling (such as, for instance, the layer one or the layer two tiling of the DI structures[1] shown in Fig. 3.29) and assign it as layer two tiling of an otherwise empty bilayer structure (this initial layer two configuration is refered to as seed tiling).
- Identify all triangles and squares in the seed tiling, considering periodic boundary conditions.
- Separate triangles in the seed tiling into two types by following the *flip*-type convention from the main text (cf. Subsection 3.1.6):
  - For flip-type $F = 0$ label all triangles sharing one edge with the vertical axes as $\mathbf{t}_m$ and all remaining triangles as $\mathbf{t}_o$ (related to $\mathbf{t}_m$ by 90° rotations).
  - For flip-type $F = 1$ exchange the above labeling convention, $\mathbf{t}_m \leftrightarrow \mathbf{t}_o$.
- Perform layer one decoration for all $\mathbf{t}_m$ and $\mathbf{t}_o$ labeled triangles as follows:
  - All triangles labeled $\mathbf{t}_m$ are *mid-edge* decorated with layer one particles.
  - All triangles labeled $\mathbf{t}_o$ are decorated *off-edge*: layer one particles are placed at the projected geometric *center* of a $\mathbf{t}_o$ triangle and *off-edge* at the projected mirror positions of the geometric center with respect to every edge of the triangle.
- Decorated squares in the following way:
  - Two parallel edges of a square, labeled $\mathbf{s}_m$, are *mid-edge decorated* if one of them is a common edge with a $\mathbf{t}_m$ triangle; the respective orthogonal edges of the square, labeled $\mathbf{s}_o$, are *off edge decorated* by placing two particles at a distance of $a_2/(2\sqrt{3})$ from the mid-edge point of $\mathbf{s}_o$ inside and outside of the square on a line orthogonal to $\mathbf{s}_o$, a_2 being the side length of the square.
  - Two parallel edges of a square, labeled $\mathbf{s}_o$, are *off-edge decorated* if one of them is a common edge with a $\mathbf{t}_o$ triangle and the respective orthogonal edges of the square, labeled $\mathbf{s}_m$, are *mid-edge decorated*.
  - Neighboring squares are iteratively decorated according to their neighboring edge decoration, *i.e.*, $\mathbf{s}_m$ and $\mathbf{t}_m$ decorated edges as well as $\mathbf{s}_o$ and $\mathbf{t}_o$ decorated edges are respectively considered as equivalent.
  - If no triangles are present an arbitrary square edge is assigned as $\mathbf{s}_m$ which uniquely defines the structure (which is symmetric with respect to $F = 0$ and $F = 1$).

[1] If the layer two tiling is used, we include central particles of dodecagons in layer one and assign them as layer two vertices such that we start with a perfect square-triangle tiling.

- Fold all decorated vertices back into the unit cell and remove multiply decorated sites (keep only unique decorations at the same position).
- Scale the unit cell to satisfy the unit-area per particle constraint we are following in this thesis (without the loss of generality [62]).
- Finally, relax the entire structure at $\eta \approx 0.226$ and $A \approx 0.765$, which deforms all squares into rectangles with a long, $|\mathbf{s}_m| \to l$, to short, $|\mathbf{s}_o| \to s$, edge-length ratio of $l/s = 2/\sqrt{3}$ and which deforms the two types of triangles accordingly such that $|\mathbf{t}_m| \to l$ and $|\mathbf{t}_o| \to s$.

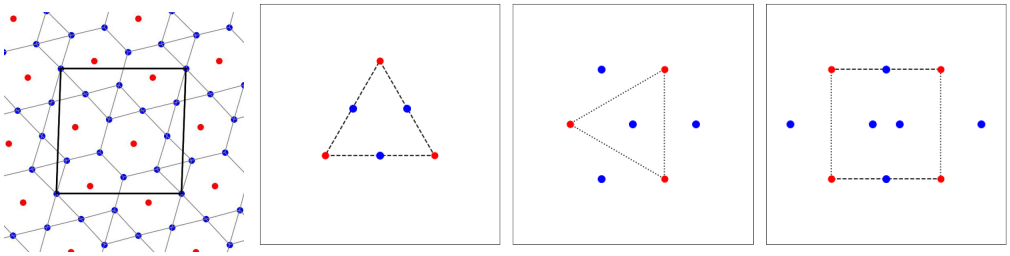


Figure A.10: Left panel: I_x-Cairo ground state structure of the asymmetric Wigner bilayer system at $\eta \approx 0.226$ and $A \approx 0.765$ [276]; red dots correspond to layer two particles, blue dots to layer one particles. From second to left to right panel: Decoration scheme of layer two triangle tiles and the rectangular tile (red vertices) from left to right: (i) *mid-edge decorated* triangle, $\mathbf{t}_m$, (ii) *center and off-edge decorated* triangle, $\mathbf{t}_o$, (iii) square with horizontal *mid-edge decorated*, $\mathbf{s}_m$, and vertical *off-edge decorated* edges, $\mathbf{s}_o$. The different edge decorations, $\mathbf{t}_m$ and $\mathbf{t}_o$ as well as $\mathbf{s}_m$ and $\mathbf{s}_o$, are additionally emphasized by dashed and dotted lines: neighboring tiles must have matching edge decorations. Triangles which are related by rotations of 60° are equally decorated in the entire tiling.

In Figs. A.11 and A.12 we present configurations where the above introduced decoration scheme is applied for the REMC simulations discussed in Figs. 3.43 and 3.44, respectively: the top panels show the decoration scheme applied on layer two square-triangle tilings, the bottom panels show the relaxed $\mathrm{RT}_l\mathrm{t}_s$ configurations of the respective configurations of the top panels (relaxed at $\eta \approx 0.1061$ and $A \approx 0.8778$). This supports our claim that not only special square-triangle tilings which feature dodecagonal motives (cf. Fig. 3.41) but also random square-triangle tilings can always be transformed into $\mathrm{RT}_l\mathrm{t}_s$ structures. In fact, we never experienced square-triangle tilings in the REMC simulations depicted in Figs. 3.43 and 3.44 which could not be transformed into $\mathrm{RT}_l\mathrm{t}_s$ structures.

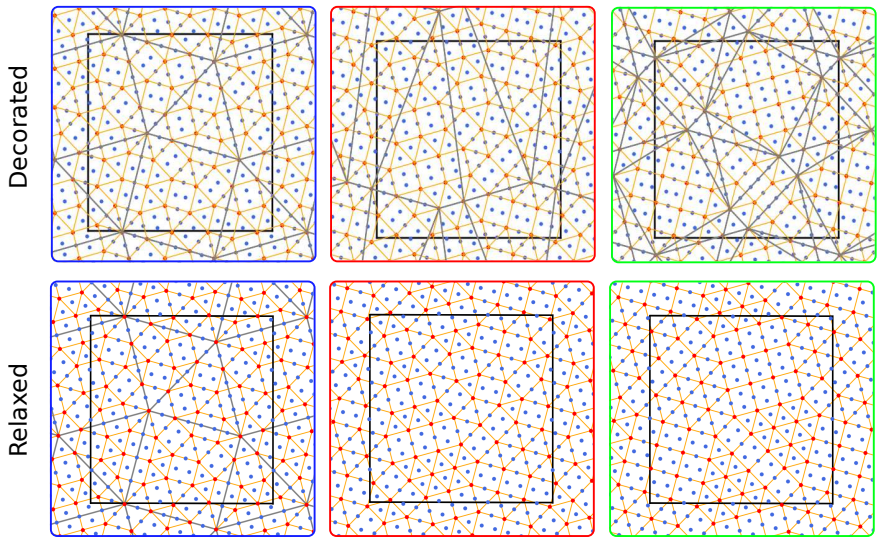


Figure A.11: Top panels: Decoration scheme of configurations shown in Fig. 3.43 (with corresponding red, green, blue color coding) using a perfect square-triangle tiling in layer two (red dots connected by orange lines indicating Voronoi nearest neighbors; layer one particles are colored blue). Bottom panels: Relaxed $\mathrm{RT}_l\mathrm{t}_s$ configurations of the respective top panel structures at $\eta \approx 0.1061$ and $A \approx 0.8778$. For the top panels also connections between hexagonal vertices in layer two are drawn via thick gray lines, in the bottom row these lines are only shown for the super-snub-rectangular structure in the bottom left panel.

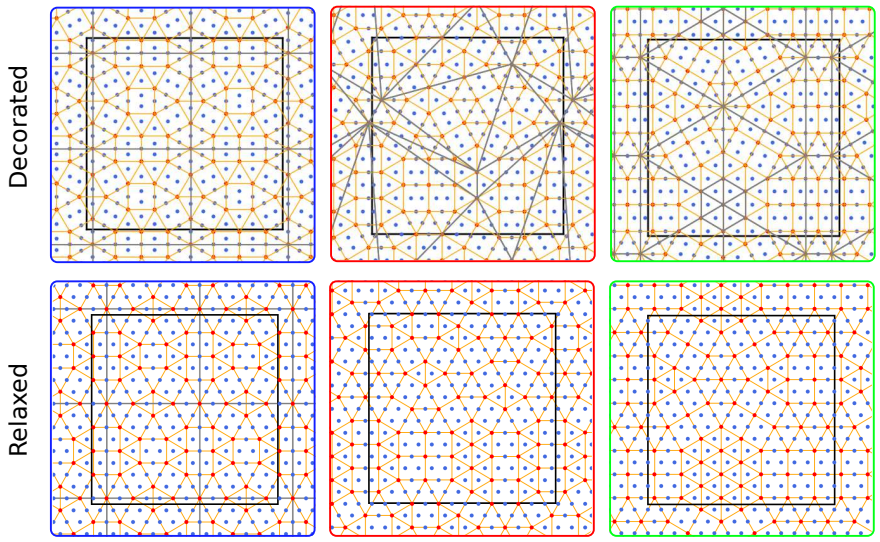


Figure A.12: Same as Fig. A.11, but for the configurations shown in Fig. 3.44.

A.1.5 Towards Constructive, Defect-Free Inflation

For some square-triangle based DI approximants (such as $DI^{(1)}_{1,1}$ and $DI^{(2)}_{2,1}$ depicted in Figs. 3.29 and 3.30 in Subsection 3.1.5) it is geometrically impossible, owed to symmetry considerations with respect to the periodic boundary conditions and the composition of layer one and two, to resolve all atomic defects of neighboring square-square pairs with a common edge in layer one. For dodecagonal *type I* structures this leads to mechanical instabilities, cf. the central left panels in Figs. 3.29, 3.31 and A.9.

Here we will investigate rules how to avoid such atomic defects directly during Stampfli-inflation [252, 254]. As suggested in the discussion of $DI^{(1)}_{2,2}$, depicted in the right panel of Fig. 3.30, a combination of Stampfli inflation and the stacking procedure (cf. Fig. 3.25) is key to accomplishing our goal: for a tiling being entirely composed of non-overlapping dodecagons which are, themselves, composed of equilateral squares and triangles as depicted in Fig. 3.24, there is only two ways how neighboring dodecagons can be arranged in order to avoid square-square neighbors, see Fig. A.13.

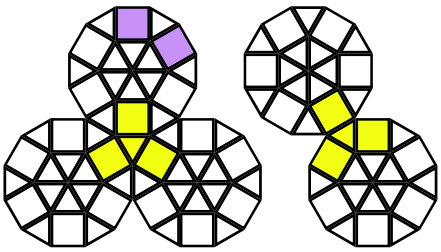


Figure A.13: The two possible ways to attach dodecagons defect free (without square-square neighbors with a common edge) in an equilateral square-triangle tiling. Problematic environments for inflation are colored in yellow and in purple, respectively.

This narrows down the number of possible vertices which are subject to Stampfli-inflation. As can be seen in the left panel of Fig. A.13 triangles can easily be inflated. The problem is decorating inflated squares with dodecagons: the orientations of such dodecagons have to alternate by 30° on the corners of the same edges of a square. In case of a environment composed of a central triangle with three neighboring squares (yellow colored environment in Fig. A.13) this is not possible since at least two dodecagons of the inflated triangle must have the same orientation, cf. left panel in Fig. A.14. This can be resolved by additionally stacking the squares and triangles directly (cf. Figs. 3.24 and 3.25) as we demonstrate in the second to left panel in Fig. A.14. Now the corners of the central triangle can be decorated with dodecagons of the same orientation and, consequently, the three neighboring squares can be inflated by dodecagons of clockwise alternating orientation of 60° and 30°.

The right two panels in Fig. A.14 depict a defect free and a defective inflation of a pair of squares which share one corner and exhibit an relative angle of 60° (cf. purple environment in the left panel of Fig. A.13). Note that the decoration with dodecagons of the right two panels in Fig. A.14 are related by a simple 30° rotation of each dodecagon, which are color-coded[2] according to Fig. 3.24. The defective inflation in the right most

[2]Dodecagons with squares on the vertical axis (on top and bottom) are colored blue, those with triangles on the vertical axis are colored green.

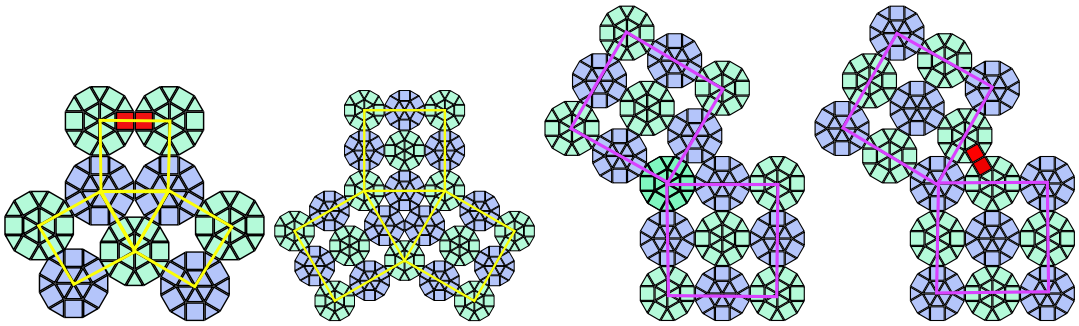


Figure A.14: Left two panels show (from left to right) a (necessarily) defective square-triangle environment (red squares) and a defect free square-triangle environment after Stampfli-inflation when additional stacking is applied to the yellow square-triangle environment (cf. Fig. A.13). The second to right panel shows the unique decoration of the purple environment (an inflated and stacked pair of squares which share one corner under an opening angle of 60°, cf. Fig. A.13) with 17 dodecagons which avoids atomic defects in the square-triangle tiling at the smallest length scale. The right most panel demonstrates a defective inflation (see red squares) of the purple environment depicted in Fig. A.13, although each purple square – for itself – is inflated defect-free.

panel of Fig. A.14 highlights that not all triangle decorations are possible in a defect-free manner but depend on the orientation (*i.e.*, $30° + n \times 60°$ or $n \times 60°$, n being an integer) of the triangles which are subject to inflation and of the orientations of the inscribed dodecagons.

With the building-blocks shown in Fig. A.14 we can compose inflated and stacked defect-free *super-stacked* dodecagons from scratch. We present in Fig. A.15 the two possible ways (which we found to be possible) for constructing such super-stacked clusters, which are themselves related by a rotation of 30°.

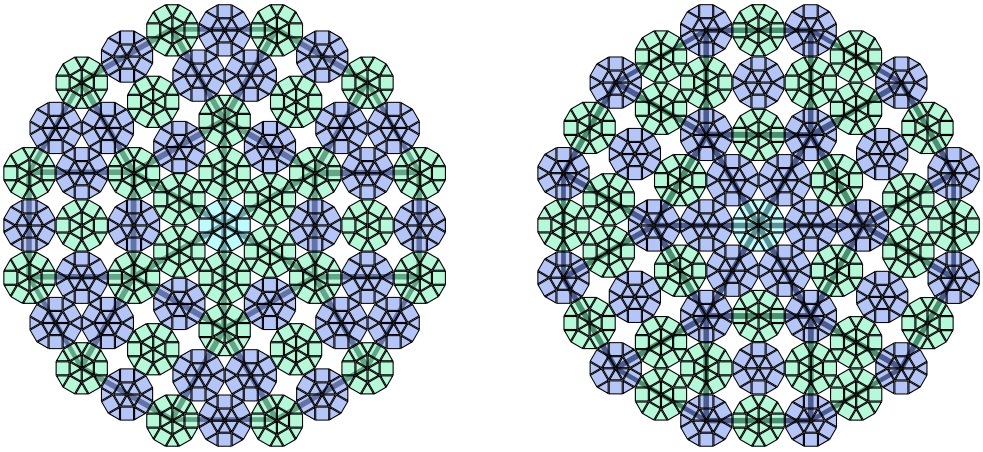


Figure A.15: Super-dodecagon with 30° rotation (left and right) with unique inflation and stacking procedure (except for the central dodecagon) thereby avoiding atomic defects by construction. Dodecagons with squares on the vertical axis (on top and bottom) are colored blue, those with triangles on the vertical axis are colored green; the central dodecagon is colored cyan, indicating that no defects can emerge by rotating this particular dodecagon by 30°. Stacked and inflated super-tiles are indicated by thick gray lines in the background.

In Fig. A.16 we present a square-triangle tiling with overlapping super-stacked dodecagons on a hexagonal lattice. Note that this tiling cannot be described by the introduced Stampfli-inflation above, which does not cover overlapping dodecagons. Here, in an initial inflation step the unit cell of a triangular tiling is inflated by a factor of $(1+\sqrt{3})$ such that the inscribed neighboring dodecagons share one square (cf. black lines in Fig. A.16); we label this structure $\mathrm{DI}^{(0)}_{I=1.5,S=2}$. This tiling can directly be used as potential candidate structure (after DI decoration) to compete for the ground state of the asymmetric Wigner bilayer system in the S_1 region (cf. Subsection 3.1.5). However, given the rather hopeless values of the energies of the DI structures investigated in Subsection 3.1.5 with respect to the ground state energies of the asymmetric Wigner bilayer system suggested in Refs. [62–64], we refrained form including this structure into our analysis.

In Fig. A.17 we present a square-triangle tiling with non-overlapping super-stacked dodecagons on a hexagonal lattice, which we label $\mathrm{DI}^{(0)}_{2,2}$. Interestingly, it is not easily possible to directly compose squares from super-stacked dodecagons without causing defects on the atomic length scale such that we cannot present super-stacked square nor snub-square configurations.

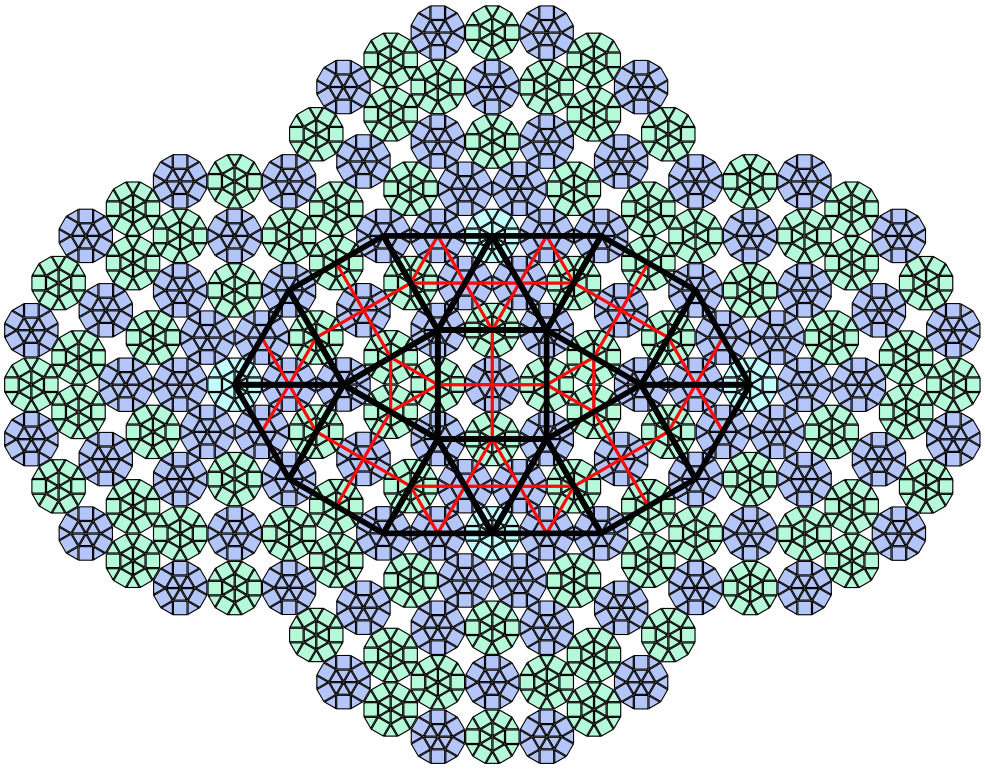


Figure A.16: Defect-free overlapping super-stacked dodecagons on a hexagonal *type 0* lattice. Black lines indicate the squares and triangles which are subject to inflation, red lines indicate stacking. Dodecagons with squares on the vertical axis (on top and bottom) are colored blue, those with triangles on the vertical axis are colored green; the central dodecagon of the super-stacked dodecagons is colored cyan, indicating that no defects can emerge by rotating this particular dodecagon by 30°.

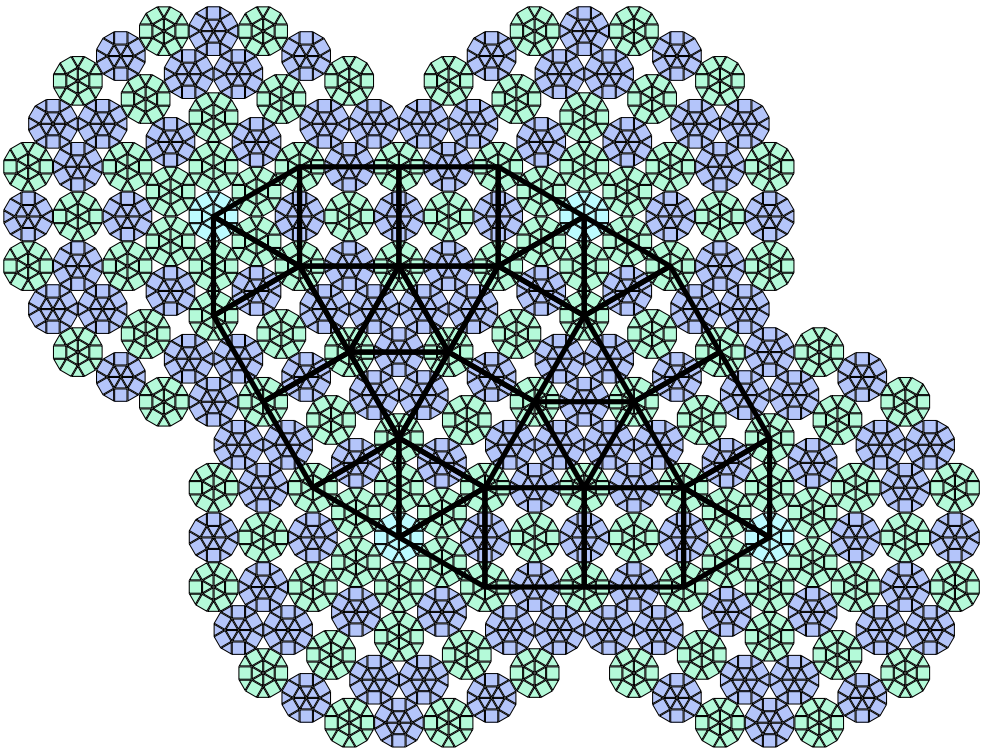


Figure A.17: Same as Fig. A.16 but for a defect-free non-overlapping super-stacked dodecagon on a hexagonal *type 0* lattice.

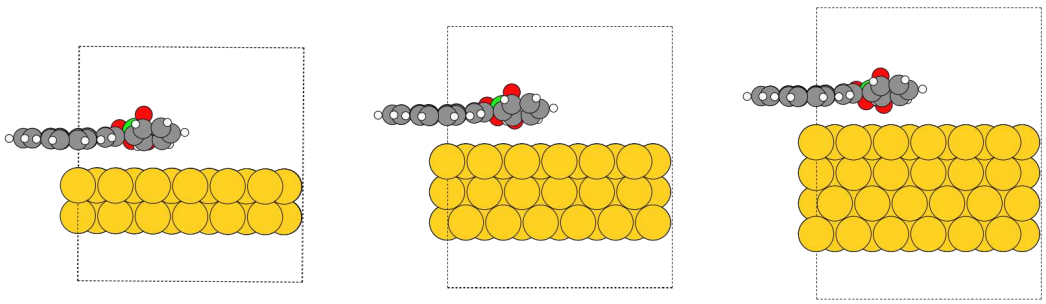
A.2 Predictive Supramolecular Self-Assembly

In the Appendix A.2 – which is largely based on the Supplementary Information of Ref. [86] – we have collected relevant information which might considerably deteriorate the readability of the main text Section 3.2 if it were placed there; still, the details presented in this document might be of relevance for an interested reader of the main text Section 3.2. For simplicity we have used in Appendix A.2 exactly the same Section-headings as in the main text Section 3.2; this will hopefully help to establish the appropriate association between the respective text passages.

A.2.1 The System and its Representations

A.2.1.1 Convergence of Theoretical Calculations

The DFT-based binding energies of the PQP^+ and ClO_4^- ions were calculated[3] for different numbers of gold layers with a constraint of fixed gold layers in an effort to study the convergence of the results with respect to the number of layers that build up the gold surface. The values for the binding energies, obtained for the different cases, are specified below the respective panels of Fig. A.18; the panels themselves provide schematic plots for the different types of gold layers. In the related DFT calculations the van der Waals scaling is not considered, *i.e.*, $\omega_S \equiv 1$ is chosen in Eq. (3.24) in Section 3.2.



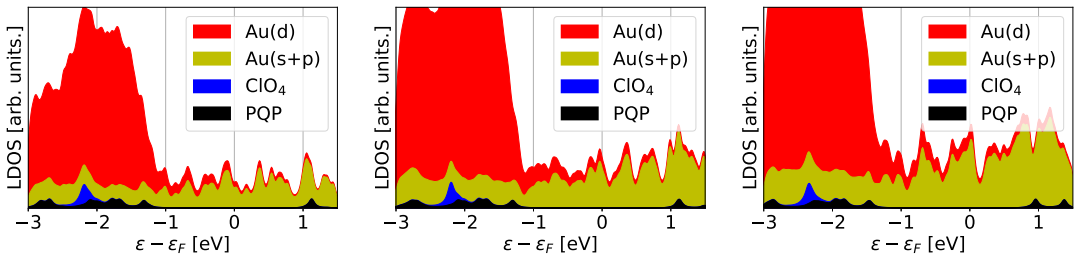
(a) BE = -2.69 eV **(b)** BE = -2.66 eV **(c)** BE = -2.71 eV

Figure A.18: Schematic views of PQP^+ and ClO_4^- ions, located above gold surfaces built up by different numbers of layers Below each panel the respective binding energies (BE) are specified.

The effect of fixing the gold atoms on the binding energy of PQP^+ and ClO_4^- ions to the surface if further examined. Allowing the uppermost gold layer to relax changed the binding energies to -2.73 eV, -2.69 eV and -2.73 eV for two, three and four layers, respectively. These values are well in accordance with the previously calculated values reported in Fig. A.18. One can therefore conclude that fixing the gold atoms has a negligible effect on the binding energy obtained.

[3] The calculations and contents of Appendix A.2.1.1 (corresponding to Subsection 3.2.2.2 of the main text) were primarily conducted by our collaborators O. Brügner, S. Sharma and M. Walter.

The calculated values for the binding energy obtained for two, three, and four layers of gold provide evidence that our choice for a two layer gold surface is sufficient to proceed with our calculation of the self-assembly scenarios of PQP^+ and ClO_4^- ions on this Au(111) surface. Local density of states were also plotted with respect to the energies of the corresponding states relative to the Fermi-level. The plotted density curves shows that the density of states for PQP^+ and ClO_4^- ions are quite similar irrespective of the number of layers of gold surface.



(a) two layers **(b)** three layers **(c)** four layers

Figure A.19: Local density of states (LDOS) relative to Fermi-level. The Kohn-Sham energies were projected on local atomic orbitals of Cl and O (ClO_4^-), C, N and H (PQP), and on gold s+p and d states, respectively.

Fig. A.19 compares the density of Kohn-Sham states projected on the atomic species for the models depicted in Fig. A.18. While the density of states of the gold part increases naturally with the number of gold layers, this has only marginal effects on the positions of the PQP^+ and ClO_4^- related states.

A.2.1.2 Angle-Axis Framework Expressing Rigid Body Orientations

In the optimization procedure put forward in Subsection 3.2.3 of the main text we rely on the angle-axis framework [336, 337] to express the orientation of rigid molecules within the lab-frame: closely related to the descriptions of rotations based on unit-quaternions [145, 358, 359], we introduce a three-component angle-axis vector, $\mathbf{P} = (P_1, P_2, P_3) = \theta\,\hat{\mathbf{P}}$, which defines an angle, $\theta = \sqrt{P_1^1 + P_2^2 + P_3^2}$, and a unit-vector, $\hat{\mathbf{P}}$, which represents the axis of the molecule; both are sufficient to describe any rotation of a rigid body in three dimensions. As discussed in Ref. [336] and following Rodrigues' rotation formula, the 3×3 rotation matrix $\mathbb{T}(\mathbf{P})$ associated with the angle-axis vector $\mathbf{P}$ is given by

$$\mathbb{T}(\mathbf{P}) = \mathbb{I} + (1 - \cos\theta)\,\tilde{\mathbb{P}}\tilde{\mathbb{P}} + (\sin\theta)\,\tilde{\mathbb{P}}; \tag{A.2}$$

here $\mathbb{I}$ is the 3×3 identity matrix and $\tilde{\mathbb{P}}$ is the skew-symmetric 3×3 matrix obtained from the components of the vector $\hat{\mathbf{P}}$ via

$$\tilde{\mathbb{P}} = \frac{1}{\theta}\begin{pmatrix} 0 & -P_3 & P_2 \\ P_3 & 0 & -P_1 \\ -P_2 & P_1 & 0 \end{pmatrix}. \tag{A.3}$$

In order to transform the coordinates of an atom, $\mathbf{r}_m^{(I)}$, defined in the center-of-mass system of molecule I to its lab-frame position, $\mathbf{r}_m$, the following transformation needs to be realized:

$$\mathbf{r}_m = \mathbf{R}_I + \mathbb{T}(\mathbf{P}_I) \cdot \mathbf{r}_m^{(I)}; \tag{A.4}$$

here $\mathbf{R}_I$ is the center-of-mass coordinate of molecule I and $\mathbb{T}(\mathbf{P}_I)$ is the rotation matrix associated with the angle-axis vector $\mathbf{P}_I$ of molecule I, as defined above.

A.2.1.3 Short-Range Potentials and Parametrization

The short-range Mie potential, defined in Eq. (3.28) of the main text Subsection 3.2.2.4, can be considered as a generalization of the Lennard-Jones (LJ) interaction [339]: if the exponents of the repulsive and attractive parts of the potential are chosen as $\gamma_{ij}^{(R)} = 12$ and $\gamma_{ij}^{(A)} = 6$ the amplitude C_{ij}, given by Eq. (3.29) of the main text, becomes $C_{ij} = 4$ and the Mie-potential reduces to the well known LJ form.

During the fitting procedure put forward in Subsection 3.2.2.5 of the main text, it occurred at some instances that $\gamma_i^{(R)} < \gamma_i^{(A)}$. In such a case the defining equation for the C_{ij} (see Eq. (3.28) of the main text Subsection 3.2.2.4) guarantees that both the repulsive and the attractive parts of the potential maintain their respective features.

In Fig. A.20 we depict the PQP^+ and the ClO_4^- ions using the actual values for the fitted Mie length parameters, σ_i (listed in Table 3.5 of the main text), as van der Waals radii. In Table A.1 we list – for comparison – also the LJ length parameters for the atomic entities of the PQP^+ and the ClO_4^- molecules as they are commonly used in literature.

	σ_H	σ_C	σ_N	σ_O	σ_{Cl}
Mendeleev	2.2	3.4	3.1	3.04	3.5
Alvarez	2.4	3.54	3.32	3.0	3.64
Bondi	2.4	3.4	3.1	3.04	3.5
Dreiding	3.195	3.8983	3.6621	3.4046	3.9503
mm3	3.24	4.08	3.86	3.64	4.14
uff	2.886	3.851	3.66	3.5	3.947
$\sigma_i^{(LJ)}$	2.243	3.658	3.743	2.865	5.953
$\sigma_i^{(Mie)}$	2.236	3.703	3.328	2.428	4.956

Table A.1: LJ length parameters σ_i (in Å) for hydrogen (H), carbon (C), nitrogen (N), oxygen (O), and chlorine (Cl) from literature as defined in the *mendeleev*-python-module (*ver. 0.5.1*) and, as comparison, model parameters $\sigma_i^{(LJ)}$ and $\sigma_i^{(Mie)}$ from Table 3.5 in the main text Subsection 3.2.2.4.

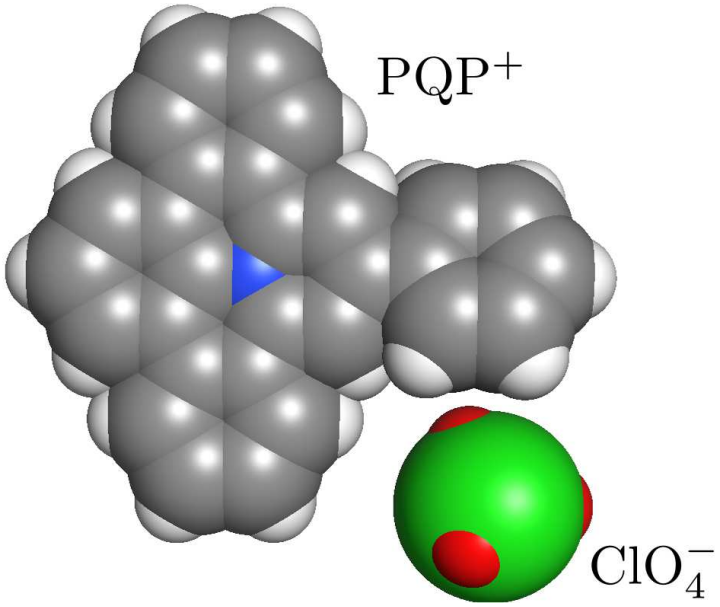


Figure A.20: Schematic representation of a PQP^+ and a ClO_4^- ion using the actual values of the fitted Mie length parameters $\sigma_i^{(\mathrm{Mie})}$ for the short-range interactions introduced in Subsection 3.2.2.4 of the main text: atomic entities are shown as spheres with their diameters fixed by their respective optimized $\sigma_i^{(\mathrm{Mie})}$-values; these entities are colored according to the following scheme: hydrogen (white), carbon (gray), nitrogen (blue), chlorine (green), and oxygen (red).

In Tables A.2 and A.3 we list the coordinates of all atomic units and their associated partial charges (extracted via a Bader analysis [342, 360]) obtained from the relaxed DFT-structures of the PQP^+ and ClO_4^- ions which serve as rigid molecular blueprints in the main text Section 3.2. Since the PQP^+ cation is built up by 48 atomic units we have supplemented Table A.2 by Fig. A.21, indicating the labeling of the different atomic entities. In contrast, as the ClO_4^- molecule (see Table A.3) is only built up by five atomic units, we have refrained in this case from a schematic presentation of the molecule.

Table A.4 compares the Bader charges of single PQP^+ and ClO_4^- pairs in the gas phase and on the gold surface. This analysis reveals that the local charges on the ions are very similar in both environments and practically unaffected by the presence of the metal surface.

i	element	x_i	y_i	z_i	q_i	i	element	x_i	y_i	z_i	q_i
1	C	0.1267	1.2051	-0.0253	0.3807	25	C	-4.8220	0.0601	-0.0341	-0.0565
2	N	-0.5941	0.0077	-0.0830	-1.1646	26	H	-4.7122	-2.0727	-0.0499	0.1101
3	C	-2.0123	0.0254	-0.0936	0.3949	27	C	-2.7268	-3.7034	-0.1427	-0.0931
4	C	0.0988	-1.2077	-0.0664	0.4209	28	H	1.1435	-3.7411	-0.1704	0.1265
5	C	-0.5685	2.4836	-0.0704	0.0642	29	C	-0.6337	-4.9026	-0.1737	-0.1096
6	C	1.5102	1.1636	0.0881	-0.0066	30	H	-3.7189	3.8240	-0.1608	0.0990
7	C	-2.7073	1.2652	-0.0766	0.0497	31	C	-1.9108	4.9519	-0.1507	-0.0561
8	C	-2.7389	-1.1964	-0.0946	0.0300	32	H	0.0658	5.8448	-0.1588	0.0863
9	C	-0.6299	-2.4673	-0.1204	-0.0401	33	C	4.4671	1.0567	-0.2093	-0.0996
10	C	1.4854	-1.2014	0.0175	-0.0272	34	C	4.3714	-1.1639	0.7541	-0.0342
11	C	-1.9791	2.5216	-0.0929	-0.0412	35	H	-5.9114	0.0745	-0.0008	0.1122
12	C	0.1514	3.7017	-0.0977	-0.0507	36	H	-3.8146	-3.7276	-0.1477	0.0984
13	H	2.0403	2.1040	0.1899	0.1396	37	C	-2.0369	-4.9037	-0.1638	0.0333
14	C	2.2322	-0.0280	0.1088	0.0058	38	H	-0.0809	-5.8419	-0.2001	0.1258
15	C	-4.1099	1.2499	-0.0437	-0.0944	39	H	-2.4371	5.9066	-0.1831	0.1140
16	C	-4.1411	-1.1474	-0.0630	-0.0382	40	H	3.9741	1.9279	-0.6429	0.0888
17	C	-2.0415	-2.4705	-0.1202	0.0066	41	C	5.8578	1.0346	-0.1202	-0.0392
18	C	0.0574	-3.7037	-0.1539	-0.0731	42	H	3.8020	-2.0184	1.1245	0.0943
19	H	1.9993	-2.1557	0.0110	0.1087	43	C	5.7627	-1.1824	0.8408	-0.1172
20	C	-2.6331	3.7708	-0.1343	-0.0780	44	H	-2.5876	-5.8452	-0.1777	0.1138
21	H	1.2385	3.7123	-0.0996	0.0676	45	H	6.4355	1.8929	-0.4673	0.1010
22	C	-0.5084	4.9176	-0.1347	0.0077	46	C	6.5117	-0.0860	0.4007	-0.0581
23	C	3.7024	-0.0450	0.2216	-0.0045	47	H	6.2672	-2.0562	1.2563	0.1042
24	H	-4.6584	2.1881	-0.0183	0.0912	48	H	7.6004	-0.1048	0.4614	0.1070

Table A.2: Atomic units building up the PQP^+ ion, labeled by the index i according to the schematic view of the molecule presented in Fig. A.21. The positions of these entities (x_i, y_i and z_i, and all in Å), as they were obtained in a relaxed DFT-based configuration, are given with respect to the center-of-mass of the molecule, marked in this figure by a cross. Furthermore, the respective charges of the atomic units, q_i (given in units of the elementary charge e), are obtained in a Bader analysis [342]; these charges are directly transferred to our classical model of the PQP^+ molecule.

i	element	x_i	y_i	z_i	q_i
1	Cl	0.0000	0.0000	0.0000	2.6996
2	O	1.4732	-0.0020	0.0000	-0.9249
3	O	-0.4916	1.3888	0.0000	-0.9249
4	O	-0.4917	-0.6933	1.2034	-0.9249
5	O	-0.4917	-0.6933	-1.2034	-0.9249

Table A.3: Atomic units building up the ClO_4^- molecule. The positions of these entities (x_i, y_i and z_i, and all in Å), as they were obtained in a relaxed DFT-based configuration, are given with respect to the center-of-mass of the molecule, which coincides with the position of the oxygen atom. Furthermore, the respective charges of the atomic units, q_i (given in units of the elementary charge e), are obtained in a Bader analysis [342]; these charges are directly transferred to our classical model of the ClO_4^- molecule.

	on top (OT)	side by side (SBS)	SBS on gold support
PQP^+	1.0	0.9	0.9
N ($M = 1$)	-1.1	-1.1	-1.2
C ($M = 29$)	0.2	0.3	0.3
H ($M = 18$)	1.9	1.7	1.8
ClO_4^-	-1.0	-0.9	-1.0
Cl ($M = 1$)	2.6	2.5	2.7
O ($M = 4$)	-3.6	-3.5	-3.7
Au (M = 72)	—	—	0.1

Table A.4: Charges obtained in Bader analysis for PQP^+, ClO_4^- ions in the gas phase and supported by the gold surface. The *on top* (OT) and *side by side* (SBS) configurations correspond to the configurations of single $PQPClO_4$ pairs as in Fig. 3.65 a) and b) of the main text Subsection 3.2.2.5, respectively. The SBS configuration on gold is depicted in Fig. A.18a). The atom specific charges are summed values of all M atoms of the same type.

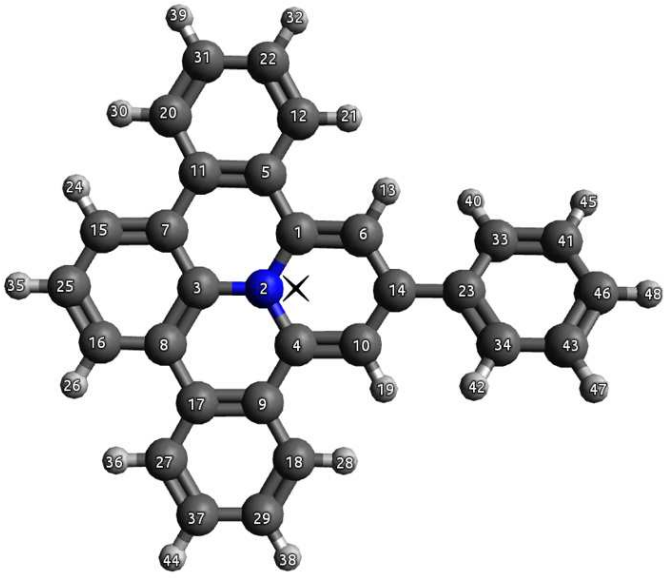


Figure A.21: Schematic view of the PQP^+ ion where its atomic constituents and the related bonds are depicted. The spheres are colored according to the respective chemical element: hydrogen (white), carbon (gray), and nitrogen (blue). The atomic constituents are labeled by indices $i = 1$ to $N_{PQP} = 48$; the positions of each of these entities (with respect to the center-of-mass of the molecule) are listed in Table A.2. The black cross marks the center of mass, $\mathbf{R}_{PQP}$, of the PQP^+ molecule.

A.2.1.4 Image Charges in a Solvent

We[4] may describe the electrostatic interaction between a charge distribution and a metallic surface by the method of image charges [158]. The electric field inside a metal is completely screened and therefore the field on the surface can only be perpendicular to the metal surface.

Inside media the Maxwell equation for the displacement field $\mathbf{D}(\mathbf{r}, t)$ reads

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) = -4\pi\epsilon_0 \rho(\mathbf{r}, t) \tag{A.5}$$

where ϵ_0 is the permittivity of the vacuum and $\rho(\mathbf{r}, t)$ a charge density. The relative permittivity ϵ_r of the medium connects to the electric field $\mathbf{E}(\mathbf{r}, t)$ to $\mathbf{D}(\mathbf{r}, t)$ via $\mathbf{D}(\mathbf{r}, t) = \epsilon_0\epsilon_r\mathbf{E}(\mathbf{r}, t)$. The Maxwell equation, Eq. (A.5), then becomes

$$\nabla \cdot \epsilon_r\mathbf{E}(\mathbf{r}, t) = -4\pi\rho(\mathbf{r}, t). \tag{A.6}$$

A spatially constant permittivity ϵ_r thus leads to

$$\nabla \cdot \mathbf{E}(\mathbf{r}, t) = -4\pi\frac{\rho(\mathbf{r}, t)}{\epsilon_r} \tag{A.7}$$

which is the Maxwell equation in vacuum with a charge density scaled by the relative permittivity.

A slab geometry with a planar, perfectly conductive lower surface of infinite extend, separating the slab region from the interior bulk region at $z = 0$ can be modeled by image charges. These are placed within the bulk according to Eq. (3.30) in the main text Subsection 3.2.2.4, *i.e.*, $\mathbf{r}_i = (x_i, y_i, z_i) \to (x_i, y_i, -z_i) = \mathbf{r}_{i'}$ and $q_{i'} \to -q_i$ and fulfill metallic boundary conditions. Including the constant electrostatic field E_z introduced in Eq. (3.33) in the main text Subsection 3.2.2.4 this setting corresponds to the electrostatic potential $\Phi(\mathbf{r})$ of the form

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\epsilon}\sum_{i=1}^{n} q_i \left(\frac{1}{|\mathbf{r} - \mathbf{r}_i|} - \frac{1}{|\mathbf{r} - \mathbf{r}_{i'}|}\right) + z\, E_z\, \Theta(z), \tag{A.8}$$

where $\Theta(z)$ is the Heaviside Theta (or step) function which is 0 if $z \leq 0$ and 1 otherwise. The potential vanishes at $z = 0$ in the entire (x, y)-plane, *i.e.*, $\Phi\left(\mathbf{r}_{xy} = (x, y, z = 0)\right) = 0$. In fact, every term in the sum vanishes separately as we have $|\mathbf{r}_{xy} - \mathbf{r}_i|^2 = (x - x_i)^2 + (y - y_i)^2 + (\pm z_i)^2 = |\mathbf{r}_{xy} - \mathbf{r}_{i'}|^2$, ensuring metallic boundary conditions [158].

[4] The content of Appendix A.2.1.4 was elaborated together with M. Walter.

A.2.2 Identifications of Self-Assembly Scenarios

A.2.2.1 Angle-Axis Gradient as Calculated from the Torque

The software package LAMMPS [93] allows to evaluate forces and torques of rigid molecules enclosed in a simulation box. Since we are interested in the gradient of the potential energy with respect to angle-axis vectors, $\mathbf{P}$, *i.e.*, $\nabla_{\mathbf{P}} U = \left(\frac{\partial U}{\partial P_1}, \frac{\partial U}{\partial P_2}, \frac{\partial U}{\partial P_3}\right)$, we present here the transformation which is required to transform a three-dimensional torque $\mathbf{T} = (T_x, T_y, T_z)$ to an angle-axis gradient, $\nabla_{\mathbf{P}} U$, or in a component-wise notation $\frac{\partial U}{\partial P_i} = \partial_i U$, using Latin indices $i = 1, 2, 3$ for three-dimensional vectors.

In LAMMPS orientations are expressed in terms of unit-quaternions via the four dimensional vector $\mathbf{Q}^{(4)}$ (using Greek indices $\nu = 0, 1, 2, 3$)

$$\mathbf{Q}^{(4)} = \left(\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\,\hat{\mathbf{P}}\right) = (Q_0, Q_1, Q_2, Q_3), \tag{A.9}$$

where, as described in Appendix A.2.1.2, $\theta = |\mathbf{P}|$ is the angle of rotation around the axis $\hat{\mathbf{P}} = \mathbf{P}/\theta$ and $|\mathbf{Q}^{(4)}| = 1$.

Following the documentation of LAMMPS [359] the resulting (four dimensional) torque vector, $\mathbf{T}^{(4)} = (0, T_x, T_y, T_z) = (0, \mathbf{T})$ on a rigid body is specified via

$$\mathbf{T}^{(4)} = -\frac{1}{2}\mathbb{S}^T_{4\times4}\nabla_{\mathbf{Q}^{(4)}} U + \mathbf{T}^{(4)}_{\text{int}}. \tag{A.10}$$

The internal torque $\mathbf{T}^{(4)}_{\text{int}}$, provided by LAMMPS [93], ensures that $T^{(4)} = (0, T_x, T_y, T_z)$, *i.e.*, $T^{(4)} = (0, \mathbf{T})$.

Henceforward the matrix index "4×4" indicates four-by-four matrices. In the above relation we have introduced the orthogonal skew-matrix $\mathbb{S}_{4\times4}$, given by

$$\mathbb{S}_{4\times4} = \begin{pmatrix} Q_0 & -Q_1 & -Q_2 & -Q_3 \\ Q_1 & Q_0 & -Q_3 & Q_2 \\ Q_2 & Q_3 & Q_0 & -Q_1 \\ Q_3 & -Q_2 & Q_1 & Q_0 \end{pmatrix}; \tag{A.11}$$

further, $\nabla_{\mathbf{Q}^{(4)}} U$ is the gradient of the potential energy with respect to the unit-quaternion vector $\mathbf{Q}^{(4)}$, or, alternatively, in a component-wise notation $\frac{\partial U}{\partial Q_\nu} = \partial_\nu U$, $\nu = 0, 1, 2, 3$.

Since we are dealing with rigid bodies and explicitly avoid intra-molecular interactions we can neglect $\mathbf{T}^{(4)}_{\text{int}}$ in Eq. (A.10); further, $\mathbb{S}_{4\times4}$ is an orthogonal matrix, thus $\mathbb{S}_{4\times4}\mathbb{S}^T_{4\times4} = \mathbb{S}^T_{4\times4}\mathbb{S}_{4\times4} = \mathbb{I}_{4\times4}$ where $\mathbb{I}_{4\times4}$ is the four dimensional unit matrix. Hence we can rewrite Eq. (A.10) as

$$\nabla_{\mathbf{Q}^{(4)}} U = -2\mathbb{S}_{4\times4}\mathbf{T}^{(4)}, \tag{A.12}$$

and further express $\partial_i U$ in terms of Eq. (A.12) using the chain rule

$$\partial_i U = \frac{\partial U}{\partial P_i} = \frac{\partial Q_\nu}{\partial P_i}\frac{\partial U}{\partial Q_\nu} = (\mathbb{Q}_{3\times4})_{i\nu}\partial_\nu U \quad i = 1,2,3 \quad \text{and} \quad \nu = 0,1,2,3; \tag{A.13}$$

above we have used the Einstein summation convention. For convenience we have introduced the 3×4 matrix $\mathbb{Q}_{3\times4}$ with components $\mathbb{Q}_{i\nu} = \frac{\partial Q_\nu}{\partial P_i}$, or equivalently, $\mathbb{Q}_{3\times4} = (\frac{\partial Q_0}{\partial \mathbf{P}}, \frac{\partial \mathbf{Q}}{\partial \mathbf{P}})$, with components that – using Eq. (A.9) – can be written as

$$\frac{\partial Q_0}{\partial \mathbf{P}} = \left(-\frac{1}{2\theta}\sin\frac{\theta}{2}\right)\mathbf{P}, \qquad \text{and} \tag{A.14}$$

$$\frac{\partial \mathbf{Q}}{\partial \mathbf{P}} = \frac{2}{\theta^2}\left(\cos\frac{\theta}{2} - \frac{1}{2\theta}\sin\frac{\theta}{2}\right)\mathbf{P}\cdot\mathbf{P}^T + \left(\theta^2\sin\frac{\theta}{2}\right)\mathbb{I}; \tag{A.15}$$

here $\mathbf{Q} = (Q_1, Q_2, Q_3)$, $\mathbf{P}\cdot\mathbf{P}^T = \theta^2(\tilde{\mathbb{P}}\tilde{\mathbb{P}} + \mathbb{I})$, where $\mathbb{I}$ is again the 3×3 identity matrix, the dot represents a dyadic product, and $\tilde{\mathbb{P}}$ is defined in Eq. (A.3).

With Eq. (A.12) and using $\mathbb{Q}_{i\nu}$, given by Eqs. (A.14) and (A.15), we can rewrite Eq. (A.13) as

$$\nabla_{\mathbf{P}} U = -2\,\mathbb{Q}\,\mathbb{S}\,T^{(4)} = \mathbb{P}\,\mathbf{T} \tag{A.16}$$

with the 3×3 matrix

$$\mathbb{P} = \frac{1}{\theta}\left[(\cos\theta - 1)\mathbb{I} + (\sin\theta - \theta)\,\tilde{\mathbb{P}}\right]\tilde{\mathbb{P}} - \mathbb{I}, \tag{A.17}$$

and with $\mathbf{T} = (T_x, T_y, T_z)$ being the torque in Cartesian coordinates in the lab-frame.

A.2.2.2 Order Parameters

In order to quantify the structural difference between configurations identified via the optimization procedure we associate a feature vector (*i.e.*, a set of order parameters), $\mathbf{x}$, to every configuration, $\mathcal{X}$, as defined by Eq. (3.31) in Subsection 3.2.2.4 of the main text. Here, we mainly rely on the so-called bond orientational order parameters [225–228], $\Psi_\nu(\mathcal{X})$, defined by Eq. (2.62) in Subsection 2.3.2 of the main text which provide information about the positional bond orientational order of ordered structures. Additionally, we use two variants of orientational order parameters [229] $\alpha(\mathcal{X})$ and $\beta(\mathcal{X})$ defined by Eq. (2.64) and Eq. (2.63) in Subsection 2.3.3 of the main text, respectively, which correlate spatial and orientational degrees of freedom. All of these order parameters describe global properties (or symmetries) of an ordered structure based on the local proximity of its atomic or molecular entities.

The evaluation of local order parameters strongly depends on the method on how to identify neighbors: here we use the well-defined method of Voronoi construction [223, 224] (cf. Subsection 2.3.1 in the main text). Further, the above order parameters are defined for two-dimensional systems. In our case of a quasi-two dimensional geometry, with a molecular self-assembly in a plane and with slightly stacked 3D structures, we

use for the calculation of the order parameters the projected coordinates of all molecules to the $z = 0$-plane.

For the particle positions, $\mathbf{r}_i$, and the orientations, $\mathbf{u}_i$, considered in $\Psi_\nu(\mathcal{X})$, $\alpha(\mathcal{X})$ and $\beta(\mathcal{X})$, given by Eqs. (2.62) to (2.64) defined in the main text Subsections 2.3.2 and 2.3.3, we used the center-of-mass coordinates and the orientations of the different molecules within a configuration $\mathcal{X}$. Henceforward, we drop the argument of $\Psi_\nu(\mathcal{X})$, $\alpha(\mathcal{X})$ and $\beta(\mathcal{X})$ and may simple write Ψ_ν, α and β.

For our PQP^+ ClO_4^- system we used a set of $N_f = 13$ order parameters to define the feature vector $\mathbf{x}$ as

$$\mathbf{x} = (f_1, \ldots, f_{N_f}); \tag{A.18}$$

with $f_1 = \Psi_4^{(PQP)}$, $f_2 = \Psi_5^{(PQP)}$, $f_3 = \Psi_6^{(PQP)}$ and $f_4 = \Psi_4^{(ClO_4)}$, $f_5 = \Psi_5^{(ClO_4)}$, $f_6 = \Psi_6^{(ClO_4)}$ quantifying the $\nu = 4, 5, 6$-fold bond-orientational order parameters, defined by Eq. (2.62) in Subsection 2.3.2 of the main text, considering only PQP^+ and ClO_4^- molecules as neighbors, respectively. $f_7 = \Psi_4^{(PQP|ClO_4)}$, $f_8 = \Psi_5^{(PQP|ClO_4)}$, $f_9 = \Psi_6^{(PQP|ClO_4)}$ describe the $\nu = 4, 5, 6$-fold bond-orientational order parameters for all PQP^+ molecules while considering only ClO_4^- molecules as neighbors. $f_{10} = \beta^{(PQP)}$, $f_{11} = \beta^{(ClO_4)}$ and $f_{12} = \alpha^{(PQP)}$, $f_{13} = \alpha^{(ClO_4)}$, defined by Eq. (2.63) and Eq. (2.64) in Subsection 2.3.3 of the main text, quantify the orientational- and spatial-orientational-correlation between neighboring PQP^+ and ClO_4^- ions, respectively.

A.2.3 Results

A.2.3.1 General Remarks and System Parameters

The following details provide an idea about the numerical costs of our calculations: in order to obtain the ground state configuration for a single state point (specified by a set of the system parameters defined in Subsection 3.2.4.1 in the main text) convergence of the full EA+LG ground-state search (based on the evolutionary algorithm and the local, steepest gradient descent procedure as specified in Subsection 3.2.3 in the main text) we require at least one to two weeks on one node on the Vienna Scientific Cluster (VSC3) [281] (equipped – per node – with either two Intel Xeon E5-2650v2, 2.6 GHz, eight core processors or two Intel Xeon E5-2660v2, 2.2 GHz, ten core processors from the Ivy Bridge-EP family). We typically used 16 to 20 asynchronous worker processes per evolutionary optimization.

A.2.3.2 Clustering of Results by Similarity

In Subsection 3.2.4.4 of the main text (see in particular Fig. 3.69) we present and discuss results which originate from an independent and separate evolutionary algorithm analysis which focuses entirely on the mobility of the perchlorate anions. The evolutionary search is done by fixing the unit cell as well as the positions and orientations of

the contained PQP^+ molecules such that we can study the local minima in the potential energy as a function of the degrees of freedom of the ClO_4^- molecules. To be more specific these investigations were performed for the structures depicted in Fig. 3.66(a), (e) and (c) of the main text Subsection 3.2.4.4. Among the solutions identified by the evolutionary algorithm we chose roughly 5000 configurations, all representing a local energy minimum in this energy landscape, for which we evaluated the set of order parameters, $\mathbf{x}$, specified in Eq. (A.18). Here, we extend this set of order parameters for the problem at hand: (i) we extend $\mathbf{x}$ by additional bond orientational order parameters $\Psi_\nu^{(ClO_4)}$ and $\Psi_\nu^{(PQP|ClO_4)}$ for $\nu = 3, 8$, and $\Psi_\nu^{(ClO_4|PQP)}$ for $\nu = 3, 4, 5, 6, 8$, as defined by Eq. (2.62) in the main text Subsection 2.3.2 following the specification of the order parameters in Appendix A.2.2.2. (ii) We further extend $\mathbf{x}$ with the orientational order parameters $\alpha^{(ClO_4|PQP)}$ and $\beta^{(ClO_4|PQP)}$ as defined by Eqs. (2.63) and (2.64) in the main text Subsection 2.3.3 but considering the orientational order of perchlorate molecules only with respect to PQP^+ neighbors and, vice versa, by $\alpha^{(PQP|ClO_4)}$ and $\beta^{(PQP|ClO_4)}$. Further, (iii) we use the minimum, mean, median and maximum value of the z coordinates of all ClO_4^- ions and (iv) the average nearest neighbor distance of the perchlorates and (v) of all molecules.

In order to identify configurations which are similar in their structure among all these molecular arrangements and to further distinguish between different collections of similar structures in these large data sets we used unsupervised clustering techniques (see Section 2.4); a very instructive review on such useful tools and many other helpful machine learning applications in physics or chemistry can be found in Ref. [119].

To be more specific, we combine here the so-called principal component analysis (PCA) [238] which reduces the dimensionality of our structural data (or better order parameters thereof) and a successive t-stochastic neighbor embedding (t-SNE) [244] in order to map high-dimensional data points to low-dimensional embedding coordinates (in two or three dimensions), while preserving the local structure in the data [119] (see also Section 2.4 in the main text). With this tool at hand we aim at representing high-dimensional data in two or three dimensions in order to unravel hidden – or hard to identify – geometries (such as structural similarities) within the data set.

Coming back to the discussion on the study of the perchlorate molecules we narrowed down the collection of different structures by admitting only such molecular arrangements whose energy is located within a certain threshold interval, ΔE, above the energy of the related best structures, respectively.

The respective set of order parameters of those structures is first scaled to unit-variance– and zero-mean coordinates considering all data points. Subsequently, this scaled set is subject to a PCA which is a projection of the N_f-dimensional feature space (in our case of the order parameters or features) to an N_l^*-dimensional latent space, see Subsection 2.4.1. PCA is performed with the intention to identify leading singular values of the correlation matrix of the data which represent directions in the feature space with large variance, *i.e.*, which contain the most relevant information [119]. Thereby we reduced the number of our order parameters from N_f=32 to N_l^*=5 latent space coordinates which we used to extract the results presented in Fig. 3.69 in Subsection 3.2.4.4 of the main text (the leading five principal components exhibit an explained variance of $\lambda^e > 5\%$, where λ^e is defined by Eq. (2.75) in Subsection 2.4.1 of the main text).

The latent space data points (each representing a molecular configuration) are then subject to a t-SNE analysis (see Subsection 2.4.4 of the main text for details) and are thereby mapped onto a two-dimensional t-SNE manifold spanned by the axes "t-SNE 1" and "t-SNE 2", see Fig. A.22. It is often easier to visually separate clusters of similar data points in a few t-SNE dimensions than in a higher dimensional PCA latent space [119].

In addition, we performed a DBSCAN [243], *i.e.*, a density-based clustering algorithm (see Subsection 2.4.3 of the main text) which automatically labels the different clusters (each representing a family of structurally similar configurations). The cluster labels are highlighted in the left panel of Fig. A.22 by color coding and the minimum energy of each cluster of structures is indicated in the right panel of Fig. A.22, also by a color code. The results clearly indicate, that several clusters are separated by only minute energy differences. Four of these configurations, separated by an energy difference of 38 meV per molecule pair, are shown in Fig. 3.69 in Subsection 3.2.4.4 of the main text.

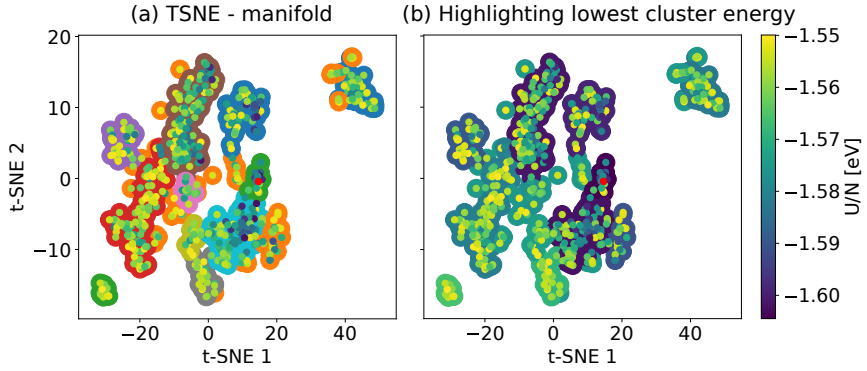


Figure A.22: Dimensional reduction of order parameters of a structural data set of supramolecular configurations into a two-dimensional manifold spanned by *t-SNE 1* and *t-SNE 2* using t-SNE [244] in order to identify the structurally different supramolecular configurations presented in Fig. 3.69 in Subsection 3.2.4.4. Panel (a) and (b): Separated clusters represent structurally different configurations. The small dots (representing the different configurations) are color-coded by the energy per PQP^+– ClO_4^- pair in the unit cell, U/N, of the respective configurations (see color bar). Panel (a): Labels of clusters are highlighted using larger, uniformly colored dots in the background (in an arbitrary color scheme). Panel (b): Same as panel (a) but the background-colors of each dot indicate the minimal energy amongst all configurations of the respective cluster to which the dots are assigned to (see color bar). The emphasized red dot indicates the structure with the lowest energy.

Bibliography

[1] P. W. Anderson, "More is different", Science **177**, 393–396 (1972).

[2] B. W. Ninham and P. L. Nostro, *Molecular Forces and Self Assembly: In Colloid, Nano Sciences and Biology* (Cambridge University Press, 2010).

[3] B. A. Grzybowski and W. T. S. Huck, "The nanotechnology of life-inspired systems", Nat. Nanotechnol. **11**, 585–592 (2016).

[4] S. I. Walker, P. C. W. Davies, and G. F. R. Ellis, *From Matter to Life: Information and Causality* (Cambridge University Press, 2017).

[5] S. T. Picraux, *Nanotechnology* (Encyclopædia Britannica, inc., Sept. 2020).

[6] R. P. Feynman, "There's Plenty of Room at the Bottom", Eng. Sci. **23**, 22–36 (1960).

[7] R. P. Feynman, *There's Plenty of Room at the Bottom*, Presentation to American Physical Society, 1959.

[8] M. Grzelczak, J. Vermant, E. M. Furst, and L. M. Liz-Marzán, "Directed self-assembly of nanoparticles", ACS Nano **4**, 3591–3605 (2010).

[9] R. Waser, *Nanoelectronics and Information Technology - Advanced Electronic Materials and Novel Devices* (John Wiley & Sons, Inc., 2012).

[10] K. T. Nam, D.-W. Kim, P. J. Yoo, C.-Y. Chiang, N. Meethong, P. T. Hammond, Y.-M. Chiang, and A. M. Belcher, "Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes", Science **312**, 885–888 (2006).

[11] J. G. Werner, G. G. Rodríguez-Calero, H. D. Abruña, and U. Wiesner, "Block copolymer derived 3-D interpenetrating multifunctional gyroidal nanohybrids for electrical energy storage", Energy Environ. Sci. **11**, 1261–1270 (2018).

[12] J. Wang, *Nanomachines* (John Wiley & Sons, Inc., 2013).

[13] M. Z. Miskin, A. J. Cortese, K. Dorsey, E. P. Esposito, M. F. Reynolds, Q. Liu, M. Cao, D. A. Muller, P. L. McEuen, and I. Cohen, "Electronically integrated, mass-manufactured, microscopic robots", Nature **584**, 557–561 (2020).

[14] J.-M. Lehn, "Perspectives in supramolecular chemistry–from molecular recognition towards molecular information processing and self-organization", Angew. Chem. Int. Ed. **29**, 1304–1319 (1990).

[15] J.-M. Lehn, "Supramolecular chemistry", Science **260**, 1762–1763 (1993).

[16] F. Biedermann and H.-J. Schneider, "Experimental binding energies in supramolecular complexes", Chem. Rev. **116**, 5216–5300 (2016).

[17] J. K. Patra, G. Das, L. F. Fraceto, E. V. R. Campos, M. d. P. Rodriguez-Torres, L. S. Acosta-Torres, L. A. Diaz-Torres, R. Grillo, M. K. Swamy, S. Sharma, S. Habtemariam, and H.-S. Shin, "Nano based drug delivery systems: recent developments and future prospects", J. Nanobiotechnology **16**, 71 (2018).

[18]K. D. Daze, T. Pinter, C. S. Beshara, A. Ibraheem, S. A. Minaker, M. C. F. Ma, R. J. M. Courtemanche, R. E. Campbell, and F. Hof, "Supramolecular hosts that recognize methyllysines and disrupt the interaction between a modified histone tail and its epigenetic reader protein", Chem. Sci. **3**, 2695–2699 (2012).

[19]S. Bureekaew, S. Shimomura, and S. Kitagawa, "Chemistry and application of flexible porous coordination polymers", Sci. Technol. Adv. Mater. **9**, 014108 (2008).

[20]A. von Zelewsky, *Stereochemistry of Coordination Compounds*, Vol. 3 (John Wiley & Sons, Inc., 1996).

[21]K. Cui, I. Dorner, and S. F. L. Mertens, "Interfacial supramolecular electrochemistry", Curr. Opin. Electrochem. **8**, 156–163 (2018).

[22]S. F. L. Mertens, "Adsorption and self-organization of organic molecules under electrochemical control", in *Encyclopedia of Interfacial Chemistry: Surface Science and Electrochemistry*, Vol. 4, edited by K. Wandelt, 1st ed. (Elsevier, 2018), pp. 13–23.

[23]R. E. Smalley, "Self-assembly of the fullerenes", Acc. Chem. Res. **25**, 98–105 (1992).

[24]T. Guo, P. Nikolaev, A. G. Rinzler, D. Tomanek, D. T. Colbert, and R. E. Smalley, "Self-assembly of tubular fullerenes", J. Phys. Chem. **99**, 10694–10697 (1995).

[25]M. Ballauff and C. N. Likos, "Dendrimers in solution: insight from theory and simulation", Angew. Chem. Int. Ed. **43**, 2998–3020 (2004).

[26]D. Astruc, E. Boisselier, and C. Ornelas, "Dendrimers designed for functions: from physical, photophysical, and supramolecular properties to applications in sensing, catalysis, molecular electronics, photonics, and nanomedicine", Chem. Rev. **110**, 1857–1959 (2010).

[27]H. Zeng, J. Li, J. P. Liu, Z. L. Wang, and S. Sun, "Exchange-coupled nanocomposite magnets by nanoparticle self-assembly", Nature **420**, 395–398 (2002).

[28]N. R. Jana, "Shape effect in nanoparticle self-assembly", Angew. Chem. Int. Ed. **43**, 1536–1540 (2004).

[29]Y. Lin, A. Böker, J. He, K. Sill, H. Xiang, C. Abetz, X. Li, J. Wang, T. Emrick, S. Long, Q. Wang, A. Balazs, and T. P. Russell, "Self-directed self-assembly of nanoparticle/copolymer mixtures", Nature **434**, 55–59 (2005).

[30]T. P. Bigioni, X.-M. Lin, T. T. Nguyen, E. I. Corwin, T. A. Witten, and H. M. Jaeger, "Kinetically driven self assembly of highly ordered nanoparticle monolayers", Nat. Mater. **5**, 265–270 (2006).

[31]S. C. Glotzer and M. J. Solomon, "Anisotropy of building blocks and their assembly into complex structures", Natu. Mat. **6**, 557–562 (2007).

[32]A. Haji-Akbari, M. Engel, A. S. Keys, X. Zheng, R. G. Petschek, P. Palffy-Muhoray, and S. C. Glotzer, "Disordered, quasicrystalline and crystalline phases of densely packed tetrahedra", Nature **462**, 773–777 (2009).

[33]B. Madivala, J. Fransaer, and J. Vermant, "Self-assembly and rheology of ellipsoidal particles at interfaces", Langmuir **25**, 2718–2728 (2009).

[34]S. Sacanna and D. J. Pine, "Shape-anisotropic colloids: building blocks for complex assemblies", Curr. Opin. Colloid Interface Sci. **16**, 96–105 (2011).

[35]P. F. Damasceno, M. Engel, and S. C. Glotzer, "Predictive self-assembly of polyhedra into complex structures", Science **337**, 453–457 (2012).

[36]L. Rossi, S. Sacanna, W. T. M. Irvine, P. M. Chaikin, D. J. Pine, and A. P. Philipse, "Cubic crystals from cubic colloids", Soft Matter **7**, 4139–4142 (2011).

[37]M. R. Jones and C. A. Mirkin, "Self-assembly gets new direction", Nature **491**, 42–43 (2012).

[38]A. B. Pawar and I. Kretzschmar, "Fabrication, assembly, and application of patchy particles", Macromol. Rapid Commun. **31**, 150–168 (2010).

[39]Q. Chen, S. C. Bae, and S. Granick, "Directed self-assembly of a colloidal kagome lattice", Nature **469**, 381–384 (2011).

[40]W. Qi, J. d. Graaf, F. Qiao, S. Marras, L. Manna, and M. Dijkstra, "Ordered two-dimensional superstructures of colloidal octapod-shaped nanocrystals on flat substrates", Nano Lett. **12**, 5299–5303 (2012).

[41]L. Di Michele, F. Varrato, J. Kotar, S. H. Nathan, G. Foffi, and E. Eiser, "Multistep kinetic self-assembly of DNA-coated colloids", Nat. Commun. **4**, 2007 (2013).

[42]R. Lifshitz and D. M. Petrich, "Theoretical model for faraday waves with multiple-frequency forcing", Phys. Rev. Lett. **79**, 1261 (1997).

[43]R. Lifshitz and H. Diamant, "Soft quasicrystals–why are they stable?", Philos. Mag. **87**, 3021–3030 (2007).

[44]K. Barkan, M. Engel, and R. Lifshitz, "Controlled self-assembly of periodic and aperiodic cluster crystals", Phys. Rev. Lett. **113**, 098304 (2014).

[45]R. Lifshitz, "Quasicrystals: a matter of definition", Found. Phys. **33**, 1703–1711 (2003).

[46]D. V. Talapin, E. V. Shevchenko, M. I. Bodnarchuk, X. Ye, J. Chen, and C. B. Murray, "Quasicrystalline order in self-assembled binary nanoparticle superlattices", Nature **461**, 964–967 (2009).

[47]U. Agarwal and F. A. Escobedo, "Mesophase behaviour of polyhedral particles", Nat. Mater. **10**, 230–235 (2011).

[48]W. R. C. Somerville, J. L. Stokes, A. M. Adawi, T. S. Horozov, A. J. Archer, and D. M. A. Buzza, "Density functional theory for the crystallization of two-dimensional dipolar colloidal alloys", J. Condens. Matter Phys. **30**, 405102 (2018).

[49]D. J. Ratliff, A. J. Archer, P. Subramanian, and A. M. Rucklidge, "Which wave numbers determine the thermodynamic stability of soft matter quasicrystals?", Phys. Rev. Lett. **123**, 148004 (2019).

[50]A. Scacchi, W. R. C. Somerville, D. M. A. Buzza, and A. J. Archer, "Quasicrystal formation in binary soft matter mixtures", Phys. Rev. Research **2**, 032043 (2020).

[51]D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, "Metallic phase with long-range orientational order and no translational symmetry", Phys. Rev. Lett. **53**, 1951 (1984).

[52]C. Janot, "The properties and applications of quasicrystals", Europhys. News **27**, 60–64 (1996).

[53]J. Fikar, "Al-Cu-Fe quasicrystalline coatings and composites studied by mechanical spectroscopy", PhD thesis (EPFL, Lausanne, 2003).

[54]V. Obersteiner, M. Scherbela, L. Hörmann, D. Wegner, and O. T. Hofmann, "Structure prediction for surface-induced phases of organic monolayers: overcoming the combinatorial bottleneck", Nano Lett. **17**, 4453–4460 (2017).

[55]N. Engheta and R. Ziolkowski, *Metamaterials: Physics and Engineering Explorations* (Wiley John Wiley & Sons, Inc., 2006).

[56]S. Guenneau, A. Movchan, G. Pétursson, and S. A. Ramakrishna, "Acoustic metamaterials for sound focusing and confinement", New J. Phys. **9**, 399 (2007).

[57]J. Page, "Neither solid nor liquid", Nat. Mater. **10**, 565–566 (2011).

[58]E. Yablonovitch, "Inhibited spontaneous emission in solid-state physics and electronics", Phys. Rev. Lett. **58**, 2059 (1987).

[59]T. F. Krauss, R. M. De La Rue, and S. Brand, "Two-dimensional photonic-bandgap structures operating at near-infrared wavelengths", Nature **383**, 699–702 (1996).

[60]R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental verification of a negative index of refraction", Science **292**, 77–79 (2001).

[61]W. Lu, J. Jin, Z. Lin, and H. Chen, "A simple design of an artificial electromagnetic black hole", J. Appl. Phys. **108**, 064517 (2010).

[62]M. Antlanger, "Ordered equilibrium structures in systems with long-range interactions", PhD thesis (TU Wien, 2015).

[63]M. Antlanger, G. Kahl, M. Mazars, L. Šamaj, and E. Trizac, "Rich polymorphic behavior of Wigner bilayers", Phys. Rev. Lett. **117**, 118002 (2016).

[64]M. Antlanger, G. Kahl, M. Mazars, L. Šamaj, and E. Trizac, "The asymmetric Wigner", J. Chem. Phys **149**, 244904 (2018).

[65]E. Wigner, "On the interaction of electrons in metals", Phys. Rev. **46**, 1002 (1934).

[66]C. C. Grimes and G. Adams, "Evidence for a liquid-to-crystal phase transition in a classical, two-dimensional sheet of electrons", Phys. Rev. Lett. **42**, 795 (1979).

[67]Y. Monarkha and K. Kono, *Two-Dimensional Coulomb Liquids and Solids* (Springer Science & Business Media, 2013).

[68]T. B. Mitchell, J. J. Bollinger, D. H. E. Dubin, X.-P. Huang, W. M. Itano, and R. H. Baughman, "Direct observations of structural phase transitions in planar crystallized ion plasmas", Science **282**, 1290–1293 (1998).

[69]T. B. Mitchell, J. J. Bollinger, X.-P. Huang, W. M. Itano, and D. H. E. Dubin, "Direct observations of the structural phases of crystallized ion plasmas", Phys. Plasmas **6**, 1751 (1999).

[70]M. J. Jensen, T. Hasegawa, and J. J. Bollinger, "Temperature and heating rate of ion crystals in penning traps", Phys. Rev. A **70**, 033401 (2004).

[71]S. Narasimhan and T.-L. Ho, "Wigner-crystal phases in bilayer quantum Hall systems", Phys. Rev. B **52**, 12291 (1995).

[72]D. Fil, "Piezoelectric mechanism of orientation of a bilayer Wigner crystal in a GaAs matrix", Low Temp. Phys. **27**, 384 (2001).

[73]Y. P. Chen, "Pinned bilayer Wigner crystals with pseudospin magnetism", Phys. Rev. B **73**, 115314 (2006).

[74]J. Ye, "Fractional charges and quantum phase transitions in imbalanced bilayer quantum Hall systems", Phys. Rev. Lett. **97**, 236803 (2006).

[75]Z. Wang, Y. P. Chen, L. Engel, D. Tsui, E. Tutuc, and M. Shayegan, "Pinning modes and interlayer correlation in high-magnetic-field bilayer Wigner solids", Phys. Rev. Lett. **99**, 136804 (2007).

[76]P. Attard and D. J. Mitchell, "The forces between surfaces of mobile, orientable dipoles: asymptotic expressions", Chem. Phys. Lett. **133**, 347–352 (1987).

[77]P. Attard and D. J. Mitchell, "The forces between surfaces of mobile, orientable dipoles. the method of reflection coefficients", J. Chem. Phys. **88**, 4391 (1988).

[78]P. Attard and G. Patey, "Continuum electrostatic interactions between planar lattices of dipoles and the possible relevance to the hydration force", Phys. Rev. A **43**, 2953 (1991).

[79]S. Leikin, "On the theory of electrostatic interaction of neutral lipid bilayers separated by thin water film", J. Chem. Phys. **95**, 5224 (1991).

[80]M. Granfeldt, B. Jönsson, and H. Wennerström, "Electrostatic interactions in lecithin-bilayer systems a Monte Carlo study", Mol. Phys. **64**, 129–142 (1988).

[81]K. Cui, K. S. Mali, D. Wu, X. Feng, K. Müllen, M. Walter, S. De Feyter, and S. F. L. Mertens, "Ambient bistable single dipole switching in a molecular monolayer", Angew. Chem. Int. Ed. **59**, 14049–14053 (2020).

[82]K. Cui, "Supramolecular templates at the solid-liquid interface", PhD thesis (KU Leuven, 2014).

[83]K. Cui, K. S. Mali, O. Ivasenko, D. Wu, X. Feng, M. Walter, K. Müllen, S. De Feyter, and S. F. L. Mertens, "Squeezing, then stacking: from breathing pores to three-dimensional ionic self-assembly under electrochemical control", Angew. Chem Int. Edit. **53**, 12951–12954 (2014).

[84]K. Cui, O. Ivasenko, K. S. Mali, D. Wu, X. Feng, K. Müllen, S. De Feyter, and S. F. L. Mertens, "Potential-driven molecular tiling of a charged polycyclic aromatic compound", Chem. Commun. **50**, 10376–10378 (2014).

[85]K. Cui, K. S. Mali, D. Wu, X. Feng, K. Müllen, M. Walter, S. De Feyter, and S. F. L. Mertens, "Reversible anion-driven switching of an organic 2D crystal at a solid–liquid interface", Small **13**, 1702379 (2017).

[86]B. Hartl, S. Sharma, O. Brügner, S. F. L. Mertens, M. Walter, and G. Kahl, "Reliable computational prediction of the supramolecular ordering of complex molecules under electrochemical conditions", J. Chem. Theory Comput. **16**, 5227–5243 (2020).

[87]C. Ambrosch-Draxl, D. Nabok, P. Puschnig, and C. Meisenbichler, "The role of polymorphism in organic thin films: oligoacenes investigated from first principles", New J. Phys. **11**, 125010 (2009).

[88]A. Jones, B. Chattopadhyay, Y. Geerts, and R. Resel, "Substrate-induced and thin-film phases: polymorphism of organic materials on surfaces", Adv. Funct. Mater. **26**, 2233–2255 (2016).

[89]M. Scherbela, L. Hörmann, A. Jeindl, V. Obersteiner, and O. T. Hofmann, "Charting the energy landscape of metal/organic interfaces via machine learning", Phys. Rev. Materials **2**, 043803 (2018).

[90]L. Hörmann, A. Jeindl, A. T. Egger, M. Scherbela, and O. T. Hofmann, "Sample: surface structure search enabled by coarse graining and statistical learning", Comput. Phys. Commun. **244**, 143–155 (2019).

[91]A. T. Egger, L. Hörmann, A. Jeindl, M. Scherbela, V. Obersteiner, M. Todorović, P. Rinke, and O. T. Hofmann, "Charge transfer into organic thin films: a deeper insight through machine-learning-assisted structure search", Adv. Sci. **7**, 2000992 (2020).

[92]M. Mazars, "Long ranged interactions in computer simulations and for quasi-2D systems", Phys. Rep. **500**, 43–116 (2011).

[93]S. Plimpton, "Fast parallel algorithms for short – range molecular dynamics", J. Comput. Phys. **117**, 1–19 (1995).

[94]P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale", Ann. Phys. **369**, 253–287 (1921).

[95]R. H. French, V. A. Parsegian, R. Podgornik, R. F. Rajter, A. Jagota, J. Luo, D. Asthagiri, M. K. Chaudhury, Y.-m. Chiang, S. Granick, S. Kalinin, M. Kardar, R. Kjellander, D. C. Langreth, J. Lewis, S. Lustig, D. Wesolowski, J. S. Wettlaufer, W.-Y. Ching, M. Finnis, F. Houlihan, O. A. von Lilienfeld, C. J. van Oss, and T. Zemb, "Long range interactions in nanoscale science", Rev. Mod. Phys. **82**, 1887 (2010).

[96]F. H. Stillinger, "Exponential multiplicity of inherent structures", Phys. Rev. E **59**, 48 (1999).

[97]J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, and V. Caignaert, "Prediction of crystal structures from crystal chemistry rules by simulated annealing", Nature **346**, 343–345 (1990).

[98]J. C. Schön and M. Jansen, "Predicting solid compounds using simulated annealing", in *Modern Methods of Crystal Structure Prediction* (John Wiley & Sons, Inc., 2010) Chap. 4, pp. 67–105.

[99]D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms", J. Phys. Chem. A **101**, 5111–5116 (1997).

[100]C. Panosetti, K. Krautgasser, D. Palagin, K. Reuter, and R. J. Maurer, "Global materials structure search with chemically motivated coordinates", Nano Lett. **15**, 8044–8048 (2015).

[101]K. Krautgasser, C. Panosetti, D. Palagin, K. Reuter, and R. J. Maurer, "Global structure search for molecules on surfaces: efficient sampling with curvilinear coordinates", J. Chem. Phys. **145**, 084117 (2016).

[102]S. Goedecker, "Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems", J. Chem. Phys. **120**, 9911 (2004).

[103]S. Goedecker, "Global optimization with the minima hopping method", in *Modern Methods of Crystal Structure Prediction* (John Wiley & Sons, Inc., 2010) Chap. 6, pp. 131–145.

[104]D. Gottwald, G. Kahl, and C. N. Likos, "Predicting equilibrium structures in freezing processes", J. Chem. Phys. **122**, 204503 (2005).

[105]J. Fornleitner, F. Lo Verso, G. Kahl, and C. N. Likos, "Genetic algorithms predict formation of exotic ordered configurations for two-component dipolar monolayers", Soft Matter **4**, 480–484 (2008).

[106]J. Fornleitner, F. Lo Verso, G. Kahl, and C. N. Likos, "Ordering in two-dimensional dipolar mixtures", Langmuir **25**, 7836–7846 (2009).

[107]G. Doppelbauer, E. Bianchi, and G. Kahl, "Self-assembly scenarios of patchy colloidal particles in two dimensions", J. Phys. Condens. Matter **22**, 104105 (2010).

[108]J. Fornleitner and G. Kahl, "Pattern formation in two-dimensional square-shoulder systems", J. Phys. Condens. Matter **22**, 104118 (2010).

[109]G. Doppelbauer, E. G. Noya, E. Bianchi, and G. Kahl, "Self-assembly scenarios of patchy colloidal particles", Soft Matter **8**, 7768–7772 (2012).

[110]N. L. Abraham and M. I. J. Probert, "A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction", Phys. Rev. B **73**, 224104 (2006).

[111]D. M. Deaven and K. M. Ho, "Molecular geometry optimization with a genetic algorithm", Phys. Rev. Lett. **75**, 288 (1995).

[112]B. Hartke, "Global cluster geometry optimization by a phenotype algorithm with niches: location of elusive minima, and low-order scaling with cluster size", J. Comput. Chem. **16**, 1752–1759 (1999).

[113]B. Hartke, "Global geometry optimization of clusters using genetic algorithms", J. Phys. Chem. **97**, 9973–9976 (1993).

[114]A. O. Lyakhov, A. R. Oganov, and M. Valle, "Crystal structure prediction using evolutionary approach", in *Modern Methods of Crystal Structure Prediction* (John Wiley & Sons, Inc., 2010) Chap. 7, pp. 147–180.

[115]A. Supady, V. Blum, and C. Baldauf, "First-principles molecular structure search with a genetic algorithm", J. Chem. Inf. Model. **55**, 2338–2348 (2015).

[116]K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science", Nature **559**, 547–555 (2018).

[117]J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science", Npj Comput. Mater. **5**, 83 (2019).

[118]J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science", InfoMat **1**, 338–358 (2019).

[119]P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to Machine Learning for physicists", Phys. Rep. **810**, 1–124 (2019).

[120]*Bayesian Optimization Structure Search*, `https://cest.aalto.fi/boss/`.

[121]M. Todorović, M. U. Gutmann, J. Corander, and P. Rinke, "Bayesian inference of atomistic structure in functional materials", Npj Comput. Mater. **5**, 35 (2019).

[122]G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning* (Springer New York, 2013).

[123]R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, 2018).

[124]S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Pearson, 2020).

[125]D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of go without human knowledge", Nature **550**, 354–359 (2017).

[126]A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Zídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, "Improved protein structure prediction using potentials from deep learning", Nature **577**, 706–710 (2020).

[127]E. T. Jaynes, "Information theory and statistical mechanics", Phys. Rev. **106**, 620 (1957).

[128]E. T. Jaynes, "Information theory and statistical mechanics. II", Phys. Rev. **108**, 171 (1957).

[129]A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes", in Adv. Neural. Inf. Process. Syst. 14, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani, NIPS'01 (2001), pp. 841–848.

[130]G. Hinton, "Boltzmann machines", in *Encyclopedia of Machine Learning*, edited by C. Sammut and G. I. Webb (Springer US, 2010), pp. 132–136.

[131]D. P. Kingma and M. Welling, "Auto-encoding variational bayes", (2013), arXiv:1312.6114 [stat.ML].

[132]I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", in Proc. 27th Int. Conf. Neural. Inf. Process. Syst. - Vol. 2, NIPS'14 (2014), pp. 2672–2680.

[133]S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing", Science **220**, 671–680 (1983).

[134]V. Cerný, "Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm", J. Optimiz. Theory App. **45**, 41–51 (1985).

[135]R. H. Swendsen and J.-S. Wang, "Replica Monte Carlo simulation of spin-glasses", Phys. Rev. Lett. **57**, 2607 (1986).

[136]C. J. Geyer, "Markov chain Monte Carlo maximum likelihood", in Computer Science and Statistics: Proc. 23rd Symp. Interface (1991), pp. 156–163.

[137]M. Falcioni and M. W. Deem, "A biased Monte Carlo scheme for zeolite structure solution", J. Chem. Phys. **110**, 1754 (1999).

[138]D. J. Earl and M. W. Deem, "Parallel tempering: theory, applications, and new perspectives", Phys. Chem. Chem. Phys. **7**, 3910–3916 (2005).

[139]P. Moscato et al., *On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms*, C3P Report (Caltech, 1989).

[140]G. Doppelbauer, "Ordered equilibrium structures of patchy particle systems", PhD thesis (TU Wien, 2012).

[141]D. H. Wolpert and W. G. Macready, *No free lunch theorems for search*, Working Papers (Santa Fe Institute, 1995).

[142]D. H. Wolpert, "The lack of *a priori* distinctions between learning algorithms", Neural Comput. **8**, 1341–1390 (1996).

[143]D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization", Trans. Evol. Comp **1**, 67–82 (1997).

[144]D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press, Inc., 2001).

[145]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford Science Publ. (Clarendon Press, 1989).

[146]L. Šamaj and E. Trizac, "Critical phenomena and phase sequence in a classical bilayer Wigner crystal at zero temperature", Phys. Rev. B **85**, 205131 (2012).

[147]L. Šamaj and E. Trizac, "Ground state of classical bilayer Wigner crystals", EPL **98**, 36004 (2012).

[148]E. R. Smith, "Electrostatic Energy in Ionic Crystals", Proc. R. Soc. Lond. A **375**, 475–505 (1981).

[149]I. C. Yeh and M. L. Berkowitz, "Ewald summation for systems with slab geometry", J. Chem. Phys. **111**, 3155 (1999).

[150]S. H. L. Klapp and M. Schoen, "Spontaneous orientational order in confined dipolar fluid films", J. Chem. Phys. **117**, 8050 (2002).

[151]V. Ballenegger, A. Arnold, and J. J. Cerdà, "Simulations of non-neutral slab systems with long-range electrostatic interactions in two-dimensional periodic boundary conditions", J. Chem. Phys. **131**, 094107 (2009).

[152]S. W. de Leeuw, J. W. Perram, E. R. Smith, and J. S. Rowlinson, "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants", Proc. R. Soc. A **373**, 27–56 (1980).

[153]V. Ballenegger, "Communication: on the origin of the surface term in the Ewald formula", J. Chem. Phys. **140**, 161102 (2014).

[154]A. Arnold, J. de Joannis, and C. Holm, "Electrostatics in periodic slab geometries. I", J. Chem. Phys. **117**, 2496 (2002).

[155]J. de Joannis, A. Arnold, and C. Holm, "Electrostatics in periodic slab geometries. II", J. Chem. Phys. **117**, 2503 (2002).

[156]E. R. Smith and J. S. Rowlinson, "Effects of surface charge on the electrostatic energy of an ionic crystal", Proc. R. Soc. Lond. A **381**, 241–247 (1982).

[157]E. R. Smith, "Electrostatic potentials in systems periodic in one, two, and three dimensions", J. Chem. Phys. **128**, 174104 (2008).

[158]J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, Inc., 1999).

[159]J. Lekner, "Summation of Coulomb fields in computer-simulated disordered systems", Physica A **176**, 485–498 (1991).

[160]T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems", J. Chem. Phys. **98**, 10089 (1993).

[161]H. G. Petersen, "Accuracy and efficiency of the particle mesh Ewald method", J. Chem. Phys. **103**, 3668 (1995).

[162]J. W. Eastwood, R. W. Hockney, and D. N. Lawrence, "P3M3DP–the three-dimensional periodic particle-particle/ particle-mesh program", Comput. Phys. Commun. **19**, 215–261 (1980).

[163]M. Deserno and C. Holm, "How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines", J. Chem. Phys. **109**, 7678 (1998).

[164]M. Deserno and C. Holm, "How to mesh up Ewald sums. II. An accurate error estimate for the particle–particle–particle-mesh algorithm", J. Chem. Phys. **109**, 7694 (1998).

[165]A. Arnold and C. Holm, "Efficient methods to compute long-range interactions for soft matter systems", in *Advanced Computer Simulation Approaches for Soft Matter Sciences II*, Vol. 185, edited by C. Holm and K. Kremer, Adv. Polym. Sci. (Springer Berlin Heidelberg, 2005), pp. 59–109.

[166]H. Lee and W. Cai, *Ewald summation for Coulomb interactions in a periodic supercell*, Lecture Notes (Stanford University, 2009).

[167]V. Elser, "Comment on "Quasicrystals: a new class of ordered structures"", Phys. Rev. Lett. **54**, 1730 (1985).

[168]C. L. Henley, "Random tiling models", in *Quasicrystals* (1991), pp. 429–524.

[169]D. Kraft, *A software package for sequential quadratic programming*, tech. rep. (DLR German Aerospace Center – Institute for Flight Mechanics, Köln, Germany, 1988), DFVLR–FB 88–28.

[170]D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization", Math. Program. **45**, 503–528 (1989).

[171]C. G. Boyden, "The convergence of a class of double-rank minimization algorithms", IMA J. Appl. Math. **6**, 222–231 (1970).

[172]R. Fletcher, "A new approach to variable metric algorithms", Comput. J. **13**, 317–322 (1970).

[173]D. Goldfarb, "A family of variable-metric methods derived by variational means", Math. Comput. **24**, 23–26 (1970).

[174]D. F. Shanno, "Conditioning of quasi-newton methods for function minimization", Math. Comput. **24**, 647–656 (1970).

[175]R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization", SIAM J. Sci. Comput. **16**, 1190–1208 (1995).

[176]P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. 1. 0. Contributors, "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python", (2019), arXiv:1907.10121 [cs.MS].

[177]J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, pathways, and the energy landscape of protein folding: a synthesis", Proteins **21**, 167–195 (1995).

[178]D. T. Leeson, F. Gai, H. M. Rodriguez, L. M. Gregoret, and R. B. Dyer, "Protein folding and unfolding on a complex energy landscape", Proc. Natl. Acad. Sci. U.S.A. **97**, 2527–2532 (2000).

[179]B. Robson and A. Vaithilingam, "Protein folding revisited", Prog. Mol. Biol. Transl. Sci. **84**, 161–202 (2008).

[180]J. Douglas and M. Douglas, *Conceptual Design of Chemical Processes*, Chemical engineering series (McGraw-Hill, 1988).

[181]M. Emmerich, B. Groß, F. Henrich, P. Roosen, and M. Schütz, "Global optimization of chemical engineering plants by means of evolutionary algorithms", Proc. Aspen World (2000).

[182]P. Schwaller, T. Gaudin, D. Lányi, C. Bekas, and T. Laino, ""Found in translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models", Chem. Sci. **9**, 6091–6098 (2018).

[183]P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, and A. A. Lee, "Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction", ACS Cent. Sci. **5**, 1572–1583 (2019).

[184]M. Hofmarcher, A. Mayr, E. Rumetshofer, P. Ruch, P. Renz, J. Schimunek, P. Seidl, A. Vall, M. Widrich, S. Hochreiter, and G. Klambauer, "Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks", SSRN Electronic J. (2020).

[185]J. M. Kleinberg and É. Tardos, *Algorithm Design* (Addison-Wesley, 2006).

[186]D. L. Applegate, R. E. Bixby, V. Chvatal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*, Princeton Series in Applied Mathematics (Princeton University Press, 2006).

[187]E. Hansen and G. W. Walster, *Global Optimization using Interval Analysis: Revised and Expanded*, 2nd ed. (CRC Press, 2004).

[188]A. Neumaier, "Complete search in continuous global optimization and constraint satisfaction", Acta Numer. **13**, 271–369 (2004).

[189]N. Metropolis and S. Ulam, "The Monte Carlo method", J. Am. Stat. Assoc. **44**, 335–341 (1949).

[190]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines", J. Chem. Phys. **21**, 1087 (1953).

[191]W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications", Biometrika **57**, 97–109 (1970).

[192]D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev, "Why the Monte Carlo method is so important today", WIREs Comput. Stat. **6**, 386–392 (2014).

[193]K. Hamacher and W. Wenzel, "Scaling behavior of stochastic minimization algorithms in a perfect funnel landscape", Phys. Rev. E **59**, 938 (1999).

[194]K. Hamacher, "Adaptation in stochastic tunneling global optimization of complex potential energy landscapes", EPL **74**, 944–950 (2006).

[195]R. Battiti and G. Tecchiolli, "The reactive tabu search", ORSA J. Comput. **6**, 126–140 (1994).

[196]F. Glover and M. Laguna, "Tabu search", in *Handbook of Combinatorial Optimization* (Springer US, 1998), pp. 2093–2229.

[197]A. Blake and A. Zisserman, *Visual Reconstruction* (MIT Press, 1987).

[198]J. Kennedy and R. Eberhart, "Particle swarm optimization", in Proc. Int. Conf. Neural Netw. Vol. 4, ICNN'95 (1995), pp. 1942–1948.

[199]Y. Shi and R. Eberhart, "A modified particle swarm optimizer", in 1998 IEEE Int. Conf. Evol. Comput. Proc., IEEE World Congr. Comput. Intell. (Cat. No.98TH8360) (1998), pp. 69–73.

[200]Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications", Math. Prob. Eng. **2015**, 931256 (2015).

[201]T. Hendtlass, "WoSP: a multi-optima particle swarm algorithm", in 2005 IEEE Congress on Evolutionary Computation (2005).

[202]S. Z. Zhao, J. J. Liang, P. N. Suganthan, and M. F. Tasgetiren, "Dynamic multi-swarm particle swarm optimizer with local search for large scale global optimization", in 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence) (2008), pp. 3845–3852.

[203]R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces", J. Global Optim. **11**, 341–359 (1997).

[204]J. H. Holland, *Adaptation in Natural and Artificial Systems*, 2nd ed. (University of Michigan Press, 1992).

[205]D. Gottwald, "Genetic algorithms in condensed matter theory", PhD thesis (TU Wien, 2005).

[206]J. Mockus, *Bayesian Approach to Global Optimization* (Springer Netherlands, 1989).

[207]B. Hartl, M. Hübl, G. Kahl, and A. Zöttl, "Microswimmers learning chemotaxis with genetic algorithms", under review in Proc. Natl. Acad. Sci. U.S.A. (2020).

[208]I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT Press, 2016).

[209]V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, and A. Jain, "Unsupervised word embeddings capture latent knowledge from materials science literature", Nature **571**, 95–98 (2019).

[210]D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines", Cogn. Sci. **9**, 147–169 (1985).

[211]C. Darwin, *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life* (J. Murray, 1859).

[212]B. Hartl, *Atuin – a Python-framework for adaptive optimization of nested datastructures*, `https://github.com/bhartl/atuin` (available on request), 2020.

[213]R. Wanzenböck, *Self-assembly of ionic liquid crystals into smectic phases*, Master's thesis (TU Wien, 2019).

[214]L. D. Dalcín, R. R. Paz, and M. Storti, "MPI for Python", J. Parallel Distr. Com. **65**, 1108–1115 (2005).

[215]L. D. Dalcín, R. R. Paz, M. Storti, and J. D. Elía, "MPI for Python : performance improvements and MPI-2 extensions", J. Parallel Distr. Com. **68**, 655–662 (2008).

[216]J. Fornleitner and G. Kahl, "Lane formation vs . cluster formation in two- dimensional square-shoulder systems – a genetic algorithm approach", EPL **82**, 18001 (2008).

[217]S. Kondrat, M. Bier, and L. Harnau, "Phase behavior of ionic liquid crystals Phase behavior of ionic liquid crystals", J. Chem. Phys. **132**, 184901 (2010).

[218]H. Bartsch, M. Bier, and S. Dietrich, "Smectic phases in ionic liquid crystals", J. Phys. Condens. Matter **29**, 464002 (2017).

[219]L. D. Dalcín, R. R. Paz, P. A. Kler, and A. Cosimo, "Parallel distributed computing using Python", Adv. Water Resour. **34**, 1124–1139 (2011).

[220]P. Chaikin and T. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, 2000).

[221]L. D. Landau, "On the theory of phase transitions", Phys. Z. Sowjet. **11**, 26 (1937).

[222]L. D. Landau, "On the theory of phase transitions", J. Exptl. Theoret. Phys. (U.S.S. R.) **7**, 19 (1937).

[223]G. Lejeune Dirichlet, "Über die Reduction der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen.", J. Reine Angew. Mathematik **40**, 209–227 (1850).

[224]G. Voronoi, "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites.", J. Reine Angew. Mathematik **133**, 97–178 (1908).

[225]K. J. Strandburg, J. A. Zollweg, and G. V. Chester, "Bond-angular order in two-dimensional Lennard-Jones and hard-disk systems", Phys. Rev. B **30**, 2755 (1984).

[226]K. J. Strandburg, "Two-dimensional melting", Rev. Mod. Phys. **60**, 161 (1988).

[227]M. Mazars, "Bond orientational order parameters in the crystalline phases of the classical Yukawa-Wigner bilayers", EPL **84**, 55002 (2008).

[228]W. Mickel, S. Kapfer, G. Schröder-Turk, and K. Mecke, "Shortcomings of the bond orientational order parameters for the analysis of disordered particulate matter", J. Chem. Phys. **138**, 044501 (2013).

[229]I. A. Georgiou, P. Ziherl, and G. Kahl, "Antinematic local order in dendrimer liquids", EPL **106**, 44004 (2014).

[230]N. W. Ashcroft and N. D. Mermin, *Solid state physics* (Holt-Saunders, 1976).

[231]W. H. Bragg and W. L. Bragg, "The reflection of X-rays by crystals", Proc. Roy. Soc. A **88**, 428–438 (1913).

[232]O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook* (Springer US, 2010).

[233]R. B. Jadrich, B. A. Lindquist, and T. M. Truskett, "Unsupervised machine learning for detection of phase transitions in off-lattice systems. I. Foundations", J. Chem. Phys. **149**, 194109 (2018).

[234]R. B. Jadrich, B. A. Lindquist, W. D. Piñeros, D. Banerjee, and T. M. Truskett, "Unsupervised machine learning for detection of phase transitions in off-lattice systems. II. Applications", J. Chem. Phys. **149**, 194110 (2018).

[235]P. Z. Moghadam, S. M. Rogge, A. Li, C. M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D. A. Gomez-Gualdron, V. Van Speybroeck, and D. Fairen-Jimenez, "Structure-mechanical stability relations of metal-organic frameworks via machine learning", Matter **1**, 219–234 (2019).

[236]E. Boattini, M. Dijkstra, and L. Filion, "Unsupervised learning for local structure detection in colloidal systems", J. Chem. Phys. **151**, 154901 (2019).

[237]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: machine learning in Python", J. Mach. Learn. Res. **12**, 2825–2830 (2011).

[238]I. Jolliffe, *Principal Component Analysis* (Springer New York, 2002).

[239]H. Steinhaus, "Sur la division des corps matériels en parties", B. Acad. Pol. Sci. - Cl. III **IV**, 801–804 (1956).

[240]E. Forgy, "Cluster analysis of multivariate data: efficiency versus interpretability of classification", Biometrics **21**, 768–769 (1965).

[241]J. MacQueen, "Some methods for classification and analysis of multivariate observations", in Proc. Fifth Berkeley Symp. Math. Stat. Prob., Vol. 1: Stat. (1967), pp. 281–297.

[242]S. P. Lloyd, "Least squares quantization in PCM", IEEE Trans. Inf. Theory **28**, 129–137 (1982).

[243]M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", in Proc. Second Int. Conf. Knowl. Discov. Data Min. KDD'96 (1996), pp. 226–231.

[244]L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE", J. Mach. Learn. Res **9**, 2579–2605 (2008).

[245]N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: is a correction for chance necessary?", in Proc. 26th Annu. Int. Conf. Mach. Learn. ICML'09 (2009), pp. 1073–1080.

[246]N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance", J. Mach. Learn. Res. **11**, 2837–2854 (2010).

[247]S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space", Phys. Chem. Chem. Phys. **18**, 13754–13769 (2016).

[248]B. A. Helfrecht, R. K. Cersonsky, G. Fraux, and M. Ceriotti, *Structure-property maps with kernel principal covariates regression*, 2020, arXiv:2002.05076 [stat.ML].

[249]C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).

[250]J. Shlens, "A tutorial on principal component analysis", (2014), arXiv:1404.1100.

[251]W. M. Rand, "Objective criteria for the evaluation of clustering methods", J. Am. Stat. Assoc. **66**, 846–850 (1971).

[252]P. Stampfli, "A dodecagonal quasiperiodic lattice in two dimensions", Helv. Phys. Acta **59**, 1260–1263 (1986).

[253]F. Gähler, "Crystallography of dodecagonal quasicrystals", in Proc. I.L.L. / CODEST Workshop, edited by C. Janot (1988), pp. 272–284.

[254]M. Oxborrow and C. L. Henley, "Random square-triangle tilings: a model for twelvefold-symmetric quasicrystals", Phys. Rev. B **48**, 6966 (1993).

[255]B. Grünbaum and G. C. Shephard, *Tilings and Patterns* (Courier Dover Publications, 1987).

[256]R. Berger, *The Undecidability of the Domino Problem* (American Mathematical Society, 1966).

[257]R. Penrose, "Pentaplexity a class of non-periodic tilings of the plane", Math. Intell. **2**, 32–37 (1979).

[258]D. Levine and P. J. Steinhardt, "Quasicrystals: a new class of ordered structures", Phys. Rev. Lett. **53**, 2477–2480 (1984).

[259]P. Kramer and R. Neri, "On periodic and non-periodic space fillings of $\mathbf{E}^m$ obtained by projection", Acta Cryst. A **40**, 580–587 (1984).

[260]R. Penrose, "The role of aesthetics in pure and applied mathematical research", Bull. Inst. Math. Appl. **10**, 266–271 (1974).

[261]F. C. Frank and J. S. Kasper, "Complex alloy structures regarded as sphere packings. I. Definitions and basic principles", Acta Cryst. **11**, 184–190 (1958).

[262]D. P. Shoemaker and C. B. Shoemaker, "Concerning the relative numbers of atomic coordination types in tetrahedrally close packed metal structures", Acta Cryst. B **42**, 3–11 (1986).

[263]A. L. Mackay, "A dense non-crystallographic packing of equal spheres", Acta Cryst. **15**, 916–918 (1962).

[264]J. Nakakura, P. Ziherl, J. Matsuzawa, and T. Dotera, "Metallic-mean quasicrystals as aperiodic approximants of periodic crystals", Nat. Commun. **10**, 4235 (2019).

[265]X. Zeng, G. Ungar, Y. Liu, V. Percec, A. E. Dulcey, and J. K. Hobbs, "Supramolecular dendritic liquid quasicrystals", Nature **428**, 157–160 (2004).

[266]K. Hayashida, T. Dotera, A. Takano, and Y. Matsushita, "Polymeric quasicrystal: mesoscopic quasicrystalline tiling in *ABC* star polymers", Phys. Rev. Lett. **98**, 195502 (2007).

[267]S. Chanpuriya, K. Kim, J. Zhang, S. Lee, A. Arora, K. D. Dorfman, K. T. Delaney, G. H. Fredrickson, and F. S. Bates, "Cornucopia of nanoscale ordered phases in sphere-forming tetrablock terpolymers", ACS Nano **10**, 4961–4972 (2016).

[268]L. Bindi, J. Pham, and P. J. Steinhardt, "Previously unknown quasicrystal periodic approximant found in space", Sci. Rep. **8**, 16271 (2018).

[269]P. Bak, "Icosahedral crystals: where are the atoms?", Phys. Rev. Lett. **56**, 861 (1986).

[270]A. Yamamoto, "Crystallography of quasiperiodic crystals", Acta Cryst. A **52**, 509–560 (1996).

[271]T. Ishimasa, S. Iwami, N. Sakaguchi, R. Oota, and M. Mihalkovič, "Phason space analysis and structure modelling of 100 Å -scale dodecagonal quasicrystal in Mn-based alloy", Philos. Mag. **95**, 3745–3767 (2015).

[272]M. Mihalkovič and M. Widom, "First-principles prediction of a decagonal quasicrystal containing boron", Phys. Rev. Lett. **93**, 095507 (2004).

[273]T. Dotera, T. Oshiro, and P. Ziherl, "Mosaic two-lengthscale quasicrystals", Nature **506**, 208–211 (2014).

[274]T. Dotera, S. Bekku, and P. Ziherl, "Bronze-mean hexagonal quasicrystal", Nat. Mater. **16**, 987–992 (2017).

[275]S. Earnshaw, "On the Nature of the Molecular Forces which Regulate the Constitution of the Luminiferous Ether", Trans. Cambridge Philos. Soc. **7**, 97 (1848).

[276]M. Martial, *Notes on the cascade and polymorphism of structures* $\boldsymbol{I}_x$, (Laboratoire de Physique Théorique (UMR 8627), Université de Paris-Sud and CNRS, 2016).

[277]T. Janssen, D. o. T. P. T. Janssen, G. Chapuis, and M. d. Boissieu, *Aperiodic Crystals - From Modulated Phases to Quasicrystals* (OUP Oxford, 2007).

[278]N. Guttenberg, N. Virgo, O. Witkowski, H. Aoki, and R. Kanai, *Permutation-equivivalent neural networks applied to dynamics prediction*, 2016, arXiv:1612.04530 [cs.CV].

[279]E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, "Learning phase transitions by confusion", Nature Physics **13**, 435–439 (2017).

[280]T. Ishimasa, M. Mihalkovič, K. Deguchi, N. K. Sato, and M. de Boissieu, "Interpretation of some Yb-based valence-fluctuating crystals as approximants to a dodecagonal quasicrystal", Philos. Mag. **98**, 2018–2034 (2018).

[281]*Vienna Scientific Cluster (VSC3)*, http://typo3.vsc.ac.at/systems/vsc-3/, 2014.

[282]M. Mihalkovič, W.-J. Zhu, C. L. Henley, and M. Oxborrow, "Icosahedral quasicrystal decoration models. I. Geometrical principles", Phys. Rev. B **53**, 9002 (1996).

[283]M. Mihalkovič, W.-J. Zhu, C. L. Henley, and R. Phillips, "Icosahedral quasicrystal decoration models. II. Optimization under realistic Al-Mn potentials", Phys. Rev. B **53**, 9021 (1996).

[284]V. W. de Spinadel, "The metallic means and design", in Nexus II: Architecture and Mathematics, edited by K. Williams (1998), pp. 141–157.

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

[286]H.-C. Jeong and P. J. Steinhardt, "Cluster approach for quasicrystals", Phys. Rev. Lett. **73**, 1943–1946 (1994).

[287]P. J. Steinhardt and H.-C. Jeong, "A simpler approach to penrose tiling with implications for quasicrystal formation", Nature **382**, 431–433 (1996).

[288]F. Gähler, "Cluster coverings as an ordering principle for quasicrystals", Mater. Sci. Eng. A **294-296**, 199–204 (2000).

[289]J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces", Phys. Rev. Lett. **98**, 146401 (2007).

[290]R. S. Becker, J. A. Golovchenko, G. S. Higashi, and B. S. Swartzentruber, "New reconstructions on silicon (111) surfaces", Phys. Rev. Lett. **57**, 1020 (1986).

[291]W. Carl and K. Schubert, "Struktur von Pt_7Zn_{12}", J. Less. Common Met. **19**, 279–283 (1969).

[292]S. Zhang, J. Zhou, Q. Wang, X. Chen, Y. Kawazoe, and P. Jena, "Penta-graphene: a new carbon allotrope", Proc. Natl. Acad. Sci. U.S.A. **112**, 2372–2377 (2015).

[293]F. Gähler, "Matching rules for quasicrystals: the composition-decomposition method", J. Non-Cryst. Solids **153-154**, 160–164 (1993).

[294]G. Ungar and X. Zeng, "Frank–kasper, quasicrystalline and related phases in liquid crystals", Soft Matter **1**, 95 (2005).

[295]J. Mikhael, J. Roth, L. Helden, and C. Bechinger, "Archimedean-like tiling on decagonal quasicrystalline surfaces", Nature **454**, 501–504 (2008).

[296]C. R. Iacovella, A. S. Keys, and S. C. Glotzer, "Self-assembly of soft-matter quasicrystals and their approximants", Proc. Natl. Acad. Sci. U.S.A. **108**, 20935–20940 (2011).

[297]A. M. Rucklidge, M. Silber, and A. C. Skeldon, "Three-wave interactions and spatiotemporal chaos", Phys. Rev. Lett. **108**, 074504 (2012).

[298]S. Förster, K. Meinel, R. Hammer, M. Trautmann, and W. Widdra, "Quasicrystalline structure formation in a classical crystalline thin-film system", Nature **502**, 215–218 (2013).

[299]T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, "Neural network models of potential energy surfaces", J. Chem. Phys. **103**, 4129 (1995).

[300]S. Lorenz, A. Groß, and M. Scheffler, "Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks", Chem. Phys. Lett. **395**, 210–215 (2004).

[301]J. Behler, S. Lorenz, and K. Reuter, "Representing molecule-surface interactions with symmetry-adapted neural networks", J. Chem. Phys. **127**, 014705 (2007).

[302]J. Behler, "Perspective: machine learning potentials for atomistic simulations", J. Chem. Phys. **145**, 170901 (2016).

[303]K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "SchNet: a continuous-filter convolutional neural network for modeling quantum interactions", in *Adv. Neural. Inf. Process. Syst. 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (NIPS, 2017), pp. 991–1001, arXiv:1706.08566 [stat.ML].

[304]O. T. Unke and M. Meuwly, "PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges", J. Chem. Theory Comput. **15**, 3678–3693 (2019).

[305]N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (Cambridge University Press, 2000).

[306]A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons", Phys. Rev. Lett. **104**, 136403 (2010).

[307]M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning", Phys. Rev. Lett. **108**, 058301 (2012).

[308]A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments", Phys. Rev. B **87**, 184115 (2013).

[309]S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space", Phys. Chem. Chem. Phys. **18**, 13754–13769 (2016).

[310]U. Kaatze, "Complex permittivity of water as a function of frequency and temperature", J. Chem. Eng. Data **34**, 371–374 (1989).

[311]J. Hamelin, J. B. Mehl, and M. R. Moldover, "The static dielectric constant of liquid water between 274 and 418 K near the saturated vapor pressure", Int. J. Thermophys. **19**, 1359–1380 (1998).

[312]N. B. Vargaftik, B. N. Volkov, and L. D. Voljak, "International tables of the surface tension of water", J. Phys. Chem. Ref. Data **12**, 817–820 (1983).

[313]J. B. Hasted, D. M. Ritson, and C. H. Collie, "Dielectric properties of aqueous ionic solutions. Parts I and II", J. Chem. Phys. **16**, 1–21 (1948).

[314]N. Gavish and K. Promislow, "Dependence of the dielectric constant of electrolyte solutions on ionic concentration: a microfield approach", Phys. Rev. E **94**, 012611 (2016).

[315]J. M. Mollerup and M. P. Breil, "Modeling the permittivity of electrolyte solutions", AIChE J. **61**, 2854–2860 (2015).

[316]L. Fumagalli, A. Esfandiar, R. Fabregas, S. Hu, P. Ares, A. Janardanan, Q. Yang, B. Radha, T. Taniguchi, K. Watanabe, G. Gomila, K. S. Novoselov, and A. K. Geim, "Anomalously low dielectric constant of confined water", Science **360**, 1339–1342 (2018).

[317]J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, "Real-space grid implementation of the projector augmented wave method", Phys. Rev. B **71**, 035109 (2005).

[318] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, "Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method", J. Phys. Condens. Matter **22**, 253202 (2010).

[319] A. H. Larsen, J. J. Mortensen, J. Blomqvist, and K. W. Jacobsen, "The atomic simulation environment–a Python library for working with atoms", J. Phys. Condens. Matter **29**, 273002 (2017).

[320] P. E. Blöchl, "Projector augmented-wave method", Phys. Rev. B **50**, 17953–17979 (1994).

[321] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple", Phys. Rev. Lett. **77**, 3865–3868 (1996).

[322] A. Tkatchenko and M. Scheffler, "Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data", Phys. Rev. Lett. **102**, 073005 (2009).

[323] A. Held and M. Walter, "Simplified continuum solvent model with a smooth cavity based on volumetric data", J. Chem. Phys. **141**, 174108 (2014).

[324] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations", Phys. Rev. B **13**, 5188–5192 (1976).

[325] R. Eisenschitz and F. London, "Über das verhältnis der van der Waalsschen kräfte zu den homöopolaren bindungskräften", Z. Physik **60**, 491–527 (1930).

[326] J. Fiedler, P. Thiyam, A. Kurumbail, F. A. Burger, M. Walter, C. Persson, I. Brevik, D. F. Parsons, M. Boström, and S. Y. Buhmann, "Effective polarizability models", J. Phys. Chem. A **121**, 9742–9751 (2017).

[327] A. Beneduci, "Which is the effective time scale of the fast debye relaxation process in water?", J. Mol. Liq. **138**, 55–60 (2008).

[328] R. Bukowski and K. Szalewicz, "Complete *ab initio* three-body nonadditive potential in Monte Carlo simulations of vapor–liquid equilibria and pure phases of argon", J. Chem. Phys. **114**, 9518–9531.

[329] J. G. McDaniel and J. R. Schmidt, "First-principles many-body force fields from the gas phase to liquid: a "universal" approach", J. Phys. Chem. B **118**, 8042–8053 (2014).

[330] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, "Accurate and efficient method for many-body van der Waals interactions", Phys. Rev. Lett. **108**, 236402 (2012).

[331] J. P. Perdew, "Energetics of charged metallic particles: from atom to bulk solid", Phys. Rev. B **37**, 6175–6180 (1988).

[332] P. A. Serena, J. M. Soler, and N. García, "Self-consistent image potential in a metal surface", Phys. Rev. B **34**, 6767–6769 (1986).

[333]N. D. Lang and W. Kohn, "Theory of metal surfaces: induced surface charge and image potential", Phys. Rev. B **7**, 3541–3550 (1973).

[334]M. W. Finnis, R. Kaschner, C. Kruse, J. Furthmüller, and M. Scheffler, "The interaction of a point charge with a metal surface: theory and calculations for (111), (100) and (110) aluminium surfaces", J. Phys. Condens. Matter **7**, 2001 (1995).

[335]A. Held, M. Moseler, and M. Walter, "Charging properties of gold clusters in different environments", Phys. Rev. B **87**, 045411 (2013).

[336]D. Chakrabarti, H. Kusumaatmaja, V. Rühle, and D. J. Wales, "Exploring energy landscapes: from molecular to mesoscopic systems", Phys. Chem. Chem. Phys. **16**, 5014–5025 (2014).

[337]D. Chakrabarti and D. J. Wales, "Simulations of rigid bodies in an angle-axis framework", Phys. Chem. Chem. Phys. **11**, 1970–1976 (2009).

[338]J.-P. Hansen and I. R. McDonald, *Theory of simple liquids: with applications to soft matter* (Academic Press, 2013).

[339]G. Mie, "Zur kinetischen Theorie der einatomigen K ö rper", Ann. Phys. **316**, 657–697 (1903).

[340]J. J. Magda, M. Tirrell, and H. T. Davis, "Molecular dynamics of narrow, liquid-filled pores", J. Chem. Phys. **83**, 1888–1901 (1985).

[341]T. Jacob and L. Sabo, "Electrochemical double layer", in *Encyclopedia of Interfacial Chemistr*, edited by K. Wandelt (Elsevier, Oxford, 2018), pp. 76–79.

[342]R. F. W. Bader, "Atoms in molecules", Acc. Chem. Res. **18**, 9–15 (1985).

[343]A. Singraber, J. Behler, and C. Dellago, "Library-based LAMMPS implementation of high-dimensional neural network potentials", J. Chem. Theory Comput. **15**, 1827–1840 (2019).

[344]F. Hanke, "Sensitivity analysis and uncertainty calculation for dispersion corrected density functional theory", J. Comput. Chem. **32**, 1424–1430 (2011).

[345]K. Berland, V. R. Cooper, K. Lee, E. Schröder, T. Thonhauser, P. Hyldgaard, and B. I. Lundqvist, "van der Waals forces in density functional theory: a review of the vdW-DF method", Rep. Prog. Phys. **78**, 066501 (2015).

[346]M. Stöhr, T. V. Voorhis, and A. Tkatchenko, "Theory and practice of modeling van der Waals interactions in electronic-structure calculations", Chem. Soc. Rev. **48**, 4118–4154 (2019).

[347]B. J. Berne and P. Pechukas, "Gaussian model potentials for molecular interactions", J. Chem. Phys. **56**, 4213–4216 (1972).

[348]M. Yoshio, T. Kagata, K. Hoshino, T. Mukai, H. Ohno, and T. Kato, "One-dimensional ion-conductive polymer films: alignment and fixation of ionic channels formed by self-organization of polymerizable columnar liquid crystals", J. Am. Chem. Soc. **128**, 5570–5577 (2006).

[349]P. Bruce, B. Scrosati, and J.-M. Tarascon, "Nanomaterials for rechargeable lithium batteries", Angew. Chem. Int. Ed. **47**, 2930–2946 (2008).

[350]T. Kato, "From nanostructured liquid crystals to polymer-based electrolytes", Angew. Chem. Int. Ed. **49**, 7847–7848 (2010).

[351]N. Yamanaka, R. Kawano, W. Kubo, T. Kitamura, Y. Wada, M. Watanabe, and S. Yanagida, "Ionic liquid crystal as a hole transport layer of dye-sensitized solar cells", Chem. Commun., 740–742 (2005).

[352]N. Yamanaka, R. Kawano, W. Kubo, N. Masaki, T. Kitamura, Y. Wada, M. Watanabe, and S. Yanagida, "Dye-sensitized TiO_2 solar cells using imidazolium-type ionic liquid crystal systems as effective electrolytes", J. Phys. Chem. B **111**, 4763–4769 (2007).

[353]F. Cichos, K. Gustavsson, B. Mehlig, and G. Volpe, "Machine learning for active matter", Nat. Mach. Intell. **2**, 94–103 (2020).

[354]M. Hübl, *(working-title:) Reinforcement learning of cellular mechanoreciprocity*, Master's thesis (TU Wien, 2020).

[355]R. Golestanian and A. Ajdari, "Analytic results for the three-sphere swimmer at low Reynolds number", Phys. Rev. E **77**, 036308 (2008).

[356]K. O. Stanley and R. Miikkulainen, "Evolving neural networks through augmenting topologies", Evol. Comput. **10**, 99–127 (2002).

[357]A. McIntyre, M. Kallada, C. G. Miguel, and C. F. da Silva, *NEAT-python*, `https://github.com/CodeReclaimers/neat-python`.

[358]D. J. Evans, "On the representatation of orientation space", Mol. Phys. **34**, 317–325 (1977).

[359]T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, and G. J. Martyna, "Symplectic quaternion scheme for biophysical molecular dynamics", J. Chem. Phys. **116**, 8649–8659 (2002).

[360]W. Tang, E. Sanville, and G. Henkelman, "A grid-based bader analysis algorithm without lattice bias", J. Phys. Condens. Matter **21**, 084204 (2009).