THE UNIVERSITY OF DUBLIN

Towards Data-Driven Magnetic Materials Discovery

by

Mario Žic

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Trinity College Dublin
School of Physics

July 2017

Declaration of Authorship

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I was involved in a number of collaborations, and where it is appropriate my collaborators are acknowledged for their contributions.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed: _____

Date: _____

THE UNIVERSITY OF DUBLIN

Abstract

Trinity College Dublin
School of Physics

Doctor of Philosophy

by Mario Žic

Magnetic materials underpin many of the technologies that define the world we live in. Despite the tremendous technological progress, the discovery of new magnetic materials has been rather slow. In this Thesis we explore and develop new, data-oriented, methods for the accelerated discovery and development of new magnetic materials. We utilize available theoretical databases of Heusler alloy properties to: develop a high-throughput (HT) screening procedure for the discovery of new permanent magnets, identify the defects in Mn-Ru-Ga thin films, and build machine learning (ML) models for predicting the structural and the magnetic properties of Heusler alloys. We identify a dozen materials, which meet all of the criteria for permanent magnet applications. The analysis of the HT data allowed us to understand the ideal composition of hard magnetic materials in the family of regular Heusler alloys, and to show how structure-property constraints affect the abundance of the potential candidates for technological applications. We find that hard permanent magnets occur with a frequency smaller than 1 in 10 000, with respect to the overall population of the regular Heusler alloys in the database. We then demonstrate that the ML techniques can be used both to improve the efficiency of the HT procedure and to perform a data-driven investigation of the material properties. In the case of the Mn-Ru-Ga thin films we show how the HT data can be utilized to guide the modelling of technologically relevant materials. The HT data was used to identify the nature of the defects that occur in the films, and hence, to obtain an accurate theoretical description of the material properties. We build ML models to investigate the magnetism of Fe in Heusler alloys. We then study how the local chemical environment affects its magnetic moment and thus address the structure-property relationship directly. We also show that new knowledge about the physics of materials can be extracted directly from the data. This work clearly demonstrates the potential that ML techniques have to offer in the analysis of a vast amount of materials data and paves the way for the future data-driven studies of magnetism.

Acknowledgements

I am in debt to a number of people who have helped me to realize this Thesis. First and foremost, I would like to thank my supervisor Prof. Stefano Sanvito for providing me the opportunity to do the PhD. in his group, for his guidance and trust, which has allowed me much freedom in the research. I also thank him for patiently reading this Thesis and providing valuable feedback. Many thanks are also due to Stefania Negro, who has kindly helped me with many administrative and financial matters throughout these four years.

This PhD. has been made an enjoyable experience by the members of the Computational Spintronics group, past and present - thank you *all*. Many thanks go to Dr. Thomas Archer, who has served as my co-supervisor, and has helped me a lot throughout the PhD. I am particularly grateful for the companionship to: Dr. Rajarshi Tiwari, Filiberto Biolcati Rinaldi, Mario Galante, Dr. Alessandro Lunghi, Dr. Carlo Motta, and Dr. Igor Popov. Special thanks go to Dr. Jacopo Simoni and Emanuele Bosoni for their friendship.

I would like to thank my collaborators from the group of Prof. Mike Coey, especially Dr. Karsten Rode. Working with him was both enjoyable and fruitful. I would also like to thank Prof. Stefan Blügel, for hosting me in his group (Forschungszentrum Jülich, Germany). I thank his group members: Dr. Daniel Wortmann, Dr. Marjana Ležaić, and Dr. Phivos Mavropoulos, for their hospitality and the professional advice. Special thanks go to Phivos, to whom I am greatly indebted. I thank him for his time, his help, and the computer codes, all of which he shared with me so generously. Without his help a big part of this Thesis would not be possible.

I would like to thank my girlfriend, for the companionship, the support, and the patience all these years. Finally, I would like to thank my family for the support and the sacrifices which had allowed me to achieve all this. I hope I will be able to repay some of that debt by making you proud.

This work would not be possible without the financial support from the FP7 project "ROMEO" and Science Foundation of Ireland. The computational resources have been provided by the Trinity Center for High Performance Computing (TCHPC) and by the Irish Centre for High-End Computing (ICHEC).

The template for this Thesis can be found at www.sunilpatel.co.uk.

Contents

Declaration of Authorship iii

Abstract iv

Acknowledgements v

List of Figures xi

List of Tables xvii

Introduction 1

1 Theoretical Background and Methods 7
1.1 The Electronic Structure Theory . . . 7
1.1.1 The Many-Body Problem . . . 9
1.1.2 The Density Functional Theory . . . 14
1.1.2.1 The Kohn-Sham Approach . . . 16
1.1.2.2 The Kohn-Sham Approach for Magnetic Systems . . . 18
1.1.2.3 Interpretation of the Kohn-Sham values . . . 22
1.1.2.4 Approximations of the Exchange Correlation Functional . 25
1.1.2.5 Pseudopotentials . . . 27
1.1.3 Electronic Structure of Solids . . . 29
1.1.3.1 Semi-Classical Theory of Electron Transport . . . 34
1.1.4 The Origin of Magnetism . . . 37
1.1.4.1 Application of SDFT to Magnetic Systems . . . 42
1.2 Machine Learning . . . 50
1.2.1 Statistical Learning Theory . . . 52
1.2.2 Bias-Variance Trade-Off and Over-Fitting . . . 55
1.2.3 Model Selection and the Error Estimation . . . 57

2 Accelerated Discovery of Permanent Magnets Using High-Throughput DFT Calculations 61
2.1 Introduction . . . 61
2.2 High-Throughput DFT Screening Approach for Permanent Magnet Discovery . . . 64

2.3 Material Selection . . . 67
2.4 T_C Calculations . . . 68
2.4.1 Method . . . 68
2.4.2 Results . . . 70
2.5 MCA Calculations . . . 76
2.5.1 Method . . . 76
2.5.2 Results . . . 77
2.6 Discovering Hard Magnets Using a Machine Learning Classification . . . 81
2.6.1 Method . . . 81
2.6.2 Results . . . 84
2.7 Conclusions . . . 88

3 Designing a Fully-Compensated Ferrimagnetic Half-Metal 91
3.1 Introduction . . . 91
3.2 Properties of Stoichiometric Mn_2Ru_xGa Heusler Alloys . . . 95
3.3 Relative Stability of the Competing Mn-Ru-Ga Heusler Phases: a High-Throughput Approach . . . 104
3.4 The Effects of the Mn-Ga Substitution on the Properties of MRG Films . 107
3.5 Conclusions . . . 114

4 Magnetism of Iron in Heusler Alloys: a Machine Learning Approach 117
4.1 Introduction . . . 117
4.2 Method . . . 120
4.2.1 General Considerations . . . 120
4.2.1.1 Defining the problem . . . 121
4.2.1.2 Data Selection . . . 122
4.2.1.3 Building the Model . . . 124
4.2.1.4 Testing the Model . . . 126
4.2.1.5 Application . . . 127
4.2.2 Machine Learning Model for Magnetism . . . 128
4.2.2.1 Heusler Alloy Volume Regression Model . . . 130
4.2.2.2 Magnetic Moment Regression Model . . . 133
4.3 Results . . . 136
4.3.1 Role of the Coordination Shells . . . 136
4.3.1.1 Next-Nearest Neighbour . . . 137
4.3.1.2 Nearest Neighbour . . . 138
4.3.1.3 Physical Origin of the Magnetic Moment Trend . . . 140
4.3.2 Application . . . 142
4.4 Conclusions . . . 144

Conclusions 147

A Mean-Field Curie Temeperature 151

B List of Publications 155

C Glossary **157**

Bibliography **163**

List of Figures

1 Simplified schematic of the computational high-throughput (HT) workflow. First a set of preselected structures is passed to the automatized calculation engine, namely material properties are calculated and post-processed for storage into the database. In the final step the data in the database is subjected to the analysis. The results of the analysis may be the final objective of the HT calculation or just a new selection of structures can be made and fed into the input of the HT cycle. The Roman numerals denote the three key steps of the HT search: I) virtual materials growth, II) rational data storage, and III) data analysis. 2

2 Crystal structure of the (X_2YZ) full Heusler alloys, also known as the $L2_1$ phase. The Z atom is typically a main group element, occupying the 4b site. The Y atom occupies the 4a site and the two X atoms occupy the 4c and 4d site. In the half-Heusler structure, $C1_b$, the 4d site remains unoccupied. 4

1.1 Flow chart of the self-consistent loop used to solve the system of Kohn-Sham (KS) equations (1.21). 18

1.2 Molecular level diagram illustrating the Kohn-Sham (KS) bandgap problem. The HOMO and the LUMO denote the highest occupied and the lowest unoccupied molecular orbital, respectively. The number of electrons in the system, N, is indicated in the brackets. The true ionization energy, I, and the affinity, A, are shown. Within the KS approach, the affinity, A_{KS}, is determined by the KS bandgap, Δ_{KS}. The latter represents a lower bound for the fundamental bandgap, Δ. 24

1.3 The four-fold cross-validation. The data is randomly distributed across four bins of equal size. For each fold, three bins are used for the training (green) and one bin is used for the validation (blue). Four iterations are performed, each time selecting a different bin for the validation. The cross-validation error is obtained as an average error over the four folds. 58

1.4 Hypothetical learning curve plot. The training and the validation accuracy, green and blue curve respectively, are plotted as a function of the training set size. The over-fitting region is indicated by the shaded background. For large training set size the asymptotic accuracy of a given model is achieved. Both curves converge to the same value. 60

2.1 Protocol for the high-throughput screening of permanent magnets. The procedure is broken down into three steps (*I* - *III*). We start from a full database of Heusler alloys [31]. The screening criteria are then applied, i.e. we look for stable magnetic materials with a tetragonal ground-state structure. This is illustrated as a two step data reduction. In the second step we calculate the Curie temperature, T_C, for the candidate materials. The materials with a $T_C < 500\,\mathrm{K}$ are filtered out. Finally, we proceed by calculating the magnetic anisotropy energy (MAE). 66

2.2 Schema of the automated KKR workflow. The user needs to provide the site occupation and the crystal structure data as an input. The workflow takes care of the rest. The ground state is obtained in two steps. First the magnetic moment is converged and then, in the second step, the electron density. Each of these steps is automatically restarted if convergence is not reached. Finally, the Monte Carlo (MC) input files are prepared and the mean-field approximation (MFA) Curie temperature, T_C, is calculated. In case the ground state calculation fails, the error handler inspects the error and decides which recovery procedure to follow. Otherwise, the calculation is stopped. 69

2.3 Comparison of the Curie temperatures, T_C, calculated using the KKR workflow presented here (HTA). *Left Panel* - shows the Monte Carlo (MC) T_C. The results obtained using the HTA are plotted against the corresponding data from the Tillmanns dataset (Tms) [144]. The line of the exact agreement and the $\pm 100\,\mathrm{K}$ error bar are indicated by the full and the dashed lines, respectively. Three of the materials show a poor agreement (diamonds), due to the truncation of the J_{ij} cutoff radius (cf. figure 2.4). *Right Panel* - compares the MC and the mean-field approximation (MFA) T_C, calculated within the HTA. 72

2.4 Comparison of the calculated Curie temperatures, T_C, and the corresponding experimental data. *Left Panel* - comparison of the Monte Carlo (MC) T_C, obtained using the KKR workflow presented here (HTA) and the values from the Tillmanns dataset (Tms) [144]. The line of the exact agreement and the $\pm 100\,\mathrm{K}$ error bar are indicated by the full and the dashed lines, respectively. *Right Panel* - the T_C obtained using the mean-field approximation (MFA), within the HTA. The materials which show poor agreement are shown with diamonds. For Au_2MnAl we show two T_C values, obtained using the cutoff radius of $R_{cut} = 1.2\ a_{lat}$ (HTA default) and $R_{cut} = 2.0\ a_{lat}$, respectively. The truncation of the Jij interaction radius thus results in a strong overestimation of the T_C. 73

2.5 The Curie temperature, T_C, for the 229 material dataset, found by the high-throughput screening procedure (cf. figure 2.1). *Left Panel* - the histogram of T_C for the whole dataset (grey) and the subset of potentially stable materials (black). *Right Panel* - the relationship between the magnetic moment and the T_C. A dashed line, having a slope of $0.003\,\mu_\mathrm{B}\,\mathrm{K}^{-1}$, is shown to guide the eye. 74

2.6 The dependence of the Curie temperature, T_C, on the structure and the composition of the (X_2YZ) Heusler alloys. The data is shown only for the potentially stable alloys, $\Delta H < 0\,\mathrm{eV}$. The T_C is plotted as a function of the magnetic moment, M, the volume, V, the valence, N_V, and the atomic number Z_i, where $i \in \{X, Y, Z\}$ denotes the respective atomic site. 75

2.7 Correlation of the magneto-crystalline anisotropy (MCA) with a number of structural and compositional parameters for the 229 material dataset (grey). The mutual dependence of all of the variables is shown. Namely, the MCA, the magnetic moment, M, the valence electron count, N_V, the volume, V, and the c/a ratio. The compounds with a negative enthalpy of formation, $\Delta H < 0\,\mathrm{eV}$ are shown in black. 78
2.8 The dependence of the magneto-crystalline anisotropy (MCA) on the valence of individual atoms, N_i, and the corresponding atomic numbers, Z_i, where $i \in \{X, Y, Z\}$. 79
2.9 The magneto-crystalline anisotropy (MCA) as a function of the Curie temperature, T_C, and the magnetic moment. The MCA$= 1\,\mathrm{MJ\,m^{-3}}$ and the $T_\mathrm{C} = 500\,\mathrm{K}$ are indicated by the dashed lines. The stable materials, $\Delta H < 0\,\mathrm{eV}$, which fall into this window make a viable candidates for permanent magnet applications. This is further discussed in the text. . . 80
2.10 Receiver operating curve (ROC) for the two hard magnet classifiers discussed in the text (left). The area under the curve (*auc*) is indicated in the brackets. The operating point of an ideal classifier, having $auc = 1$, is indicated by the arrow. In the right panel (bottom), a probability density for the true positive, $p_i(x)$, and the true negative samples, $n(x)$, is shown, for two hypothetical classifiers. In the upper right panel we show the corresponding ROC curves. 83
2.11 *Top panel* - histogram of the site occupation for the 114 material training set. The magnets have been labeled as "true soft", "soft" and "mixed". The first group is made from materials which have MCA $\leq 0.8\,\mathrm{MJ\,m^{-3}}$, as predicted by DFT. The second two groups are labeled according to the CLF_2 prediction. The "soft" magnets are those, which are classified as positive, when the classifier operating point is set just before the knee in figure 2.10. In this case there is no misclassified samples. The negative samples are labelled as "mixed", since they are a mixture of the hard and the soft magnets, that can not be resolved by the classifier. *Bottom panel* - a closer look at the trends in the mixed magnet group. Here the materials have been labeled as the "true hard" and the "true soft", according to the MCA value predicted by the DFT. 86
2.12 The learning curve for the CLF_2 classifier. 87

3.1 Illustrated density of states, $D(E)$, of common magnetic metals. The absolute magnetic moment, M, and the spin polarization, P, are shown. The spin up and the spin down channel are denoted by arrows. a) paramagnet ($M = 0$ and $P = 0$), b) ferromagnet/ferrimagnet ($M \geq 0$ and $0\,\% \leq P \leq 100\,\%$), and c) half-metal ($M \geq 0$ and $P = 100\,\%$) 92
3.2 The SDFT ground state energy of Mn_2Ga for three different magnetic orders: the diamagnetic (diam), the ferromagnetic (ferrom), and the ferrimagnetic (ferrim). The equilibrium lattice constants are indicated by the minima of the quadratic curves, obtained by fitting the energy as a function of the lattice constant. 97
3.3 Total energy and magnetization for Mn_2Ga and Mn_2RuGa, top and bottom pannel respectively, as a function of c/a lattice parameter. For each energy curve the volume is constant, and determined by the cubic lattice constant a. The lowest energy curve, and the corresponding magnetization, is shown in blue. 99

3.4 Magnetism in cubic Mn_2Ru_xGa compounds. a) Site projected magnetic moment for the two Mn sites, 4a and 4c, as a function of the lattice constant, a, and the Ru fraction, x. The vertical lines denote the Mn_2Ga lattice constant as determined by the theory, 5.7 Å, and the experiment, 5.97 Å. b) Difference of the Mn magnetic moments, ΔM, as a function of the lattice constant (left) and the Ru concentration (right). The dashed line shows a linear interpolaton of the experimentally observed magnetization slope. 100
3.5 Electronic structure for stoichiometric Mn_2Ru_xGa compounds at the experimental lattice constant, 5.97 Å. The unit cell is cubic. The spin up/down channels are shown in the upper/lower panel, respectively. . . . 101
3.6 Transport properties for three stoichiometric Mn_2Ru_xGa compositions, x = 0.33, 0.66 and 1.0. Fermi level position, E_f, is determined by the valence electron number, N_V. *Left panel* - Spin resolved resistivity (black) and the corresponding DOS (blue). The data corespondingto the spin down channel is represented with negative values. *Right panel* - Calculated transport spin polarization. 103
3.7 Isovolumetric energy profiles for stoichiometric Mn_2Ga calculated at the observed experimental and the relaxed DFT primitive unit cell volume, $V_{EXP} \approx 52.9$ Å and $V_{DFT} \approx 46.4$ Å, respectively. The theoretically expected c/a ratios are denoted by the vertical dashed lines. ΔE is the energy difference between the experimental and the relaxed DFT structure. E_d is the energy gain due to the formation of Ga defects. The net energy gain by forming the defects and taking the experimental structure is denoted by E_g. 106
3.8 Calculated pressure for cubic MRG compounds at experimental (in-plane) lattice constant, $a = 5.97$ Å, as a function of electronic doping, n_{el}. . . . 108
3.9 Magnetic moment of Mn_2Ru_xGa compounds. The DFT values calculated for stoichiometric compounds are shown by the green line. Correction to the DFT magnetization due to the formation of Ga defects is shown by the black points. The corresponding estimate of the Mn/Ga ratio is shown in the inset. The dashed line, equivalent to doping of 2 e per Ru, is shown to guide the eye. 111
3.10 The spin resolved resistivity, $\rho(E)$, and the corresponding spin polarization, $P(E)$, as a function of energy. A scaled DOS is shown as a shaded background. The position of the spin gap is marked the divergence of the resistivity in the majority spin channel. 112
3.11 Calculated Mn_2Ru_xGa spin polarization, for different levels of electron doping, n_{el}. The strain induced polarization is shown by the thick black line. The experimentally reported range of values is shown by a shaded background. 113

4.1 The local coordination of atomic sites in Heulser alloys. The neighbours of the central atom form two shells of different symmetry. Atoms belonging to the nearest neighbour shell, shown in blue and magenta, coordinate the central atom tetrahedrally. The next nearest neighbour shell is made out of 6 (green) atoms and has octahedral symmetry. 119

4.2 Supervised learning. The ML approach consists of 5 steps: **a)** defining the problem, **b)** selecting and pre-processing the data, **c)** building the ML model, **d)** testing the model, and, finally, **e)** applying the ML model to the problem of interest. 122
4.3 The learning curve for the volume regressor. The score is plotted as a function of the training set size. For each case the training and the validation score are shown. The feature space is spanned by a 9-dimensional input vector of Eq. 4.1. 132
4.4 Accuracy of the volume regressor. *Left Panel:* The ML estimate is plotted against the reference DFT data. The points which belong to the training set (189 points) are shown smaller (green). The test data (82 points) are shown in red. The line of the exact agreement with the reference data is shown in blue. *Right Panel:* The frequency at which the regression error occurs, as a function of its relative value (ΔV / V_{DFT}). The distribution was estimated from the test data. 132
4.5 Accuracy of the magnetic moment regressor. *Left Panel:* The ML estimate is plotted against the reference DFT data. The line of the exact agreement data is shown in red. *Right Panel:* The frequency at which the regression error occurs, as a function of its relative value (ΔM / M_{DFT}). The distribution was estimated from the test data. 135
4.6 a) Magnetic moment of the Fe atom as a function of the next-nearest-neighbour atomic number, Z_3, for a Wigner-Seitz radius of $2.7\,a_0$. The c/a ratio of the parent Heusler alloy is color coded. A machine learning estimate is shown by a solid blue line. b) The same data as previously, only now the radius of the points is proportional to the calculated enthalpy of formation, ΔH [210]. Large circles correspond to more stable alloys. . 137
4.7 *Left Panel:* Magnetic moment of Fe for a wide range of nearest neighbours, at a constant Wigner-Seitz volume of $\approx 12\,\text{Å}^3$. The atomic number of the next nearest neighbour, Z_3, is color coded. We can notice a linear increase of the magnetic moment in the transition metal series which does not depend on Z_3. The corresponding machine learning trend is shown as a blue line. *Right Panel:* a data sample containing a wider range of main group elements. The data elucidates the origin of the oscillation in the machine learning trend throughout the main group series in the left panel, indicated by a shaded ellipse. 139
4.8 The integrated charge (blue) and the magnetization (green) of Fe in the TM_2FeZ clusters, as a function of the transition metal (TM) atomic number, Z_{TM}. **a)** the total charge and the magnetization. **b)** the orbital resolved charge and the magnetization. The t_{2g} and the e_g orbitals are shown as full and doted line, respectively. **c)** The orbital resolved charge for the spin up and the spin down channel represented with positive and negative values, respectively. 141
4.9 Projected density of states of Fe in TM_2FeZ clusters, where the selected TMs are the Sc ($Z_1 = 21$) and the V ($Z_1 = 23$). The t_{2g} band is shown in the left panel and the e_g in the right. The spin up and down channels are shown as the positive and the negative values, respectively. 141

4.10 The magnetic moments in the LTM_2FeMG Heusler alloys. The machine learning regression results are compared to the corresponding DFT values. The magnetic moment of Fe is predicted directly by using the regression (green dots). The total magnetic moment of the alloy (blue dots) is estimated by using the empirical correction scheme described in the text. The exact agreement line (red) and the $\pm$ 0.5 μ_B error bars are shown to guide the eye. 143

List of Tables

2.1 Summary of minimal required properties of a permanent magnet, namely the saturation magnetization, M_s, the uniaxial magnetic anisotropy constant, K_u, and the Curie temperature, T_C. 63

2.2 Comparison of the mean-field Curie temperature, T_C, calculated using the KKR, T_C^{KKR}, and the FLEUR code, T_C^{FLEUR}, with the experimental, T_C^{exp}, and the previously reported theoretical values, T_C^{ref} [26, 137]. The temperatures are expressed in units of K. The calculation details are described in [138]. All calcualtions were performed for the regular cubic Heusler structure. The respective lattice constants, a, are shown in the table. 68

2.3 A list of candidate materials for permanent magnet application found using the high-throughput screening approach described in this work. . . 80

3.1 A comparison of the experimental lattice constants, a, and magnetic moments, M, for the Mn_2Ru_xGa compounds with the corresponding DFT values. The respective units are Å and $\mu_B\, f.u.^{-1}$. 98

3.2 Calculated enthalpies of formation, ΔH, for the most stable competing Heusler phases of the 1221 structural and magnetic Mn-Ru-Ga cells investigated. The configurations investigated are limited to the primitive 4-atom Heusler cell (3-atom for half-Heuslers). 105

3.3 The electron doping estimate obtained using equation (3.6). The in-plane lattice constant, $a_{exp} = 5.97$ Å, and $c/a = 1.02$ are taken as parameters. . 110

Dedicated to my family.

Introduction

The prosperity of modern society is mainly based on the constant development of new technologies, allowing us to solve everyday tasks and problems more efficiently. Such a development is enabled by the discovery of new materials, which underpin the new technologies. For example, the development of the spin-valve [1] and the magnetic tunnel junction [2, 3] has revolutionized the computer storage technology by making high-capacity magnetic storage cheap and affordable to a wider range of consumers. Online storage services have consequently become widely used by both companies and individuals, transforming the way people live and work. Many other material classes, like room-temperature superconductors, photovoltaics, thermoelectrics, and catalysts, to name a few, have game changing potential. The desired properties are sometimes intrinsic to a particular material, and hence easier to discover, but many require a suitable combination of materials and environment conditions to manifest itself and that is where the combinatorial possibilities explode. In such situations, classical research methods e.g. experimental work, requiring man-work, time, and other resources, become a bottleneck because they are hard to scale-up. Modern society requires that the new discoveries and technological advancements are continuously achieved, putting a lot of pressure on the research community. In order to accelerate discoveries new research methods need to be developed.

The technological developments brought a significant advancement to the computer technology and computational science. These technologies nowadays offer a scalable solution for the combinatorial bottleneck that impedes a fast material development. An exponential increase in the available computational power over a period of 40 years has made computers affordable and accessible to a large community of researchers. The continuous development of computers has also resulted in the appearance of a number of computer codes, which are used to calculate a variety of material properties. Many of these codes are freely and easily accessible to the research community, thus enabling researchers to simulate complex materials and phenomena with very little overhead, associated to planning and setting up the experiment. One of the key advantages of computational research is that one and the same computer can run an unlimited number of codes and

thus support the work of many research groups. This makes the computational research relatively cheap, as compared to experimental work, and the production capacity can be easily scaled-up.

The high-throughput (HT) methods are nowadays actively developed by the research community, as a way to accelerate the material discovery. The defining trait of the method is the exploration of a large, possibly opened, space of parameters [4]. To exemplify what is meant by an open parameter space, one can imagine a HT search for new stable structures, where the set of structures to be explored is not assigned in advance but determined automatically during the search, e.g. by means of a heuristic algorithm. Unlike the HT method, a combinatorial search has a limited and predefined parameter space. Although there is no strict and generally accepted definition of the HT method, it is worthy to mention the definition given by Curtarolo et al. [5]. According to the authors, the HT method can be defined as: " ... the throughput of data that is way too high to be produced or analysed by the researcher's direct intervention, and must therefore be performed automatically: HT implies an automatic flow from ideas to results". The HT infrastructure involves both hardware and software, which facilitate the HT workflow. The computational HT workflow begins with a candidate material selection, followed by a large scale calculations of predefined properties, and finishes with data post-processing (see figure 1). The previously described loop can be repeated, if necessary, in order to refine the search using newly obtained results. The HT workflow is also discussed in reference [6].

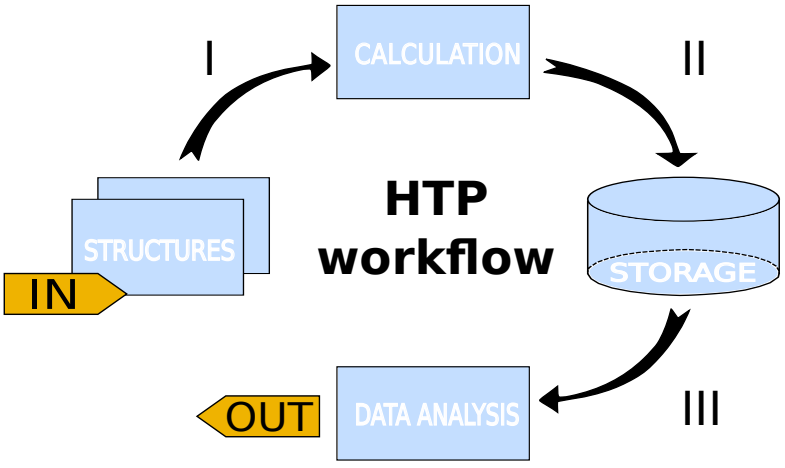

FIGURE 1: Simplified schematic of the computational high-throughput (HT) workflow. First a set of preselected structures is passed to the automatized calculation engine, namely material properties are calculated and post-processed for storage into the database. In the final step the data in the database is subjected to the analysis. The results of the analysis may be the final objective of the HT calculation or just a new selection of structures can be made and fed into the input of the HT cycle. The Roman numerals denote the three key steps of the HT search: I) virtual materials growth, II) rational data storage, and III) data analysis.

The HT method has been applied to a wide range of investigations, namely in the search of the thermodynamic stability of binary and ternary compounds [7–9], photovoltaics [10], water photo-splitting [11] and catalysis [12, 13], carbon capture and gas storage [14–16], scintilators [17], piezoelectrics [18, 19], thermoelectrics [20] and many more. The need to process and interpret the HT results has driven the introduction of data-mining and the machine learning (ML) methods into the material science [4, 21]. The amount of data that is produced in such studies is overwhelming and precludes the traditional, case study, investigations. The large amount of data, however, allows one to adopt a completely new approach to materials science. The data produced by the HT search can be analysed using a range of ML methods, e.g. principal component analysis, clustering techniques, artificial neural networks, genetic algorithms, regression methods, decision trees, self-organizing maps, etc. [22, 23]. The main aim of such studies is to establish a correlation, or ideally a statistical model, between the properties of the material and its structure, composition, etc., namely to extract the knowledge about the structure-property relationship (SPR). The latter can then be used to optimize the material properties. For example, the Slater-Pauling rule [24, 25] establishes a linear dependence between the valence number of electrons and the magnetic moment of a magnetic alloy. In some cases, simple (linear) relations can be found even for the highly non-trivial properties, like the Curie temperature [26]. In general, the SPR will be complex and multidimensional, and it can not be easily identified nor comprehended by the human. In order to fully appreciate the problem, we need to consider that HT data samples the (approximate) solutions for a large number of complex quantum mechanical systems. The researchers task then is to find an "effective model", which can condense and distil these observations into a useful knowledge about the materials. The statistical and the machine learning methods offer us new ways to do this, namely to identify patterns, to extract knowledge, and to build the models in a data-driven fashion. These methods thus have the potential to change how science is being done. However, we are now in the midst of an early adoption stage, where the full potential of these new tools still needs to be discovered and their utility demonstrated.

In this Thesis we will apply the HT and the ML methodology to the problem of finding and engineering new magnetic materials. Magnetism and the magnetic devices are important for both applied and theoretical materials science. The latter being related to the inherent complexity associated with the accurate description, and the understanding, of the properties of the magnetic systems. The practical importance of magnetic systems can not be overemphasized. We can freely say that the magnetic devices underpin modern society. Magnets can be found in all segments of everyday life, from high-performance electric motors and generators, to sensors, to electronics devices and computer memories, to trivial objects like toys and fridge magnets. Although there

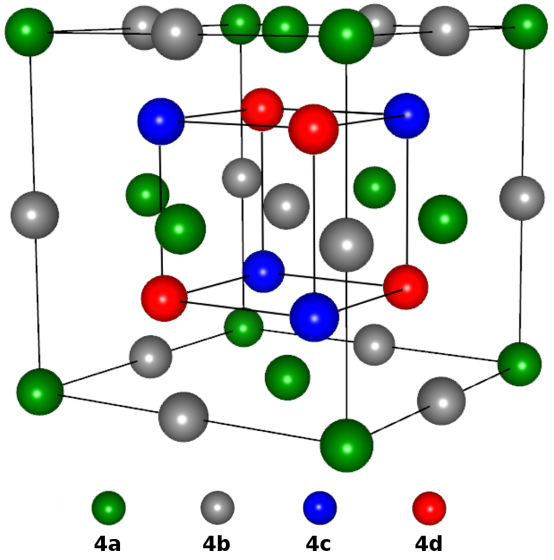

FIGURE 2: Crystal structure of the (X_2YZ) full Heusler alloys, also known as the $L2_1$ phase. The Z atom is typically a main group element, occupying the 4b site. The Y atom occupies the 4a site and the two X atoms occupy the 4c and 4d site. In the half-Heusler structure, $C1_b$, the 4d site remains unoccupied.

are a few thousands of materials, which are known to order magnetically, perhaps only a dozen of them have found practical applications [27]. The variety of magnets used today, in the sense of performance and characteristics, has been obtained by material engineering, by optimizing the composition and the production process. The discovery of new magnetic phases, especially those suitable for the permanent magnet applications, has been scarce [28]. Here we investigate the magnetism in Heusler alloys. The Heusler alloys are a versatile class of intermetallic materials with more than 1500 known members [29]. The range of properties exhibited by these materials: metals, half-metals, semiconductors, topological insulators, and many more, is extraordinary. The fact that these materials crystallize in only two, fcc based, structures ($C1_b$ and $L2_1$) makes them extremely appealing for applications. The properties of these materials can be tuned by changing the chemical composition and there is no mismatch of crystal structures. The crystal structure of the Heusler alloys is shown in figure 2. Since the unit cell consists of 3 to 4 atoms the number of material combinations to be explored is huge. All this makes it an ideal set of materials for a HT study. In addition to the great potential that these alloys have to offer, there is a strong (practical) reason for this choice. Namely, two large HT databases [30, 31], containing information about 200 000 to 300 000 Heusler alloys, are readily available. In particular, we use the data found in [31] as a starting point for all our investigations.

The Thesis is structured as follows. In Chapter 1 we give an overview of density functional theory (DFT), in particular its extension to magnetic systems. DFT is used throughout the work as a primary tool for obtaining the electronic structure of materials.

The basics of the electronic structure theory are reviewed. Our goal is to demonstrate how the DFT information is used to describe the properties of real materials. Special attention is devoted to the origin of magnetism in solids. We then discuss the methods for calculating the magneto-crystalline anisotropy and the Curie temperature, which are later used to characterize magnetic materials. Finally we give a gentle overview of the statistical learning theory and discuss the practical aspects of the ML approach, viewed as a function fitting problem. In Chapter 2 we design and test a HT screening procedure for finding new permanent magnets within the family of the regular Heusler alloys. The results of the HT investigation are used to identify the optimal composition of a hard permanent magnet. Finally, a list of candidate materials, which meet all of the HT screening criteria, is compiled. In addition, a simple statistical analysis allows us to estimate the likelihood of finding new high-performance magnets. We then utilize the ML classification in order to enhance the efficiency of the HT screening procedure. In Chapter 3 we use the high-throughput data to perform a stability analysis of the competing phases in Mn-Ru-Ga thin films, which show a great promise for realising a fully-compensated ferrimagnetic half-metal [32, 33]. This is a material, which holds a great promise for the application in spintronics [34]. By combining DFT investigation and the HT data we identify the defects in these materials and investigate their effect on the intrinsic material properties. We then construct models, which capture all the key properties of the nominal Mn_2Ru_xGa films, thus bridging a gap between the theoretical (DFT) description and the experimental results. In Chapter 4, we build ML models capable of describing the magnetism of Fe in regular Heusler alloys. In comparison to other relevant materials science topics, magnetism has received very little attention in terms of the application of the ML techniques [35, 36]. Due to the novelty of the ML approach, we discuss in great detail the practical aspects of building a ML model. These considerations are then followed, step-by-step, in the construction of ML models for magnetism. We use these models to identify and understand the physical origin of the magnetic moment trends, with respect to the local chemical environment of Fe. Based on the acquired knowledge, the ideal prototype for an Fe-based (X_2FeZ) Heusler alloy is postulated. We thus address the SPR directly. Using the ML models developed here, we are able to characterize the magnetization for all alloys belonging to this prototype. The most promising material candidates for the technological applications are identified. The entire study is purely data-driven. No auxiliary DFT calculations have been performed. We thus demonstrate the potential that the ML techniques have to offer in the analysis of vast amounts of materials data and pave the way for the future data-driven studies of magnetism.

Chapter 1

Theoretical Background and Methods

1.1 The Electronic Structure Theory

The importance of materials for the development of the human civilization can not be overstated. Ever since ancient history mankind has been shaping materials to aid them in overcoming their daily challenges. The development of science, Physics and Chemistry in particular, has allowed us to understand what makes certain, naturally occurring, materials special; how to classify them, and how to make completely new functional materials. This knowledge makes one of the pillars of the modern society. Today we understand that all these properties arise from the interaction between the atoms. More precisely, electrons interact with the atomic nuclei and with other electrons to give rise to a wide spectrum of material properties. These interactions are described in full by the laws of quantum mechanics. Since we know what are the relevant interactions in any material, we can always write down the system of equations whose solution would allow us to calculate any imaginable property of that material. In other words, today we know how to write down the Hamiltonian for an arbitrary material and how to obtain the solution, i.e. the wavefunction. Within the framework of quantum mechanics the wavefunction contains all the information about the system. However, obtaining an exact solution often represents an intractable problem.

Electronic structure theory (EST) is concerned with two issues that arise naturally. The first is the necessity to establish the mathematical formalism and methods, which can be used to solve the aforementioned problem in practice. The second is to provide a formal connection between the quantum mechanical solution and the material properties, which are sometimes phenomenological in character. For example: “Why does a material

have a silver or yellow color, or why is it transparent ?!" Likewise, one would hope to provide an explanation for all the rules and the phenomena observed in Chemistry. The core of the problem, which precludes finding an exact solution, in most of the cases, is the interaction between a large number of electrons. In the absence of the latter the material physics problem would be a rather simple one, namely it would be reduced to an electrostatic problem. At the same time we would loose the concept of chemical bond and life as we know would not be possible. Therefore, the description of electron-electron (e-e) interactions is crucial for understanding of the properties of matter, which in turn affect many aspects of our daily life.

Unfortunately, the interaction between a large number of electrons makes it one of the most difficult problems in the entire physics. The crux of the problem lies in the fact that the electrons are fermions and thus need to obey the Pauli exclusion principle. This makes the problem intractable for more then a dozen of electrons. Any many-electron wavefunction can, in principle, be represented as a linear combination of Slater determinants [37], which satisfy the Pauli principle by construction. However, for moderately large systems such a direct approach will likely never be feasible, due to the factorial scaling. In practice, exact numerical solutions can be found only for very small systems, containing tens of electrons. Luckily, relatively simple approximations, like the independent-particle approach, can be used to obtain a reasonably good description of the electronic structure for a wide range of materials. Here the electrons are considered as independent particles moving under the influence of an effective one-body potential, which describes the combined interaction with all other electrons in the system in some approximate way. This approach was first pioneered by Hartree [38, 39] and later developed by Fock [40]. Nowadays it is known as the Hartree-Fock method and it represents the simplest wavefunction-based approach for solving many-body problems. Density functional theory (DFT) is another example of such an approach, which is nowadays widely used as a universal tool to describe realistic materials [41]. Although DFT is in principle an exact method, within the Kohn-Sham scheme it is reduced to an independent-particle theory, where the e-e interactions are treated in an approximate way via an effective "mean-field" potential. Nevertheless, it is simple and accurate enough to be considered as the most important method in electronic structure theory today. The approaches which go beyond the independent-particle description are known as the many-body methods. Here we can distinguish two main approaches, namely the wavefunction-based and the Green's function-based methods. The former are often employed in chemistry and hence are known as the quantum-chemical methods. The approach relies on a systematic expansion of the true many-body wavefunction in terms of Slater determinants [42, 43]. This expansion in principle leads to a converging and formally exact solution. In addition, the procedure allows for a simple interpretation

in terms of the admixture of different electronic configurations, which is important for gaining chemical intuition, e.g. about the nature of the chemical bonds. The principal shortcoming is that these methods can only be applied to small systems. The investigation of extended systems, e.g. solids, is not practically feasible using such an approach. Since the physics of solids was, and still is, an active area of research, alternative approaches mainly based on the Green's function methods were developed. The Green's functions allow an elegant interpretation of the many-body processes in terms of the motion of particles and their interactions. These processes can be visualized in terms of the Feynman diagrams [44], which makes the approach appealing to physicists. The perturbation approach can be efficiently represented in terms of the Green's functions. The many-body problem can then be solved using an iterative approach, where at each iteration the description of the e-e interaction can be refined. However, an important point is that the e-e interaction is often large enough to make the perturbation approach invalid. Therefore, the many-body methods often do not work with electrons directly but instead use Green's functions to define the equation of motion for quasiparticles, i.e. for long lived excitations, which appear under the influence of the external perturbation. By assumption, these quasiparticles are interacting via a weak force, namely the screened Coulomb interaction, which makes the perturbation approach possible. The problem here is to obtain the reasonable initial ground state for which the perturbation can be applied. Lower level methods, like DFT or Hartree-Fock, are commonly used for this. The second issue is related to the truncation of the interaction potential. More specifically, it is hard to define a priori which diagrams will yield a rapidly converging perturbation series. This makes the details of the approach problem specific. An important point is that the effective screened potential depends on the definition of the quasiparticle, and hence the equations of motion need to be solved self-consistently. The computational effort needed for this procedure is often prohibitive. The non-triviality of the mathematical formalism, which underpins this approach, together with the aforementioned practical limitations, confine the use of these methods to a number of specialist areas, e.g. for studying optical excitations. DFT is at the present the only method which provides practically useful results while requiring a modest computational effort. In addition, DFT has a low entry barrier, i.e. it easy for the beginners to start using it. These are probably the main reasons why DFT is used so often in the condensed matter physics today, despite all its shortcomings.

1.1.1 The Many-Body Problem

The properties of an arbitrary material can be determined from the solution of the Schrödinger equation, which is defined by a Hamiltonian of the following form

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i,j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,l}\frac{Z_l}{|\mathbf{r}_i - \mathbf{R}_l|} + \hat{E}_{ext}^{(0)} + \hat{H}_I \ . \tag{1.1}$$

Here the index i goes over all the electrons and l over all the nuclei of the system. Hartree units, $\hbar = e = \frac{4\pi}{\epsilon_0} = 1$ were assumed, with e being the electron charge and ϵ_0 the free space permittivity. The atomic number of atom l is denoted as Z_l. The first two terms in Eq. (1.1) describe the electronic degrees of freedom, i.e. the kinetic energy and the e-e interaction, respectively. $\hat{H}_I$ denotes the Hamiltonian for the positively charged nuclei, or simply ions. This term becomes relevant only when the ion dynamics is studied, e.g. in phonon calculations. If the atomic centres are not allowed to move this term becomes just an additive constant, which redefines the zero energy. In this case it does not play an important role for the electron dynamics. The third term describes the electron-ion interaction.

The operator $\hat{E}_{ext}^{(0)}$ represents an arbitrary external potential, e.g. a magnetic or an electric field. It is often convenient to treat the electron-ion interaction as a part of the external potential, in which case we can define

$$\hat{E}_{ext} = \hat{E}_{ext}^{(0)} - \sum_{i,l}\frac{Z_l}{|\mathbf{r}_i - \mathbf{R}_l|} \ , \tag{1.2}$$

and the Hamiltonian can be symbolically expressed as

$$\hat{H} = \hat{T} + \hat{H}_{ee} + \hat{E}_{ext} + \hat{H}_I. \tag{1.3}$$

We note that the original Hamiltonian given by equation (1.1) is non-relativistic. The leading relativistic corrections, namely the spin-orbit coupling, the mass-velocity and the Darwin term [45], are often added as single-particle terms, after the many-body problem has been replaced by an effective independent-particle system. The stationary Schrödinger equation, defined by the Hamiltonian (1.3), then becomes

$$\hat{H}\Psi(\{\mathbf{r}_i\};\{\mathbf{R}_l\}) = E\ \Psi(\{\mathbf{r}_i\};\{\mathbf{R}_l\}) \ , \tag{1.4}$$

where E is the total energy of the system and $\Psi(\{\mathbf{r}_i\};\{\mathbf{R}_l\})$ is the wavefunction, which contains all the information about the system. The latter depends explicitly on both the electronic and the ionic degrees of freedom. In most of the situations of interest the ions move slowly enough compared to the characteristic electron time-scale, $\approx 10\,\text{fs}$, or do not move at all. The characteristic time-scale for the vibrations is of the order of ps. Therefore, the electronic degrees of freedom are in many cases practically decoupled from the ionic motion. The wavefunction $\Psi(\{\mathbf{r}_i\};\{\mathbf{R}_l\})$ can then be separated into a

product of two terms

$$\Psi(\{\mathbf{r}_i\};\{\mathbf{R}_l\}) = \psi(\{\mathbf{r}_i\}) \cdot \Phi(\{\mathbf{R}_l\}) \ , \tag{1.5}$$

where each term now depends exclusively on either the electronic or the ionic degrees of freedom. The Schrödinger equation then decomposes into two equations, coupled by the electron-ion interaction. This is known as the Born-Oppenheimer approximation. The electronic part is solved first for a fixed ionic configuration, and then the force on the ions can be evaluated and the ion positions updated. Since the ions have a large mass, the quantum-mechanical phenomena associated with the wave nature of matter does not usually manifest itself for them. For this reason the ions are often described using semi-classical equation of motion. This further simplifies the problem. The non-classical part involves the action of the electrons on the nuclei, i.e. the electronic wavefunction is used to evaluate the force on the nuclei. With the exception of hydrogen, this approximation works very well for the usual range of temperatures and pressures. Matter under extreme conditions requires special attention, which is out of the scope of this thesis. We note that the ionic positions in the Born-Oppenheimer approximation define an effective energy surface for the electrons. If the ions are allowed to move, these surfaces can intersect and the approximation will break down. In this case a small displacement of the nuclei can cause a large change in the electronic wavefunction. Therefore, as long as the potentially energy surfaces are well separated and the motion of the nuclei can be considered adiabatic the Born-Oppenheimer approximation will remain valid.

The independent-electron approximation, where the motion of individual particles is unaffected by the other ones, at least not explicitly, is of great practical importance. The most simple approach, which would result in such an approximation would be to neglect the e-e interaction term altogether. In this case the motion of every electron would be affected only by the electric potential of the ions and the individual electrons would move in a completely uncorrelated manner. However, this approximation is often not realistic. A more common approach is to take the e-e interaction into account in some approximate way, e.g. via an effective one-body potential. An example of such a model is the nearly-free electron model (see Ch. 7 in Ref. [46]), which was successfully used in the early days of solid state physics to study the properties of metals. Despite its simplicity, the approach allowed the development of many concepts in electronic structure theory, like the concept of bands, Brillouin zone, etc. The opposite limit is given by tight-binding models. Here one starts from an atomic reference state and treats the electron hopping between the different sites as short-ranged. Although the physical picture in these models is grossly simplified, they elucidate why the notion of atoms is still relevant in solids and, more importantly, represent a great pedagogical tool. When equipped with a proper parametrization, these models can provide a reasonably accurate

description of the electronic structure [47]. In this sense, they are still used in the solid state physics research.

The most important extension of the previously discussed independent-particle approach is the Hartree-Fock method (HFM). The crux of the method is to take the full interacting Hamiltonian and to approximate the many-body wavefunction with a single Slater determinant. In this way the Pauli principle is satisfied by construction. Since the wavefunction is expressed in terms of the single-particle orbitals, the method is unable to describe completely the e-e correlations. By including additional degrees of freedom in the wavefunction it is possible to further lower the energy of the system [48]. The correlation induced by Pauli principle is known as the Fermi correlation. However, it acts only between particles of the same spin (cf. equation 1.10). The correlation between the electrons of different spin, due to the Coulomb interaction, is not captured by the HF approximation. This is known as the Coulomb correlation. In order to capture the latter, approaches that go beyond the HF are needed. In quantum-chemistry these are known as the post-Hartree-Fock methods, e.g. the Møller-Plesset perturbation theory, the configuration interaction and the coupled-cluster methods [42, 43, 49]. Within many-body perturbation theory particle correlations are captured by the quasi-particle self-energy operator [50] (see also Ch. A4 in Ref. [51]), which plays the role of an effective, non-local, single-particle potential.

In order to gain a clearer insight into the problem of correlation, it is instructive to express the 2-particle density, $n(\mathbf{x}, \mathbf{x}')$, using the normalized pair distribution function, $g(\mathbf{r}, \mathbf{r}')$, which is defined as

$$g(\mathbf{x}, \mathbf{x}') = \frac{n(\mathbf{x}, \mathbf{x}')}{n(\mathbf{x})n(\mathbf{x}')} \ . \tag{1.6}$$

The single-particle density, $n(\boldsymbol{x})$, and the two-particle density, $n(\mathbf{x}, \mathbf{x}')$, are defined as

$$\begin{aligned} n(\boldsymbol{x}) &= N \int d\boldsymbol{x}_2 \ d\boldsymbol{x}_3 \ \ldots \ d\boldsymbol{x}_N \ |\Psi(\boldsymbol{x}, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_N)|^2, \\ n(\boldsymbol{x}, \boldsymbol{x}') &= N(N-1) \int d\boldsymbol{x}_3 \ d\boldsymbol{x}_4 \ \ldots \ d\boldsymbol{x}_N \ |\Psi(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{x}_3, \ldots, \boldsymbol{x}_N)|^2 \ , \end{aligned} \tag{1.7}$$

and $\boldsymbol{x}$ stands for a collective coordinate of the position and the spin, $(\boldsymbol{r}, s)$. The "integral" over the spin is just a shorthand notation for a summation. Naturally, the definition of the pair distribution function makes sense only if we have two or more electrons in the system. The Pauli principle then implies that, in a N-particle system, the electrons form a single wavefunction, which is given in the simplest case by a single Slater determinant. At all times the motion of electrons need to be such that the antisymmetry of the wavefunction is preserved under particle permutation. This implies that the 2-particle electron density, $n(\mathbf{x}, \mathbf{x}')$, which completely determines the Coulomb

interaction, can not be simply written as the product of the two single-particle densities, i.e. $n(\mathbf{x},\mathbf{x}') \neq n(\mathbf{x}) \cdot n(\mathbf{x}')$. The latter can be interpreted as the conditional probability of finding one electron, having spin s', at $\mathbf{r}'$ if the other one, having spin s, is found at $\mathbf{r}$. By writing this as a product of two probabilities would mean that the two events are statistically independent, i.e. uncorrelated. However, this is not true. For fermions the probability to find two particles in the same state is exactly zero. Using the definition (1.6) we can write the e-e interaction energy, E_{ee}, as

$$
\begin{aligned}
E_{ee} &= \frac{1}{2}\int d\mathbf{x}\; d\mathbf{x}'\; \frac{n(\mathbf{x},\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} = \frac{1}{2}\int d\mathbf{x}\; d\mathbf{x}'\; \frac{n(\mathbf{x})n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\; \frac{n(\mathbf{x},\mathbf{x}')}{n(\mathbf{x})n(\mathbf{x}')} \;= \\
&= \frac{1}{2}\int d\mathbf{x}\; d\mathbf{x}'\; \frac{n(\mathbf{x})n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\; g(\mathbf{x},\mathbf{x}') = \\
&= \frac{1}{2}\int d\mathbf{x}\; d\mathbf{x}'\; \frac{n(\mathbf{x})n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + \frac{1}{2}\int d\mathbf{x}\; d\mathbf{x}'\; \frac{n(\mathbf{x})n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\; \big(g(\mathbf{x},\mathbf{x}')-1\big) = \\
&= \frac{1}{2}\int d\mathbf{x}\; n(\mathbf{x}) \int d\mathbf{x}' \frac{n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + \frac{1}{2}\int d\mathbf{x}\; n(\mathbf{x}) \int d\mathbf{x}' \frac{n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\; \big(g(\mathbf{x},\mathbf{x}')-1\big) = \\
&= \frac{1}{2}\int d\mathbf{x}\; n(\mathbf{x}) V_H[n(\mathbf{x})] + \frac{1}{2}\int d\mathbf{x}\; n(\mathbf{x}) V_{XC}[n(\mathbf{x})] = \\
&= E_H + E_{XC}\;. \qquad (1.8)
\end{aligned}
$$

The first term is the Hartree energy, which represents the classical electrostatic interaction, and the second one is the exchange-correlation energy. Therefore, the e-e interaction term can be formally rewritten as a sum of single-particle terms, where the electrons interact only with the effective potential field, $V^{eff} = V_H + V_{XC}$. This is an exact expression. Note that only V_{XC} is non-local and depends on the electron correlation, now given by the unknown function $g(\mathbf{x},\mathbf{x}')$, which itself implicitly depends on the electron density. The practical problem is then to find a reasonable approximation for the V_{XC}. DFT, which is discussed in the next section, tells us that this potential can, in principle, be chosen such that the exact many-body density, $n(\mathbf{x})$, is obtained as a solution. However, in practice, the potential is always approximated.

From equation (1.8) we can see that when there is no correlation, $g(\mathbf{x},\mathbf{x}') = 1$, the second term vanishes, and we are left only with the electrostatic interactions. It is also important to notice that the Hartree energy is divergent, due to the self-interaction. However, this divergence is exactly cancelled out by the contribution from the exchange-correlation term. In other words, the Pauli principle ensures that the e-e interaction remains finite. Any approximation to the exchange-correlation term will leave the divergences unbalanced and, consequently, give rise to a self-interaction error. The V_{XC} term lends itself to an interesting interpretation. Using the 2-electron density we can define the Coulomb

hole as

$$\Delta n(\mathbf{x}, \mathbf{x}') = n(\mathbf{x}, \mathbf{x}') - n(\mathbf{x})n(\mathbf{x}') \ , \tag{1.9}$$

which reduces to the exchange hole, $\Delta n_X(\mathbf{x}, \mathbf{x}')$, within the HF theory

$$\Delta n_X(\mathbf{x}, \mathbf{x}') = -\delta_{ss'} \left| \sum_i \phi_i^s(\mathbf{r}) \phi_i^s(\mathbf{r}') \right|^2 \ . \tag{1.10}$$

Where $\phi_i^s(\mathbf{r})$ are the single-particle spin-orbitals, which minimize the HF energy. It can be shown that the exchange hole integrates to one missing electron per electron at position $\mathbf{r}$. In general the Coulomb hole is a sum of the exchange and the correlation hole, where the latter only appears in approximations that go beyond the HF approach. The effect of the correlation hole is to redistribute the electron density (see Ch. 3 in Ref. [52]). Therefore, V_{XC} describes the interaction of an electron with the Coulomb hole that surrounds it. The "hole" corresponds to a decreased probability of finding a second electron at the same position where the first one is located. In other words, every electron is screened by the collective motion of all other electrons. If one electron moves, the whole electronic cloud around it has to move, in such a way that the Coulomb hole always tends to be centred at the position of the electron. Therefore, the electron and the Coulomb hole behave in many respects like a single object, i.e. like a quasi-particle. This behaviour is a direct consequence of the Pauli principle and makes the crux of the many-body problem. Namely, the motion of a single electron is fundamentally inseparable from the motion of all the other electrons. As a result the long range electrostatic interaction is screened and has a much shorter effective range. This is the reason why the independent-electron picture works so well for metals, where the screening is strong.

1.1.2 The Density Functional Theory

The density functional theory (DFT) is one of the most popular methods for the calculation of the electronic structure today. The origin of the method can be traced back to the seminal paper of Hohenberg and Kohn (1964) [53], which proved that the ground state of an interacting many-body system is uniquely determined by the electronic density. The theory was later extended for a number of cases, e.g. to finite temperatures [54], to time-dependent problems [55], to spin polarized and relativistic systems [56–58], etc. Quantum mechanics teaches us that the information about the state of a quantum system is contained entirely in the N-body wavefunction, which is formally a function of $3N$ spatial coordinates. For the time being we will omit discussing the spin degrees of freedom. The Hohenberg and Kohn (HK) theorem tells us that the same information is contained in the electron density, which is a function of only three coordinates. Even more so, it establishes a 1-to-1 correspondence between the potential of the system, $V(\mathbf{r})$,

which uniquely defines the Hamiltonian, the many-body wavefunction, $\Psi(\mathbf{r}^N)$, and the ground state electronic density, $n_0(\mathbf{r})$, namely

$$V(\mathbf{r}) \longleftrightarrow \Psi(\mathbf{r}^N) \longleftrightarrow n_0(\mathbf{r}) \ . \tag{1.11}$$

The Hamiltonian is assumed to be of the form

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i,j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \int d\mathbf{r}\ n(\mathbf{r})V_0(\mathbf{r}) \ , \tag{1.12}$$

which is completely general, aside from a particular form of the last term, involving the external potential. From a practical point of view, a Hamiltonian of this form covers most of the known many-body problems. In other words, the HK theorem establishes that the wavefunction is a functional of the electron density, $\Psi(\mathbf{r}^N) \equiv \Psi\left[n(\mathbf{r})\right]$, and the isomorphism between the potential and the wavefunction implies that the Hamiltonian is also a functional of the density, $\hat{H}(\mathbf{r}^N) \equiv \hat{H}_{v_0}\left[n(\mathbf{r})\right]$. In effect, the energy of the system is a functional of the ground-state density

$$E_{v_o}\left[n_0(\mathbf{r})\right] = \langle \Psi\left[n_0(\mathbf{r})\right] | \hat{H}_{v_o}\left[n_0(\mathbf{r})\right] | \Psi\left[n_0(\mathbf{r})\right] \rangle \ . \tag{1.13}$$

By virtue of the Rayleigh-Ritz variational principle it can then be shown that the energy functional (1.13) is variational, namely that an arbitrary, V-representable, electron density, $n(\boldsymbol{r}) \neq n_0(\boldsymbol{r})$, always satisfies the inequality

$$E_0 = E_{v_o}\left[n_0(\mathbf{r})\right] < E_{v_o}\left[n(\mathbf{r})\right] \ , \tag{1.14}$$

where n_0 is the true ground state density and E_{v_o} is the energy functional defined by it. Provided that the latter is known, the ground state can in principle be obtained by performing an energy minimization over all permissible densities, as implied by Eq. (1.14). Unfortunately the HK theorem does not provide any indication on how to construct such a functional. Even more so, despite 50 years of theoretical efforts, the form of the exact functional is still unknown today. In order to implement the energy minimization of Eq. (1.14) one needs to ensure that the trial densities are V-representable, namely that a 1-to-1 mapping with the potential, in the spirit of Eq. (1.11), can be established for all of them. Initially the hope was that all electron densities are V-representable, since they represent a solution of a physical problem. However, counter-examples disproving this assumption have been found [41]. In other words, not all electron densities are V-representable.

The V-representablity problem is circumvented by the two step minimization procedure, proposed by Levy [59] and Lieb [60]. Instead of performing the minimization over the set

of V-representable densities, the Levi-Lieb (LL) approach suggests that the minimization is first performed over a set of all N-electron wavefunctions, $|\Psi^N\rangle$, which have density, $n(\mathbf{r})$. In the second step the minimization is performed over the established set of N-representable densities.

$$\begin{aligned} E_0 &= \inf_n \{ \inf_{\Psi \to n} \langle \Psi^N | \hat{T} + \hat{H}_{ee} + \hat{V} | \Psi^N \rangle \} = \\ &= \inf_n \{ \left(\inf_{\Psi \to n} \langle \Psi^N | \hat{T} + \hat{H}_{ee} | \Psi^N \rangle \right) + \int d\mathbf{r}\ n(\mathbf{r}) V(\mathbf{r}) \} = \\ &= \inf_n \{ F_{LL}[n(\mathbf{r})] + \int d\mathbf{r}\ n(\mathbf{r}) V(\mathbf{r}) \}\ , \end{aligned} \tag{1.15}$$

where $F_{LL}[n(\mathbf{r})]$ is the Levi-Lieb functional, which is universal, i.e. it is independent of the external potential V. Here inf denotes the *infimum*, i.e. the greatest lower bound. The energy functional above becomes minimal for the same density that minimizes the HK functional, and only for that density, which makes it a valid extension of the true energy functional beyond the ground state. It turns out that any smooth non-negative function $n(\mathbf{r})$, which is normalized to the total number of electrons, N, is N-representable. However, despite such exact theoretical results a practical formulation of the energy functional has proved elusive.

1.1.2.1 The Kohn-Sham Approach

The success of DFT is mainly due to the ansatz made by Kohn and Sham (KS), which asserted that a true many-body electron density can be obtained as a ground state of an equivalent independent-particle system, where the effective potential has been chosen appropriately. This system is now known as the KS system. It was latter proved that the KS approach makes an exact scheme (see Ch. 4 in Ref. [41]), provided that the energy functional that generates the effective single-particle potential is known exactly. In fact, this problem is equivalent to the original difficulties with the DFT approach and hence, an exact functional was never obtained. Nevertheless, practically useful approximations based on the KS scheme have allowed calculating the electronic structure for large many-body systems, with sufficient accuracy needed to calculate, e.g. the magnetic moments, the electric current under bias in nano- and molecular-junctions and even to perform *ab initio* molecular dynamics. The latter is crucial to obtain relaxed structures, which are a prerequisite for studying the electronic structure of realistic systems, e.g. in the presence of defects, adsorbates on surfaces, etc. In addition, DFT often represents a starting point for the more sophisticated many-body methods, which were discussed in the preceding section. The method has revolutionized applied Physics

and Chemistry. For this reason the 1998 Nobel prize in Chemistry has been awarded to W. Kohn (together with J. A. Pople) for his contribution to developing DFT.

The KS approach asserts that the ground-state density of a fully interacting, many-body system, can be obtained as the ground-state density of an equivalent system of non-interacting electrons, the KS system. The N-electron density is defined as

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \ , \tag{1.16}$$

where $\phi_i(\boldsymbol{r})$ are the KS single-particle states.

The total energy functional

$$E[n(\mathbf{r})] = T[n(\mathbf{r})] + E_{ee}[n(\mathbf{r})] + \int d\mathbf{r}\ n(\mathbf{r}) V_{ext}(\mathbf{r}) \ , \tag{1.17}$$

where both the kinetic energy, $T[n(\mathbf{r})]$, and the e-e interaction energy, $E_{ee}[n(\mathbf{r})]$, functional are unknown, can be rewritten as

$$E[n(\mathbf{r})] = T_S[\{\phi_i(\mathbf{r})\}] + E_H[n(\mathbf{r})] + E_{XC}[n(\mathbf{r})] + \int d\mathbf{r}\ n(\mathbf{r}) V(\mathbf{r}) \ . \tag{1.18}$$

Here the KS kinetic energy functional is defined as

$$T_S[\{\phi_i(\mathbf{r})\}] = -\frac{1}{2} \sum_{i=1}^{N} \langle \phi_i | \nabla^2 | \phi_i \rangle \ , \tag{1.19}$$

and represents a sum of single-particle kinetic energies. The Hartree energy and the associated potential V_H, are defined in the same way as in equation (1.8). The exchange-correlation (XC) functional, E_{XC}, is formally given by

$$E_{XC}[n(\mathbf{r})] = T[n(\mathbf{r})] - T_S[\{\phi_i(\mathbf{r})\}] + E_{ee}[n(\mathbf{r})] - E_H[n(\mathbf{r})] \ . \tag{1.20}$$

We can see that it includes the many-body corrections to the kinetic and the e-e interaction energy. Therefore, all the many-body effects remain implicit in the XC functional. The advantage of the KS scheme is that the KS kinetic energy functional is given explicitly, while useful approximations can be obtained for the XC. The Rayleigh-Ritz principle then leads to a set of N single-particle equations of motion (EOM) for the KS orbitals

$$-\frac{1}{2}\nabla^2 \phi_i(\boldsymbol{r}) + V_S[n(\boldsymbol{r})]\phi_i(\boldsymbol{r}) = \epsilon_i \phi_i(\boldsymbol{r}) \ , \tag{1.21}$$

where $V_S[n(\boldsymbol{r})] = V_H[n(\boldsymbol{r})] + V_{XC}[n(\boldsymbol{r})] + V_{ext}(\boldsymbol{r})$.

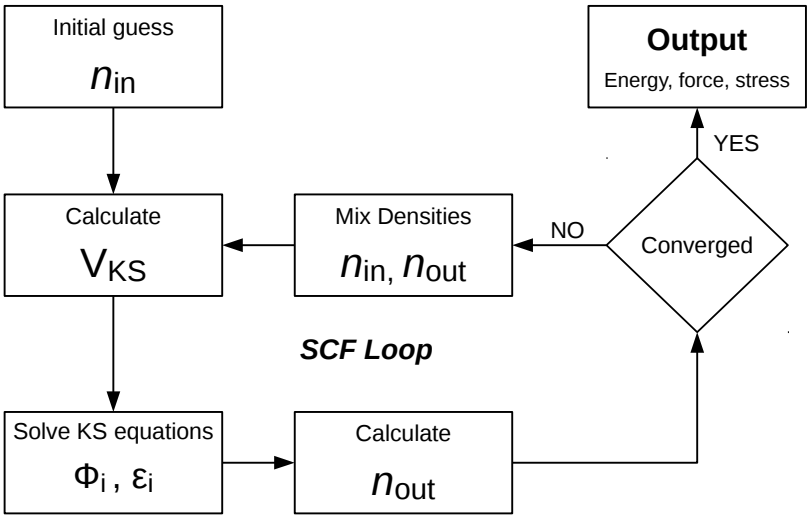

FIGURE 1.1: Flow chart of the self-consistent loop used to solve the system of Kohn-Sham (KS) equations (1.21).

The XC potential is given formally by the functional derivative of E_{XC},

$$V_{XC}[n(\mathbf{r})] = \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \ . \tag{1.22}$$

Since the KS potential, V_S, depends implicitly on the KS orbitals, the KS system of equations (1.21) needs to be solved using a self-consistent procedure. This is known as the self-consistent-field (SCF) loop. The scheme of the procedure is shown in figure 1.1. The loop is starts by making a guess for the density, n_{in}, usually from the superposition of atomic densities. Then the KS potential can be constructed and the system of equations (1.21) solved, to obtain a set of KS orbitals. The latter define the new electron density, n_{out}, via Eq. (1.16). The new and the old densities are then compared. The iteration stops when they differ by less then a specified tolerance. In this case the self-consistency is reached. If the difference is greater than that, the two densities are mixed using an appropriate scheme and a new trial density is constructed [61]. A new iteration of the SCF loop can then be started. As a corollary of the HK theorem, all observable ground-state quantities can, in principle, be calculated from the converged density. In practice, the KS eigenstates are used to calculate the quantities of interest.

1.1.2.2 The Kohn-Sham Approach for Magnetic Systems

The previously discussed DFT formulation is also valid for magnetic systems, since the magnetization is also a functional of the density. However, a more practical way to

formulate the DFT problem for the magnetic case is to explicitly use the magnetization as an additional variable. In general, this leaves us with 4 independent scalar fields that need to be varied in order to minimize the energy functional, $E[n, \boldsymbol{m}]$. The extension of DFT to magnetic systems, namely spin-DFT (SDFT), was first provided by von Barth and Hedin (1972) [56]. In this case the scalar potential is replaced by a 2×2 matrix, $w_{\alpha\beta}(\boldsymbol{r})$, acting on a 2-component spinor wavefunction, Ψ, where $\alpha, \beta \in \{\uparrow, \downarrow\}$, are the spin labels. The spinor components are the KS spin-orbitals, $\phi_{i\alpha}(\mathbf{r})$. The density matrix, $n_{\alpha\beta}(\mathbf{r})$, can be defined [56, 62] in terms of the KS spin-orbitals as

$$n_{\alpha\beta}(\mathbf{r}) = \sum_{i}^{\text{occ}} \phi^{*}_{i\beta}(\mathbf{r})\phi_{i\alpha}(\mathbf{r}) \ . \tag{1.23}$$

Similarly, the total energy functional is given as

$$E[n_{\alpha\beta}] = \langle \Psi[n_{\alpha\beta}]\hat{T} + \hat{H}_{ee}|\Psi[n_{\alpha\beta}]\rangle + \sum_{\alpha\beta}\int d\boldsymbol{r}\ n_{\alpha\beta}(\boldsymbol{r})w_{\alpha\beta}(\boldsymbol{r}) \ . \tag{1.24}$$

The electron density, $n(\mathbf{r})$, can be obtained as

$$n(\mathbf{r}) = Tr\left(n_{\alpha\beta}(\mathbf{r})\right) \ , \tag{1.25}$$

where the trace is taken over the spin indices and the magnetic moment density, $\boldsymbol{m}(\boldsymbol{r})$, as

$$\boldsymbol{m}(\boldsymbol{r}) = -\mu_B\ Tr\left(\sum_{\beta} n_{\alpha\beta}(\boldsymbol{r})\boldsymbol{\sigma}_{\beta\gamma}\right) \ . \tag{1.26}$$

Here $\boldsymbol{\sigma}_{\beta\gamma}$ represents a vector of Pauli matrices and $\mu_B = \frac{\hbar e}{2m_e} \equiv 0.5\,(a.u.)$.

More explicitly, the four components of the density are

$$\begin{aligned} m^x(\boldsymbol{r}) &= -\left(n_{\uparrow\downarrow}(\boldsymbol{r}) + n_{\downarrow\uparrow}(\boldsymbol{r})\right)\cdot\mu_B \\ m^y(\boldsymbol{r}) &= -i\left(n_{\uparrow\downarrow}(\boldsymbol{r}) - n_{\downarrow\uparrow}(\boldsymbol{r})\right)\cdot\mu_B \\ m^z(\boldsymbol{r}) &= -\left(n_{\uparrow\uparrow}(\boldsymbol{r}) - n_{\downarrow\downarrow}(\boldsymbol{r})\right)\cdot\mu_B \\ n(\boldsymbol{r}) &= n_{\uparrow\uparrow}(\boldsymbol{r}) + n_{\downarrow\downarrow}(\boldsymbol{r}) \ . \end{aligned} \tag{1.27}$$

Therefore, the density matrix of Eq. (1.23) can be rewritten as

$$n_{\alpha\beta}(\boldsymbol{r}) = \frac{1}{2}n(\boldsymbol{r})\delta_{\alpha\beta} - \frac{1}{2\mu_B}\boldsymbol{m}\cdot\boldsymbol{\sigma}_{\alpha\beta} \ . \tag{1.28}$$

The external potential, $w_{\alpha\beta}(\boldsymbol{r})$, can then be rewritten in a similar form as

$$w_{\alpha\beta}(\boldsymbol{r}) = v(\boldsymbol{r})\delta_{\alpha\beta} - \mu_B \ \boldsymbol{B}(\boldsymbol{r}) \cdot \boldsymbol{\sigma}_{\alpha\beta} \ , \tag{1.29}$$

where it is decomposed into a scalar, $v(\boldsymbol{r})$, and a vector component, $\boldsymbol{B}(\boldsymbol{r})$. The former couples to the electron density and the latter to the magnetization. This can be clearly seen if we recognize that the last term in Eq. (1.24), i.e. the interaction of the four-component density with the external potential, can be written as a trace of the matrix product. By using the definitions (1.26) and (1.29) we can then write

$$\begin{aligned}
\sum_{\alpha\beta} \int d\boldsymbol{r} \ n_{\alpha\beta}(\boldsymbol{r}) w_{\beta\alpha}(\boldsymbol{r}) = & \int d\boldsymbol{r} \ Tr\left(n_{\alpha\beta}(\boldsymbol{r}) w_{\beta\gamma}(\boldsymbol{r})\right) = \\
& = \int d\boldsymbol{r} \ \left[Tr\left(n_{\alpha\gamma}(\boldsymbol{r}) v(\boldsymbol{r})\right) - \mu_0 \ \boldsymbol{B}(\boldsymbol{r}) \cdot \ Tr\left(n_{\alpha\beta}\boldsymbol{\sigma}_{\beta\gamma}\right)\right] = \\
& = \int d\boldsymbol{r} \ \left[n(\boldsymbol{r}) v(\boldsymbol{r}) + \boldsymbol{B}(\boldsymbol{r}) \cdot \boldsymbol{m}(\boldsymbol{r})\right] \ .
\end{aligned} \tag{1.30}$$

Now it is obvious that Eq. (1.24) is a functional of the density and the magnetization.

Following the KS ansatz, now employing an orthonormal set of spin-orbitals, $\phi_{i\alpha}$, the Rayleigh-Ritz variation principle yields the following system of equations:

$$\sum_{\beta} \left(-\frac{1}{2}\nabla^2 \delta_{\alpha\beta} + \ \sum_{\gamma} \int d\boldsymbol{r}' \ n_{\gamma\gamma}(\boldsymbol{r}') v(\boldsymbol{r} - \boldsymbol{r}') \delta_{\alpha\beta} + \ w_{\alpha\beta}(\boldsymbol{r}) + V^{XC}_{\alpha\beta} \right) \phi_{i\beta}(\boldsymbol{r}) = \epsilon_{i\alpha} \phi_{i\alpha}(\boldsymbol{r}) \ , \tag{1.31}$$

which couples all the 4 components of the density. Naturally, the equations need to be solved self-consistently, following a procedure similar to the one shown in figure 1.1. The XC potential, $V^{XC}_{\alpha\beta}$, is formally defined as a functional derivative of the XC energy with respect to the spin density

$$V^{XC}_{\gamma\delta} = \frac{\delta E_{XC}[n_{\alpha\beta}]}{\delta n_{\gamma\delta}} \ . \tag{1.32}$$

We note that most of the approximations for V^{XC} utilize a parametrization, which is obtained from calculations where the spin configuration is restricted to the collinear situation. This is most commonly known as the spin-polarized case. Only recently a local XC functional, which explicitly considers a non-collinear magnetization was proposed [63]. In other words, the XC functional in practice depends only on two components, the electron density, $n(\boldsymbol{r})$, and the z-component of the magnetization, $m^z(\boldsymbol{r})$. Spin-polarized calculations are also the most commonly used DFT approach for studying magnetic systems. In total, only a small number of DFT calculations considers non-collinear spin configurations explicitly. Needless to say that solving the system (1.31) is at least 4 times more demanding than solving the non-magnetic case. In addition, the phase-space

available for realizing the ground state becomes much larger. The different non-collinear magnetic configurations are also very close in energy, typically $\leq 100\,\text{meV}$ per atom, which makes the convergence procedure much more demanding from a technical point of view. All these issues limit a more widespread use of the non-collinear approach. The most common way to get around the problem with the restricted XC functional is to assume that the magnetization is collinear within the atomic sphere. The ab initio calculations suggest that this is a reasonable approximation [63, 64]. A local coordinate frame, parallel to the local magnetic moment, where the XC functional depends only on $n(\boldsymbol{r})$ and $m^z(\boldsymbol{r})$, can then be considered. This approach was first proposed by Kübler et al. [62]. After the XC potential is calculated, an inverse transformation, from the local to the global reference frame, can be performed.

Now we will restrict ourself to the collinear case and look at the XC potential as a function of the density and the magnetization rather than the density matrix components. In this case the energy functional can be, quite generally, written as $E[n_\uparrow, n_\downarrow]$, where $n_\uparrow \equiv n_{\uparrow\uparrow}$ and $n_\downarrow \equiv n_{\downarrow\downarrow}$. The equation (1.28) for this special case reduces to

$$\begin{aligned} n_\uparrow &= \frac{1}{2}n - \frac{1}{2\mu_B}m^z(\boldsymbol{r}) \ , \\ n_\downarrow &= \frac{1}{2}n + \frac{1}{2\mu_B}m^z(\boldsymbol{r}) \ . \end{aligned} \tag{1.33}$$

We can then define the z-component of the XC magnetic field, B^z_{XC}, as

$$\begin{aligned} B^z_{XC} &= \frac{\delta E_{XC}[n_\uparrow, n_\downarrow]}{\delta m^z} = \\ &= \frac{\delta E_{XC}[n_\uparrow, n_\downarrow]}{\delta n_\uparrow}\frac{\delta n_\uparrow}{\delta m^z} + \frac{\delta E_{XC}[n_\uparrow, n_\downarrow]}{\delta n_\downarrow}\frac{\delta n_\downarrow}{\delta m^z} = \\ &= -\frac{1}{2\mu_B}\left(\frac{\delta E_{XC}[n_\uparrow, n_\downarrow]}{\delta n_\uparrow} - \frac{\delta E_{XC}[n_\uparrow, n_\downarrow]}{\delta n_\downarrow}\right) \ , \end{aligned} \tag{1.34}$$

where we have used Eq. (1.33) in the last step. This effective magnetic field, which arises from the XC energy, enters the KS EOM in complete analogy to the external field, $\boldsymbol{B}$ of Eq. (1.29). We note that the external field, $\boldsymbol{B}$, is usually taken to be zero. Therefore, the magnetic moment in DFT is entirely due to B_{XC}, which is the only component of the effective KS potential that couples to the magnetic moment. We note in passing that the B_{XC} of Eq. (1.34) is collinear with $\boldsymbol{m}(\boldsymbol{r})$, at every point in space, by construction. From (1.34) it is clear that B_{XC} depends on the occupation number of the spin channels. The total energy can thus be lowered by redistributing the electrons between the two spin channels. Since $m^z \propto n_\uparrow - n_\downarrow$, this implies adjusting the magnetic moment of the system. Magnetism thus makes a sensitive test for the XC functional. Luckily, most of the interesting magnetic atoms, like Mn, Fe, and Co, have a robust magnetic moment, which

reduces their sensitivity to the choice of XC functional. However, for weak itinerant magnets, where the moment is highly susceptible to external perturbations, or where the magnetic response properties are being calculated, the approximations commonly used for the XC may prove to be inadequate, since the magnetic susceptibility of the system will depend on higher derivatives of the XC energy.

1.1.2.3 Interpretation of the Kohn-Sham values

The fundamental objects of the KS approach are the KS spin-orbitals and the KS energies, which are obtained as the eigenvectors and the eigenvalues of the system of KS equations (1.21). We recall that the KS system was constructed to reproduce the correct ground state density of an interacting many-body system via Eq. (1.16). For now we will ignore the approximations that are made to construct the XC functional and treat the KS approach as exact. As such the KS eigenvalues are only mathematical tools that allow us to construct the density. In other words, we can not grant physical meaning to the KS spin-orbitals. In practice, this fact is regularly ignored and the KS spin-orbitals are treated as quasi-particle states, which in this case have an infinite lifetime, since they are obtained as the eigenvectors of the KS system. Results, which are obtained in this way, are usually in good agreement with experiment. For example, a practical application of the GW approach relies directly on the KS eigenstates to obtain a correction for the gap (see Ch. A4 in Ref. [51]). In other words, the KS eigenstates are often a good starting point to describe the system quasi-patricle spectrum. An explanation for this fact can be offered using the same line of arguments that are given to justify the independent-particle approach. The expected lifetime of the quasi-particles in, e.g. normal metals, is much larger than the relaxation time due to other sources of scattering (see Ch. 17 in Ref. [65]), especially at higher temperatures. Therefore, the decay of the quasi-particle state, on the relevant timescale, does not play an important role and the KS spin-orbitals can be safely used. Obviously, there will be systems where this approximation breaks down, but the problem is not directly related to DFT.

Similarly, the KS eigenvalues can not be given a direct physical significance, with the exception of the highest occupied energy level. In the latter case Janak's theorem [66] demonstrates that this eigenvalue corresponds to the first ionization energy of the system. By combining the expressions (1.18) and (1.21), the total energy of the KS system can be expressed as

$$E = \sum_i \varepsilon_i - \frac{1}{2}E_H + E_{XC} - \int d\boldsymbol{r}\ n(\boldsymbol{r})V_{xc}[n(\boldsymbol{r})]\ . \tag{1.35}$$

From here it is obvious that the sum of the KS eigenvalues does not yield the energy of the system. The additional terms are known as the double counting correction. From an empirical point of view, one can observe that the variation procedure employed to obtain the ground state usually yields structural parameters of molecules and solids which are within 1 % to 2 % of the experimental results. Similarly, the derivative of the energy, with respect to the position of the atoms in the Born-Oppenheimer approximation, i.e. the forces on the atoms, are also valid. This allows for structural relaxation to be performed, which is crucial for practical applications. For example, the bulk modulus, describing the elasticity of solids, is usually captured with an error smaller than 10 % [67]. Therefore, the energies obtained from DFT are valid, although the KS eigenvalues can not be directly interpreted.

There are however situations when the KS eigenvalues are directly used, e.g. to estimate the bandgap of a material. Most of the cases implicitly treat DTF as an effective "mean-field" theory, which gives a sufficiently good description of the quasi-particle states. Some of these cases are discussed in more detail in section 1.1.3. Here we will shortly discuss the bandgap problem to illustrate some of the limitations of DFT, which have to be kept in mind. We start from the definition of the conduction and the valence band energy, ϵ_c and ϵ_v, respectively. These can be expressed as the difference in the total energy of a true many-body system of N particles, E^N, as

$$\begin{aligned} \epsilon_c &= E^{N+1} - E^N \\ \epsilon_v &= E^N - E^{N-1} \ . \end{aligned} \tag{1.36}$$

The fundamental bandgap is defined as

$$\begin{aligned} \Delta = \epsilon_c - \epsilon_v = \\ = E^{N+1} + E^{N-1} - 2E^N \ , \end{aligned} \tag{1.37}$$

where (1.36) was used in the last step. Therefore, the fundamental bandgap can be calculated directly, via Eq. (1.37), by calculating the energies of the systems with N+1, N-1, and N electrons. The former two correspond to the negatively and positively ionized systems hence, the energy differences in Eq. (1.36) can also be expressed in terms of the ionization energy, I, and the chemical affinity, A (cf. figure 1.2). It was demonstrated [68] that within the KS approach the fundamental gap is related to the

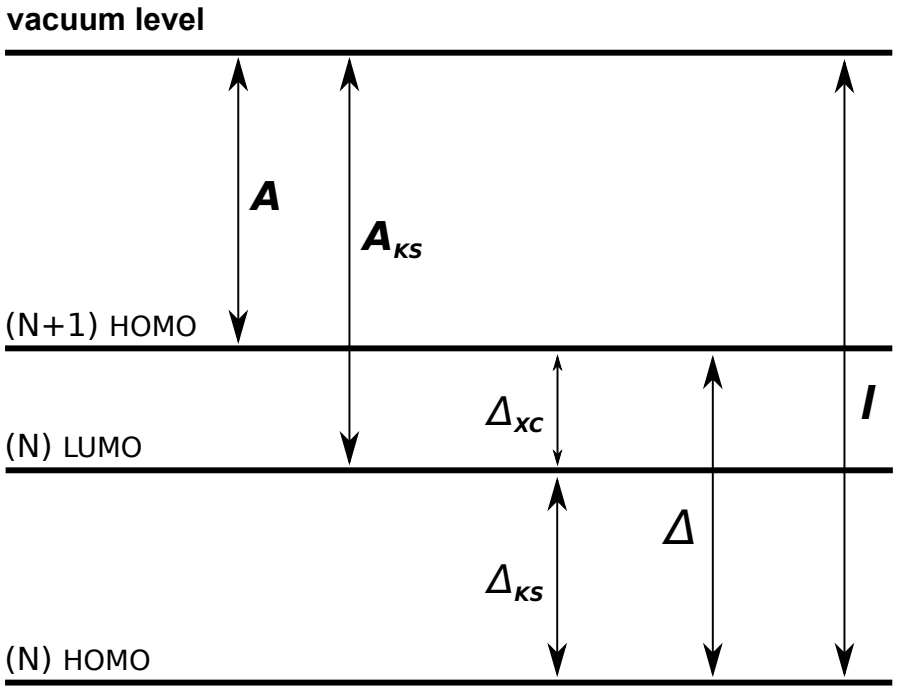

FIGURE 1.2: Molecular level diagram illustrating the Kohn-Sham (KS) bandgap problem. The HOMO and the LUMO denote the highest occupied and the lowest unoccupied molecular orbital, respectively. The number of electrons in the system, N, is indicated in the brackets. The true ionization energy, I, and the affinity, A, are shown. Within the KS approach, the affinity, A_{KS}, is determined by the KS bandgap, Δ_{KS}. The latter represents a lower bound for the fundamental bandgap, Δ.

derivative of the energy functional as

$$\begin{aligned}\Delta &= \left.\frac{\delta E[n]}{\delta n}\right|_{N=N^+} - \left.\frac{\delta E[n]}{\delta n}\right|_{N=N^-} = \\ &= \left(\left.\frac{\delta T_{KS}[n]}{\delta n}\right|_{N=N^+} - \left.\frac{\delta T_{KS}[n]}{\delta n}\right|_{N=N^-}\right) + \left(\left.\frac{\delta E_{XC}[n]}{\delta n}\right|_{N=N^+} - \left.\frac{\delta E_{XC}[n]}{\delta n}\right|_{N=N^+}\right) = \\ &= \Delta_{KS} + \Delta_{XC} \ ,\end{aligned} \tag{1.38}$$

where Δ_{KS} corresponds to the difference between the lowest unoccupied and the highest occupied energy level in the ground state calculation. Here we have used a shorthand notation: $N^{\pm} \equiv N \pm \delta$. Therefore, the KS bandgap will always underestimate Δ. The relation is illustrated in figure 1.2. The calculations based on the local density approximations (cf. 1.1.2.4) show that the Δ_{KS} systematically underestimates the bandgap by roughly 0.5 eV to 1.0 eV [69].

It is a well known property of the true XC functional that the derivative of the energy functional in Eq. (1.38) has to be discontinuous, since it is determined by the difference in the chemical potential for N and $N\pm1$ particle system. For a fractional occupancy the total energy is piecewise linear, since it can only be defined in terms of an ensemble of

systems, which have an integer number of electrons (see Ch. 2 in Ref. [41]). Therefore, the functional derivative has to change discontinuously at the points with integer occupancy. The local density approximations of the XC functional, which are most commonly used in practice, do not describe this feature. Furthermore, from Eq. (1.38) we can infer that any approximation to the XC functional will result in an error when estimating the bandgap, even in the case when the total energies for a system with N, $N \pm 1$ electrons are directly calculated. This is due to the underestimation of the correlation effects that should be contained in Δ_{xc}. A common way to practically obtain a correction for the bandgap is to employ the one-shot GW approach [51, Ch. A4], as a post-processing step.

1.1.2.4 Approximations of the Exchange Correlation Functional

In order to make DFT implementation practical an approximation to the XC functional needs to be made. The quality of the approximation will depend on the amount of many-body effects that are included in the functional. Unfortunately, the quality is paid in price of the complexity hence, the most sophisticated approximations are rarely used in practice. In fact, the success of the DFT is mostly due to the simplest possible approximation, namely the local density approximation (LDA) and its generalization, the generalized gradient approximation (GGA). Both approximations are based on the exact results obtained for the uniform electron gas (UEG). The central idea is to split the XC energy density into the exchange, ϵ_X, and the correlation contribution, ϵ_C. The former is obtained using the HF approximation for the UEG [52, Ch. 5], where an analytical solution is known, and the latter is fitted from the results of MC calculations, e.g. [70], which are used as a reference. For spin polarized systems the LDA XC functional is given by:

$$E_{XC}^{(LDA)}[n_\uparrow, n_\downarrow] = \int d\boldsymbol{r}\; n(\boldsymbol{r}) \left[\epsilon_X\left(n_\uparrow(\boldsymbol{r}), n_\downarrow(\boldsymbol{r})\right) + \epsilon_C\left(n_\uparrow(\boldsymbol{r}), n_\downarrow(\boldsymbol{r})\right)\right] \ . \tag{1.39}$$

The non-polarized case is obtained by setting $n_\uparrow = n_\downarrow$. The functional is a simple integral over all points in space, where the contribution at each point depends only on the density at that point. The LDA XC functional is thus said to be local. Similarly, the GGA is obtained by rescaling the ϵ_{XC} with a dimensionless function $F_{XC}\left(n_\uparrow(\boldsymbol{r}), n_\downarrow(\boldsymbol{r}), \nabla n_\uparrow(\boldsymbol{r}), \nabla n_\downarrow(\boldsymbol{r})\right)$, where the density gradient appears through a dimensionless parameter, $s = \frac{1}{2k_F}\frac{|\nabla n|}{n}$, known as the generalized gradient [71]. The Fermi wavevector, $k_F = (3\pi^2 n)^{1/3}$, represents the inverse of the characteristic screening length of the system. The main improvement of the GGA over the LDA is to reduce the binding energy and thus to improve the bond lengths and the crystal lattice parameters in

molecules and solids, respectively. Strictly speaking the GGA XC functional is semi-local, in the sense that it depends on the gradient of the density. However, this is not a severe limitation and an efficient numerical implementation of the method is possible.

Beyond the LDA and the GGA, which are by far the two most commonly used approximations, there exists an entire hierarchy of the XC functionals [72]. By continuing with the gradient expansion we arrive at the meta-GGA functionals [73], which include explicitly the dependence on $\nabla^2 n$ and on the KS kinetic energy density, $\tau_\sigma(\boldsymbol{r})$, where

$$\tau_\sigma(\boldsymbol{r}) = \sum_{i}^{occup} \frac{1}{2} |\nabla \phi_{i\sigma}(\boldsymbol{r})|^2 \ .$$

Next the orbital dependent functionals follow, where the XC energy is expressed as a functional of orbitals, instead of the density. Despite the increased computational requirements, orbital dependent functionals are of particular interest, since they naturally include the derivative discontinuity, with respect to the particle number, and thus yield better bandgaps. A different type of orbital dependent approaches are the self-interaction corrected (SIC) and the DFT+U. Both approaches are aimed at systems where the e-e interaction is strong, due to electron confinement. The former tries to improve the functional by explicitly removing the self-interaction error. The approach is straightforward to implement in finite systems, however, in solids the definition of the SIC becomes somewhat arbitrary, due to de-localized states (see Ch. 8 in Ref. [52]). The self-interaction error becomes smaller in the latter case. The LDA+U approach attempts to address the strong on-site correlations in localized orbitals, which are relevant in certain classes of materials, e.g. transition metal-oxides. In this case the on-site interaction, which is described at a "mean-field" level in DFT, is replaced by an effective Hubbard model for a given band, usually the transition-metal d band (see Ch. A8 in Ref. [51]). In order to make the approach truly *ab initio*, the parameters for the model need to be obtained from a constrained DFT calculation. Finally, we have the hybrid functionals [74, 75]. They are based on an exact expression for the XC energy, in terms of the adiabatic connection formula [76]. The adiabatic connection allows one to express the XC energy using an integral over the e-e coupling-strength constant, λ, as

$$E_{XC} = \frac{1}{2} \int d\boldsymbol{r}\ d\boldsymbol{r}'\ \frac{n(\boldsymbol{r}) n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \int_0^1 d\lambda \left[g_\lambda(\boldsymbol{r}, \boldsymbol{r}') - 1 \right] \ , \tag{1.40}$$

where $g_\lambda(\boldsymbol{r}, \boldsymbol{r'})$ is the λ-dependent pair distribution function. The functional dependence on the coupling constant, which is used to scale the e-e interaction, is then approximated in between the end points of the integral, for which the exact result is known. The most widely used functional of this family is the HSE06 [75, 77]. The hybrid functionals provide the most accurate energies and systematically improve the description of the

bandgap [78]. However, in metals they tend to overestimate the exchange contribution, which in turn results in too narrow bandwidth [79].

1.1.2.5 Pseudopotentials

The independent-particle approaches rely on single-particle orbitals to form a complete basis set in which the many-body wavefunction can be represented. The unfavorable scaling of the problem with the number of electrons then leads to numerical difficulties for large systems. Therefore, an important aspect of practical electronic structure calculations is to reduce the number of electrons to be explicitly described by the EOM. Conceptually this is not a problem. Since all material properties are determined by the valence electrons it would be enough to consider a subspace of the Hamiltonian involving only these states. The remaining electrons, occupying energy levels below the valence band, are called "core" electrons. The total electron density, n, can thus be split into the valence, n_{val}, and the core contribution, n_{core},

$$n(\boldsymbol{r}) = n_{val}(\boldsymbol{r}) + n_{core}(\boldsymbol{r}) \tag{1.41}$$

where the latter is assumed to remain fixed when the atom forms chemical bonds, and only n_{val} needs to be determined. In order to have a complete description of the system, an appropriate, reduced, Hamiltonian needs to be constructed. The latter also needs to include the effect of the core states on the valence band. For a single atom, stripped of the valence electrons, this results in an effective ionic potential, which is known as the pseudopotential (PP). To appreciate the simplification, one can consider the case of Pb, which has 82 electrons in total, of which only 3 are in the valence band.

The objective when constructing a good PP is the following: the number of valence electrons should be as small as possible and the resulting effective potential should be as simple as possible, while accurate and transferable. The latter condition guarantees that the PP will remain accurate when used in real systems, e.g. molecules and solids. However, there is no way to enforce the transferability, except by careful testing. With respect to the PP complexity, one can talk about hard and soft PPs, where the latter have a more compact representation in Fourier space. The "hardness" of a PP is measured by the number of Fourier components that are needed to represent the potential. The idea that a soft PP can be employed instead of the true hard potential comes from scattering theory. It was shown that, far from the nucleus, the true potential can be replaced by a simpler effective potential (see Ch. 11 in Ref. [52]). The phase of the scattered wave, however, remains the same. This means that the radial part of a valence orbital can be split into two parts, the core and the tail, where only the latter

is relevant for the formation of the chemical bond. In addition, the atomic potential is spherically symmetric and hence, it is sufficient to consider the PP for the radial part of the wavefunction for a given angular momentum, l. For a smooth pseudo-wavefunction (PWF), $\tilde{\psi}(r)$, the radial Schrödinger equation,

$$\left(-\frac{1}{2}\frac{d^2}{dr^2}+\frac{l(l+1)}{2r^2}+\tilde{V}_l(r)\right) r\tilde{\psi}_l(r)=\epsilon r\tilde{\psi}_l(r)\ , \tag{1.42}$$

can be inverted to obtain the l-dependent PP, $\tilde{V}_l(r)$, which satisfies Eq. (1.42) for a corresponding all-electron (AE) eigenvalue, ϵ. The tail of the PWF, $\tilde{\psi}$, is always chosen to match the AE result, while the core part is replaced by a smooth function, which matches the AE part at the core boundary, r_{core}. In general, the PP is constructed to reproduce the PWF for a given range of energies. However, if the chemical environment causes a large enough change in the true wavefunction, the PWF produced by the PP may become a bad approximation.

There are three different types of PPs, which are commonly used today. These are the norm-conserving (NCPP), the ultra-soft (USPP), and the projector augmented wave (PAW) PP [80, 81]. The difference between these PPs is in the way they achieve accuracy and transferability. The NCPP defines a set of PWF for which the appropriate potential is constructed so that a set of predefined rules is satisfied. One of these is that the norm of the PWF within the core, i.e. the core charge, Q_{core}, has to be conserved in the construction of the PP. Q_{core} is defined as

$$Q_{core}=\int_{|\boldsymbol{r}|\leq r_{core}} d\boldsymbol{r}\ |\tilde{\psi}(\boldsymbol{r})|^2\ . \tag{1.43}$$

The conservation implies that the change of the PWF at the core boundary will be the same as the change of the AE wavefunction, up to linear order in energy. The expansion is valid only close to the atomic eigenvalue corresponding to the wavefunction. In other words, the matching condition at the boundary is preserved for small perturbations. This ensures the transferability of the PP. The norm conservation imposes a limitation on the shape of the PWF in the core region, which is a strong constraint for the already nodeless wavefunctions, e.g. the $3d$ orbitals. The resulting shape of the PWF has to be similar to the AE wavefunction, which precludes the construction of a potential that is both smooth and accurate (see Ch. 11 in Ref. [52]). This constraint is loosened by USPPs, leading to a smoother, but more complex PP. The PWF are no longer orthogonal, which in turn leads to a generalized eigenvalue problem. The accuracy is simultaneously improved by allowing the PWF to reproduce the AE results for an arbitrary energy range. The PAW approach shares some similarities with the USPP, however, the starting point for the construction of the potential is different. In the PAW approach [81] the AE wavefunction,

$|\psi\rangle$, is replaced by a smooth wavefunction, $|\tilde{\psi}\rangle$, using a linear transformation, T, as

$$|\psi\rangle = T|\tilde{\psi}\rangle \ . \tag{1.44}$$

The transformation leaves the wavefunction identical to the AE solution outside the core region, while in the core a set of support, i.e. augmentation, wavefunctions is used to represent the fast oscillations. The advantage is that in this approach all electrons are kept in the PP and the AE Hamiltonian, H, is related to the pseudo-Hamiltonian, $\tilde{H}$, via the same linear transformation, T, that acts on the wavefunctions, namely

$$H = T\tilde{H}T^{\dagger} \ . \tag{1.45}$$

The transformation then leads to the generalized eigenvalue problem. The accuracy is, in principle, limited only by the completeness of the projectors, which are needed to represent the AE wavefunction on the augmentation basis set. In practice, the core electron levels are not allowed to change, which is known as the "frozen core" approximation. The PAW PP are nowadays the method of choice, being both soft and accurate. The results that can be obtained using these PPs are comparable to those of AE calculations [82]. However, the computational cost is significantly lower in the former case.

1.1.3 Electronic Structure of Solids

In the preceding two sections we have shown that a many-body problem can be recast, within a certain degree of approximation, into an equivalent problem for non-interacting particles. The approximation in reality deals with the motion of fermionic quasi-particles, nevertheless, in the discussion of the electronic structure it is more common to talk about independent electrons. We will also do this in the remainder of the text, keeping in mind, however, that the independent-particle picture is strictly valid only for quasi-particles. We note that the discussion presented here is far from exhaustive due to space constraints. Our aim is merely to introduce the basic concepts of the electronic structure theory and to provide connection with the commonly used methods later in the thesis. Unfortunately, this means that many relevant and interesting aspects of the solid state physics had to be omitted. Interested readers should consult some of the standard textbooks [46, 65] for a more complete overview.

The Hamiltonian that describes the motion of an independent-electron in a crystal can be written as

$$\hat{H} = -\frac{1}{2}\nabla^2 + \hat{V}_{crys}(\mathbf{r}) + \hat{V}^{\sigma}_{eff}(\mathbf{r}) \ , \tag{1.46}$$

where the second term is the spin independent crystal potential and the third term is the effective, spin-dependent, potential, which originates from the e-e interaction. Therefore, we can separate the spin up and spin down electrons, according to their spin projection, σ. The equation (1.46) then implies that electrons of opposite spin may exhibit different properties. We note that in the general case, the spin projection can be defined only locally, at every point $\mathbf{r}$. This is the case, for instance when the magnetic ground state is non-collinear. For now we will limit ourself to the collinear case and assume that a global spin-quantization axis can be defined. Since the periodic potential V_{crys} is the dominant interaction, it is usually assumed that the electron density also has the periodicity of the crystal. The crystal itself can be thought as the convolution of the basis, i.e. a set of vectors, $\{\boldsymbol{\tau}_i\}$ describing the position of the ions within the unit cell, and the lattice, which is defined by a set of vectors

$$\mathbf{R}(\{n_i\}) = \sum_{i=1}^{N_D} n_i \mathbf{a}_i \ , \tag{1.47}$$

where $n_i \in \mathbb{Z}$ and $\mathbf{a}_i$ are the basis vectors defining the unit cell. Here N_D denotes the dimensionality of the crystal, i.e. $N_D \in \{1, 2, 3\}$. Therefore, the combination of the basis and the lattice defines the crystal structure. The periodicity of the electron density implies that the electron wavefunction after a translation by an arbitrary vector $\mathbf{R}(\{n_i\})$ changes only up to a phase factor. This is formalized by the Bloch theorem (see Ch. 8 in Ref. [65]), which states that the wavefunction in a crystal has the following form

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) \ u(\mathbf{r}) \ , \tag{1.48}$$

where $u(\mathbf{r})$ is a function that has the same periodicity of the crystal and $\mathbf{k}$ is the quantum number associated to the commutation of the Hamiltonian and the translation operator. It can be shown that, under the Born-von Karman boundary conditions, the wave vector $\mathbf{k}$, takes a discrete set of values, which can be chosen from the first Brillouin zone [46]. The latter can be constructed for every crystal. When the system size is taken to infinity, the wave vector becomes a continuous variable and the eigenvalues become a smooth functions of $\mathbf{k}$. The Bloch theorem allows us to classify the eigenstates of the crystal Hamiltonian by a new quantum number. The states with different value of $\mathbf{k}$ are decoupled. This is of immense practical importance. The problem where we would need to solve a large, theoretically infinite, Hamiltonian describing to the whole crystal, can then be replaced by an easier problem, where we have to solve a reasonably sized Hamiltonian, for an infinite number of k-points. The size of the Hamiltonian, in the latter case, is determined by the number of electrons per unit cell. The Schrödinger

equation, which explicitly considers only the unit cell, can then be written as

$$\hat{H}_{\mathbf{k}}(\mathbf{r})u_{n,\mathbf{k}}(\mathbf{r}) = \epsilon_{n,\mathbf{k}}u_{n,\mathbf{k}}(\mathbf{r}) \ . \tag{1.49}$$

The solutions of (1.49) are labelled by the k point, $\mathbf{k}$, and the band index, n, which corresponds to the number of the electronic states, i.e. the number of single-particle orbitals, per unit cell, that are included in the calculation. In general, other labels associated with the operators that commute with the Hamiltonian may appear in addition to $\mathbf{k}$ and n, e.g. the spin label σ. Since the eigenvalues and the eigenfunctions are analytical with respect to $\mathbf{k}$, in practice it is sufficient to consider a finite set of kpoints in equation (1.49). The k-point dependence of the eigenvalues, $\epsilon_{n,\mathbf{k}} \xrightarrow{\Delta\mathbf{k} \to 0} \epsilon_n(\mathbf{k})$, is known as the band dispersion relation. The minimal and the maximal value define the edges of the single-particle band and hence the bandwidth. From a physical point of view, as the atoms of an infinite crystal are brought closer together, their atomic orbitals start to hybridize, and initially a discrete set of atomic energy levels starts to form a continuum of electronic states, which is then known as the energy band. We note that the expression "bandwidth" comes over-loaded here, since it may be mentioned in the context of a dispersion of a single-particle eigenvalue or, alternatively, it may denote the energy interval for a whole set of states, e.g. a d band. However, it should be clear from the context what is meant. We note that in practice we usually do not solve equations similar to (1.49), instead a DFT calculation provides us with single-particle KS eigenstates, $\Psi^{KS,\sigma}_{\mathbf{k}n}(\mathbf{r})$, and the corresponding KS eigenvalues, $\epsilon^{KS,\sigma}_{n,\mathbf{k}}$. These quantities already define the entire electronic structure from which all other representations and quantities can be derived. The main problem then is to understand the physics that leads to these solutions. In this respect, the KS system is interpreted as an effective, "mean field", description of the real many-body system.

With the knowledge of the eigenvalues we can define the spin resolved density of states (DOS), $g(E,\sigma)$, as

$$g(E,\sigma) = \sum_{n,\mathbf{k}} \delta(E - \epsilon^{\sigma}_{n,\mathbf{k}}) \ , \tag{1.50}$$

which represents the number of electronic states in the infinitesimal interval centered around E. In other words, it is the distribution of the electronic states as a function of energy. When studying real materials it is often convenient to understand how an individual atoms or an orbital contribute to the electronic structure. For this the projected density of states (PDOS) is commonly used. There is no unique way to do the projection, however, the following two approaches are commonly used. Namely, the atomic PDOS, $g_a(E,\mathbf{x})$, and the orbital resolved PDOS, $g_\Phi(E,\mathbf{x})$. The former is obtained by projecting the space- and energy-resolved DOS on the sphere S_a, centered on the atom a, while for the latter the projection is done on a particular orbital, e.g. on an orbital

having a particual angular symmetry. One can also choose to project the DOS of a molecule adsorbed on a substrate onto the molecular orbitals of an isolated molecule. Useful information about the chemical bonding can then be inferred by looking at how the charge redistributes after the adsorption. The starting point for both approaches is the local density of states (LDOS), which is defined as

$$g(E, \mathbf{x}) = \sum_{n,\mathbf{k}} \left| \langle \ \mathbf{r} \ | \ \Psi^{\sigma}_{n,\mathbf{k}} \ \rangle \right|^2 \delta(E - \epsilon^{\sigma}_{n,\mathbf{k}}) \ , \tag{1.51}$$

where $| \ \Psi^{\sigma}_{n,\mathbf{k}} \ \rangle$ are the KS spin-orbitals and $\mathbf{x}$ denotes a collective coordinate of the position and the spin $(\mathbf{r}, \sigma)$.

The atomic PDOS, $g_a(E, \mathbf{x})$, can then be obtained directly by integrating the LDOS

$$g_a(E, \sigma) = \int_{S_a} d\mathbf{r} \ g(E, \mathbf{x}) \ . \tag{1.52}$$

Similarly, the orbital resolved PDOS, $g_{\Phi}(E, \mathbf{x})$, can be obtained by formally replacing the position ket $|\mathbf{r}\rangle$ in Eq. (1.51) with a molecular orbital ket, $|\Phi\rangle$, to obtain

$$g_{\Phi}(E, \sigma) = \sum_{n,\mathbf{k}} \left| \langle \ \Phi \ | \ \Psi^{\sigma}_{n,\mathbf{k}} \ \rangle \right|^2 \delta(E - \epsilon^{\sigma}_{n,\mathbf{k}}) \ . \tag{1.53}$$

The density of states has a great importance in condensed matter physics, since it appears in almost all integrals that are used to estimate the physical observables of a system. For example, to calculate the total energy we need to evaluate

$$\sum_{\sigma} \int dE \ f(E; \mu, T) \ g(E, \sigma) \ E \ .$$

Similarly, the Fermi level DOS plays a decisive role for the magnetism and the transport properties. These will be discussed later on. Here $f(E; \mu, T)$ is the occupation of the energy level at the temperature T and μ is the chemical potential. Since the electrons are fermions this is just the Fermi-Dirac distribution

$$f(E; \mu, T) = [1 + exp(\beta \, (E - \mu))]^{-1} \ , \tag{1.54}$$

where $\beta^{-1} = k_B T$ and k_B is the Boltzmann constant. At $T = 0$ the chemical potential corresponds to the highest occupied energy level, which is the Fermi energy, E_F. However, the latter can only be defined in a strict sense for metallic systems, where the Fermi level falls within an energy band. The set of states at the Fermi energy forms the Fermi surface, which plays a central role in the case of transport properties.

The electronic states in a solid do not form an uninterrupted continuum of states, instead there are energy ranges where their existence is forbidden. These energy intervals are known as gaps. In practice the only energy gap that matters is the one that (potentialy) exists just above the highest occupied state. Similarly, for most of the electronic properties the energy bands that matter the most are the highest occupied and the lowest unoccupied energy band, which are more commonly known as the valence and the conduction band, respectively. For metals the Fermi level falls within the valence band, which is partially occupied. This is the reason why metals conduct electricity. Since the closest excited states, in which the electrons are able to move, are infinitesimally close to the Fermi surface, a small perturbation, in this case the electric field, can set the electrons into motion. For semiconductors and insulators the valence band is fully occupied and the closest excited state, i.e. the bottom of the conduction band, is separated by an energy gap from the highest occupied state. Consequently, a large electric field is required to excite the electrons into the conduction band and thus drive the electric current. Insulators have a large energy gap, of the order of 10 eV, which precludes the flow of current under moderate electric fields. In comparison, semiconductors have a gap of the order of 1 eV. At zero temperature they behave like insulators, however, at higher temperatures, e.g. at the room temperature, a small fraction of the conduction states is occupied and the semiconductors become poor conductors (usually). However, this is exactly what makes them interesting for technological applications. The occupation of the conduction band can be modified by the electric field, i.e. via a shift of the chemical potential, which allows for an electric control of the conductance. In other words, the conduction in the semiconductors can be switched on and off, at will, making them behave like an electric switch. This simple property has made possible the development of modern electronics, namely the computer and information technology. The gap thus plays an important role in the electronic structure. In solid state physics we define the fundamental gap, Δ, as the difference between the bottom of the conduction band, ϵ_c, and the top of the valence band, ϵ_v

$$\Delta = \epsilon_c - \epsilon_v \ . \tag{1.55}$$

As shown in Section 1.1.2.3 this quantity can be estimated from DFT calculations. One must keep in mind though that the KS gap, which is directly accessible from the eigenvalues, is always smaller, typically by 0.5 eV to 1.0 eV, than the fundamental gap [69]. We note that the gap can be directly probed in optical experiments. However, the measured value is called the "optical" gap, to distinguish it from the fundamental gap. The optical gap is smaller than the latter, due to the attraction between the excited electron, now occupying the conduction band, and the hole that is left in the valence band. In spin polarized systems the gap may occur for only one spin channel. This is

a direct consequence of the spin dependence of the effective single-particle potential, as indicated in equation (1.46). This gives rise to an interesting new class of materials, namely the half-metals [83], which are discussed in more detail in Chapter 3.

As mentioned previously, the electronic structure determines the transport properties of the system. In reality, the electron transport is a non-equilibrium statistical problem. This is best illustrated by the Kubo's linear response approach [51, Ch. 9]. Nevertheless, independent-particle approaches, usually based on the DFT, are commonly used to describe the motion of electrons under the external perturbation. In most cases the driving force is the electric field. However, a gradient of the temperature or the chemical potential, μ, are also important. The temperature gradient-driven currents and their coupling to the electrical currents have received a lot of attention lately, due to the issues with thermal transport in nano-devices and as an alternative for improving energy efficiency in devices and industry processes [84]. The chemical potential plays a central role in the conductivity in condensed matter. From a mesoscopic point-of-view it can be argued that the gradient of the electro-chemical potential (ECP), defined as $\mu_{ecm} = \mu + e\Phi$, where Φ is the electric potential, is more fundamental than the electric field alone. The electrons will only flow from the region that has a high ECP towards the region where it is lower, even if the electric field is pointing in the opposite direction. In this sense, applying an electric field between the two ends of a conduction channel, i.e. between the electrodes, corresponds to creating the difference in the ECP. This viewpoint is best illustrated by the Landauer's approach to the electron transport (see Ch. 9 in Ref. [51]).

1.1.3.1 Semi-Classical Theory of Electron Transport

The semi-classical theory of the electron transport, more commonly known as the Boltzmann transport (see Ch. 12 in Ref. [65]), describes the motion of the electrons in a crystal starting from the phase-space distribution function. The motion of electrons in this approach is intimately connected to the band dispersion relation. The scattering processes are included explicitly in the picture, where their main role is to maintain the local thermodynamic equilibrium. The local equilibrium is specified, for each band, by the Fermi-Dirac-like distribution (cf. 1.54), $g_n^0(\mathbf{r}, \mathbf{k})$, where the chemical potential and the temperature may vary from point-to-point. The electrons are treated as independent particles, whose trajectories between the collisions are given by the following equations

of motion (EOM)

$$\dot{\boldsymbol{r}} = \frac{1}{\hbar}\frac{\partial \epsilon_{n\boldsymbol{k}}}{\partial k}$$
$$\hbar\dot{\boldsymbol{k}} = -e\left[\boldsymbol{E}(\boldsymbol{r},t) + \frac{e}{c}\boldsymbol{v}_{n\boldsymbol{k}} \times \boldsymbol{H}(\boldsymbol{r},t)\right] . \tag{1.56}$$

Here $\boldsymbol{E}$ and $\boldsymbol{H}$ are the electric and the magnetic field, respectively. The first equation tells us that the classical velocity of an electron in the crystal, $\dot{\boldsymbol{r}}$, is given by the energy dispersion of the band. As a direct consequence, narrow bands, like in $3d$ metals, are associated with "slow", mostly localized, electons. In contrast, the s and the p band, which have bandwidth larger than 10 eV, are associated with highly mobile electrons. We note that the Boltzmann approach neglects inter-band transitions, hence, the electron which is found in the band ϵ_n at the beginning of the time evolution will be found there at the end. The second equation is the EOM for the crystal momentum of the electron, $\hbar\boldsymbol{k}$. The latter should be distinguished from the actual momentum of the electron. The electrons move under the influence of the crystal potential, which is not explicitly included in these equations. Instead, it is implicit in the band dispersion relation, which determines the effective mass of the electron. The bands which have a small curvature, i.e. flat bands, result in a large effective mass of the electron. This is easily shown from the equation (1.56). Assuming $\boldsymbol{H} = 0$, we can express the second time derivative of the position, $\ddot{\boldsymbol{r}}$, as

$$\ddot{\boldsymbol{r}} = \frac{1}{\hbar}\frac{\partial^2 \epsilon_{n\boldsymbol{k}}}{\partial \boldsymbol{k}^2}\dot{\boldsymbol{k}} = = \left(\frac{1}{\hbar^2}\frac{\partial^2 \epsilon_{n\boldsymbol{k}}}{\partial \boldsymbol{k}^2}\right)\boldsymbol{F}_{el} , \tag{1.57}$$

where $\boldsymbol{F}_{el} = \hbar\dot{\boldsymbol{k}} = -e\boldsymbol{E}$ is just the electric force. Therefore, the term in the bracket, in analogy with the Newton EOM, is just the inverse mass. It is important to note that this object is a tensor which can take negative as well as positive values. This often happens at the Brillouin zone boundary, where the electron gets reflected by the crystal potential, although the electric field is trying to accelerate it in the opposite direction. Strictly speaking, the equations above do not apply directly to electrons, instead they should be interpreted as the EOM for a wave packet, which is sufficiently localized in the reciprocal space to be assigned a well defined momentum, $\boldsymbol{k}$. This implies that the corresponding wave packet is spread over many unit cells in real space. In light of this, the first equation can be interpreted as the group velocity of the packet, $\frac{d\omega}{dk}$. Therefore, this approach can not be directly applied at the atomic scale. Despite the simplifications, these equations are often in the back of the mind when the electronic structure is analysed, since they provide a clear picture about the mobility of the electrons in a particular energy band.

The scattering processes, which by assumption occur on a length scale larger than the

mean-free path of an electron, are usually included through the relaxation-time approximation [65, p. 245]

$$dg(\boldsymbol{r},\boldsymbol{k},t) = \frac{dt}{\tau_n(\boldsymbol{r},\boldsymbol{k})} g_n^0(\boldsymbol{r},\boldsymbol{k}) \ , \tag{1.58}$$

where dg denotes the change of the distribution function due to the scattering in the time interval dt. The probability of scattering is given by the ratio $dt/\tau_n(\boldsymbol{r},\boldsymbol{k})$, where τ_n is the characteristic relaxation time for the n-th band. The relaxation time approximation of Eq. (1.58) allows us to write a closed-form solution for the non-equilibrium distribution [65, p. 248]:

$$g(t) = g^0 + \int_{-\infty}^{t} dt' P(t,t') \left[\left(-\frac{\partial f}{\partial \epsilon} \right) \boldsymbol{v} \cdot \left(-\nabla \mu_{ecm} - \left(\frac{\epsilon - \mu}{T} \right) \nabla T \right) \right] \ , \tag{1.59}$$

where $P(t,t')$ is the fraction of electrons, which are not scattered during the evolution from t to t'. The Fermi-Dirac distribution function, f, and the (electro-)chemical potential, μ (μ_{ecm}), were previously defined in the discussion of Eq. (1.54) and in the paragraph preceding section 1.1.3.1. We note that in equation (1.59) the temperature T and the electro-chemical potential μ_{ecm} are allowed to be position dependent, namely $T \equiv T(\boldsymbol{r})$ and $\mu_{\mathrm{ecm}} \equiv \mu_{\mathrm{ecm}}(\boldsymbol{r})$. Hence the function f of Eq. (1.59) needs to be evaluated using local values for the temperature and the electro-chemical potential. Here it is implicitly assumed that we are interested in the non-equilibrium distribution, $g(t)$, at a fixed point in the phase space, $(\boldsymbol{r},\boldsymbol{k})$, and a given band, n. The indices have been suppressed for clarity. With a few further simplifications the integral above can be evaluated [65, p. 250], to obtain:

$$g(t) = g^0 + \left(-\frac{\partial f}{\partial \epsilon} \right) \boldsymbol{v} \cdot (-e\boldsymbol{E}) \ . \tag{1.60}$$

Here the relaxation time has been approximated as having only an energy dependence, i.e. $\tau_n(\boldsymbol{r},\boldsymbol{k}) \approx \tau(\epsilon_{n\boldsymbol{k}})$. This approximation can not be justified for anisotropic systems, for example, where one would expect to have $\boldsymbol{k}$-dependent scattering. Similarly, if the defects in a material are relevant and have well defined positions the $\boldsymbol{r}$ dependence could not be ignored. By using equation (1.60) the current density, $\boldsymbol{j}_n$, associated with the band n, can be calculated as

$$\boldsymbol{j}_n = -e \int \frac{d\boldsymbol{k}}{4\pi^3} \boldsymbol{v}_{n\boldsymbol{k}} \ g_{n\boldsymbol{k}} \ , \tag{1.61}$$

from which the band-resolved conductivity tensor, $\sigma^{(n)}$, defined as $\boldsymbol{j}_n = \sigma^{(n)}\boldsymbol{E}$, can be expressed

$$\sigma^{(n)} = e^2 \int \frac{d\boldsymbol{k}}{4\pi^3} \tau_n \boldsymbol{v}_n(\boldsymbol{k}) \boldsymbol{v}_n(\boldsymbol{k}) \left(-\frac{\partial f}{\partial \epsilon} \right)_{\epsilon = \epsilon_n(\boldsymbol{k})} \ . \tag{1.62}$$

The total conductivity is then given by the sum over all bands. If the relaxation time

is further approximated to be energy independent, i.e. it is assumed to be a constant, we arrive at the constant relaxation time approximation, which is used in Chapter 3. We note that for spin polarized systems it is possible to calculate the conductivity independently for each spin channel. This allows one to obtain an estimate of the spin polarization in a material. However, all sources of spin scattering, e.g. the spin non-collinearity, are neglected.

1.1.4 The Origin of Magnetism

The origin of magnetism represented a long standing problem for classical physics. Electrodynamics showed that the origin of magnetism has to be due to electric currents in the materials, however, the required strength of the current was too large to be plausible, $\sim 1 \times 10^6$ A for Fe [27]. Even if the magnetic dipole of an atom is postulated in an *ad hoc* manner, the dipole-dipole interactions are simply not strong enough to explain the magnetic ordering at the room temperature. For example, the dipole interaction would be proportional to $\mu_0 \mu_B^2 / a_0^3 \approx 1 \times 10^{-4}$ eV, where μ_0 is the permeability of the vacuum, μ_B is the Bohr magneton and a_0 is the Bohr radius. This implies critical temperatures of the order of 1 K. Even more so, an exact theorem, due to Bohr and van Leuwen [85], demonstrates that within classical physics the average magnetic moment of a non-rotating body needs to be zero in an arbitrary external field. This forbids any notion of magnetic order within the framework of classical physics. A remarkable success in explaining the critical temperature in a ferromagnet was achieved later within the framework of the Weiss molecular theory [86][27, Ch. 5], which postulated the existence of a molecular field, proportional to the magnetization of the material, that is responsibe for the magnetic ordering. In addition, the concept of a magnetic domain was first introduced to explain the hysteresis of a magnet. Although this model was successful in explaining the temperature behaviour of the magnetic moment, it could not offer any justification for the molecular field, which had to be of the order of ~ 100 T. Such magnetic fields were never measured experimentally before at a macroscopic scale. With the advent of quantum mechanics it was soon realized that magnetism has a purely quantum origin, and the molecular field proposed by Weiss turned out to be of electrostatic nature, namely arising from the Pauli exclusion principle.

Today it is known that magnetism originates from the exchange interaction, which provides the necessary effective magnetic field needed by the Weiss theory. Since the wavefunction has to be antisymmetric with respect to the permutation of the electron coordinates, i.e. with respect to the electron exchange, an additional term, alongside the classical electrostatic interaction, emerges in the expression for the e-e interaction energy, E_{ee}. Using a Slater determinant for the 2 electron system, Ψ, it is easy to show

that E_{ee} becomes

$$\begin{aligned} E_{ee} &= \langle\Psi|\hat{V}|\Psi\rangle = \\ &= \langle\phi_1\phi_2|\hat{V}|\phi_1\phi_2\rangle - \langle\phi_2\phi_1|\hat{V}|\phi_1\phi_2\rangle\delta_{\sigma_1\sigma_2} \ , \end{aligned} \tag{1.63}$$

where the first term is just the classical Hartree energy (cf. 1.8) and the second term is called the exchange integral, J. A typical exchange energy is of the order of $\lesssim 1\,\mathrm{eV}$. It is important to note that this term acts only between the electrons having the same spin and always lowers the energy of the system. More explicitly, the exchange term can be written as

$$\begin{aligned} E_X &= -\delta_{\sigma_1\sigma_2}\langle\phi_2\phi_1|\hat{V}|\phi_1\phi_2\rangle = \\ &= -\delta_{\sigma_1\sigma_2}\ \frac{1}{2}\iint d\boldsymbol{r}_1 d\boldsymbol{r}_2\ \phi_2^*(\boldsymbol{r}_1)\phi_1^*(\boldsymbol{r}_2)\frac{1}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_1(\boldsymbol{r}_1)\phi_2(\boldsymbol{r}_2)\ . \end{aligned} \tag{1.64}$$

The exchange energy favours spins aligned parallel, which in turn leads to ferromagnetic order. In real systems, there are usually other competing mechanisms, which may lead to different magnetic order. But, bottom line, it can be said that the magnetic order emerges from the competition between the exchange and the kinetic energy. A qualitative argument can be provided based on the symmetry of the wavefunction. Parallel spins imply symmetric spin part of the wavefunction under exchange, which means that the real space wavefunction needs to be antisymmetric. This, in turn, implies a larger gradient, i.e. nodes in the wavefunction, and thus a higher kinetic energy. By using the same arguments it can be shown that for antiparallel spins a lower kinetic energy is expected.

Equation (1.63) lends itself to an interesting interpretation. If the orbitals ϕ_1 and ϕ_2 are occupied by electrons with antiparallel spin, the exchange term is identically zero, while the first term is independent of the relative spin oriantation. Therefore, if a spin flip occurs in one of the orbitals, the total energy of the system changes by J. This allows us to write an effective Hamiltonian for the system, involving only spin degrees of freedom as

$$H_{\text{eff}} = -2J\boldsymbol{S}_1\cdot\boldsymbol{S}_2\ , \tag{1.65}$$

where $\boldsymbol{S}_i$ is the spin of the electron occupying the orbital i. Here $\boldsymbol{S}_i$ can take only two values $\pm\frac{1}{2}$. Therefore, for $\uparrow\uparrow$ spin configuration the energy is $-J/2$ and for, the $\uparrow\downarrow$ (or $\downarrow\uparrow$) is $+J/2$. The energy corresponding to a spin flip is thus equal to J. This is known as the Heisenberg spin-$\frac{1}{2}$ Hamiltonian and J is known as the exchange coupling constant. A generalization of this model (cf. 1.73) is widely used for studying the thermodynamic properties of magnetic systems.

The microscopic origin of the exchange term in real systems can vary significantly. The exchange that originates from the direct overlap of orbitals, given by (1.64), is known as the *direct (or potential) exchange*. It is relevant when the overlap of orbitals is large, e.g. in atomic systems, where it drives the parallel alignment of electron spins in less-than-half filled shells. This is the rationale for the Hund's first rule, which explains the origin of the magnetic moment of the atom. In systems where the electrons can delocalise to a neighbouring site, the direct exchange competes with the kinetic energy term, which favours symmetric real-space wavefunction and antiparallel spins. The delocalization is driven by the reduction of the Hartree energy, which is usually an order of magnitude larger than J.

In insulators, where the electron hopping can be treated as a perturbation, another mechanism, called the *kinetic exchange*, is responsible for the magnetic order. It can be shown that the virtual electron hopping leads to an effective Heisenberg model, where the exchange depends on the hopping strength, t, and the on-site e-e interaction energy, U, as $J \propto t^2/U$ [87, Ch. 2]. A generalization of this exchange, where the hopping between the two magnetic atoms occurs through an intermediate non-magnetic atom is known as the *superexchange* mechanism. It was shown that this interaction in most of the cases leads to an antiparallel spin order [88] (antiferromagnetism). Another mechanism, known as the *s-d exchange*, which favours the ferromagnetic order, was proposed by Zener [89]. This is a particular interaction mechanism involving two species of electrons, where the mobile s-electrons serve as a mediator of the coupling between the on-site magnetic moments due to the localized d-electrons.

In metals, the exchange interaction is mediated by itinerant electrons, occupying states close to the Fermi level. Away from the atomic centre, the electrons in a metal can be well approximated by freely propagating plane waves, which scatter on the atomic centre and exhibit interference effects. The exchange in metals is thus long ranged and oscillatory in character. This is best exemplified by the Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction, which was derived to describe the exchange in $4f$ transition metals (see Ch. 2 in Ref. [87]). Unlike insulators, where the interaction with the few nearest neighbours completely determines the magnetic order, in metals even the sites which are several lattice constant away from a particular site can be relevant. It is important to note here that although the strength of the interaction with distant neighbours decreases quickly ($\propto 1/R^3$ for the RKKY interaction), their number also grows progressively, which makes an effective contribution from such a distant shell non-negligible. This makes it hard to predict the magnetic order without the information obtained from *ab initio* calculations.

The diversity of exchange mechanisms and possible geometric ordering of magnetic atoms in various crystal structures results in a number of magnetic orders. Due to small energy

differences these often compete, making the collection of possible magnetic textures extremely rich. Here we will briefly mention only a few common cases. Paramagnetism is the most simple type of magnetic order, where the atomic moments are randomly oriented and the magnetic moment averages to zero. However, under the influence of an external magnetic field the moments will get oriented, yielding a small net magnetic moment. A special case is the Pauli paramagnetism, which is due to the polarizability of the electron spins (cf. 1.70). Some atoms are very close to becoming magnetic, e.g. Pt, and hence they exhibit a particularly strong paramagnetic response. For this reason they may even become magnetic in a suitable chemical environment. Diamagnetism is another type of magnetism, which is exhibited by all materials. The effect arises due to the interaction of the external magnetic field with the orbital currents in a material. This leads to the quantization of the energy levels and consequently, induces a magnetic moment. The defining feature of the diamagnetic response is that the induced magnetic moment is antiparallel to the external field. In comparison, the moment induced in paramagnetic materials is aligned parallel to the magnetic field. Ferromagnetism (FM) is another characteristic type of magnetism, where the magnetic moments are spontaneously aligned, exhibiting a net moment below a certain critical temperature. The latter is known as the Curie temperature. Above this temperature the material becomes paramagnetic. This is an example of a second order phase transition, where the magnetization is the order parameter. A more complex magnetic pattern is displayed in antiferromagnets (AFM). In this case the crystal lattice has two symmetry equivalent magnetic sites, for which the moments are identical but oppositely oriented. From a macroscopic point of view the AFM appears non-magnetic. However, by applying a magnetic field, transverse to the magnetization axis, the AFM will exhibit a magnetic response. Since there is more than one way to arrange the spins in a staggered pattern, there are different types of AFM order depending on the crystal structure. The critical temperature above which an AFM becomes a paramagnet is called the Néel temperature. In case a material has two, or more, symmetry inequivalent magnetic sites, which are spontaneously ordered below a certain temperature, we have a ferrimagnet. Furthermore, there can exist a number of complex non-collinear magnetic textures, the simplest of which is the single spin-spiral wave. These are discussed in more detail in the next section. A more exhaustive discussion about the complex magnetism can be found in [27] and Ch. C4 of Ref. [51].

For the properties of metals of particular interest is the magnetism, which arises due to itinerant electrons. Here a picture of interacting localized moments may be completely inappropriate. However, even in such a case the magnetism arises due to the competition between the electrostatic interaction and the kinetic energy. In order to describe this, it is useful to consider the Stoner model, which gives a strict condition for the existence

of magnetism in a system of interacting electrons. Here we follow Fazekas [90, p. 352] and consider an onset of ferromagnetism in the Hubbard model

$$\hat{H} = \sum_{\boldsymbol{k},\sigma} \epsilon_{\boldsymbol{k}} \hat{n}_{\boldsymbol{k},\sigma} + U \sum_{j} \hat{n}_{j\uparrow}\hat{n}_{j\downarrow} - \frac{g\mu_B H}{2} \sum_{j} (\hat{n}_{j\uparrow} - \hat{n}_{j\downarrow}) \ , \tag{1.66}$$

where the first term is the single-particle band energy and the second term is the Hubbard interaction. The latter represents a pure electrostatic repulsion and acts by penalizing the double occupation of an energy level on a given atomic site j. Therefore, its role is to promote the on-site magnetism. The last term is the coupling of the magnetic moment to the external magnetic field H. Within the mean-field approximation the change in energy density due to an uneven occupation of the two spin channels, m, is given by

$$\varepsilon(m) - \varepsilon(0) = \int_{\mu_0}^{\mu_\uparrow} d\epsilon \ D(\epsilon)\epsilon + \int_{\mu_0}^{\mu_\downarrow} d\epsilon \ D(\epsilon)\epsilon - Um^2 - g\mu_B H m \ , \tag{1.67}$$

where m is defined by

$$\langle \hat{n}_{j\uparrow} \rangle = \frac{n}{2} + m \qquad\qquad \langle \hat{n}_{j\downarrow} \rangle = \frac{n}{2} - m \ . \tag{1.68}$$

For small m we can approximate the integrals with the value of the DOS at the Fermi level, which gives us the change in band energy, $\Delta E_{\text{band}} = \frac{m^2}{D(\mu_0)}$. The energy $\varepsilon(m)$ can then be minimized with respect to m to obtain

$$m = \frac{g\mu_B D(\mu_0)}{1 - UD(\mu_0)} H \ . \tag{1.69}$$

It follows that the susceptibility is given by

$$\begin{aligned} \chi &= \frac{(g\mu_B)^2 D(\mu_0)}{1 - UD(\mu_0)} = \\ &= \frac{\chi_{Pauli}}{1 - UD(\mu_0)} \ , \end{aligned} \tag{1.70}$$

where χ_{Pauli} is the susceptibility of the non-interacting electron gas. The denominator corresponds to an enhancement factor, which is entirely due to the e-e interactions. The Stoner criteria for the onset of ferromagnetism in itinerant magnets is given by

$$UD(\mu_0) > 1 \ , \tag{1.71}$$

which implies that the susceptibility becomes divergent if the density of states at the Fermi level is large enough. If this criterion is satisfied the magnetic ground state becomes energetically more favourable. This mechanism gives a natural explanation why the magnetism appears in $3d$ and $4f$ atoms. Therefore, a material can reduce its energy

by undergoing a spin split, which is often referred to as the Stoner split. Since the derivation does not depend on the explicit details of the bandstructure, the Stoner criterion is quite general and describes the onset of magnetism in real materials surprisingly well. However, the model seriously fails to predict the critical temperature of ferromagnets, overestimating the value typically by a few thousand K. The reason for this is that the only way to reduce the magnetic moment is by means of a spin flip, which entails a high energy price, of the order of eV. This gives a fundamentally wrong energy scale for the critical temperature. The failure arises because the transversal excitations, related to the formation of spin-waves, are completely neglected by the model. The transversal susceptibility is typically much larger than the longitudinal, which is associated with the Stoner spin-flip demagnetization mechanism. In fact, in ferromagnets the excitation energy corresponding to a long wavelength spin-waves vanishes $\propto 1/\lambda^2$, where λ is the wavelength. The spin-waves and the critical temperatures are discussed further in the next section.

In conclusion, the electronic structure plays an important role in the formation of magnetism in real materials. For example, in order to evaluate the Stoner criteria the Fermi level DOS needs to be known. In addition, the magnetic phenomena and the competition between different magnetic textures is associated with a meV energy scale. Therefore, *ab initio* methods are needed to provide sufficiently accurate information. This makes the SDFT a crucial tool for studying magnetism.

1.1.4.1 Application of SDFT to Magnetic Systems

From a general perspective, SDFT calculations provide access to the spin-resolved electronic structure. This makes it the starting point for understanding all the material properties. In practice, the validity of DFT is limited to zero Kelvin temperature. Nevertheless, the properties of the materials at room temperature are usually well described, providing that the electronic excitations do not cause a strong departure from the 0 K state. At elevated temperatures different transitions can occur, both structural and magnetic. The magnetic ordering and the related critical temperature, at which the magnetic order disappears, are in most cases studied theoretically for a fixed crystal structure. Otherwise the problem becomes very difficult due to magneto-elastic coupling [91]. Here we will focus only on the spin degrees of freedom.

Critical temperature calculations: At finite temperature the most common magnetic excitations are magnons. These are quasi-particles corresponding to a collective spin excitation that can be viewed as a spin flip "distributed" over the entire lattice. To a good approximation, magnons are non-interacting and behave like bosons [92]. A

classical analogue of this state is the spin-spiral, defined by Eq. (1.72), where the magnetic moment, $\boldsymbol{M}$, at a given instance of time, smoothly rotates around the equilibrium magnetization axis (here z-axis) as we move along the spin spiral direction,

$$\begin{aligned} M_x &= |\boldsymbol{M}| \sin(\theta) \cos\left(\boldsymbol{q} \cdot \boldsymbol{r} - \omega t\right) \ , \\ M_y &= |\boldsymbol{M}| \sin(\theta) \sin\left(\boldsymbol{q} \cdot \boldsymbol{r} - \omega t\right) \ , \\ M_z &= |\boldsymbol{M}| \cos(\theta) \ . \end{aligned} \tag{1.72}$$

The spiral is periodic in space, with the corresponding wavelength $\lambda = 2\pi/|\boldsymbol{q}|$. Likewise, at a given point in space, the magnetic moment precesses around the same axis with frequency ω. Therefore, every spin-spiral is characterized by three parameters: the canting angle, θ, the precession frequency, ω, and the wavevector, $\boldsymbol{q}$, which determines both the rate of magnetization precession with respect to the translation and the spin spiral direction. Since the magnitude of the local magnetic moment, $|\boldsymbol{M}|$, remains constant, the spin-wave is related to the transverse susceptibility of the magnetization. Within the adiabatic approximation, where the precession of the magnetic moment can be considered slow compared to the electron hopping times, the electrons are considered to be in equilibrium with respect to the spin-spiral configuration. In this sense, the spin-spiral represents an instantaneous snapshot of a magnon propagating through the system, i.e. represents a "frozen-magnon". The energy of a magnon depends on the wavevector $\boldsymbol{q}$, which represents its momentum. It can be shown that at low energies the spin waves have a quadratic dispersion relation for ferromagnets (see Ch. 5 in Ref. [27]). The complexity involved in systems with many fermions makes the study of magnons, especially their dynamics, very demanding using an *ab initio* approach. It is therefore more common to employ effective spin Hamiltonians, like the Heiesnberg model and its extensions (discussed below), to capture the main features of the excited magnetic states in realistic materials. The electronic structrure calculations performed using SDFT make such a mapping, which reflects the properties of a real material, possible. At higher energies the longitudinal fluctuations, i.e. the Stoner excitations, become relevant and the spectrum of the excited states becomes more complex. In this case the spin-spiral picture is not sufficient any more to describe the material properties [93].

In order to study finite temperature effects on the spin configuration it is practically inevitable to map the many-body problem onto a model spin Hamiltonian. The most common choice is the the classical Heisenberg model:

$$H\left(\{\boldsymbol{M_i}\}\right) = -\frac{1}{2} \sum_{i,j} J_{ij} \boldsymbol{M}_i \cdot \boldsymbol{M}_j \ , \tag{1.73}$$

where the interaction between the magnetic moment, $\boldsymbol{M}_i$, is given by a set of exchange parameters, J_{ij}. In a periodic solid, the J_{ij} are invariant under translations, hence the Heisenberg Hamiltonian of Eq. (1.73) can be Fourier transformed and the energy becomes a function of the wavevector $\boldsymbol{q}$. It is then easy to show that a spin-spiral represents an exact solution in the absence of spin-orbit coupling. Variation over the wavevector leads to an energy minimum, which represents the ground state. In addition, an energy value can be associated to each spin configuration in Eq. (1.73), which lends the spin-spiral states, given by equation (1.72), to be used as a practical mean to explore the excited states of realistic materials within the SDFT formalism. The energy corresponding to a given spin-configuration can therefore be obtained using SDFT by constraining the magnetization to a corresponding spin-spiral state. By calculating a number of spin-spin spiral energies, a system of linear equations, which allows one to extract the effective J_{ij} matrix, can be written down. This approach has been followed in Refs. [94, 95]. An alternative is to use the multiple-scattering Korringa-Kohn-Rostoker (KKR) method, based on Green's functions, to extract the spin-spin interaction energies [96, 97]. In this approach, the J_{ij}'s are extracted from the real space representation (cf. Eq. (1.73)) by using the method of infinitesimal rotations [98]. This represents a perturbation of the system for which the corresponding energy can be easily obtained by solving the Dyson equation. Therefore, both approaches utilize energy corresponding to an excited state calcuated within SDFT to obtain J_{ij}'s. We note that in both approaches it is crucial that the magnitude of the local magnetic moment is not changed by the perturbation, $|\boldsymbol{M}_i| = const.$

Both methods have comparative advantages and can be considered as complementary. The spin-spiral method is numerically more demanding since it requires performing a large number of q-point calculations in order to obtain the J_{ij} matrix. For each q-point a constrained SDFT calculation needs to be performed. Even when all the crystal symmetries are exploited up to 100 to 1000 calculations needs to be performed. The computational cost increases further, at a combinatorial rate, with the number of magnetic atoms in the unit cell. This makes the approach limited to small unit cells. Nevertheless, since the calculations are performed in reciprocal space, magnon dispersion relations can be obtained straightforwardly. The computational cost is now significantly lower than before, since a much smaller number of q-point calculation needs to be performed. In contrast, the KKR method works in real space and the J_{ij} matrix can be obtained with little effort. The perturbation approach makes the calculation significantly faster than the spin-spiral method. This makes it the method of choice if only the J_{ij}'s are needed. However, the KKR scheme usually provides a short range J_{ij}'s, up to a few neighbouring shells. This also implies that the Hamiltonian of Eq. (1.73) is truncated, which leads to a renormalization of the J_{ij} constants, depending on the cutoff

radius. The short range of the J_{ij}'s makes it difficult to perform a Fourier transform and thus to obtain the magnon dispersion in this way. It is worth to note that KKR is more sensitive to the reference ground-state magnetic configuration, for which the perturbation is calculated. In comparison, the spin-spiral approach uses an exact energy for each spin-wave and the J_{ij}'s are, in a sense, averaged over a set of spin configurations. This is made more explicit in the approach followed by Jacobsson et al. [95]. The resulting J_{ij}'s are thus expected to be more robust, in particular when the resulting Heisenberg model is solved using a Monte Carlo approach.

The access to the J_{ij} constants allows one to estimate the critical temperature of a ferromagnet, i.e. the Curie temperature, using three different methods. These are the mean-field approximation (MFA), the random-phase approximation (RPA) and the Monte Carlo approach [94, 96]. The most simple method is the MFA. Here the temperature dependence of the magnetization is approximated by the behaviour of the magnetic moment of a single atom coupled to an average effective magnetic field that arises from the neighbours. The magnetic moment is given in this case by the Brillouin function, which can be linearized for high temepratures, leading to a closed set of equations for the magnetic moment as a function of the temperature. From there the susceptibility can be expressed and the Curie temperature extracted. For a single magnetic atom in the unit cell the expression is particularly simple [65, p. 715] and reads

$$T_C^{(MF)} = \frac{M^2}{3k_B} \sum_j J_{0j} \ , \tag{1.74}$$

where M is the magnitude of the magnetic moment. It is possible to extend the equation (1.74) to unit cells containing multiple magnetic sites. A full derivation of the expression for the MFA Curie temperature in this case is given in Appendix A. We state only the final result here, namely that the Curie temperature is obtained as the largest eigenvalue of the Θ matrix [99], where

$$\boldsymbol{\Theta}_{\alpha\beta} = M_\alpha \left(\frac{\sum_{\boldsymbol{R}_j} J_{0\alpha,\boldsymbol{R}_j\beta}}{3k_B} \right) M_\beta \ . \tag{1.75}$$

Here the M_α is the magnitude of the magnetic moment at site α, and the $J_{0\alpha,\boldsymbol{R}_j\beta}$ are the magnetic exchange constants, i.e. the J_{ij}'s obtained from *ab initio* calculations. $\boldsymbol{R}_j$ denotes the unit cell position and k_B is the Boltzmann constant. We can see that in the case of a single magnetic atom in the unit cell ($\alpha = \beta = 1$) the expression (1.75) reduces to that of equation (1.74).

The equation (1.74) can be rewritten using the Fourier components $J(\boldsymbol{q})$, where the summation over the neighbours gets replaced by a summation over the wavevectors, i.e.

$T_C^{(MF)} \propto \sum_{\boldsymbol{q}} E(\boldsymbol{q})$, where

$$E(\boldsymbol{q}) = \frac{2\mu_B}{M} \left[J(0) - J(\boldsymbol{q}) \right] \ , \tag{1.76}$$

is the spin-wave dispersion relation [96]. From Eq. (1.76) it is obvious that the MFA assigns equal probability to all excitations $J(\boldsymbol{q})$. The RPA uses essentially the same information as the MF, however the magnonic excitations are weighted in a different way. Namely, within the RPA: $T_C^{(RPA)^{-1}} \propto \sum_{\boldsymbol{q}} E(\boldsymbol{q})^{-1}$. Therefore, the low energy excitations are given more weight, as one would expect from a thermodynamic average. The MF approximation is valid only for high temperatures, due to the linearisation of the Brillouin function. For this reason Eq. (1.74) is also known as the paramagnetic Curie temperature. In comparison, the RPA is valid for low and high temperatures. For the intermediate temperatures, $T \sim T_C$, it can be viewed as an interpolation between these two limits. It therefore tends to give more accurate estimate of the Curie temperature [96]. Beyond these methods lies the Monte Carlo approach, which provides the most accurate estimate of T_C [97]. Here, the Heisenberg model is simulated on a finite periodic grid of classical spins interacting via J_{ij}'s obtained from the *ab initio* results. The spin fluctuations and the clustering of spins, essentially describing the correlations, can be captured more systematically using this approach. It provides information which is independent of the approximations, beyond the assumption of a classical Hamiltonian. Unlike the previous two methods, the Monte Carlo method can accurately treat the region of temperatures close to the T_C, thus providing an accurate estimate for the latter. For this reason it is considered as a reference method for determining T_C.

Magnetic Anisotropy Calculations: For magnet applications three material parameters play a crucial role. These are the critical temperature, the saturation magnetization and the magnetic anisotropy. In order to preserve the ferromagnetic order the critical temperature needs to be much higher than the operating temperature, namely than room temperature. The saturation magnetization is relatively trivial, it can be obtained directly from a SDFT calculation. A full temperature profile of the magnetization, $M(T)$, can be obtained, e.g. using the Monte Carlo approach, as discussed earlier. However, in systems where the orbital currents may play an important role, for example in systems involving $3d$ and $4f$ elements, the SDFT moment may happen to deviate substantially from the experimental one. In such case a special treatment is needed [100, 101]. The final parameter is the anisotropy, which has two main contributions. One contribution comes from the dipolar interactions and has extrinsic character, namely it depends on the boundary conditions. It is therefore known as the shape anisotropy. This contribution is more relevant when dealing with magnetic samples whose geometry is known. The second, intrinsic, contribution is known as the magneto-crystalline anisotropy (MCA)

and it originates from the spin-orbit coupling (SOC). This was first recognized by van Vleck [102]. The energy scale of the SOC, $\sim 50\,\text{meV}$ in $3d$ atoms, is much smaller than the exchange interaction. This is additionally suppressed in solids, due to the "quenching" of the angular momentum due to the crystal field, so that the energy associated with the MCA is typically well below 1 meV. The exception are the rare-earth based alloys, like $Nd_2Fe_{14}B$, which exhibit a large anisotropy due to localized $4f$ electrons. Therefore, the MCA in $3d$ metal alloys will not strongly affect the critical temperature, however, it is vital for making a practically useful magnet. The role of the anisotropy is to pin the magnetization along a particular crystal axis direction or plane. In this way the magnetic moment is protected against demagnetizing fields and temperature fluctuations. In the absence of SOC the magnetic ground state is infinitely degenerate, as any global rotation in the spin space leaves the energy unchanged. Under such circumstances the magnetization would be free to precess and the information could not be stored and preserved using the spin degrees of freedom. Pinning of the magnetization provided by the anisotropy is what makes possible practical application of magnets.

Besides the MCA the SOC can give rise to another interesting interaction in systems with broken inversion symmetry, namely the Dzyaloshinskii-Moriya (DM) interaction [103–106]. At the level of the Heisenberg Hamiltonian, where two spins in general interact in a bi-linear manner, the DM interaction would correspond to a traceless antisymmetric part of the J_{ij} matrix, $J_{ij}^A = \frac{1}{2}\left(J_{ij} - J_{ij}^T\right)$. Hence, it is also known as the antisymmetric exchange. This interaction favours a non-collinear magnetism, i.e. spin canting and spin-spiral magnetism. However, it also lifts the degeneracy between the left, $-\boldsymbol{q}$, and the right, $\boldsymbol{q}$, propagating spin spirals, thus introducing the notion of the chirality in the magnetic texture (see Ch. C4 in Ref. [51]). This interaction becomes particularly important at interfaces, where the inversion symmetry is broken and the surface SOC is strong. We will not consider this interaction further in this work.

The microscopic origin of the MCA can be best understood by following the perturbation approach of Bruno [107]. The role of the spin-orbit is to unquench the angular momentum, which in turn gives a non-vanishing magneto-crystalline anisotropy energy (MAE) contribution. Within the 2nd order perturbation theory the energy change due to the SOC is given by

$$\delta E(\hat{\boldsymbol{n}}) = \sum_{\text{exc}} \frac{|\langle \text{gr}|H_{\text{SO}}(\hat{\boldsymbol{n}})|\text{exc}\rangle|^2}{\varepsilon_{\text{gr}} - \varepsilon_{\text{exc}}} \,, \tag{1.77}$$

where "gr" and "exc" denote the ground state and the excited state, respectively. The SOC Hamiltonian, $H_{\text{SO}}(\hat{\boldsymbol{n}})$, explicitly depends on the orientation of the spin quantization axis, $\hat{\boldsymbol{n}}$, with respect to the reference frame of the crystal. The expression (1.77) was later generalized to systems with more than one atom in the unit cell [108]. In the

same work, Bruno has also shown that for a strong magnet the MAE is proportional to the anisotropy of the angular momentum. Therefore, the SO scattering gives rise to orbital currents, which in turn couple to the spin. Since the orientation of the angular momentum will depend on the type of orbitals that contribute to the scattering, the crystal will pin the angular momentum along a particular axis in the ground state. The latter corresponds to a minimum of the total energy of the system. In other words, the SO energy is dependent on the relative orientation of the spin with respect to the crystal axes. Hence, the total energy will change as the global spin quantization axis is rotated.

An axis along which the total energy is minimal is called the easy axis. It may happen that the energy of the system takes minima along an entire plane, in which case we can talk about an easy plane. Likewise, there will also exist an axis such that the energy of the system becomes maximal, which is known as the hard axis. These two axes are experimentally distinguishable as it is much easier to magnetize the sample along an easy axis than along the hard one (see Ch. 5 in Ref. [27]). The MAE is defined as the difference $\delta E_{\text{hard}} - \delta E_{\text{easy}}$, which correspond to $\hat{\boldsymbol{n}}$ pointing along the hard and the easy axis, respectively. For systems with uniaxial anisotropy, for instance tetragonal crystals, the energy $\delta E(\hat{\boldsymbol{n}})$ becomes just a function of the angle, θ, between the easy axis and $\hat{\boldsymbol{n}}$. Therefore the MAE can be defined as

$$MAE = \delta E(\theta = 90^o) - \delta E(\theta = 0^o) \ , \tag{1.78}$$

where $\theta = 0^o$ corresponds to the easy axis. From a macroscopic point of view the total energy density, E, for a system with an uniaxial anisotropy can be expressed as

$$E = K_0 + K_1 \sin^2(\theta) + K_2 \sin^4(\theta) + ... \ . \tag{1.79}$$

The series is converging rapidly because, the nth order contribution is proportional to ξ^{2n}, where ξ is the SOC parameter. This follows directly from the perturbation expression (1.77). In practice, it is often assumed that the leading term is the only relevant contribution hence, the approximation $K_1 \equiv K_u$ is usually known as the uniaxial anisotropy constant. Therefore, the connection between the macroscopic and the microscopic anisotropy is given by

$$K_1 = \frac{MAE}{V} \ , \tag{1.80}$$

where the MAE is given by Eq. (1.78) and V is the volume of the unit cell for which the MAE was calculated.

As a consequence of the anisotropy one needs to perform work in order to reverse the magnetization direction. This allows the magnet to store energy as well as to preserve

information. In the latter case, the anisotropy prevents the magnet from randomly switching the magnetization direction under an external perturbation. This is the foundation of all magnetic recording devices. Therefore, an accurate calculation of the MCA become of fundamental interest in the field of applied magnetism.

In this thesis we will consider calculations of MAE for $3d$ transition metal alloys. Since the MCA depends in great detail on the electronic structure of the material, accurate *ab initio* methods need to be employed for this task. In practice this means performing SDFT calculations, which include the SO coupling in the KS equations [45, 109]. However, the calculation of the MAE represents a numerical challenge. For cubic crystals, where the symmetry requires that the MCA appears only in the 4th order perturbation expression, even the most accurate all-electron DFT method fails to give a trustworthy result. It is much easier to obtain the results in less stringent cases, like that of tetragonal crystals with uniaxial symmetry or thin film surfaces. In the latter case the MAE is of the order of 1 meV [110]. Typical MAE values, for instance in bulk Heusler alloys the MAE is usually around 100 µeV per atom, are at the limit of numerical accuracy. Unfortunately, this does not mean that a good agreement with the experiment can be achieved, mainly because the MCA is often very sensitive to the imperfections in real crystals. It is therefore difficult to perform a direct comparison between theory and experiment, unless the experiment was performed for a defect-free single crystal.

A practical approach to calculate the MCA is based on the "force theorem" [98, 111]. The theorem considers the change of the SDFT total energy under the influence of a small perturbation. The main result is that it is possible to substitute the total energy difference of an interacting many-body system with the difference of the corresponding band energies, provided that the perturbation is small. In other words, the quasi-particle interactions can be neglected within the linear response. Here we will assume that the perturbation is the SOC. Following Daalderop et al. [112], we can express the change in energy, $\Delta E(\hat{\boldsymbol{n}})$, as

$$\Delta E(\hat{\boldsymbol{n}}) = \sum_{i,\boldsymbol{k}}^{\text{occ}} \epsilon_i(\hat{\boldsymbol{n}}, \boldsymbol{k}) - \sum_{i,\boldsymbol{k}}^{\text{occ}} \epsilon_i(\boldsymbol{k}) \ , \tag{1.81}$$

where the first term is the band energy including the SO and the second one is the scalar-relativistic band energy. It then follows that the MAE can be expressed as

$$MAE \equiv \Delta E(\hat{\boldsymbol{n}}_{\boldsymbol{1}}, \hat{\boldsymbol{n}}_{\boldsymbol{2}}) = \sum_{i,\boldsymbol{k}}^{\text{occ}} \epsilon_i(\hat{\boldsymbol{n}}_{\boldsymbol{1}}, \boldsymbol{k}) - \sum_{i,\boldsymbol{k}}^{\text{occ}} \epsilon_i(\hat{\boldsymbol{n}}_{\boldsymbol{2}}, \boldsymbol{k}) \ . \tag{1.82}$$

Here the MAE is given as a difference of band energies, calculated for two different directions of the spin quantization axis, $\hat{\boldsymbol{n}}$. In order to appreciate this result, let us consider a MAE energy, which is of the order of 1×10^{-4} eV. In comparison, the total

energy of the system is of the order of 1×10^4 eV. Hence, the MAE is roughly eight orders of magnitude smaller than the scale of the total energy. In order to obtain the MAE one would in principle need to subtract two energies with a precision of at least 10 decimal places. This is clearly a very difficult task. The single-particle energies are of the order of 10 eV, which makes the subtraction practically feasible. It is important to note that the theorem is completely general and can be extended to any linear response calculation [113]. For example, the same trick is used to calculate the energy differences needed to extract the J_{ij}'s from the SDFT.

In conclusion, the calculations of MAE are numerically challenging, however, still feasible for systems with uniaxial symmetry. The practical calculations are performed within the SDFT by employing the force theorem (cf. 1.82). Here the SOC is treated as a perturbation and the difference of band energies for different spin directions gives the MAE. The connection with the macroscopic anisotropy constant, K_u, is achieved via Eq. (1.80). Such a result can then be compared to the experimental data in case of defect-free single crystals. In general, the agreement is hard to obtain due to sensitivity of the MAE to the details of the electronic structure. In particular the sensitivity to defect induced electron/hole doping or strain, which is often present in poly-crystalline samples. Even a small difference in the electron count, of the order of $1 \times 10^{-4}\, e$, due to the shift of the Fermi level, may result in the MAE error of the order of 1 meV. This value is considered to be a large, for MAE, and is similar to the values obtained in thin films, where the MCA is enhanced due to surface induced symmetry breaking. For this reason the theory may only strive to identify potentially good candidate materials and to study the trends, which may help to engineer the MCA in real materials.

1.2 Machine Learning

The scientific process can be summarised as a series of observations of natural phenomena, followed by their interpretation, through a logically self-consistent theory, and finally its verification through new observations. If the latter fails, the theory needs to be improved. Hence, the interpretation of data, embodied in statistics and statistical learning, has a prominent place in modern science. Generally speaking, there are different ways to build a theory, which offers an understanding of the observed phenomena. For example, the theory can be descriptive or qualitative. However, modern natural science prefers quantitative mathematical models, which can be rigorously tested. Although some phenomena allow interpretation in terms of causal laws, e.g. the motion of stars is deterministic over a long period of time, there exists a plethora of phenomena and processes in nature which can be addressed only using a statistical description. For

such processes we can deduce something about the probability distribution of a given property, but nothing about the outcome of an individual measurement. The statistical features emerge naturally when the parameters, which drive the process are unknown, or the system has many degrees of freedom, making a causal interpretation practically impossible. For example, one can think of the properties of the ideal gas [114]. Although the motion of each individual particle is trivial, describing the motion of all particles is intractable. Nevertheless, useful properties can be obtained using the statistical approach and the knowledge of how an average particle behaves. The statistical approach is thus a valid foundation for building scientific theories.

Statistical learning attempts to formalize the previously discussed scientific process and to generalize it to an arbitrary learning problem. It aims to provide a mathematical framework, which can be used to deduce the optimal model to describe a given data distribution. Machine learning (ML) has been developed as a more practical approach to the same problem. Namely, the goal of the ML approach is to construct algorithms that can use a set of training data to make prediction for future input, for instance the neural networks [115]. In materials science the ML techniques have been successfully used to classify and to quantitatively estimate the material properties [116–119], thus showing a great potential for further application. The field of statistical learning is nowadays closely related to that of ML, since the theoretical foundations of statistical learning theory provide a firm ground to build robust ML algorithms. However, while statistical learning theory focuses on exact mathematical properties, the ML approach is more focused on creating practical algorithms, which are able to provide good results for datasets of finite size.

In general, a learning problem falls under one of the following categories: the unsupervised, the supervised, the online, and the reinforcement learning [120, 121]. The main difference between these methods comes down to our prior knowledge of the data. In the first case, the training data is unlabelled and the algorithm needs to find a pattern in the data without external guidance. The second case assumes that the label for each data point can be determined by an expert system, i.e. by a supervisor. The learning process is thus guided and the accuracy of the prediction can be determined objectively. In the online learning approach the training data is received sequentially and the prediction model has to be continuously updated. This problem arises naturally when a large stream of data is available, for instance when analysing the usage of a popular internet site. Finally, in the reinforcement learning approach the ML algorithm actively learns through an interaction with the environment. A decision leading to a successful action is rewarded and the wrong one penalized. Therefore, this is a trial-and-error approach to learning, requiring the algorithm to be able to interact with the source of data. Here we will be exclusively concerned with the supervised learning approach.

The supervised learning problem in practice comes down to finding an unknown function, $\hat{f}(\boldsymbol{x}, \alpha_0)$, which matches the response of the supervisor as closely as possible. Here $\boldsymbol{x}$ is the vector of input features and α_0 is a parameter, which uniquely identifies a function belonging to a family of functions, Λ. For example, a normal distribution is characterised by two, real valued, parameters. However, in principle, this can be a quite general class of functions [122], i.e. the parameter α is not necessarily real valued. The function $\hat{f}$ can only approximate the true supervisor response, f, and hence, the learning problem can be understood as a "fitting" problem. In a general case, the function f does not need to be deterministic. The range of f can either be continuous or discrete. One can thus talk about the regression or the classification task, respectively. In both cases, the supervisor allows us to objectively estimate how close $\hat{f}$ is to the true function f. However, it is not obvious what metric is the best to use, in order to estimate the "distance" for an arbitrary function, f. In fact, the "no free lunch" theorem tells us that the optimal learning algorithm is distribution dependent [123, Ch. 5].

Here we will try to provide a brief introduction to statistical learning theory in order to expose the mathematical backbone of the ML method (Sec. 1.2.1). We will then proceed with the discussion of the bias-variance trade-off, which is important to understand the limitations of the ML approach. In section 1.2.2 we will illustrate this problem using the simplest elementary algorithms, namely the linear regression and the k-nearest-neighbour model. The extension of these simple models makes the basis for many modern algorithms. Finally, we will give an overview of the practical aspects of the ML approach in section 1.2.3.

1.2.1 Statistical Learning Theory

Statistical learning theory (SLT) provides a rigorous mathematical framework to formulate the supervised learning task as a risk functional minimization problem. In a nutshell, the task is to minimize the risk of making a bad prediction. A full exposition of the topic would need to go far beyond the scope of this thesis, therefore, we will only provide a brief summary, following Vapnik [124], and omitting many of the technical detail from the discussion. Hopefully, this will be sufficient to provide a clear picture of the underlying mathematical problem.

The SLT formulates the learning task as a problem of finding an optimal mapping, $\hat{f} : X \to Y$, based on a set of input, $\{\boldsymbol{x}_i \mid \boldsymbol{x} \in X\}$, and output, $\{y_i \mid y \in Y\}$, data. The set of ordered pairs $\{(\boldsymbol{x}_i, y_i)\}$ is know as the training set. If the mapping is known, for any future input $\boldsymbol{x}$ it is possible to predict the output, y. The input data is assumed to be drawn from an unknown but fixed probability distribution, $P(\boldsymbol{x})$. Therefore, the theory

assumes that the training set is made of independent and identically distributed data (i.i.d.), drawn from the joint probability distribution $P(\boldsymbol{x}, y) = P(\boldsymbol{x})P(y|\boldsymbol{x})$. Where $P(y|\boldsymbol{x})$ is the conditional probability of seeing y if the input is $\boldsymbol{x}$. We note that in practice the assumption of i.i.d. data may not be true. The learning is facilitated by the expert system, i.e. the supervisor, which represents the true response of the system and will return y based on the exact probability distribution $P(y|\boldsymbol{x})$, for any $\boldsymbol{x}$. We note that, due to the probabilistic mapping, the supervisor need not reproduce the exact responses, y_i, that are realized in the training set. Finally, an algorithm which can implement a family of functions, $f(\boldsymbol{x}, \alpha)$, where $\alpha \in \Lambda$, needs to be provided. The learning task then is to identify an optimal function $\hat{f}(\boldsymbol{x}) \equiv f(\boldsymbol{x}, \alpha_0)$, which minimizes the risk of making a bad prediction.

The *risk mimimization* is not uniquely defined. In general, one can write down the risk functional, $R(\alpha)$, once the loss function, $L(y, \boldsymbol{x})$, is given. The latter measures the distance between the predicted output value, $\hat{y} = \hat{f}(\boldsymbol{x})$ and the true output value y. The risk functional is defined as an expectation value of the loss function

$$R(\alpha) = \int dP(\boldsymbol{x}, y) L(y, \boldsymbol{x}) \ . \tag{1.83}$$

For the real valued functions $f_\alpha(\boldsymbol{x}) \equiv f(\boldsymbol{x}, \alpha)$, the most common choice of the loss function is the squared error loss, $L_2(y, \boldsymbol{x}) = [y - f_\alpha(\boldsymbol{x})]^2$. In this case it is known that the optimal function, $\hat{f}$, which minimizes $R(\alpha)$, is the expectation value of y

$$\hat{f}(\boldsymbol{x}) = \int y \ dP(y|\boldsymbol{x}) \ . \tag{1.84}$$

If the response of the supervisor is not deterministic, there will always be some residual variance in predicting the output. This is an intrinsic error, which is out of our control. In other words, a ML algorithm, which predicts without any loss on the training set should be regarded as suspicious. An alternative loss function is the L_1-loss, given by $L_1(y, \boldsymbol{x}) = |y - f_\alpha(\boldsymbol{x})|$. There is no a priori reason to assume that one loss function, i.e. distance metric, is better than the other one. The best choice needs to be determined for a particular problem, i.e. for a particular distribution, $P(\boldsymbol{x}, y)$.

Learning problems can be classified into three main categories: the pattern recognition (classification), the regression estimate, and the probability density estimation [124]. In case of the classification task supervisor's output is described by an indicator function. In the simplest case of binary classification, the output of the indicator function will take only values 0 and 1, for example. The most common loss function to use here is the zero-one loss [23], which is a $K \times K$ matrix with zeros on the diagonal and ones elsewhere. Here K is the number of different classes in the set of possible outcomes,

e.g. $K = 2$ for binary classification. Therefore, every misclassification is penalized by a unit loss. Minimizing the risk functional leads to the Bayesian classifier, which tells that the optimal class is the one that is the most probable, for a given $\boldsymbol{x}$. The regression estimation was already discussed, in the preceding paragraph. Finally, the problem of density estimation assumes that we have at our disposal a set of observations, $\{\boldsymbol{x}_i\}$, which are randomly drawn from an unknown distribution function. The task then is to choose an optimal density, from a set of densities $p(\boldsymbol{x}, \alpha)$, $\alpha \in \Lambda$, which best describes the observations. The loss function that is used here is the log-loss one, $L\left(p(\boldsymbol{x}, \alpha)\right) = -\log p(\boldsymbol{x}, \alpha)$. Therefore, the risk minimization approach encompasses all of the aforementioned learning tasks, only the loss function differs.

Although the risk functional (1.83) provides a satisfactory theoretical way to minimize the risk of making a bad prediction, in practice, all samples are finite. Hence, a practical learning theory needs to provide means to minimize the risk using only a finite set of training data. For this purpose in statistical learning theory a concept of "empirical risk functional" is introduced. This is defined as

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^{l} L(y_i, \boldsymbol{x}_i) \ , \tag{1.85}$$

where l is the size of the training set. Therefore, the empirical risk of Eq. (1.85) is completely analogous to that in Eq. (1.83), only defined for a finite sample. The task of the SLT then is to answer under which conditions the empirical risk converges to the true risk and how to control the rate of convergence. Finally, the theory needs to provide guidelines for constructing practical ML algorithms. Without going into technical details it turns out that the empirical risk can converge to the true risk, if certain conditions are met. However, the convergence is guaranteed only in the asymptotic case, namely when $l \to \infty$ [122]. The strict criteria is given in terms of the Vapnik-Chervonenkis (VC) entropy of the set of functions $f(\boldsymbol{x}, \alpha)$ and the corresponding VC-dimension, h [124]. These measures, loosely speaking, describe the capacity of a set of functions, i.e. its ability of fitting different data. If the training set size, l, is much larger than h then the empirical risk provides a good estimate for the true risk. However, for small sample set sizes this is no longer true. Unfortunately, the "curse of dimensionality" implies that in high dimensional space, any finite sample is small [23]. The results of the theoretical analysis suggest that for finite samples a new risk minimization principle needs to be employed, namely the *structural risk minimization*. Here the objective is to construct a structure of nested subsets of functions, S_k, having a finite VC-dimension, h_k, such that

$$S_1 \subset S_2 \subset \cdots \subset S_n \cdots \ .$$

For each subset S_k we choose the optimal function $\hat{f}_k$, by minimizing the empirical risk, so that the true risk is bounded by

$$R(\alpha_l^k) \leq R_{\text{emp}}(\alpha_l^k) + \Omega(\frac{l}{h_k}) \ . \tag{1.86}$$

The second term is the penalty functional, which goes to zero when $l/h \to \infty$. It corresponds to the confidence interval, which increases as h_k gets larger (for a fixed l). A more complete set of functions S_k has a larger confidence interval and hence, the risk is less bounded. In other words, we can always minimize the empirical risk by choosing a large enough set of functions, however, the uncertainty of the fit simultaneously increases, namely we have over-fitting. Therefore, the structural risk minimization implies that we need to simultaneously minimize the empirical risk and the capacity of the family of functions S_k.

1.2.2 Bias-Variance Trade-Off and Over-Fitting

In this section we will consider the key aspects of SLT, from a function approximation point of view. We will show, following Hastie et al. [23], that the bias-variance decomposition of the error plays a central role in limiting the accuracy of a ML model. Using as an example the additive linear model, we will identify three different sources of error that can arise in the fitting process. A simultaneous minimization of all three errors is impossible. This will provide us with a conceptual connection between the theoretical method of structural risk minimization and its practical counterpart, the over-fitting problem. Finally, we will use the simple algorithms like the least squares fit and the k-nearest-neighbour (KNN) to illustrate these concepts.

Consider an additive linear model, $Y = f(X)+\varepsilon$, where X and Y are the set of input and output values, respectively, and ε is a random noise, independent of $\boldsymbol{x} \in X$. We have, by assumption, that the expectation value, $E[\varepsilon] = 0$, and the variance, $Var[\varepsilon] = \sigma^2$. The task of the function approximation is to determine a $\hat{f}(\boldsymbol{x})$, which minimizes the L_2-loss function, given a (test) set of i.i.d. data. Since the inputs are statistically independent, the expectation value of the error, Err, can be evaluated for each $\boldsymbol{x}$ independently. We can write

$$\begin{aligned} Err(\boldsymbol{x}_0) &= E[(Y - \hat{f}(\boldsymbol{x}_0))^2 \mid X = \boldsymbol{x}_0] = \\ &= \sigma^2 + Bias^2(\hat{f}(\boldsymbol{x}_0)) + Var(\hat{f}(\boldsymbol{x}_0)) \ , \end{aligned} \tag{1.87}$$

where $Bias(\hat{f}(\boldsymbol{x}_0)) = E[\hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)]$. It is important to note that the expectation value, E, takes average over all random variables, including all possible realizations of the training set.

The first term in Eq. (1.87) is the intrinsic error, arising from the noise in the measurement of the output, y. This error can not be controlled nor reduced. The second term is the *bias*, which tells us how accurately our model approximates the true mean, for a given data point, $\boldsymbol{x}_0$. The third term is the *variance*, i.e. the expected square deviation of $\hat{f}(\boldsymbol{x}_0)$ around its mean. Typically, more complex models will have a lower bias, however, they will also be more sensitive to the actual values of the training points, y_i, and will thus have a higher variance. Similarly, simple models will have a lower variance but a higher bias. For example, for the KNN model, the expected error is given by, [23, p. 223],

$$Err(\boldsymbol{x}_0) = \sigma^2 + \left[f(\boldsymbol{x}_0) - \frac{1}{k} \sum_{l=1}^{k} f(\boldsymbol{x}_l) \right]^2 + \frac{\sigma^2}{k} \ . \tag{1.88}$$

Here the model complexity is inversely proportional to k, i.e. the most complex model is $k = 1$. For small k we can describe a more complex function f, while for large k we tend to describe its average value. We note here that the crux of the KNN model is to approximate the true function by a set of, locally, constant functions. Hence, when the number of neighbours is large, the approximating function becomes smoother, and less sensitive to the actual value of an individual neighbour. From Eq. (1.88) it is clear that the third term, the variance, decreases with k. Therefore, a simple model has a low variance. On the other hand, the second term, the bias, will tend to increase as the number of neighbours increases. Similar results can be shown to hold for the p-dimensional linear model fit [23]. In this case the variance is directly proportional to the number of parameters p.

In general, the competition between the bias and the variance reflects the implications of the structural risk minimization principle. The more complex methods can locally adapt better to the actual function and thus decrease the bias, i.e. the empirical risk, however, the uncertainty of the fit becomes larger. This is known as over-fitting. Models that have a small bias and a large variance tend to give a large error when applied on the new data. The total error thus needs to be minimized with respect to both the model complexity and the fitting error. This means that an optimal model parameter, α_0, needs to be chosen from the training data. Using the entire dataset to estimate both the optimal parameters and the error would give us the *in-sample error*, which is usually too optimistic and does not reflect the true accuracy of the ML model. A much better estimate of the accuracy is given by the *extra-sample error*, i.e. the error estimated from the previously unseen data. This is also known as the generalization error. The methods

that are employed in order to select the optimal model and to estimate the error are discussed in the next section.

1.2.3 Model Selection and the Error Estimation

So far we have discussed the supervised learning approach, viewing it as a function approximation problem. We have shown, quite generally, that the error can be split into three components. Besides the irreducible error, which is out of our control, there are the bias and the variance components, which depend on the complexity of the ML model. Simultaneous minimization of both errors is impossible and hence the optimal model needs to be selected based on the training data. However, this precludes making an objective estimate of the error, in the case the entire dataset is used for this task. Since the same data is used to train the ML model and to estimate the error. A standard method for circumventing this issue is to perform a blind split of the dataset into two parts, namely the training and the test set. The test set is kept aside and used only to estimate the error of the model. Since the data is never used for model selection, the error that is estimated on this dataset gives a good estimate of the generalization capability of the model. In other words, the test set provides us with an objective estimate of the error. The training set is used for model selection. Here we wish to expose in more detail some of the more important elements of the model selection process. Namely, we will discuss the k-fold cross-validation (CV) procedure and the learning curves as the most common, general methods, used to assess the validity of a ML model. We note that bootstrapping [23, 125] is another generally applicable method, which we will not consider here.

The model selection implies choosing an optimal function to fit the training data. This may include selecting a subset of input features, $\boldsymbol{x}$, and a suitable ML algorithm. Finally, optimal values of the parameters which control the complexity of the ML model need to be determined. The latter are known as the hyper-parameters. The selection of the input features depends on the task at hand and the available data. In general, there is no recipe on how to do this task except by trial-and-error. However, we will see later that it is important to do this task within the k-fold CV loop, in order to avoid introducing bias into the error estimate. The selection of a suitable ML algorithm is also a trial-and-error procedure. This is a consequence of the "no free lunch" theorem, which states that without any prior knowledge about the data distribution, we can not chose an optimal learning algorithm. In practice, especially when little is known about the properties of the underlying data distribution, this implies that we need to try out different algorithms to establish how well they perform for a given task. Finally, when an algorithm is chosen, we need to optimize the hyper-parameters in order to minimize

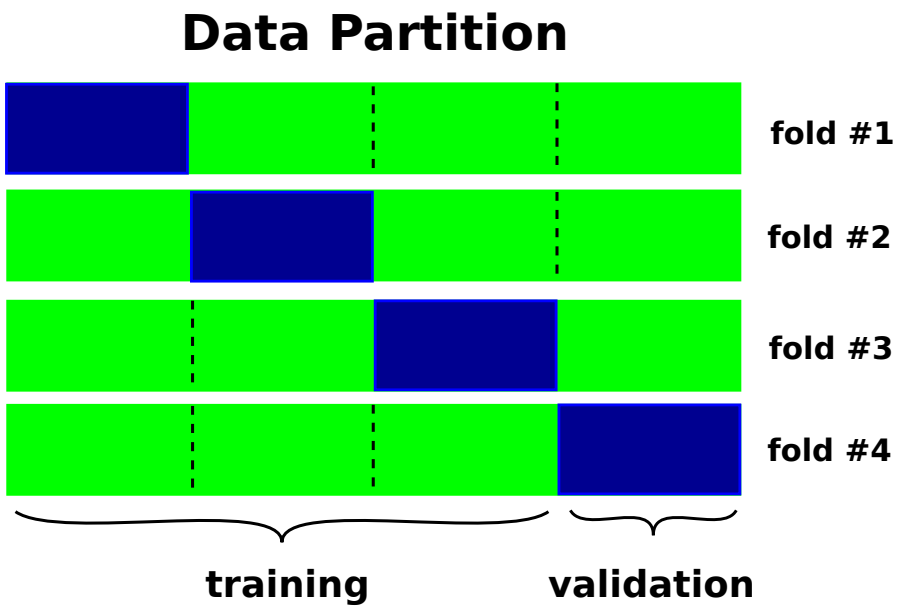

FIGURE 1.3: The four-fold cross-validation. The data is randomly distributed across four bins of equal size. For each fold, three bins are used for the training (green) and one bin is used for the validation (blue). Four iterations are performed, each time selecting a different bin for the validation. The cross-validation error is obtained as an average error over the four folds.

the generalization error. Henceforth, we will focus only on the latter and assume that the first two steps have been performed.

The simplest way to obtain an estimate of the extra-sample error, within the training set, is to hold out a part of the data from the model fitting process. For this reason, the training set is split into two parts, the train and the validation set. The latter is used to estimate the accuracy of candidate models and to choose the best one. For each model, i.e. for each value of α, the model is fitted and the error is calculated on the validation set. Finally, the model that produces the minimal error, $f(\boldsymbol{x}, \alpha_0)$, is kept and the generalization error is determined on the test set, which is kept aside during the entire model selection procedure. In many cases the datasets are not large enough to have both the validation and the test set unused. Keeping the latter aside it is crucial, however, to perform the validation in a more data efficient way. Namely, we can perform the k-fold CV, which allows us to obtain a reasonable estimate of the extra-sample error and, at the same time, to use the whole dataset for the training.

The idea of the k-fold CV is to repeat the above discussed procedure a number of times, and to calculate the average error. Namely, the dataset is split into k subsets, or bins, of which $k-1$ is used for training and one is used as a validation set (see figure 1.3). The data is assigned randomly to each subset. The train-validate procedure can then be performed k times, each time holding out a different bin. For each iteration the validation error, $err^{(i)}(\alpha)$, is estimated. The CV error is then given by the average

validation error

$$CV(\alpha) = \frac{1}{k}\sum_{i=1}^{k} err^{(i)}(\alpha) \ . \tag{1.89}$$

It is important to note here that the procedure is entirely general, and does not depend on the learning task or the choice of the loss function. The best model is chosen by minimizing $CV(\alpha)$. Since the procedure is performed k times, the entire dataset is used for both training and validation. However, one needs to keep in mind that the k-fold CV yields a biased estimate of the variance [126], which can result in selecting a less than optimal model [127]. We note that there exist other methods to estimate the extra-sample error from the in-sample value, e.g. the Akaike or the Bayesian information criteria [23]. However, these methods depend on the properties of the model used. We will not consider these methods here. The final issue concerns choosing an optimal value of k for the k-fold CV procedure. It turns out that the bias-variance competition again plays an important role in this discussion. For the case when $k = N$, where N is the size of the training set, the model will have a large variance, since the errors, $err^{(i)}(\alpha)$, are correlated. The $k = N$ case is known as the "leave one out" CV. On the other hand, for small k the training set size becomes too small and, consequently, the error increases. In practice $k = 5$ and 10 are the most commonly used values, which provide balance between the variance error and the training set size [23, p. 243]. However, the error in this case will depend on how the ML algorithm performs for a given data set size. For this reason it is always instructive to plot the learning curve.

The learning curve shows the accuracy of a ML algorithm as a function of a training set size, as shown in figure 1.4. This can be done for both the training and the validation data. Both of these curves are very informative. We assume here that the complexity of the algorithm is fixed, i.e. that the hyper-parameters do not change. The training curve is expected to be a monotonically decreasing function, which converges to an asymptotic value, i.e. the true precision, for a large training set size. For small datasets the error is small, since the algorithm can "memorize" all the examples. In other words, over-fitting occurs. In this case the training error gives a poor estimate of the generalization error. This is reflected in a large validation error. As the dataset size increases, the training accuracy gradually decreases. The validation error, however, becomes smaller as the training set is increased. For a large dataset, it should, in principle, converge to the same asymptotic value as the training error. However, for finite size datasets this may not be the case. The training error will always be smaller that the validation error and the algorithm will exhibit some degree of over-fitting. If this difference is significant, for a relatively large training set size, it may be beneficial to reduce the complexity of the algorithm. Nevertheless, this is not a fundamental problem, since the validation curve gives us a more realistic estimate of the true error. If the slope of the validation

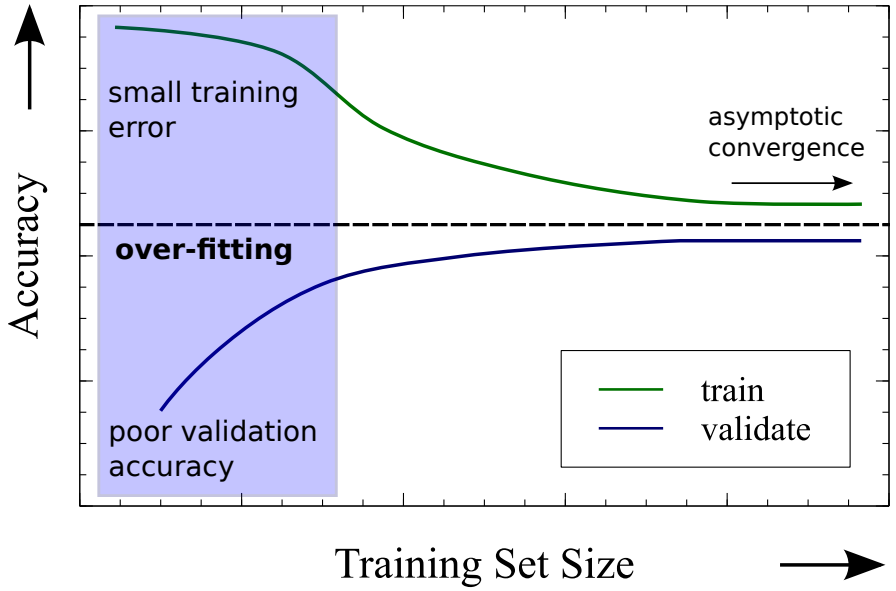

FIGURE 1.4: Hypothetical learning curve plot. The training and the validation accuracy, green and blue curve respectively, are plotted as a function of the training set size. The over-fitting region is indicated by the shaded background. For large training set size the asymptotic accuracy of a given model is achieved. Both curves converge to the same value.

curve is large, for a given training set size, the algorithm is not well trained and the error will depend on the details of the learning procedure, namely how the training set was sampled, the choice of k in k-fold CV, etc. If the curve is flat, it means that we have enough data and the algorithm can not benefit from more training data. One thus needs to ensure that a plateau of the learning curve has been reached. In conclusion, this technique helps us to decide if the model is sufficiently converged and to determine if we have enough data for the training. This is a crucial information, in situations where acquiring new data is expensive.

Finally, it is worth noting that k-fold CV does not guarantee an unbiased error estimate, if the input features are selected prior to the validation step, using the entire dataset. For example, exploratory analysis can be used to deduce which features best correlate with the data, and then used to train the model. However, choosing the features in this way will likely introduce bias. Since the choice of the features depends on the peculiarity of the training data distribution. Therefore, poor generalization capability can be expected. This is discussed in more detail in Ref. [23]. In order to avoid this, feature reduction should ideally be performed within the CV loop.

Chapter 2

Accelerated Discovery of Permanent Magnets Using High-Throughput DFT Calculations

2.1 Introduction

Magnetism and magnetic devices are omnipresent in modern technology. From low power electronics and digital storage [1–3, 128], to high power applications and devices, for example electric power generators and motors, wind turbines, medical devices, sensors etc., the need for magnets is continuously increasing. Despite the long history of magnetism the discovery of new magnetic materials is rare, at best. The majority of the technological progress related to magnetism in the last 40 years has concerned the engineering of the properties of the existing materials, namely the coercivity [27]. Here we will distinguish two categories of magnets. The first of these comprises magnetic thin films, which are of interest for spintronic applications and high-density integration of devices. These magnets are usually tailored to a particular application, and grown at the microscopic scale, $\sim 100\,\text{nm}$. The ability to characterize and control their properties is the focal point of the research in this field. They are typically in physical contact with other materials, e.g. forming interfaces and material stacks, and thus strain and defects can have a significant impact on the device performance. Combined theoretical and experimental investigations are needed to understand the properties of these materials. This is discussed in more detail in Chapter 3. The second category includes "bulk" ferromagnetic magnets, where the intrinsic material properties play a

crucial role. The application of such magnets is more common in high-power devices, e.g. power transformers and motors. These magnets will be the focus of the present chapter.

The technological requirements for the two aforementioned permanent magnet applications are substantially different. While power transformers require a "soft" magnet, which can be easily (de)magnetized in order to reduce the energy loss, electric motor applications require hard permanent magnets. These are characterized by their resilience to external magnetic fields and to high temperatures, namely they maintain their magnetic field in rough operating conditions. Typical materials representing these two groups are ferrites and rare-earth magnets [27, 28]. While ferrites have a low production cost and their constituent raw materials are abundant, high-performance permanent magnets usually require, at least, small amounts of rare-earth elements. The production of rare-earth-based magnets is exposed to a number of economical and ecological issues [28, 129]. Namely, the mining is currently controlled by a handful of countries, making the prices of the raw materials volatile. This in turn affects the whole production chain and ultimately the availability of the products. In addition, the ability to recycle the small amounts of these elements, diluted in large quantities of iron, is limited. This represents a long-term sustainability issue. For these reasons there is a great need to find suitable replacement materials. While there is no possibility to replace rare-earth magnets entirely, especially for applications where extreme performances are required, there is a lot more room to find suitable magnets, which have the properties intermediate between those of ferrites and of rare-earth magnets [28, 130].

Our focus here will be the search for new permanent magnets. In order to understand why these are so rare one needs to consider the requirements of a hard permanent magnet, which stem from its operating conditions, i.e. the need for stability at high temperatures. A hard magnet needs to be able to store a large amount of magnetic energy. This is measured by the energy product [28], which is proportional to the surface area under the hysteresis curve. It can be shown that for a magnet having an uniaxial anisotropy, K_u, the energy product is proportional to $2K_u$ [27, 131]. Therefore, the determination of the magnetic anisotropy becomes of key importance for finding permanent magnets. A target value for the latter is set to $K_u > 500\,\mathrm{kJ\,m^{-3}}$ [129]. In comparison, a replacement magnet for the best performing rare-earth based compound, $Nd_2Fe_{17}B$, would have to have a saturation magnetization, M_s, in excess of $1.2\,\mathrm{MA\,m^{-1}}$, and K_u of at least $2\,\mathrm{MJ\,m^{-3}}$. None of the known magnetic materials can simultaneously meet all of these requirements [28]. Similarly, an efficient room temperature operation requires that the magnetic order is stable, i.e. that the Curie temperature is at least $\sim 500\,\mathrm{K}$. A list of minimal requirements for a permanent magnet is summarised in

TABLE 2.1: Summary of minimal required properties of a permanent magnet, namely the saturation magnetization, M_s, the uniaxial magnetic anisotropy constant, K_u, and the Curie temperature, T_C.

Property	Min. Required
M_s	$1.2\,\mathrm{MA\,m^{-1}}$
K_u	$500\,\mathrm{kJ\,m^{-3}}$
T_C	$500\,\mathrm{K}$

table 2.1. Clearly there is a great interest in finding new magnetic compositions which could meet these goals.

However, a stable magnetic order is not common. In fact, out of $\sim$ 100000 inorganic compounds known to man, only $\sim$ 2000 show any kind of magnetic order [132]. Of these, barely a dozen have found practical applications [27]. The magnetic order is sensitive to details of the crystal structure, the composition and the texture of the microscopic grains. This makes the production of an optimized magnet a lengthy and time-consuming process. It is probably impossible to discover a new magnet using a one-shot procedure, given that magnetism is sensitive to details over a number of length-scales. A first step in this process, however, would be to identify a structure or a composition which, at least in theory, has a good chance to produce a useful magnet. This will be the aim here. Since the magnetism is a phenomenon associated with the meV energy scale, DFT *ab initio* calculations need to be employed in order to provide an accurate information about the material properties. Only by doing so we can be assured that the screening process is systematic, robust and accurate enough; limited only by the precision of the currently available computational methods.

Here we will describe and perform a three-stage high-throughput screening procedure, which combines data-mining techniques and the *ab initio* computation methods, with the aim of identifying candidate materials for permanent magnet applications. First in section 2.2 we discuss the material requirements and then we proceed with the initial screening (Sec. 2.3). A set of 229 materials is selected from the Materials Mine database [31, 133] as input for the second and the third step. In section 2.4 we discuss the automated workflow for the calculation of the Curie temperature. The performance is discussed and the results analysed. Finally, we calculate the magneto-crystalline anisotropy (MCA), for all of the materials, and identify the Heusler prototype, which maximizes it in section 2.5. We list the materials which meet all of the requirements for permanent magnet application and discuss the findings. In the last section we use the generated MCA data to investigate the prospect of applying machine learning methods, in this case the classification, as a computationally light and scalable method to be used in the high-throughput screening procedure.

2.2 High-Throughput DFT Screening Approach for Permanent Magnet Discovery

A material development typically begins with the discovery of a new interesting compound, which is subsequently studied and improved over time to yield better properties. This process becomes a bottleneck in a world of fast developing technology, which has a constant need for tailored functional materials. The high-throughput computational approach (HTA) [5] takes a different perspective. Modern computers and DFT codes can offer a fast and cost-effective characterization of the material properties. In this way a large portion of the chemical space can be explored much faster than before. The goal of the HTA is to characterize an entire family of compounds at once using a completely automatized process. The properties are stored in a searchable database so that the material screening can be performed as a post-processing step. The problem of material discovery then becomes a data-mining problem. In general, one needs to define search criteria, namely to identify the crucial material properties and the range of permissible values, which can be obtained from the database and used to identify interesting materials. This is often a challenging and non-trivial task. A more in-depth discussion of the related problems is deferred to Chapter 4. Here we are able to utilize a simple search criteria to select potentially interesting magnetic materials.

Finding a new permanent magnet requires a number of criteria to be satisfied. These are derived from both the underlying physics and the application requirements. Due to the availability of the data, we will focus solely on the family of full Heusler alloys [29]. These will serve as a testing ground to assess the feasibility of the high-throughput approach. Although the high-throughput and the combinatorial investigations are becoming ever more widespread, we know of only two works, namely Drebov et al. [134] and Sanvito et al. [132], where the discovery of permanent magnets was the main objective of the high-throughput search. The work of Drebov et al. focused on a number of rare-earth transition metal structures, for which the Curie temperature and the magneto-crystalline anisotropy was determined. The work of Sanvito et al. was a data-mining oriented investigation, where the perspective candidates were selected from a pre-computed database of materials, using descriptors. The Curie temperature was estimated using a regression technique and a couple of new magnets were synthesized, clearly demonstrating the advantage of performing an *ab initio* thermodynamic stability analysis. The work presented here can be understood as a combination of the two aforementioned approaches. We combine the data-mining candidate selection with *ab initio* DFT techniques, in order to achieve an accurate characterization of the magnetic properties, namely the anisotropy and the Curie temperature. These parameters are then used for further selection. In this way we incorporate the competitive advantages of the aforementioned methods. The

main search criteria that we employ to identify viable candidate materials are discussed below.

Thermodynamic stability - entails calculating all possible decomposition paths for a given compound. In other words a full phase diagram needs to be calculated. Obviously, the effort scales combinatorially with the number of atomic species in the unit cell. This is an incredibly demanding computational task. Nevertheless, the AFLOW database [135] implements just such a protocol [8, 132], offering robust stability estimates for the materials. We note that the temperature effects and, in the case of the magnetic materials, the full phase-space of spin configurations, are not taken into account here. In the latter case this means that an error associated with the magnetic order needs to be considered, typically between $\sim 100\,\mathrm{meV}$ and $400\,\mathrm{meV}$ per formula unit. These values are sufficiently large to affect the accuracy of the stability assessment. Since the AFLOW database is not yet completed, a complete stability analysis is not possible for all materials at this point. A computationally cheaper but less robust estimate of the stability is given by the enthalpy of formation calculated with respect to decomposition into the most stable elemental phases. Although it can not tell us, which materials will exhibit a full thermodynamic stability, it can be used to eliminate materials that are likely unstable. This leads to a significant reduction of the potential candidates. In addition a relative stability of different Heusler phases can be compared using this method. This approach was used, for example, in Chapter 3.

Tetragonal crystal structure - guarantees that the magneto-crystalline anisotropy is not quenched, as discussed in section 1.1.4.1. This criterion represents another computational challenge. Since the calculations in the AFLOW database are performed for cubic Heusler cells, additional structural relaxation needs to be performed for the magnetic systems. In this case an interplay between the crystal structure, namely the c/a ratio of the tetragonal unit cell and the magnetic moment may lead to different energy minima. Both these mechanisms tend to reduce the Fermi level DOS and thus, to lower the kinetic energy. *Ab initio* calculations need to be used here in order to establish what is the lowest energy structure. Here we do not perform such calculations, instead we identify the lowest energy structures from the data available in the Materials Mine database [31, 133]. The database provides both the enthalpy of formation, calculated with respect to the decomposition into elemental phases, magnetic moments and the c/a ratio for each structure. We will therefore make this database our default source of data.

Saturation Magnetization - needs to be high enough in order to achieve a high energy product and to prevent the magnet from undergoing spontaneous demagnetization. The target values for the magnetization are in the range $1\,\mathrm{MA\,m^{-1}}$ to $1.5\,\mathrm{MA\,m^{-1}}$ [28]. In

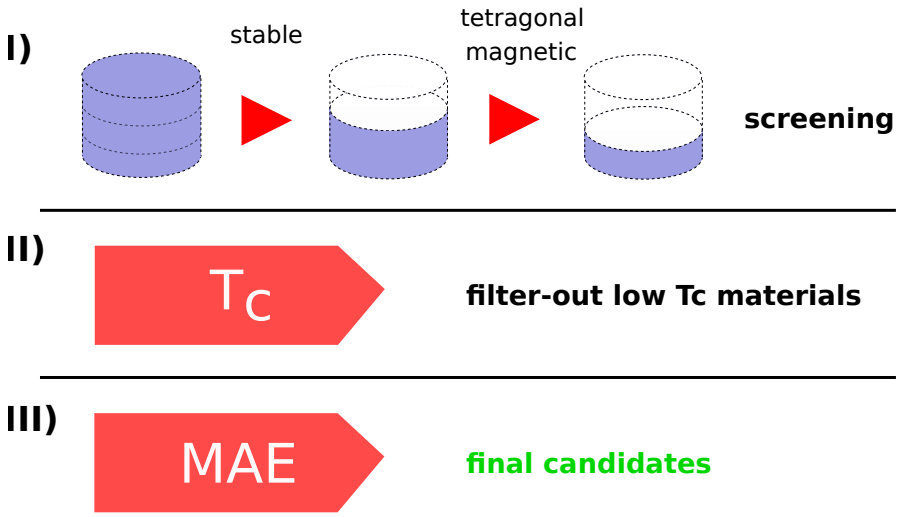

FIGURE 2.1: Protocol for the high-throughput screening of permanent magnets. The procedure is broken down into three steps (*I* - *III*). We start from a full database of Heusler alloys [31]. The screening criteria are then applied, i.e. we look for stable magnetic materials with a tetragonal ground-state structure. This is illustrated as a two step data reduction. In the second step we calculate the Curie temperature, T_C, for the candidate materials. The materials with a $T_C < 500\,\mathrm{K}$ are filtered out. Finally, we proceed by calculating the magnetic anisotropy energy (MAE).

the case of the Heusler alloys, this roughly corresponds to $6\,\mu_\mathrm{B}\,f.u.^{-1}$ to $9\,\mu_\mathrm{B}\,f.u.^{-1}$. However, slightly lower values should be expected for the Heusler alloys. Here we will consider any magnet that exhibits a magnetic moment $> 0.5\,\mu_\mathrm{B}\,f.u.^{-1}$. In effect we aim at considering all tetragonal magnetic materials. In doing so we wish to achieve two objectives. The first is to provide the best magnets for the application and the second, to create a large enough dataset, which can be used to study the trends in the magneto-crystalline anisotropy, with respect to the composition, the structure, etc.

Critical temperature - guarantees that the magnet will maintain its properties at elevated temperatures. This quantity is not available in any of the databases and should be estimated via direct calculation for the ground state structure.

Magneto-crystalline anisotropy - is required in order for a magnet to qualify for real-world use. A minimum value required in practical applications is $K_u > 500\,\mathrm{kJ\,m^{-3}}$. However, here we will target a value of $1\,\mathrm{MJ\,m^{-3}}$, since the former one is close to the established numerical precision of the calculation method. Similarly to the case of the critical temperature, this quantity needs to be calculated for candidate materials explicitly. However, due to a high sensitivity of this quantity to the experimental growth conditions, it is unclear if it can be used as a reliable descriptor.

2.3 Material Selection

In order to perform a high-throughput search for permanent magnets, a protocol for the candidate material selection needs to be defined. This is shown in figure 2.1. We start from the Materials Mine database (MMDB), which contains information about the relative stability of Heusler alloys, the magnetic moments and the relaxed crystal structure parameters. The latter is of vital importance. As previously mentioned, MMDB contains calculations for both cubic and tetragonal phases of a large number of Heusler alloys. This allows us to easily identify tetragonal materials. Since all of this information is already contained in the database, it is sufficient to perform a SQL query [136] in order to screen the materials.

The screening is performed in two steps. First the lowest energy phase for a given Heusler composition is identified and then we check if the material is magnetic and tetragonal. We consider all materials with $M_S > 0.5\,\mu_\mathrm{B}\,f.u.^{-1}$ and $c/a \neq 1.0$. This has left us with approximately 290 Heusler alloys, which satisfy this criteria. In this work we report the results for 229 of them. In the first screening step, the materials with a negative enthalpy of formation (EOF), $\Delta H < 500\,\mathrm{meV}$, have been selected. An exploratory data analysis had revealed that the EOF data obtained at the time of the acquisition, was not accurate, due to the problems with the reference, elemental phase, energies. This problem was corrected at a latter stage. We note that incorrect EOF data may affect the accuracy of the procedure with respect to experiments. However, the primary goal here was to explore the computational feasibility of the HTA and to perform an investigation of the data trends. We have therefore decided to proceed with the available data. As a consequence, our test dataset is a mixture of the stable (159) and the unstable (70) materials. In the remainder of the work we will always refer to the correct EOF.

The screening procedure leaves us with a list of potentially stable, magnetic and tetragonal materials. These are the prerequisites for finding a viable permanent magnet. Further screening, i.e. the second and the third step in figure 2.1, require *ab initio* calculations to be performed. Since these calculation methods are not a part of the mainstream set of DFT techniques we will have to assess the performance and the accuracy of such an approach. Especially the HTA requires a certain degree of automation, and not all computational methods can be safely used without a human supervision. This will be discussed in more detail in the following sections.

TABLE 2.2: Comparison of the mean-field Curie temperature, T_C, calculated using the KKR, T_C^{KKR}, and the FLEUR code, T_C^{FLEUR}, with the experimental, T_C^{exp}, and the previously reported theoretical values, T_C^{ref} [26, 137]. The temperatures are expressed in units of K. The calculation details are described in [138]. All calcualtions were performed for the regular cubic Heusler structure. The respective lattice constants, a, are shown in the table.

Material	a [Å]	T_C^{exp}	T_C^{KKR}	T_C^{FLEUR}	T_C^{ref}
Co_2FeSi	5.64	1100	1074	1187	1120
Co_2MnAl	5.76	697	563	697	808
Cu_2MnAl	5.95	630, 685	1028	1000	691
Cu_2MnSn	6.17	530	-	700	220

2.4 T_C Calculations

The calculation of the Curie temperature is, in general, a non-trivial matter. Its value is sensitive to the details in the crystal structure, i.e. the exact arrangement of the atoms, the distances between them and the spin configuration. These determine the values of the magnetic exchange constants, J_{ij}'s. Here we make a number of simplifying assumptions, namely, we use the four atom Heusler cell, which limits the possible number of spin configurations and calculate J_{ij}'s for a reference state obtained from the MMDB. We note that in addition to screening the potential crystal structures, MMDB also performs the search over a limited number of spin configurations, by initializing the spins of the magnetic atoms in different directions. Only the lowest energy states are then stored in the database. In comparison, the AFLOW database probes only the ferromagnetic order, thus making a big simplification for systems having multiple magnetic sublattices. The latter case is quite common in the family of Heusler alloys. The reader should note here that only collinear magnetic orders are considered in both databases. The database data thus provides us with both the ground-state crystal structures and the initial magnetic moments, which greatly simplifies all further calculations.

2.4.1 Method

We calculate the J_{ij}'s using the KKR method, discussed in section 1.1.4.1, and implemented in the Münich SPR-KKR code [139, 140]. The LDA approximation is used, as implemented in the code, together with a $20 \times 20 \times 20$ kpoint grid. For the J_{ij} calculation we consider a maximal interaction distance of 1.2 a_{lat}. This is a relatively short distance for metals, however, in this way we are able to calculate all the J_{ij}'s in a single calculation. We note that we have also tested the J_{ij} calculations using the spin-spiral method, as implemented in FLEUR [94, 141]. We have found the results to be comparable (see table 2.2). Since the latter method involves a significant computational cost we find it

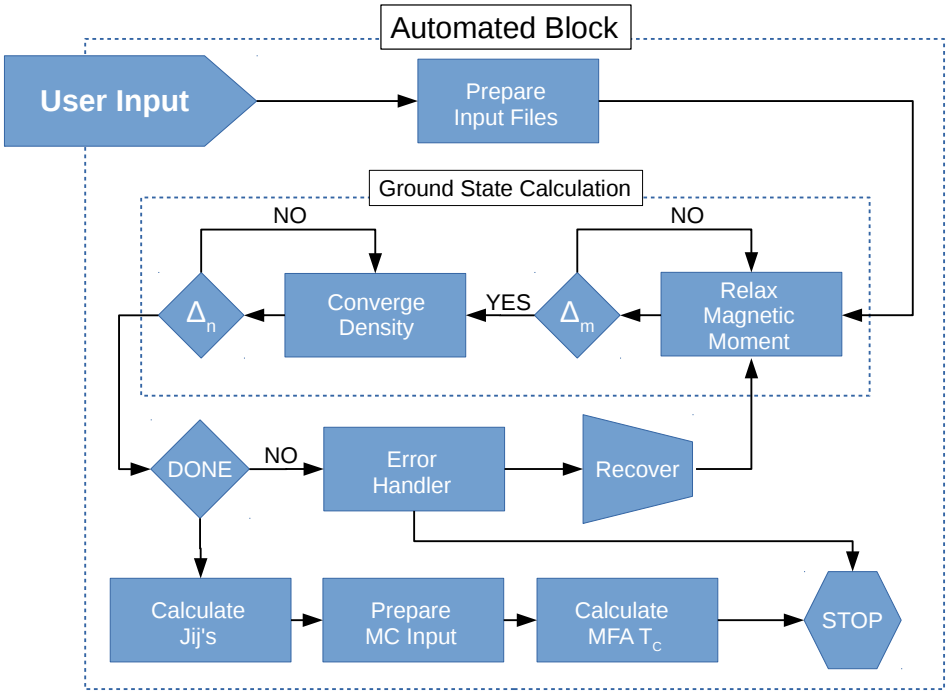

FIGURE 2.2: Schema of the automated KKR workflow. The user needs to provide the site occupation and the crystal structure data as an input. The workflow takes care of the rest. The ground state is obtained in two steps. First the magnetic moment is converged and then, in the second step, the electron density. Each of these steps is automatically restarted if convergence is not reached. Finally, the Monte Carlo (MC) input files are prepared and the mean-field approximation (MFA) Curie temperature, T_C, is calculated. In case the ground state calculation fails, the error handler inspects the error and decides which recovery procedure to follow. Otherwise, the calculation is stopped.

inappropriate for a high-throughput application. In comparison, the computational cost of the KKR approach is negligible, being equivalent to a single ground-state DFT calculation. However, due to technical reasons, the KKR calculations need to be performed in several steps. This complicates implementing the high-throughput process but, the computational gain makes the effort worthwhile. We have found that it is best to first relax the magnetic moments, e.g. by performing 20 SCF iterations using a linear mixing method, and then to restart the calculation using the Broyden mixing [142, 143]. The second step may be repeated a number of times until the convergence is reached. After the grounds state is obtained, the J_{ij}'s are computed in a single iteration. Therefore, the bulk of the work is to done to reproduce the ground state.

The whole procedure was automated using Python scripts, making the human intervention minimal. Namely, the Heusler composition and the crystal cell structure are the only parameters that need to be specified. These can be obtained from any of the available material database. The scripts allow calculation restarts, in case the convergence

is not reached or the calculation crashes. Before the input can be set, or in the case convergence is not achieved, an additional DOS calculation needs to be performed, and the DOS analysed, to determine the energy range for the complex contour integration. This is done automatically by the workflow. A log file of the entire process is written to the output for debugging. As a final step, the Curie temperature is calculated within the mean-field approximation by calling an auxiliary script. In addition, the input files that are needed to perform the Monte Carlo T_{C} calculation are prepared [94, 97]. The workflow thus fully automates the procedure, enabling a large number of calculations to be performed with a minimal human effort.

For the calculation of the MFA T_{C} we have developed a separate Python script, which can be executed manually, provided some user input, or automatically, by the KKR workflow. In the latter case there is no need for the user input. The method is general and independent of the workflow itself. The script uses the Monte Carlo input file, which is parsed to extract all the information needed for the T_{C} calculation. In this sense, the same method can be used in conjunction with other DFT codes, e.g. FLEUR [141]. The T_{C} is obtained as the largest eigenvalue of the $\mathbf{\Theta}$ matrix, where

$$\mathbf{\Theta}_{\alpha\beta} = \bar{M}_{\alpha} \left(\frac{\sum_{\boldsymbol{R}_j} J_{0\alpha,\boldsymbol{R}_j\beta}}{3k_B} \right) \bar{M}_{\beta} \ . \tag{2.1}$$

Here the $\bar{M}_{\alpha}$ is the magnitude of the magnetic moment at site α, and the $J_{0\alpha,\boldsymbol{R}_j\beta}$ are the magnetic exchange constants, i.e. the J_{ij}'s obtained from *ab initio* calculations. $\boldsymbol{R}_j$ denotes the unit cell position and k_B is the Boltzmann constant. A complete derivation of the expression (2.1) is given in Appendix A.

2.4.2 Results

Our goal here is two-fold. First, we wish to test the automated T_{C} calculation procedure in order to estimate the efficiency and the accuracy of the method. Second, we want to perform the second HTA screening step, as described in figure 2.1, on a set of 229 materials that were selected as candidates for permanent magnet applications. Finally, we will assess the feasibility of the T_{C} screening based on these results.

Following the previously described approach, the Curie temperature can be estimated using two methods, the mean-field approximation (MFA) and the Monte Carlo (MC) method. The latter is used as a reference, while the MFA allows a virtually effortless estimate of the "paramagnetic" Curie temperature. If successful, this would provide us with a light-weight and scalable method for evaluating the T_{C}. One of our goals here is to assess the applicability of the MFA when dealing with a large number of materials.

In this case the Heusler alloys, which have multiple magnetic sublattices and a varying ferrimagnetic order. To the best of our knowledge, only a few studies [144–146] have investigated the Curie temperature for a large number of Heusler alloys in a single work. The obtained estimates of T_C, compared to the experiment, are usually satisfactory. We have aggregated the experimental and the theoretical data from Refs. [132, 144], to obtain a sample of 43 Heusler alloys. The dataset is then used as a test for the workflow. We choose to make a close comparison with Tillmanns [144], for two reasons. The first is that the test dataset is large and the method that is employed is identical to ours, apart from the convergence criteria and the calculation procedure. The second reason is that Tillmanns has performed MC T_C calculations for all of the materials in his work. This provides us a good reference for assessing the precision of the MFA.

Benchmark Results - for the aforementioned 43 test materials we have calculated the J_{ij}'s, using the automated KKR workflow. We find that 29 calculations have converged and 14 have failed. Of these, 11 failures have occurred on the Tillmanns dataset (34), which leaves us with 23 materials, which can be used to compare the MC results. This corresponds to 67 % success rate, which is not satisfactory for an efficient high-throughput process. We note that the failures have nothing to do with the calculation of the J_{ij}'s. Instead, the automated KKR code fails to reach the ground state, i.e. to achieve the self-consistency. We did not follow up with a systematic investigation of the failures. Nonetheless, we can imagine two reasons for this. The first is the initialization of the magnetic moments, which is currently not performed, and the second could be the mixing of the semi-core states and the valence band. The latter could be alleviated by including semi-core states into the integration contour. The workflow performance could definitively be improved, however, a significant amount of (pure) programming work would be needed, so we have chosen to postpone this for a later date.

The accuracy of the automated calculation procedure is shown in figure 2.3. The theoretical results obtained using the KKR workflow are compared with Tillmanns' dataset [144]. The overall agreement is satisfactory, however, we notice that in a few cases, namely for the Au_2MnAl, Cu_2MnAl, and Cu_2MnSn, a strong overestimation of the T_C occurs. This is found to be due to the truncation of the J_{ij} interaction radius, as clearly shown in figure 2.4, below. We note here that the aforementioned materials are very similar, from a chemical point of view. Namely, the Au_2MnAl and the Cu_2MnAl are isoelectronic, with the Mn-Mn interaction being entirely responsible for the magnetic order. Cu_2MnSn is similar, having only 1 additional valence electron. Therefore, the Mn atoms exhibit a long range interaction, where the J_{ij} tail obviously reduces the critical temperature, by $\sim 200\,\mathrm{K}$ to $400\,\mathrm{K}$. This is a very good example of how important are the long range interactions in metals. It also implies that models, which consider only a few interacting neighbours are not a good way to reduce the computational load of

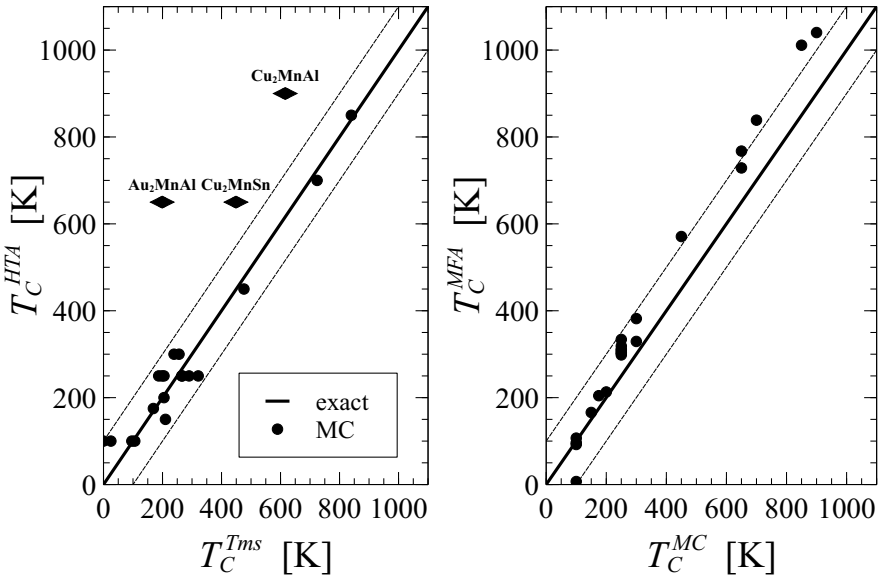

FIGURE 2.3: Comparison of the Curie temperatures, T_{C}, calculated using the KKR workflow presented here (HTA). *Left Panel* - shows the Monte Carlo (MC) T_{C}. The results obtained using the HTA are plotted against the corresponding data from the Tillmanns dataset (Tms) [144]. The line of the exact agreement and the $\pm 100\,\mathrm{K}$ error bar are indicated by the full and the dashed lines, respectively. Three of the materials show a poor agreement (diamonds), due to the truncation of the J_{ij} cutoff radius (cf. figure 2.4). *Right Panel* - compares the MC and the mean-field approximation (MFA) T_{C}, calculated within the HTA.

the T_{C} calculations, e.g. for the purpose of a high-throughput screening. In the right panel we compare the T_{C} obtained using the MFA and the MC method. The J_{ij}'s here are identical and the methods can be directly compared. We see that the MFA on average slightly overestimates the MC T_{C}, especially for the high-T_{C} materials. However, the correlation is strong, exhibiting a very weak variance and a noticeable bias (different slope). The latter can be empirically corrected in order to obtain a closer estimate of the MC results. We note that this only implies that the MFA holds very well for the metallic Heusler alloys. It is therefore "safe" to use the MFA as a quick and computationally light method for the T_{C} estimation.

In figure 2.4 we compare the theoretical and the experimental values of the Curie temperature, T_{C}. The left panel shows a comparison of the T_{C} values obtained using the KKR workflow (HTA) and the respective results from the Tillmanns dataset. We can see that MC on average underestimates the T_{C} by $\sim 100\,\mathrm{K}$. This holds true both for the HTA and the Tillmanns approach. The exceptions are the three aforementioned materials, for which we have already established that T_{C} is overestimated, due to the truncation of the J_{ij} interaction radius. This is clearly illustrated in the right panel. Here the MFA T_{C} is compared to experiments. It can be seen that the MFA captures the experimental T_{C} quite accurately. The variance is of the order of $100\,\mathrm{K}$ and no bias

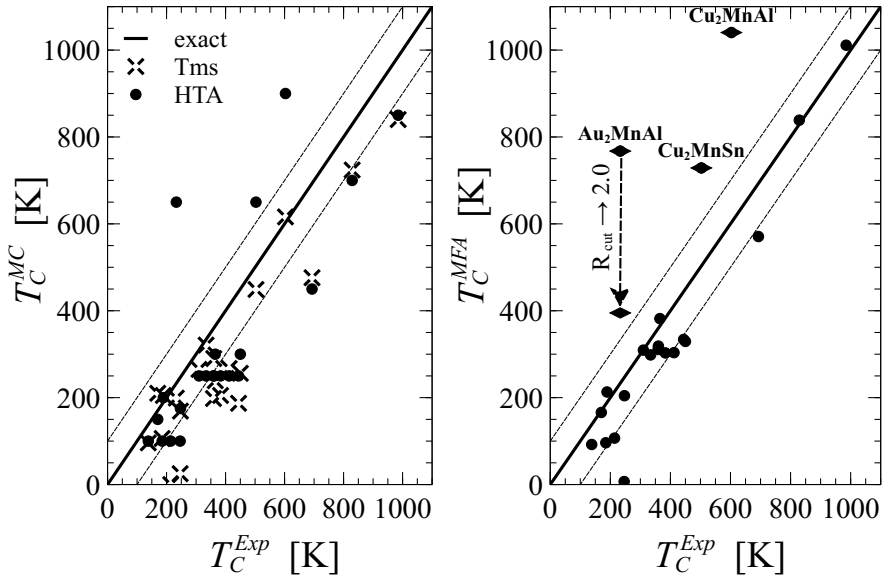

FIGURE 2.4: Comparison of the calculated Curie temperatures, T_{C}, and the corresponding experimental data. *Left Panel* - comparison of the Monte Carlo (MC) T_{C}, obtained using the KKR workflow presented here (HTA) and the values from the Tillmanns dataset (Tms) [144]. The line of the exact agreement and the $\pm 100\,\mathrm{K}$ error bar are indicated by the full and the dashed lines, respectively. *Right Panel* - the T_{C} obtained using the mean-field approximation (MFA), within the HTA. The materials which show poor agreement are shown with diamonds. For Au_2MnAl we show two T_{C} values, obtained using the cutoff radius of $R_{cut} = 1.2\ a_{lat}$ (HTA default) and $R_{cut} = 2.0\ a_{lat}$, respectively. The truncation of the Jij interaction radius thus results in a strong overestimation of the T_{C}.

is noticeable. In other words it is the MC T_{C} that exhibits the bias, a wrong slope, with respect to the experiment. It is not clear why this is the case. The longitudinal excitations, which are not correctly described by the classical spins assumed in the MC approach, may play an important role at high temperatures. However, one would expect this would further reduce T_{C}. Another possibility is that the exchange interactions that are extracted for the ground state are not adequate for very high temperatures. Our study can not give an answer to this question. The effect of the truncation of the J_{ij} interaction radius, R_{cut}, is shown in the case of Au_2MnAl alloy. The T_{C} is found to be reduced from 768 K to 395 K when R_{cut} is increased from 1.2 a_{lat} to 2.0 a_{lat}. In conclusion, we find that the MFA estimate of the Curie temperature is preferable, with respect to both the required computational effort and the accuracy. On the other hand, an accurate ab initio estimate of the J_{ij}'s is necessary.

HTA T_{C} Screening - was performed for the 229 materials, previously selected from the database. A convergence success rate of 67 %, similar to the benchmark results, was obtained. This gives a strong confirmation that the convergence problems are a general issue, which needs to be solved in order to meet the HTA efficiency requirement. For the 154 materials that we managed to converge, the Curie temperature was estimated

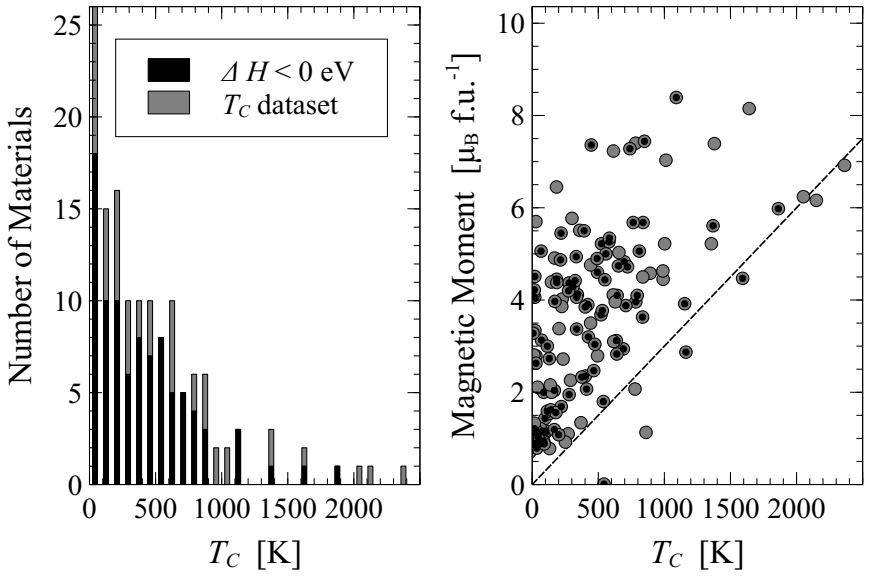

FIGURE 2.5: The Curie temperature, T_C, for the 229 material dataset, found by the high-throughput screening procedure (cf. figure 2.1). *Left Panel* - the histogram of T_C for the whole dataset (grey) and the subset of potentially stable materials (black). *Right Panel* - the relationship between the magnetic moment and the T_C. A dashed line, having a slope of $0.003\,\mu_B\,K^{-1}$, is shown to guide the eye.

using the MFA. We have previously shown that this is a good approximation. The results are summarized in figure 2.5. We find 31 materials with T_C above the required 500 K threshold, and for 28 of these T_C is between 500 K and 1200 K. Here we consider only the 146 materials which have a negative EOF, $\Delta H < 0$ eV. The results for the entire dataset are shown in the histogram in figure 2.5. This is a similar percentage, $\approx 20\,\%$, to the one that can be estimated from Coey (2010) [27, Fig. 1.8, p. 15]. In the latter case the estimate includes the experimental critical temperatures for both the ferromagnetic and the antiferromagnetic ordering. Therefore, it is reasonable to expect that the HTA estimate is slightly optimistic. Namely, we could expect that some of the materials would loose the ferromagnetic order if larger simulation cells were used. In the right-hand panel we show a distribution of the magnetic moment as a function of the calculated T_C. We see that there is no strong correlation, except that the magnetic moment seems to be bounded from below. This is indicated by the dashed line. Therefore, we find that a minimum moment of $3\,\mu_B\,f.u.^{-1}$ is required to yield $T_C \approx 1000$ K. This gives a useful rule-of-thumb for screening high-T_C materials. However, a large magnetic moment does not automatically guarantee a large T_C. We can also note that about a dozen materials has, simultaneously, a large magnetic moment, $M_S > 6\,\mu_B\,f.u.^{-1}$, and a large T_C, of which only 4 to 5 have a negative EOF, ΔH. The probability of having a high-performance magnet thus appears to be less than 0.5 %, with respect to the total number of candidates selected here. A full stability analysis would probably reduce this estimate further. It thus becomes clear why there is only a

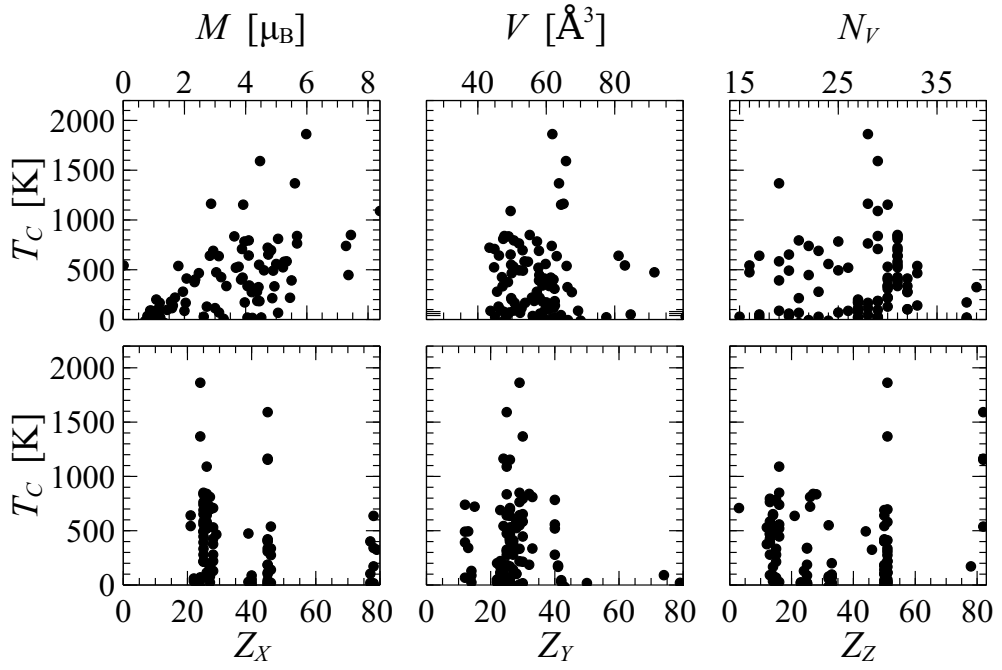

FIGURE 2.6: The dependence of the Curie temperature, T_C, on the structure and the composition of the (X_2YZ) Heusler alloys. The data is shown only for the potentially stable alloys, $\Delta H < 0\,\text{eV}$. The T_C is plotted as a function of the magnetic moment, M, the volume, V, the valence, N_V, and the atomic number Z_i, where $i \in \{X, Y, Z\}$ denotes the respective atomic site.

handful of high-performance magnets around.

As a special point of interest, in figure 2.6 we show the dependence of the T_C on a number of parameters, defined by the structure and the chemical composition of the Heusler alloys. In the upper row we show the dependence of T_C on the magnetic moment, the volume and the valence. These parameters can be related to the J_{ij}'s, which in turn decide the magnetic order. We can see that, besides the previously discussed lower bound on the magnetic moment, there are no obvious trends in the data. In the lower row we show the T_C as a function of the composition parameters. These graphs reveal what is well known for the Heusler alloys. The X and the Y elements are predominantly $3d$ and $4d$ transition metals (TM), while the Z are mostly main group elements. It can also be seen that some early and late $3d$ TMs make a potential candidate for this site. This kind of composition, free of main group elements and involving only TMs, was unknown to the chemistry of Heusler alloys up until recently, when the Co_2MnTi alloy was successfully synthesized after a HTA prediction [132]. In conclusion, the exploratory analysis indicates that it will be very difficult to gain intuition, or to identify new trends, in large datasets purely based on a visual inspection. A large scatter in the data tells us that the underlying problem is likely to be multi-dimensional and thus hard to represent in a humanly comprehensible format. The exception to this is the linear scaling of the

T_C with the valence in Co_2XY Heusler alloys [26]. However, any indication of such a simple rule is missing in figure 2.6.

2.5 MCA Calculations

As a final step of the HTA screening we perform magneto-crystalline anisotropy (MCA) calculations, for the previously discussed material dataset. We consider all the materials and not only those that pass the $T_C > 500\,\mathrm{K}$ screening step. We do this in order to maintain the dataset size, which will be used later, to explore the potential of machine-learning methods for the HTA screening application. This will be discussed in the next section. We will first briefly consider the accuracy and the computational requirements of the approach presented here. We will then look at trends in the MCA data and, finally, identify the materials, which meet all the criteria for permanent magnet applications.

2.5.1 Method

The magneto-crystalline anisotropy (MCA) was calculated using VASP [61, 147–150], and the PBE GGA density functional approximation [71]. PAW pseudopotentials [81] are used throughout the work. The magnetic anisotropy energy (MAE) was calculated using the magnetic force theorem, which was discussed in section 1.1.4.1. In this approach three calculations need to be performed, one self-consistent ground-state calculation and two non-self-consistent calculations, where spin-orbit coupling is included as a perturbation.

Since the MAE is a very small quantity, it is imperative to assure a tight convergence of the energy differences. In comparison the total energy converges much slower. A very dense k-point mesh should be used to obtain well converged results. We find that a $14 \times 14 \times 14$ grid allows us to achieve a $100\,\mu\mathrm{eV}\ f.u.^{-1}$ to $200\,\mu\mathrm{eV}\ f.u.^{-1}$ accuracy. Below this limit we find that the Fermi level contribution, due to the change of the occupation of the electronic states, in the presence of the perturbation, becomes significant. Given an average Heusler alloy volume of $50\,\mathrm{Å}^3$ to $60\,\mathrm{Å}^3$, we are able to achieve an MCA precision of about $0.3\,\mathrm{MJ\,m^{-3}}$. We find this sufficient to characterise the tetragonal Heusler alloys. Our goal is to identify those which have MCA $> 1\,\mathrm{MJ\,m^{-3}}$. We note that this approach is not suitable for cubic alloys, where the MAE is of the order of µeV. In addition, we have found that a smearing parameter, which is used to smooth out the Fermi level occupation, is very important for achieving a well converged result. Typical values used to obtain the ground state are $\approx 0.1\,\mathrm{eV}$, however, for the MCA calculations we converge the ground state by setting this parameter to 1 meV. This slows down the convergence,

and in some cases appreciably corrects the ground-state magnetic moment, compared to the results that can be found in the database [31]. The latter serve as a starting point for our calculations. The value of the MCA is very sensitive to the details of crystal structure and the convergence procedure, so we have performed a number of calculations to compare the results with those found in the literature. We have calculated the MCA for the Mn_2{Cr, Mn, Mo}Ga in the regular tetragonal Heusler phase, finding the MCA to be: $2.41\,\mathrm{MJ\,m^{-3}}$, $2.68\,\mathrm{MJ\,m^{-3}}$ and $1.80\,\mathrm{MJ\,m^{-3}}$, respectively. In comparison to the work of Wollmann et al. [146], we find the difference in the MAE to be below $40\,\mu\mathrm{eV}$ ($\Delta\mathrm{MCA} \lesssim 0.06\,\mathrm{MJ\,m^{-3}}$), for all the compounds. This is well within the established limits of accuracy. We note that we have used the same set of lattice parameters for the comparison. However, we did not modify the convergence parameters in order to match the reference results. For Mn_3Ga an anisotropy of $2.35\,\mathrm{MJ\,m^{-3}}$ was reported by Rode et al. [151]. In this case the geometry was slightly different from ours and the MAE was calculated as a difference of the total energies, using the all electron code FLEUR. Nevertheless, the results agree within the required $0.3\,\mathrm{MJ\,m^{-3}}$. In conclusion, we are able to estimate the MAE in a very good agreement with the literature available.

2.5.2 Results

We begin our discussion with the analysis of the trends in the MCA data. The first step is the exploratory analysis. The correlation of the MCA with a number of structural and compositional parameters is shown in figure 2.7. Namely, the magnetic moment, M, the valence electron count, N_V, the volume, V, and the c/a ratio. We can observe that the materials with a large MCA are correlated with the magnetic moment, the valence, and the volume. The magnetic moment, however, does not show an obvious dependence on the latter two parameters. We also note that the large volume alloys, $V \gtrsim 70\,\mathrm{\AA}^3$, have a low valence electron count, $\lesssim 25$. The alloys with a high N_V are centered around $60\,\mathrm{\AA}^3$. These findings suggest that the alloys with a high MCA have a magetic moment determined by a $3d$ transition metal (TM) and, in addition, contain late TMs. This is further investigated in figure 2.8.

In figure 2.8 we break down the dependence of the MCA on the valence, N_i, and the atomic numbers, Z_i, of the atomic site $i \in \{X, Y, Z\}$ of the X_2YZ Heusler alloy. We can confirm the aforementioned hypothesis. The MCA-N_V correlation is determined by the X element. The MCA peaks at $N_X \approx 10$, i.e. when X is a late TM. Similarly, the correlation with the magnetic moment is reflected in the valence of the Y site. The MCA peaks for $N_Y \approx 7$, i.e. when the Y site is occupied by the Mn or Fe. The large-MCA materials also have $N_Z \approx 4$, i.e. contain elements like Sn and Pb. All this is further confirmed by the dependence of the MCA on the atomic numbers for each site, shown

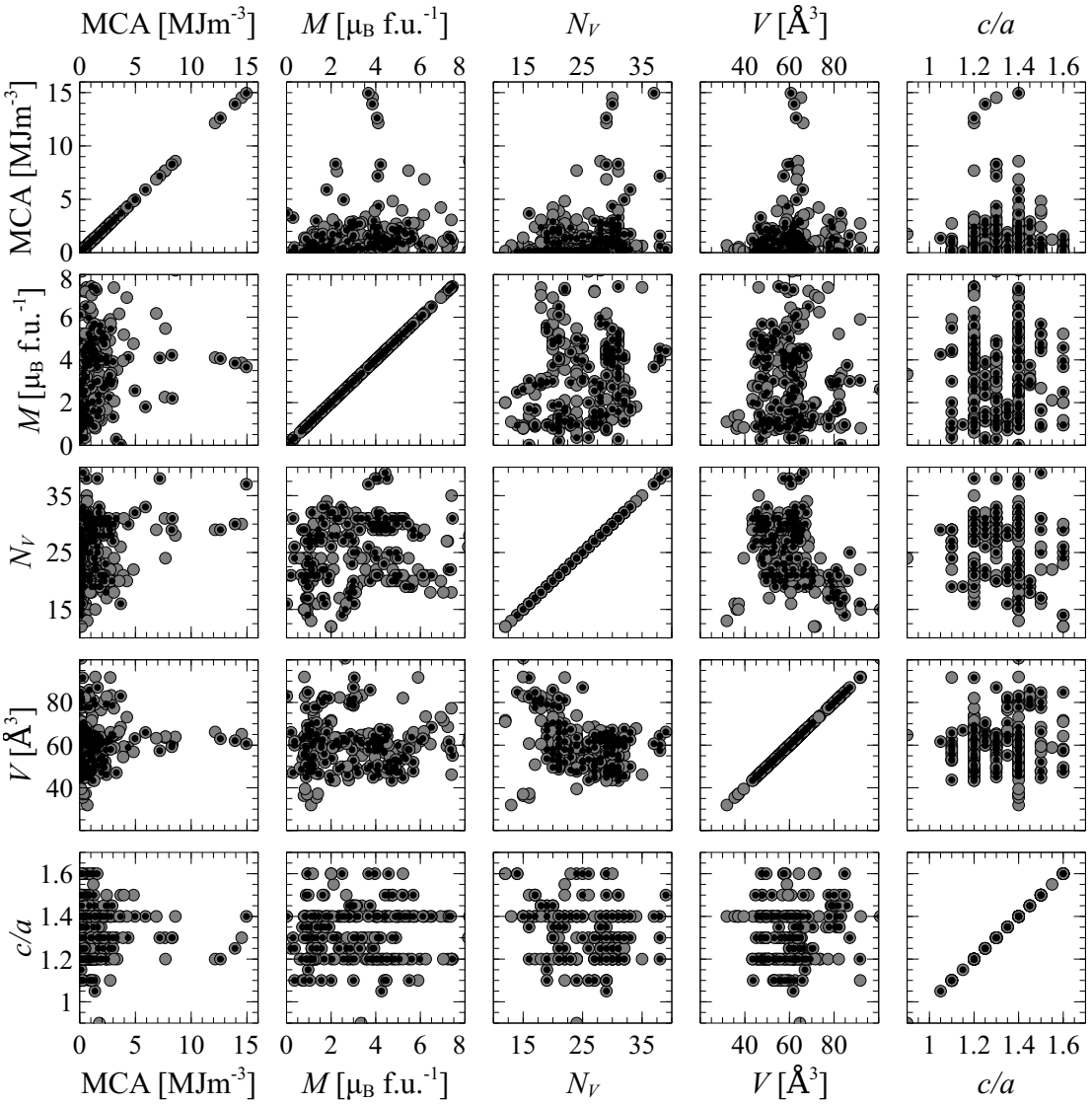

FIGURE 2.7: Correlation of the magneto-crystalline anisotropy (MCA) with a number of structural and compositional parameters for the 229 material dataset (grey). The mutual dependence of all of the variables is shown. Namely, the MCA, the magnetic moment, M, the valence electron count, N_V, the volume, V, and the c/a ratio. The compounds with a negative enthalpy of formation, $\Delta H < 0\,\mathrm{eV}$ are shown in black.

in the bottom row. A large MCA, $\geq 6\,\mathrm{MJ\,m^{-3}}$, can be achieved only when a $5d$ TM is occupying the X site. The Y site has to be simultaneously occupied by a $3d$ TM, giving a maximal magnetic moment of $\sim 5\,\mu_\mathrm{B}\,f.u.^{-1}$. The viable Z elements are the main group elements and a small number of TMs. It is interesting to note that a handful of elements, where the Z site is occupied by the TM yields a sizeable MCA. We can identify these materials as Pt_2ZnMn, Pt_2CuMn, Pt_2HfMn, Pt_2AlMn, and Pt_2ZnMn, ordered by a decreasing value of the MCA (MCA$\geq 2\,\mathrm{MJ\,m^{-3}}$). We can exclude Pt_2AlMn because Al and Mn occupy symmetry equivalent sites, hence the standard notation would put the Al on the Z site. The remaining 4 compounds are indeed all TM Heusler alloys and have a negative EOF. We note that the arbitrariness of the Y and the Z site order

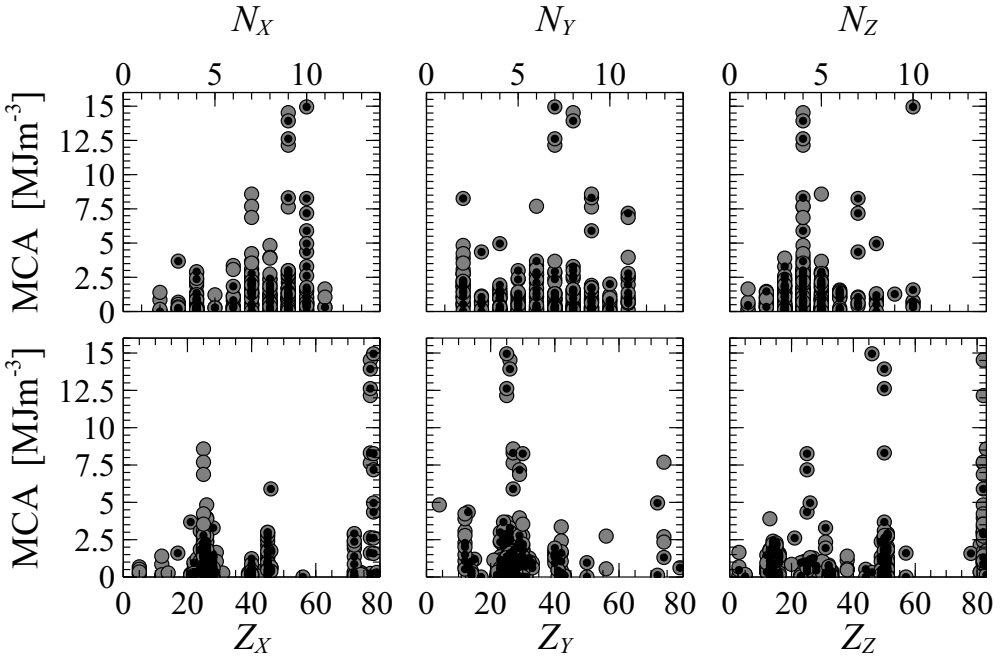

FIGURE 2.8: The dependence of the magneto-crystalline anisotropy (MCA) on the valence of individual atoms, N_i, and the corresponding atomic numbers, Z_i, where $i \in \{X, Y, Z\}$.

introduces artificial scattering in the data. A standardized site order would be helpful in performing data analysis. However, we will not pursue this issue further.

Finally in figure 2.9 we investigate the materials suitable for permanent magnet applications. Here, the MCA is plotted as a function of the Curie temperature. The magnetic moment is colour coded. We recall that the criteria for a permanent magnet application were set as follows: $T_{\mathrm{C}} > 500\,\mathrm{K}$, MCA$> 1\,\mathrm{MJ\,m^{-3}}$, and $M_S > 6\,\mu_{\mathrm{B}}\,f.u.^{-1}$. The first two criteria are indicated by the dashed lines in the figure. We can observe that a dozen (potentially) stable materials enter this widow and only 3 to 5 of them have a magnetic moment that approaches $6\,\mu_{\mathrm{B}}\,f.u.^{-1}$. This is determined by the compositional constraint found above. Namely, large MCA requires heavy $5d$ elements, however, these are not magnetic, so the total moment is limited to that of a single TM. The materials, which are found to satisfy all of the criteria above are listed in table 2.3. We note that the stability of the materials is determined only with respect to the decomposition into the elemental phases. In reality, a full thermodynamic stability analysis needs to be performed, involving all possible decomposition paths. In this case all the binary and the ternary alloys. After the possible candidates have been shortlisted to a handful of materials, this becomes a feasible computational task. This is routinely performed in the AFLOW database [8, 132].

As a final point we wish to briefly comment on the list of candidate materials shown in table 2.3. For the sake of the argument we will split these materials into 4 groups,

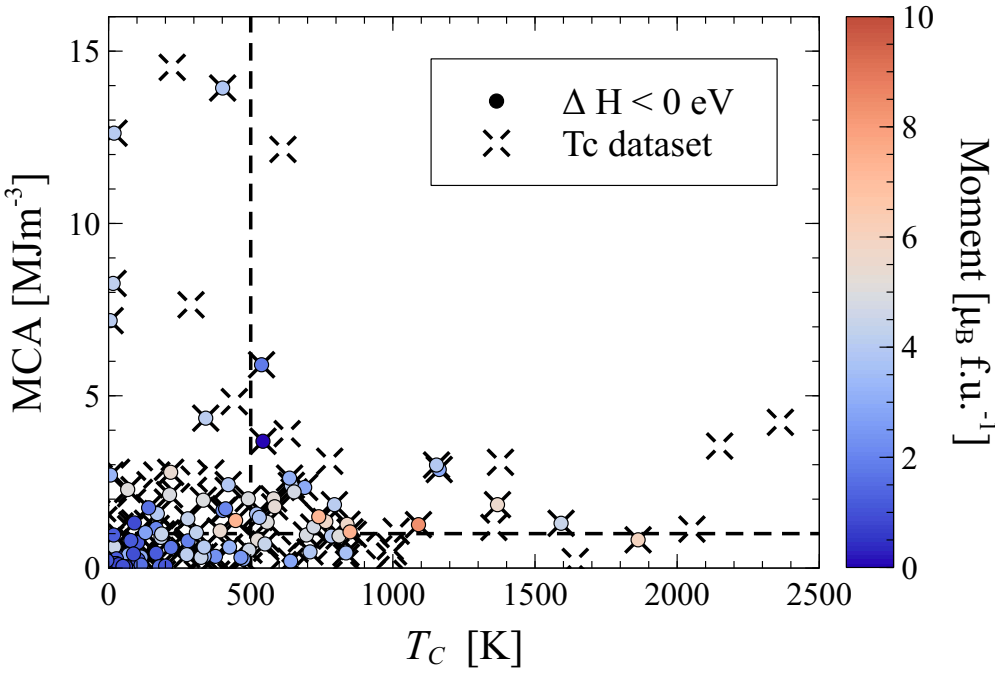

FIGURE 2.9: The magneto-crystalline anisotropy (MCA) as a function of the Curie temperature, T_C, and the magnetic moment. The MCA$= 1\,\mathrm{MJ\,m^{-3}}$ and the $T_C = 500\,\mathrm{K}$ are indicated by the dashed lines. The stable materials, $\Delta H < 0\,\mathrm{eV}$, which fall into this window make a viable candidates for permanent magnet applications. This is further discussed in the text.

TABLE 2.3: A list of candidate materials for permanent magnet application found using the high-throughput screening approach described in this work.

Material	MCA [MJ m^{-3}]	Tc [K]	Moment [μ_B $f.u.^{-1}$]
Fe_2MnS	1.26	1090	8.39
Mn_2CuS	1.05	849	7.44
Mn_2MgS	1.49	739	7.28
Fe_2GeCo	1.27	839	5.68
Mn_2CuAl	1.35	764	5.68
Cr_2ZnSb	1.84	1368	5.61
Mn_2ZnAl	1.78	584	5.34
Fe_2NiSb	2.01	581	5.26

namely, { Mn_2CuS, Mn_2MgS, Fe_2MnS }, { Mn_2CuAl, Mn_2ZnAl }, { Cr_2ZnSb }, and { Fe_2CoGe, Fe_2NiSb }. The first group is characterised by the Mn-S, the second by the Mn-Al, and the third by the Zn-Sb combination, while the last group are the Fe_2YZ compounds. Fe_2CoGe is known experimentally and grows in a cubic inverse-Heusler phase [152]. No data is available for the Fe_2NiSb, however, it can also be expected to grow in this phase, since the valence of the Y element is larger than that of the X [153]. This is a well known empirical rule. For example, the Fe_2NiGa and the Fe_2NiGe were predicted theoretically to grow in a tetragonal Heusler structure [154], however, they were synthesized in an inverse cubic phase experimentally [152]. Similar argument can be applied to: Mn_2CuS, Mn_2CuAl and Mn_2ZnAl, which are found here in the tetragonal

Heulser phase. An additional argument regarding the thermodynamic stability of the first three groups can be put forward, by considering their composition. It is known that Mn makes stable binary alloys with both sulphur [155] and aluminium [156], which open a number of possible reaction paths for the Heulser alloys to decompose into binaries. We note that MgS_2, FeS, and ZnSb are also known to exist. In addition zinc-antimonide, ZnSb, is known to be toxic and dangerous for the environment [157].

In conclusion, none of the short-listed materials, shown in table 2.3, makes a convincing candidate for permanent magnet applications. Nonetheless, we have demonstrated that a high-throughput approach can be used to screen a large number of materials. A predefined set of criteria was used to narrow down the list of 229 candidates to a dozen, which were then inspected in more detail. This is the primary goal of the high-throughput approach.

2.6 Discovering Hard Magnets Using a Machine Learning Classification

In the preceding sections we have shown that a high-throughput screening of the magnetic materials is feasible, however, the computational requirements are substantial. In particular, the majority of the effort is spent to estimate the MCA. For screening purposes it is sufficient to know if the MCA is larger than some threshold value. It would be very convenient if the materials with a large MCA, i.e. the hard magnets, could be identified using another approach, which requires less computational effort. Machine learning (ML) methods seem an obvious choice for this task. In this section we will explore the potential of the ML classification approach as a computationally light method for distinguishing hard and soft magnets. We will show that the elimination of the soft magnets is a more viable strategy, which allows us to circumvent the finite accuracy of the ML model and to maintain the DFT level of the accuracy. Finally, the performance of the classifiers will be investigated and the computational benefits of this approach will be discussed.

2.6.1 Method

We start with the previously discussed Heusler dataset, containing 229 materials, and extend the MCA data to include Rh_2- and Zr_2-based Heusler alloys. For these alloys the MCA has been computed using the same method that was described in the previous section. Finally, limited by the available Stoner parameter data [158], we keep only the materials, which contain elements with the atomic number $Z \leq 50$. This has left us with

a 153 material dataset, which we then split into the train and the test set, containing 114 and 39 materials, respectively. The test set is used only to estimate the accuracy of the final ML model.

Although the main objective is to find hard magnets, we find it more advantageous to train a classifier that identifies soft magnets. There are two reasons for this. The first is that the MCA for a potential hard magnet still needs to be calculated, eventually. So there is no direct computational gain in identifying a hard magnet. The second reason is related to the limited accuracy of a realistic classifier. Hard permanent magnets are very scarce and the cost of unintentionally dismissing a viable candidate is very high. Therefore, we need to be sure that all of the materials have been properly characterized. Any less-than-ideal classifier, targeting hard magnets, will never be able to achieve this goal. The problem is avoided if the soft magnets are the classification target. In other words, it is advantageous to eliminate the soft magnets from the dataset, and to calculate the MCA for the rest. Since we are not interested in the properties of soft magnets, these can be safely dismissed from further investigation. This implicitly requires that the classifier has a large specificity. We will come back to this point later. The remaining mixture of hard and soft magnets can then be analysed using *ab initio* methods, but now the computational effort will be invested more rationally. For these reasons we build the ML models to search for the soft magnets.

Two classifier models were trained, using the "Ridge Classifier" algorithm, implemented in the scikit-learn package [159]. The general procedure followed for building the classification models, as well as the auxiliary ML regression models (for the site specific magnetic moment and the volume of the Heusler alloys) whose inputs are used here, are discussed in great detail in Chapter 4. We only note that the latter two allow a quantitative prediction of the respective properties, using only the knowledge of the Heusler alloy chemical formula. A regular Heusler structure is assumed.

The first classifier, CLF_1, uses a simple input vector

$$\boldsymbol{v}_1 = (Z_X, Z_Y, Z_Z, M, V) \ , \tag{2.2}$$

where Z_i, M, and V are the atomic number, the total magnetic moment and the Heusler alloy volume, respectively. The latter can be taken from a database or estimated via the volume regression (see Chapter 4). Both approaches yield similar accuracy, since the volume regression is quite accurate. The magnetic moment needs to be obtained from a ground-state DFT calculation. This makes the classifier dependent on the DFT input. However, for many materials the required data is already available in the database. The

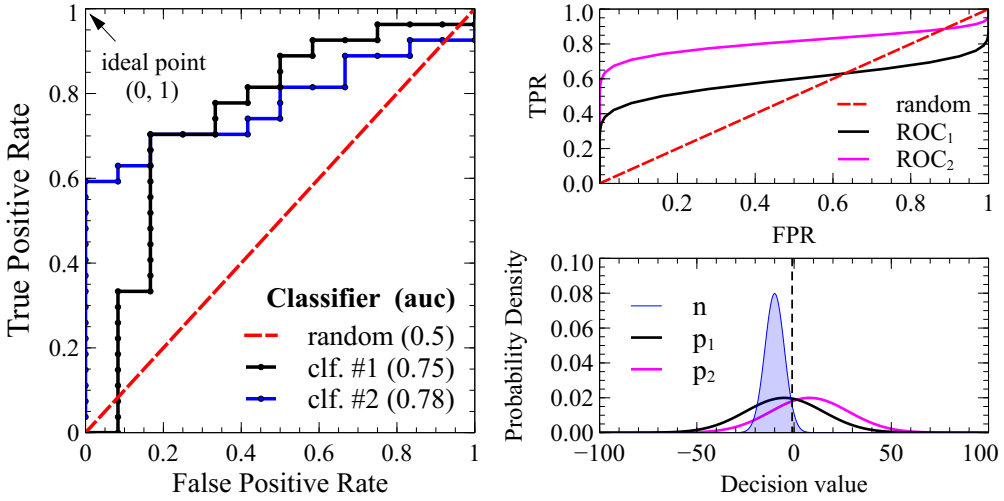

FIGURE 2.10: Receiver operating curve (ROC) for the two hard magnet classifiers discussed in the text (left). The area under the curve (*auc*) is indicated in the brackets. The operating point of an ideal classifier, having $auc = 1$, is indicated by the arrow. In the right panel (bottom), a probability density for the true positive, $p_i(x)$, and the true negative samples, $n(x)$, is shown, for two hypothetical classifiers. In the upper right panel we show the corresponding ROC curves.

second classifier, CLF_2, uses a larger input vector

$$\boldsymbol{v}_2 = (\{Z_i\}, \{N_i\}, \{\zeta_i\}, \{M_i\}, V) \ , \tag{2.3}$$

where N_i and M_i are the atomic valence and the magetic moment, for site $i \in \{X, Y, Z\}$. The Z_i and V are the same as above. We also introduce an additional descriptor, $\zeta_i = Z_i^4/V_i$, which is related to the density of the spin-orbit coupling strength [160]. Here V_i is the atomic volume. The site magnetic moments, M_i, are estimated using a ML regression. The CLF_2 is completely free of *ab initio* input. We can use it to identify interesting materials, which may not yet be present in the database. This opens up a possibility of guiding the high-throughput DFT calculations, thus performing an informed search over the chemical space. The true MCA labels, "high" and "low", were assigned by setting the threshold value of $0.8\,\mathrm{MJ\,m^{-3}}$. This value is convenient for 2 reasons. First, the value is close to the HTA screening criteria of $1\,\mathrm{MJ\,m^{-3}}$. Second, by setting the threshold at $0.8\,\mathrm{MJ\,m^{-3}}$ the two classes are of equal size, i.e. there is an equal number of hard and soft magnets in the training dataset. This eliminates the problems with the dataset bias, namely, both classes should be equally represented in the training data.

2.6.2 Results

The accuracy of the classifiers was established on the test set. We find that CLF_1 achieves a score of 74 % and CLF_2 of 69 %. These results suggest that CLF_1 is superior, although the overall accuracy improvement is not significant. We were not able to improve upon this quality by changing the ML algorithm or the input features. However, the possibility of achieving such an improvement is not excluded. A more in-depth analysis of the classifier performance was obtained by plotting the "receiver operating curve" (ROC) [161, 162], shown in figure 2.10. In the left-hand panel we plot the true positive (classification) rate (TPR) as a function of the false positive rate (FPR), for each classifier. A "random classifier", which assigns the labels arbitrarily, but with equal probability, is shown by a red dashed line. It is important to note here that the operating point of the classifier can be moved along the ROC curve, by changing the decision value cutoff. In this way we can choose if we find it more important to have a classifier which is highly sensitive or one which is specific. The former maximizes the TPR, at the expense of having a larger amount of FP predictions, while the latter tries to minimize the FPR, at the expense of a lower TPR. Usually a compromise between these two goals needs to be made. A good classifier has a large area under the curve (*auc*), ideally 1. The ideal operating point is thus located at $(0, 1)$, as indicated in the figure, where all the samples are correctly classified. We find that both MCA classifiers have a similar *auc*. The CLF_2 performs slightly better, having $auc = 0.78$, compared to the $auc = 0.75$, achieved by CLF_1. The latter were estimated on the test set.

Although the test scores for both classifiers are comparable, the shape of the respective ROC curves differs significantly. This implies different classification capabilities of the two models. The FPR does not vanish for any point on the CLF_1 ROC curve, i.e. this classifier will always yield a certain amount of misclassified samples. The CLF_2, on the other hand, has the initial part of the ROC curve along the y axis, where FPR = 0, followed by a sharp turn and a sudden increase of the FPR. The operating region just before this knee is very interesting, since it means that the classifier will never mistake a hard magnet for a soft one. In other words, CLF_2 can be very specific, FPR = 0, although the sensitivity, TPR $\sim$ 0.6, is far from the maximal, TPR = 1. From this point of view CLF_2 is clearly superior to CLF_1. Therefore, using CLF_2, in the best case scenario, about 60 % of the soft magnets in the dataset could be identified based solely on the Heusler chemical formula. We note that this would cover only 30 % of the entire dataset, and the MCA would have to be calculated for the rest. We will return to this point later.

In the right-hand panel of figure 2.10 we give a qualitative interpretation of the CLF_2 ROC curve. We note here that a binary classification can be understood as a problem of

finding an optimal hyperplane, which separates the two classes [23]. For every sample, a distance from the separating hyperplane, d, can be calculated and, based on this distance, the algorithm decides if the sample belongs to a given class. The latter is given by the decision function. In the case of the *RidgeClassifier* the decision function returns a signed distance of the sample to the hyperplane. For $d > 0$ the class is predicted as positive [159]. The true positive samples will follow a distribution, $p(x)$, and the true negative, $n(x)$, as indicated in the figure [162]. There we compare two hypothetical classifiers, which have slightly different distribution functions for the positive samples, p_1 and p_2 (bottom right panel), but a similar shape of the ROC curve to the one found for CLF_2 (top right panel). We assume that the distributions $p_i(x)$ and $n(x)$ are normalized Gaussian functions. The TPR and the FPR, are then given as

$$TPR(T) = \int_T^{+\infty} dx\ p(x) \qquad\qquad FPR(T) = \int_T^{+\infty} dx\ n(x)\ . \tag{2.4}$$

Here T denotes the cutoff value for the decision function. All samples for which $d \geq T$ are classified as positive. More specifically, for the *RidgeClassifier* the cutoff is given by $T = 0$. By changing the value of T we choose a different operating point on the ROC curve. We see that a conservative choice of T, in this example $T \gtrsim 0$, leads to a correct classification of the samples, i.e. FPR = 0. For smaller values of T, i.e. as we move along the ROC curve (to the right), the integration over a sharply peaked distribution $n(x)$, leads to a significant increase of the FPR, while the TPR changes a little. This makes the ROC curve to appear almost flat. The limiting case $T \to -\infty$ leads us to the upper right corner of the plot. We can see that the misclassification is related to the overlap of the distributions $n(x)$ and $p(x)$. In the case of $p_2(x)$, the overlap with $n(x)$ is smaller and thus a higher initial TPR can be achieved. This case is a close match to the ROC curve observed for CLF_2. In conclusion, the reason why CLF_2 manages to achieve FPR = 0, in the initial part of the ROC curve, is because the corresponding distributions $n(x)$ and $p(x)$ are reasonably well separated by the model. This implies that the vector of features of Eq. (2.3) can be used to fully segregate a part of the soft magnet population ($\sim 60\,\%$) from the rest. This is an interesting result, which is analysed more closely in figure 2.11.

As previously discussed, the ideal operating point is located just before the knee on the CLF_2 ROC curve (TPR = 0.6 and FPR = 0.0). In this case the classifier can identify soft magnets without making a mistake, although not all of them are detected. As we have already mentioned, this indicates that a part of the soft magnet population is well separated in the feature space defined by Eq. (2.3). We are interested here if this group of materials has a specific chemical composition, which should be avoided in the design of hard magnets. In figure 2.11, top panel, we assign a descriptive label

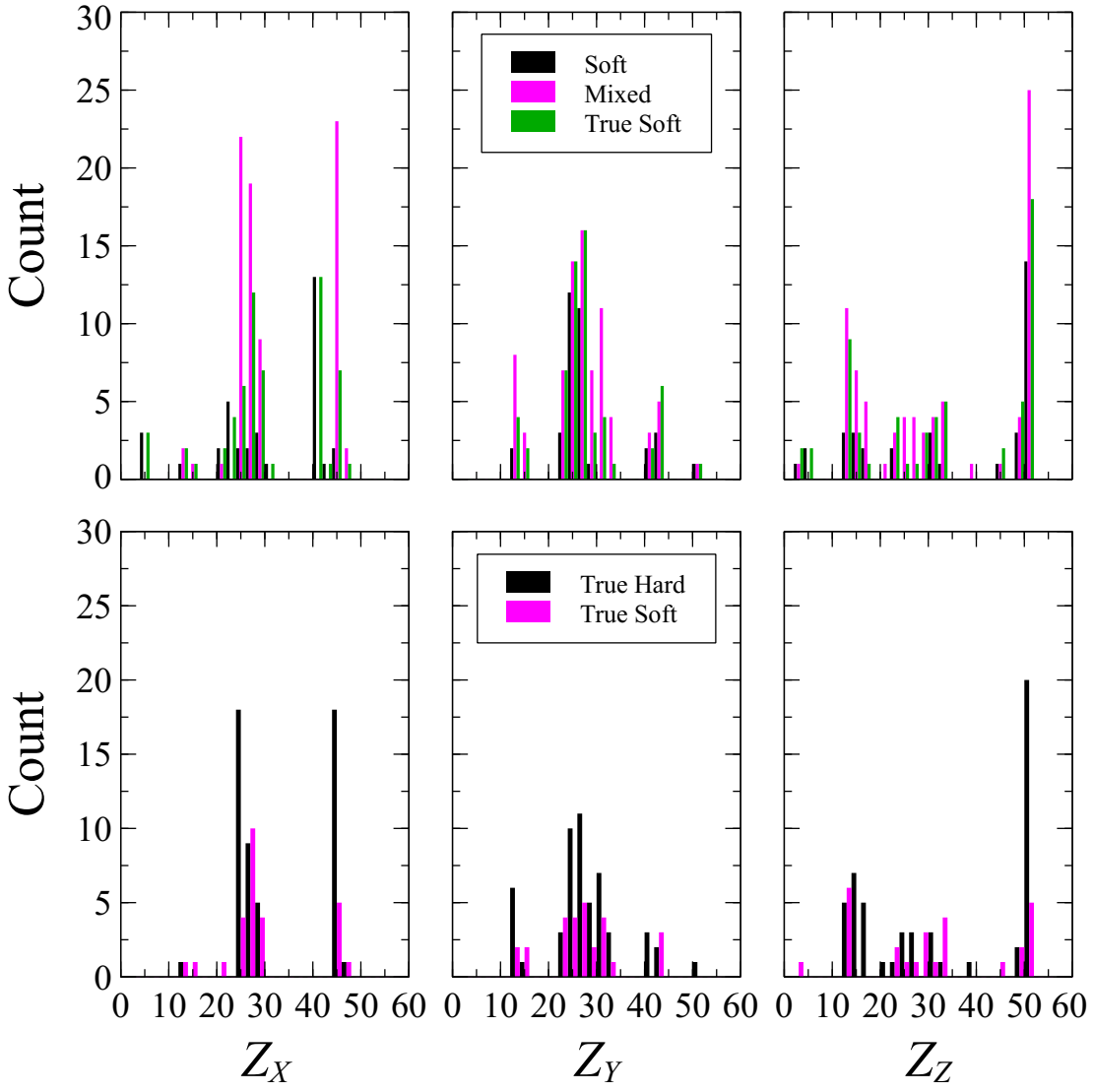

FIGURE 2.11: *Top panel* - histogram of the site occupation for the 114 material training set. The magnets have been labeled as "true soft", "soft" and "mixed". The first group is made from materials which have MCA $\leq 0.8\,\mathrm{MJ\,m^{-3}}$, as predicted by DFT. The second two groups are labeled according to the CLF_2 prediction. The "soft" magnets are those, which are classified as positive, when the classifier operating point is set just before the knee in figure 2.10. In this case there is no misclassified samples. The negative samples are labelled as "mixed", since they are a mixture of the hard and the soft magnets, that can not be resolved by the classifier. *Bottom panel* - a closer look at the trends in the mixed magnet group. Here the materials have been labeled as the "true hard" and the "true soft", according to the MCA value predicted by the DFT.

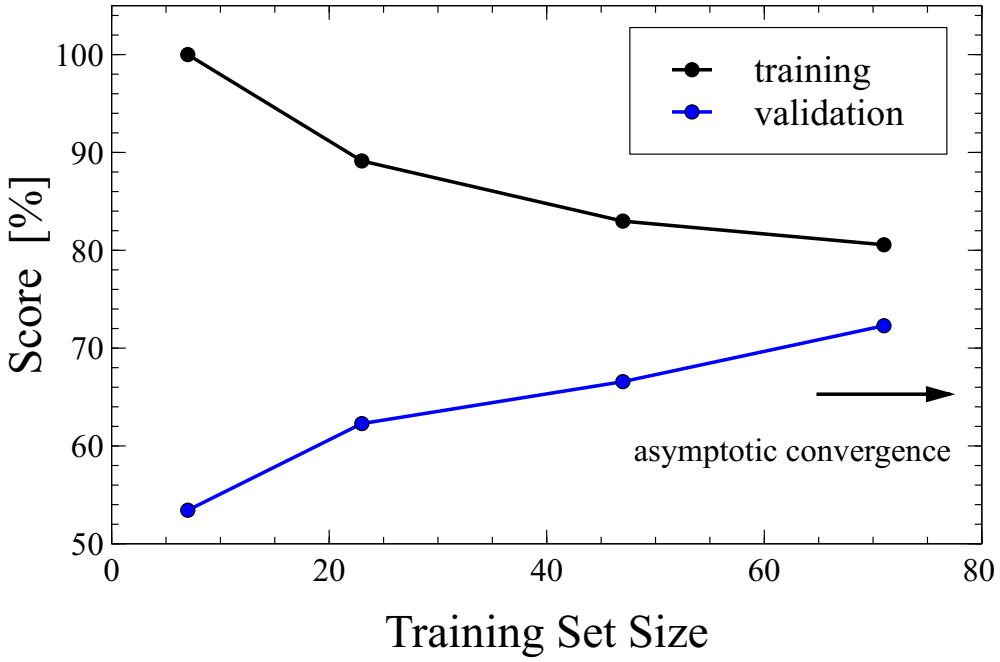

FIGURE 2.12: The learning curve for the CLF_2 classifier.

to the output of CLF_2, namely we identify the magnets as soft or mixed. The former are guaranteed to have the MCA $\leq 0.8\,\mathrm{MJ\,m^{-3}}$, while the latter are a mixture of hard and soft magnets, that can not be distinguished by the classifier. The population of the "true soft" magnets, as predicted by DFT, is shown for comparison. It is important to note that the population of the mixed magnets has a strong overlap with the population of the "true soft" magnets. For this reason, the two populations will share similar traits in the analysis. We can observe that the site occupation for all groups overlaps, over the entire composition space, defined by (Z_X, Z_Y, Z_Z), and thus a clear design rule can not be inferred. The strongest correlation between "soft" and "true soft" magnets can be made with respect to the occupation of the X site. We can observe that a soft magnet occurs with a high probability when the X site is occupied by an early transition element ($3d$ or $4d$), or by a very light element, $Z_X < 10$. In contrast, the mixed magnets, which include the group of hard magnets, are more likely when the X site is occupied by a late transition element, although the frequency of the "true soft" magnets in this case is not negligible. For other sites the composition appears to be very similar for all groups. This implies that the X site plays an important role in establishing a large MCA. Namely, the Y site determines the magnetic moment, while the X site provides a large SOC. The MCA classifier thus leads us to the same conclusion that was found in the previous section, by direct MCA calculations. The group of mixed magnets is analysed more closely in the bottom panel. We can see that the "true hard" and the "true soft" magnets follow very similar trends. This may be a part of the reason why the classifier has difficulties distinguishing the soft magnets that fall into this group.

Finally, we would like to estimate the computational benefits of following the ML approach presented here. In figure 2.12 we show the learning curve for CLF_2. We can see that approximately 80 materials are needed to achieve a converged result. Therefore, our training set of 114 materials was a reasonable choice. In comparison, this makes 75 % of the MCA dataset. The classifier could be applied to ~ 40 materials in the test set. Given that CLF_2 can effortlessly identify 60 % of the total soft magnet population, this gives us approximately 28 candidate magnets for which only the Curie temperature would need to be calculated. The 12 soft magnets can be safely eliminated. We recall here that the threshold for the hard magnet, MCA $\geq 0.8\,\mathrm{MJ\,m^{-3}}$, was chosen such that approximately 50 % of all the materials in the training set fall into this class. Assuming that the training set is representative of the entire population of magnets in the database, using the ML classification we can reduce the computational effort by 30 %. Given these considerations, we can conclude that our ML model becomes useful only when dealing with large datasets, which are of the order of 1000 materials or more. For small datasets it is better to proceed with the direct calculation. In comparison, the expected number of magnetic (full) Heusler alloys in the MMDB, which satisfy the first HTA screening step (cf. figure 2.1), is of the order of 1000. The ML approach could thus save us the trouble of calculating ~ 300 materials. Although the overall computational gain in this case is moderate, it is encouraging that a dataset of about 100 materials was sufficient to train the classifier, which means that a ML approach can be attempted with a rather low initial computational resource investment. We do not consider here the number of calculations that are needed to provide the input for the classifier. In conclusion, we find that the ML techniques can be used in conjunction with the HTA screening procedure (cf. figure 2.1), leading to an expected 30 % savings in the computational effort, while maintaining the DFT level of accuracy.

2.7 Conclusions

In this Chapter we have performed a high-throughput search for permanent magnets, within the family of full Heusler alloys. The requirements for a permanent magnet, having properties intermediate to the ferrites and the rare-earth, were discussed. Following these requirements a high-throughput screening procedure was designed. Starting from the available materials database [31], a data-mining search for suitable candidates was performed. We have selected ~ 230, potentially stable, magnetic materials, for which we then calculate the Curie temperature, T_C, and the magneto-crystalline anisotropy, MCA. We find that the T_C is bounded (from below) by the magnitude of the magnetic moment. A minimum of $3\,\mu_\mathrm{B}\,f.u.^{-1}$ is required for a material to have the $T_\mathrm{C} \geq 1000\,\mathrm{K}$. The results suggest that hard magnets, exhibiting the MCA $\geq 1\,\mathrm{MJ\,m^{-3}}$, are realized

when the X site is occupied by the heavy late transition metal element ($4d$ or $5d$), while the Y site is, ideally, occupied by Fe or Mn. This limits the maximal attainable magnetic moment to $\approx 4\,\mu_B\ f.u.^{-1}$, which is not sufficient for the permanent magnet applications. Only a dozen of the materials listed in table 2.3 have met all of the screening criteria. The probability of finding a new permanent magnet, within a pool of preselected tetragonal magnetic compounds can thus be estimated to approximately 5 %. Without any doubt, a full thermodynamic stability analysis would further reduce this probability. Therefore, magnets suitable for technological applications are indeed scarce and the high-throughput screening procedures need to be designed having this in mind. The final list of candidate materials was qualitatively analysed. We have identified a number of possible decomposition paths into binary alloys for most of the materials. A full phase diagram should be calculated to ascertain the stability of each individual candidate. In addition, a number of the materials is suspected to form in the inverse Heusler phase, despite the theoretical predictions. For example, Fe_2CoGe was shown experimentally to grow in this phase. In summary, none of the shortlisted materials makes a convincing candidate for the permanent magnet applications, however, at least one of the materials found here actually exists [152].

In the final section we have investigated the potential of machine learning (ML) techniques for accelerating the discovery of hard magnets. We have constructed two MCA classifiers using a training dataset of $\sim$ 100 Heusler alloys. While the first classifier, CLF_1, uses DFT information as an input, the second, CLF_2, does not. The test scores were comparable. However, the investigation of the operating characteristic had shown that CLF_2 is more advantageous to use, since it can operate with a zero false positive rate, FPR = 0, i.e. it can achieve the ideal specificity. The sensitivity of the classification was however moderate, TPR = 0.6. We have used the high specificity to eliminate the soft magnets from the population, without losing any of the hard magnets. In this way we had circumvented the problem of the limited accuracy of the classifier. This is a very important point, since permanent magnets are extremely rare and the cost of neglecting viable candidate materials is very high. In summary, the ML approach had allowed us to improve the computational efficiency of the high-throughput screening by 30 %. The DFT level of accuracy was not sacrificed in the process and the combined method is expected to discover all the candidate materials as the direct calculation approach. We have therefore shown that ML methods can be a valuable asset for the high-throughput investigations.

Chapter 3

Designing a Fully-Compensated Ferrimagnetic Half-Metal

3.1 Introduction

Power consumption represents a serious limitation for large-scale integration in electronic devices. The energy required to manipulate electric charges is of the order of an electronvolt. While such a technology is appropriate for devices incorporating millions of transistors operating at GHz frequencies, contemporary devices, which incorporate billions of transistors, operate at the very edge of the maximally allowed thermal dissipation. The reduction of the characteristic feature size, L, of the transistor, from the micrometer to the nanometer scale, has increased the power density of the device by roughly one million times ($\propto 1/L^2$). For this reason thermal conductivity investigations are currently very focus of much ongoing materials research [84, 163]. At the same time, alternative technologies, commonly branded as the "green" electronics, are receiving increasing attention from the research community.

Spintronics is a branch of electronics, which aims at utilizing the spin of an electron, rather than its charge, to store and manipulate information [128]. Spintronic devices intrinsically operate on the millielectronvolt scale, which offers a great perspective to alleviate the thermal dissipation issues. The most tantalizing aspect of spintronics is the possibility to operate devices using a pure spin-current. The latter corresponds to a flux of the magnetic (spin) moment, where the net flux of the electric charge remains zero. In principle, there is no Joule heating associated with a pure spin-current. Although pure spin-currents have been demonstrated experimentally [164], devices utilizing them remain much to be desired. Nevertheless, the successful implementation of spin valves [1] and magnetic tunnel junction devices [2, 3] has triggered a revolution in the data storage

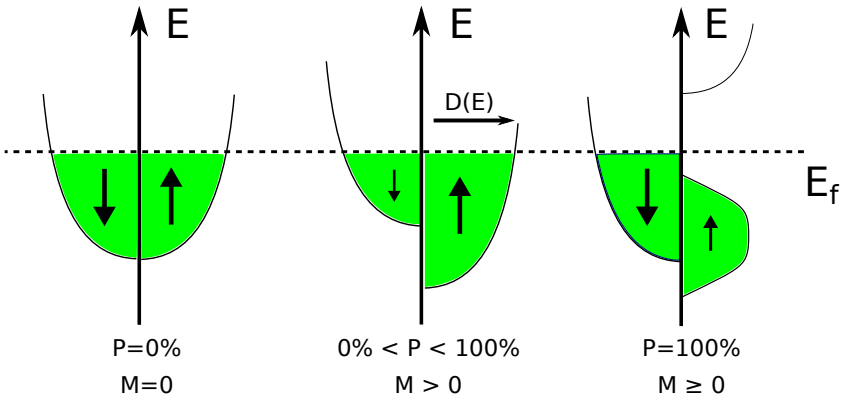

FIGURE 3.1: Illustrated density of states, $D(E)$, of common magnetic metals. The absolute magnetic moment, M, and the spin polarization, P, are shown. The spin up and the spin down channel are denoted by arrows. a) paramagnet ($M = 0$ and $P = 0$), b) ferromagnet/ferrimagnet ($M \geq 0$ and $0\,\% \leq P \leq 100\,\%$), and c) half-metal ($M \geq 0$ and $P = 100\,\%$)

technology. The existing spintronic devices are operated using a spin-polarized electric current, i.e. using a flux of electric charge where there exists a difference in the number of spin up/down electrons. This is often seen as a two component, spin up and spin down, electric current [165]. One can thus distinguish the spin up and spin down conduction channel and the electric conductance of an ordinary conductor can be broken down into two parallel-connected resistors, each conducting only one of the spin-current components. The spin polarization of the conductivity is often defined as

$$P = \frac{\sigma_\uparrow - \sigma_\downarrow}{\sigma_\uparrow + \sigma_\downarrow} \times 100\,\%, \tag{3.1}$$

where σ_s is the conductivity of the spin channel $s \in \{\uparrow, \downarrow\}$. In the case where the conductivity of one of the channels vanishes, e.g. in equation (3.1) let $\sigma_\downarrow \to 0$, we have a full spin polarization ($P = 100\,\%$). We note that this is not a pure spin-current, since the electric charge still flows. The advantage over conventional electronic devices is that a spintronic device efficiency depends on the spin-polarization of the current, in addition to its magnitude. Namely, the efficiency can be achieved by having a low current and a large spin polarization. In practice a large current density, $j \approx 1 \times 10^6\,\mathrm{A\,m^{-2}}$ to $1 \times 10^8\,\mathrm{A\,m^{-2}}$, is often required for the writing operation. The amount of magnetic moment transfer, namely the torque needed to change the state of the device is proportional to the current, as we can transfer at most $2\,\mu_\mathrm{B}\,e^{-1}$, which corresponds to the electron flipping its spin and transferring the magnetic moment onto the device. The efficiency can be maximized by having a low resistance device, which offers a full spin polarization. These requirements can be met by using a special class of metals, called half-metals.

Half-metals were first envisioned by de Groot [83]. They are defined by their electronic structure, shown in figure 3.1-c. In comparison with a normal metal, where both spin

channels are conductive, the half-metal behaves like a metal in one spin channel and like a semiconductor (or an insulator) in the other. In other words, a half-metal exhibits a gap in one spin channel, often called the spin gap. A half-metal thus offers a full spin polarization and the electric conductance of a metal, which makes it desirable for application in spintronics, as discussed earlier. The peculiar electronic structure of a half-metal makes it rare, as any kind of disorder or spin-orbit coupling tend to close the spin gap. The most direct way to prove the existence of the spin gap is by spin-resolved positron annihilation spectroscopy [166]. The half-Heusler alloy NiMnSn is the only material where the half-metallicity has been experimentally confirmed [167, 168], up to the precision of the experiment, $\approx 1/100\,e$. In order to have half-metallicity the spin order needs to be collinear. At finite temperatures transverse spin fluctuations (formally) destroy the spin gap. This makes the true half-metal a theoretical concept, which could be only approximately realized at finite temperature [168, 169]. However, for temperatures much lower than the critical ordering temperature, $T \ll T_{\mathrm{C}}$, the spin gap will provide protection against the spin-flip excitations and thus, work to preserve the spin polarization. From a practical point of view any material, which offers a spin polarization approaching 100 % can be regarded as a half-metal. We note here that the spin polarization in a half-metal can be defined using the Fermi level density of states, as an alternative to the definition presented in equation (3.1). Therefore, a half-metal should exhibit a full polarization regardless if we probe the transport properties or the density of states, e.g., in an optical experiment. The equation (3.1) allows us to make a more loose definition of a half-metal. It may happen that a material has no true spin gap, but if the mobility of one spin channel is practically zero, the current will still be fully spin polarized. Here we will refer to these materials as transport half-metals. The concept of a transport half-metal was also discussed in Ref. [170], where it was classified as $M_{\alpha\beta'}$-type half-metal. The reason for the distinction is that an optical measurement will give polarization which is less than 100 % for these materials. From a spintronics point of view a half-metal and a transport half-metal are equivalent, since the device operation depends on the torque, which is produced by the current.

Half-metals are not ideal spintronic elements. Usually a half-metal has a net magnetic moment, which in turn produces a stray field. Since the magnetic field is not localized, it will cause interference in devices, which tightly integrate many such elements. The operation of spintronic devices inevitably involves magnetization dynamics, which in turn leads to energy losses in the device, generated by the magnetic field induced currents. This is particularly true for high-frequency operation devices. The elimination of the stray field thus becomes an engineering concern. The stray field is naturally absent in antiferromagnets (AFM), but for a long time it was not clear how to switch the magnetic state of an antiferromagnet. Recently it has been demonstrated that the

spin-orbit torque can provide means to achieve this task [171, 172]. This realization has led to a new research field called antiferromagnetic spintronics [173]. The AFM-based devices certainly offer an interesting perspective, however such a technology is currently a long way from practical applications. It is important to note that AFMs can not act as spin filters, since by symmetry the spin up and spin down density of states needs to be identical. In other words, the Fermi level DOS of an AFM lacks the asymmetry needed to produce a spin polarized current. An alternative to an AFM is a fully-compensated ferrimagnet. Ferrimagnets have at least two symmetry-inequivalent magnetic sublattices, which can, in principle, compensate each other. Such materials are free of stray fields and have an asymmetric Fermi level DOS, which is required to use them as spin filters. In addition, the ferrimagneitic order is quite common in materials containing multiple magnetic atoms, which offers more options from an engineering viewpoint. In conclusion, an ideal combination of desired properties would be provided by a (fully-)compensated ferrimagnetic half-metal (CFHM). Therefore, these materials offer a great perspective for engineering spintronic devices [34, 174].

The class of CFHMs was first envisaged by van Leuken and de Groot [33] in 1995, but despite significant effort [175–180] the goal of a CFHM had proved elusive [174]. Recent experimental work by Kurt et al. [32] reported that a CFHM can be realized in Mn_2Ru_xGa (MRG) thin films, mid-way between the half-Heusler Mn_2Ga ($C1_b$) and the full-Heusler Mn_2RuGa ($L2_1$) structure. Both compounds have an underlying fcc structure, where the 4a and 4c sites are occupied by the two Mn atoms and the Ga occupies the 4b site. The additional Ru is assumed to occupy the empty 4d sites. Kurt et al. have found that the addition of Ru leads to a full magnetization compensation in $Mn_2Ru_{0.5}Ga$. At this composition there exists an ordered spin state, which is stable up to approximately 550 K and exhibits a small net magnetic moment, The onset of the compensation is thus in agreement with the generalized Slater-Pauling rule [181], since $Mn_2Ru_{0.5}Ga$ has exactly 21 valence electrons. This is further supported by the linear variation of the magnetization with the Ru content, which exhibits a slope of $2\,e$ per Ru, as would be expected from the Slater-Pauling rule. Following the same rule, it is expected that the compound is half-metallic. In the same work Andréev reflection (PCAR) spectroscopy showed a 54 % spin polarization at the Fermi level. Although this is far from the ideally expected 100 %, it shows a good agreement with the values measured in other thin film half-metals and most importantly demonstrates spin polarization of the current. In addition, anomalous Hall angle measurements yield an order of magnitude higher value than those observed in the $3d$-transition metals [32, 182]. All this indicates a high level of spin polarization due to the spin gap at the Fermi level. Despite the experimental evidence, the theoretical work of Galanakis et al. [183] predicted that the Ru doping does not lead to magnetization compensation for the

proposed site occupation. The main conclusion of the study was that these compounds are not half-metallic and likely exhibit some sort of disorder, in a strong contradiction to the experimental findings. Therefore, the current experimental [32, 182, 184] and theoretical understanding [183] of the MRG films has left a number of open questions regarding their properties. In this work we conduct a systematic investigation of the properties of the MRG films. We first summarise and discuss the main discrepancies between the theoretical and the experimental description of the MRG thin films in section 3.2. Following a high-throughput analysis, we establish the nature of the defects that occur in the MRG (Sec. 3.3). We then establish that these defects can fully account for the aforementioned discrepancies. This is investigated in section 3.4. Finally, we discuss the prospect and the strategy to engineer a CFHM in this class of materials. We note that the summary of the discussion presented here was already published in [185].

3.2 Properties of Stoichiometric Mn_2Ru_xGa Heusler Alloys

The previous section had outlined the current state of understanding of MRG thin films, from both the theoretical and the experimental standpoint. In our view, several important questions have not been addressed so far: I) the validity of the proposed Slater-Pauling-like mechanism for magnetization compensation, II) the effects of the substrate-induced strain on the electronic and the magnetic properties of the films, and finally, III) the relative stability of the competing disordered phases. In this section we will address the first two questions and the phase stability will be analysed in the next section.

In MRG thin films the magnetization compensation is a function of the Ru concentration, the film thickness, and the temperature. The dependence of the magnetization compensation on the film thickness is a complex problem. Strain, originating from both the substrate and the caping layer, plays a dominant role in thin films. It determines the lattice distortion and may induce a number of crystal defects. An appearance of meta-stable structures, namely of long-lived but thermodynamically unstable crystals, is also common. All this makes studying realistic thin films using *ab initio* methods practically unfeasible. For this reason we will limit ourself here to studying bulk MRG structures, whose properties can be later compared to those of the real thin films. The agreement is expected to be good for films which are thick, i.e. for films with thickness greater than 50 nm. The temperature dependence of the magnetic moments for the two symmetry inequivalent Mn sites is, in general case, expected to be different. This, in turn, leads to the temperature dependence of the net magnetization. Consequently, the

compensation can be achieved only at a particular temperature. SDFT calculations allow us to study directly the zero temperature compensation, which is determined by the material composition and crystal structure. Capturing the magnetization temperature dependence would require us to go beyond the SDFT description. For example, one could use a Monte Carlo approach [97, 186] or atomistic spin dynamics [187–189] to study the temperature dependence of the MRG sublatice magnetization. In both approaches the original system is mapped to an equivalent Heisenberg spin-system. Here we focus on understanding the zero-temperature compensation mechanism, arising from the electronic structure properties. We are mainly concerned with the validity of the proposed, Slater-Pauling-like, electron filling mechanism, through which the compensated ferrimagnetic ground state in MRG thin films is supposedly realized. If the proposed mechanism is valid the ideal film composition at which the compensation occurs would be constrained to a stoichiometry with the Ru fraction close to $x = 0.5$. Otherwise there should exist an alternative condition for achieving the compensation. It is therefore important to understand how the electronic structure is affected by the variation of the Ru fraction.

We have performed SDFT calculations for a series of bulk MRG crystals. Supercells having the Ru fraction, x: 0, 1/3, 1/2, 2/3 and 1 have been simulated. By performing calculations for small supercells we can obtain a qualitative picture of the Ru effect on the electronic structure. For example, the compositon with x = 0.5 is represented by a $2 \times 1 \times 1$ supercell, where the formula unit is repeated twice in the direction of the first real-space basis vector. We note that the size of the supercells should ideally be as large as possible in order to average out the effects originating from a particular choice of the atom distribution within the supercell. However, limited computational resources make this approach unfeasible. For each MRG composition we look at different volumes and c/a ratios of the unit cell. The corresponding total energies and magnetic moments are obtained directly from DFT calculations, performed using the VASP package [61, 147–150] and the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional [71]. Where needed, we estimate the electronic transport properties of a compound using the BoltzTraP code [190], in a post-processing step. The electrical conductivity is determined by solving the Boltzmann equations in the constant relaxation time approximation. We employ this method to study the intrinsic transport properties of the MRG systems and to understand how these are affected by the Ru doping and the mechanical strain. However, the method has certain limitations. In this approach, the electrical conductivity is scaled by the relaxation time, which can be arbitrarily chosen. A typical value for the relaxation time, τ, in metals is of the order of 1×10^{-14} s [65, p. 9]. In reality, the relaxation time is an anisotropic function of both the conduction electron energy and its spin [191]. We are thus unable to directly compare the results with the

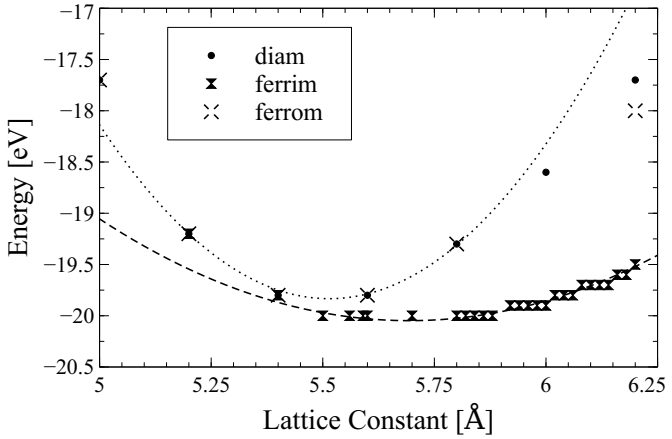

FIGURE 3.2: The SDFT ground state energy of Mn_2Ga for three different magnetic orders: the diamagnetic (diam), the ferromagnetic (ferrom), and the ferrimagnetic (ferrim). The equilibrium lattice constants are indicated by the minima of the quadratic curves, obtained by fitting the energy as a function of the lattice constant.

experiment. In addition, in the transport calculations the two spin channels are treated independently and hence, all spin scattering is neglected. The source of the scattering is the spin non-collinearity, which is always present to some extent, namely the non-collinearity arising from the thermal fluctuations. For this reason, the estimated spin polarization is expected to be larger than the corresponding experimental value.

Determining the ground state crystal structure of MRG compounds is a complex task. These systems contain multiple magnetic atoms in the unit cell and hence, the energy of the system should be simultaneously minimised with respect to the crystal structure parameters and the spin configuration. The two are intimately connected, a phenomenon known as the magneto-elastic coupling [91]. The latter makes the problem practically intractable, as there can be many local energy minima for each crystal structure considered. We take a more practical approach. The ground state crystal structures for all compositions are determined by performing a systematic scan over the volume and the c/a parameter of the unit cell. The atomic positions are kept fixed to the crystallographic positions. The experimental X-ray analysis [32] suggests that the crystal symmetry of the Heusler phase, $C1_b$ for Mn_2Ga and $L2_1$ for Mn_2RuGa, is perserved in the MRG films. The magneto-elastic coupling is tackled by performing constrained SDFT calculations for a given set of lattice parameters. During each calculation the spin configuration is kept fixed but the magnitude of the magnetic moment is allowed to relax. We repeat this procedure for a number of high-symmetry spin configurations. A combination of the crystal structure parameters and the spin configuration having the lowest energy is taken as the ground state.

x	a_{exp}	a_{dft}	M_{exp}	M_{dft}
0.00	5.96	5.69	-1.5	-0.4
0.50	6.01	5.93	0.0	0.2
1.00	6.06	5.97	0.4	1.0

TABLE 3.1: A comparison of the experimental lattice constants, a, and magnetic moments, M, for the Mn_2Ru_xGa compounds with the corresponding DFT values. The respective units are Å and $\mu_B\, f.u.^{-1}$.

In figure 3.2 we show the total energy of cubic Mn_2Ga as a function of the lattice parameter, a. We compare the energies for the diamagnetic, the ferromagnetic and the ferrimagnetic order of the Mn magnetic moments. The Ga ion is found to be non-magnetic in all cases. The Mn atoms are found to be in a high-spin state, with a magnetic moment of $\approx 3\,\mu_B$, in agreement with the work of Kubler et al. [192]. The total magnetization arises mainly due to the difference of the Mn magnetic moments, located at the symmetry inequivalent sites 4a and 4c. We find the ferrimagnetic ordering to have the lowest energy, in agreement with the previous studies of the Mn_2XY Heusler alloys [193–196]. The energy difference between the ferromagnetic and the ferrimagnetic configuration is of the order of $0.3\,eV/f.u.$. This makes the ferrimagnetic order stable with respect to the thermal fluctuations. We find similar results for cubic Mn_2RuGa. The Ru ion is found to have a negligible magnetic moment, which depends on the crystal structure, usually $\ll 0.5\,\mu_B$. For the intermediate compounds, which are represented by a supercell, we can explore a larger set of ferrimagnetic configurations. For the $Mn_2Ru_{0.5}Ga$ supercell we find that the lowest energy spin configuration is ferrimagnetically ordered with the all Mn 4a spins pointing in the same direction, opposite to that of the the Mn 4c site. For stoichiometries with x =1/3 and 2/3 the number of possible spin configurations grows combinatorially. We therefore adopt the aforementioned ferrimagnetic spin configuration as the lowest energy state.

A summary of the ground state DFT calculations is presented in table 3.1. The crystal structure parameters and the unit cell magnetic moments are compared to the experimental values. Significant discrepancies in the material properties are observed. A DFT calculation employing the PBE exchange-correlation functional is expected to predict the lattice parameters of a metal structure with an accuracy of 1 % to 2 %. In the case of MRG the error is of the order of 5 % and 10 %, for the Mn_2Ga and the Mn_2RuGa, respectively. Such a difference in the crystal structure is significant if we consider its potential effects on the magnetic ordering. The calculated magnetization fails to reproduce the experimental values, even qualitatively, with the absence of compensation being the most striking difference.

In thin films the substrate can exert a strain on the crystal, which in turn leads to structural deformations. The deformation energies involved are large compared to the

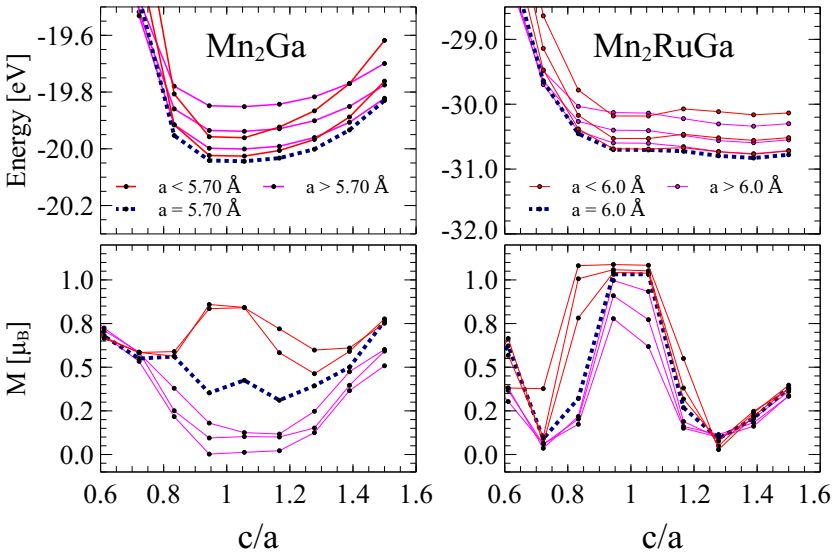

FIGURE 3.3: Total energy and magnetization for Mn_2Ga and Mn_2RuGa, top and bottom pannel respectively, as a function of c/a lattice parameter. For each energy curve the volume is constant, and determined by the cubic lattice constant a. The lowest energy curve, and the corresponding magnetization, is shown in blue.

magnetic energy. A change in the lattice parameters of the order of 1 % to 2 % can be sufficient to change the magnetic ordering in the material. In order to obtain a complete picture of the effects of the substrate strain we have performed a detailed scan over the volume and c/a parameter for the ferrimagnetic ground state. The results for Mn_2Ga and Mn_2RuGa are shown in figure 3.3. The structures with the lowest energy correspond to the ones reported in table 3.1. We find the ground state crystal structure of Mn_2Ga to be cubic, with a lattice parameter of 5.7 Å. For the Mn_2RuGa the lowest energy structure is tetragonal, with $c/a = 1.4$ and an in-plane lattice constant $a^{||} = 5.36$ Å. Mn_2Ga exhibits a transition from a high- to a low-magnetization state as the volume is increased. The magnetization of the ground state is located in the middle of the high-to-low transition with a magnetic moment $m = 0.4\,\mu_B\ f.u.^{-1}$. For the experimentally reported structure Mn_2Ga should become non-magnetic, i.e. the sublattice magnetizatic moments should become equal. This is in a stark constrast with the experimentally measured magnetization, which ranges from $1.0\,\mu_B\ f.u.^{-1}$ to $1.5\,\mu_B\ f.u.^{-1}$, depending on the film growth conditions. In case of Mn_2RuGa we find the magnetization to be almost entirely quenched in the ground state. For the cubic structure the magnetization is roughly $1.0\,\mu_B\ f.u.^{-1}$. This value is larger than the one experimentally measured, but in a good agreement with the neutron-diffraction study of Mn_2RuGa by Hori et al. [177]. We note that the Mn_2RuGa cubic structure, with a lattice constant $a = 6.0$ Å, is almost degenerate in energy to the tetragonal one, but agrees much better with the experimentally reported crystal structure. These results clearly indicate that the experimental Mn_2Ga films are not well described by the theory, as we find large differences in both the crystal structure and the magnetization values.

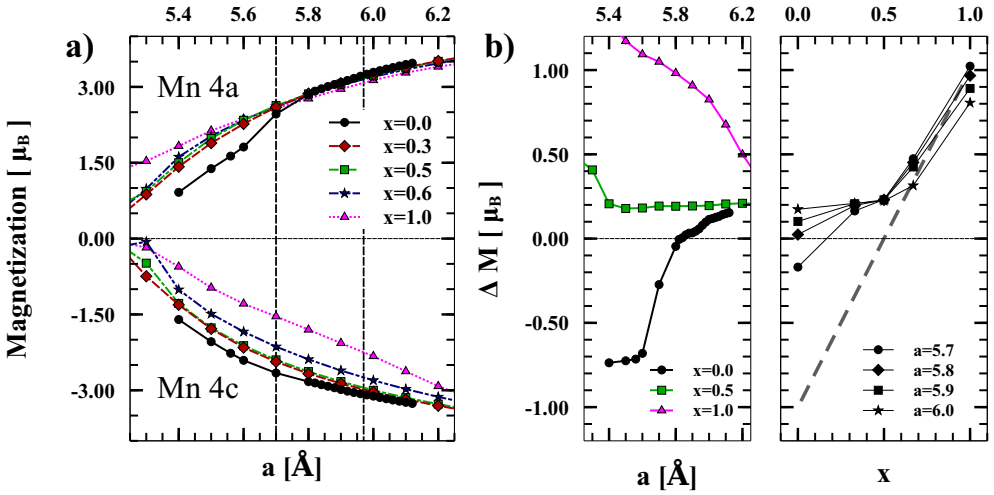

FIGURE 3.4: Magnetism in cubic Mn_2Ru_xGa compounds. a) Site projected magnetic moment for the two Mn sites, 4a and 4c, as a function of the lattice constant, a, and the Ru fraction, x. The vertical lines denote the Mn_2Ga lattice constant as determined by the theory, 5.7 Å, and the experiment, 5.97 Å. b) Difference of the Mn magnetic moments, ΔM, as a function of the lattice constant (left) and the Ru concentration (right). The dashed line shows a linear interpolaton of the experimentally observed magnetization slope.

The results for Mn_2RuGa can be understood if we assume that a (nearly) cubic metastable crystal structure forms in the presence of the substrate strain. The discrepancy in the magnetic moment may be due to the presence of the antiferromagnetic Ru_2MnGa phase, which has a similar lattice constant to Mn_2RuGa [177]. The presence of this phase would lead to an incorrect normalization of the magnetic moment in the experimental measurements, namely the Mn_2RuGa magnetization would be underestimated. We defer further discussion of the relative phase stability of Mn-Ru-Ga alloys to the next section.

Now we turn our attention to the magnetization compensation mechanism in ideal MRG compounds. In figure 3.4 we show the site projected magnetization for the two Mn sites, 4a and 4c. We look at cubic compounds of different volumes. The magnetic moment of the 4a site is found to be invariant with respect to the the Ru doping and decreases smoothly as the volume is decreased. The 4c site shows the same variation of the moment with respect to the volume but, unlike the 4a site, it is affected by the hybridization with Ru atoms. We find that increasing the Ru content decreases the 4c-moment for all volumes, which leads to an overall increase of the MRG magnetization. The rate of the magnetization change is volume dependent, as shown in figure 3.4-b. The magnetization compensation is observerved for lattice constants below $a = 5.8$ Å, but not within the range of experimentally measured values. Therefore, the theoretical results are unable to explain the experimentally observed compensation at $x \approx 0.6$. The calculated magnetization slope, with respect to the Ru doping, significantly differs from the experimental value. In particular, for the compounds having a Ru fraction $x \leq 0.5$ (see figure 3.4-b).

The microscopic origin of the magnetization compensation mechanism can be understood by inspecting the electronic structure of the MRG compounds, shown in figure 3.5. For each MRG composition we show the spin-resolved band structure and the corresponding density of states (DOS), projected over the Mn and Ru atoms.

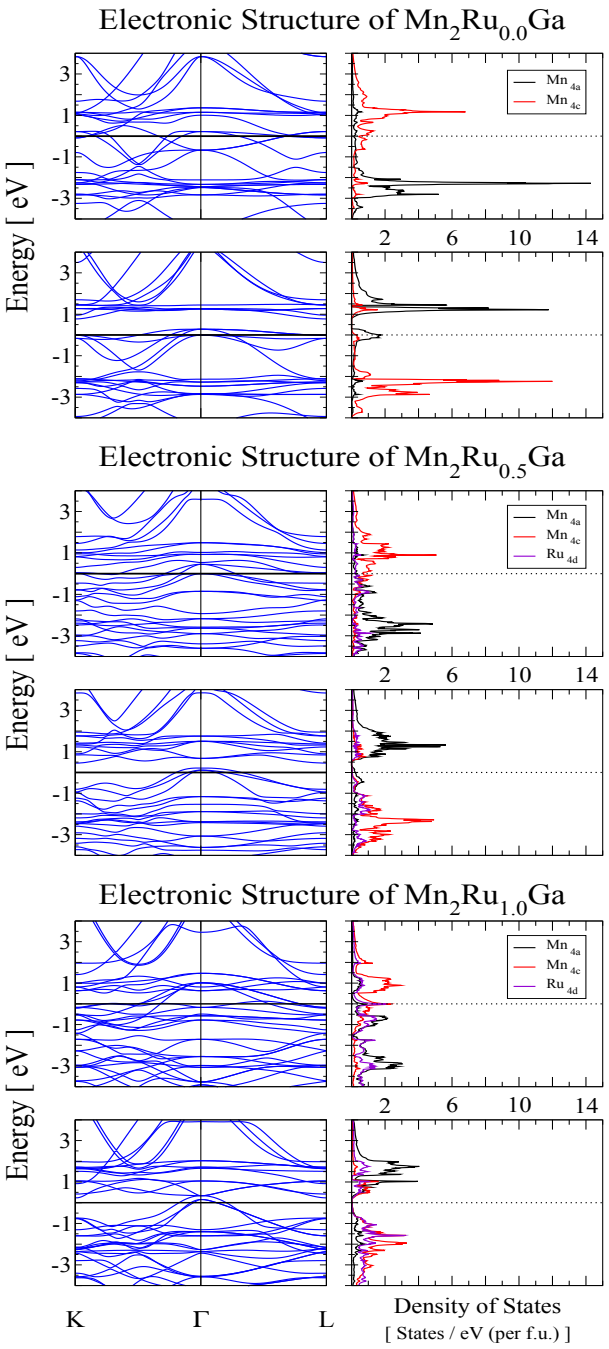

FIGURE 3.5: Electronic structure for stoichiometric Mn_2Ru_xGa compounds at the experimental lattice constant, 5.97 Å. The unit cell is cubic. The spin up/down channels are shown in the upper/lower panel, respectively.

Starting from Mn_2Ga, we see the strong exchange splitting in the projected density of states (PDOS) of the Mn atoms. The tails of the PDOS that cross the Fermi level hybridize the two Mn atoms. The hybridization can be infered from the single-particle energy bands, which have a large bandwidth and connect the PDOS of the two Mn atoms. This is most clearly seen for the minority DOS. The addition of Ru at the 4d site introduces additional electronic states around the Fermi level. The Ru states overlap with states from both Mn atoms. However, the hybridization with the 4c site is stronger. This is understandable as Ru coordinates octahedrally the nearest-neighbour Mn-4c site and tetrahedrally the next-nearest-neighbor Mn-4a site. The Mn-Ru hybridization increases the occupation of the Mn 4c spin-up band and thus increases the net magnetization. The local magnetization of the 4c site is reduced. This is consistent with the magnetization variation due to the Ru doping described previously (cf. figure 3.4-a).

We find that stoichiometric MRG compounds are not half-metallic. This is valid for the entire range of compositions. The non-linear variation of the MRG magnetization with

Ru doping arises due to the simultaneous filling of both spin-bands. Consequently, the magnetization increase is not proportional to the valence of the compound. In other words the magnetization of stoichiometric MRG compounds does not follow the generalized Slater-Pauling rule. This is particularly pronounced in compounds having a low Ru content, where the minority spin band has appreciable density of states close to the Fermi level. For compounds with $x > 0.5$ the majority of the spin down states is already occupied and, in turn, the magnetization slope becomes proportional to the Ru electron doping, $\approx 2\,e$/Ru. In conclusion, we find that stoichiometric MRG compounds do not follow the generalized Slater-Pauling rule due to the absence of half-metallicity.

Finally we look at the transport properties of stoichiometric MRG compounds, calculated by solving the Boltzman transport equations in the constant relaxation time approximation, as discussed earlier. In figure 3.6 we show the spin resolved resistivity and the spin polarization for three MRG compositions, $x = 1/3$, $2/3$ and 1.0, i.e. for x below, close to, and above the experimentally reported magnetization compensation point, $x_c \approx 0.6$. The spin resolved resistivity is shown as a function of energy, E, in the left panel (black). The resistivity at a given energy, $\rho(E)$, is calculated by rigidly shifting the Fermi level so that the bandstructure never changes. For each compound we place the reference Fermi level, E_f, at the energy for which the occupied DOS integrates to the correct number of valence electrons, N_V. The total density of states, interpolated by the BoltzTraP code, is shown in blue. In the right panel we show the transport spin polarization, calculated using the equation

$$P(E) = \frac{\sigma_\uparrow(E) - \sigma_\downarrow(E)}{\sigma_\uparrow(E) + \sigma_\downarrow(E)}, \tag{3.2}$$

where $\sigma_s(E) = \rho_s^{-1}(E)$ is the conductivity of the spin channel $s \in \{\uparrow, \downarrow\}$.

For all the compositions considered MRG is devoid of half-metallicity. The resistivity of the spin up channel is roughly constant with respect to the Fermi level position and the composition. This indicates a strong metallic character of the spin up band. For compounds with $x < 0.6$ the spin down band exhibits a divergence in the resistivity approximately 0.4 eV above the Fermi level. This corresponds to the position of the true spin gap. For $x = 1.0$ the divergence of the resistivity is less pronounced and located $\approx$0.2 eV above the Fermi level. The shift of the spin gap can be understood from figure 3.5, where we see that the gap in the minority spin band is practically closed for $x = 1.0$. The absence of spin gap at the Fermi level is reflected in a low spin polarization, which we find to be $\approx$40 % for $x = 2/3$. At the true position of the spin gap we find the spin polarization to be 100 %. As the spin gap gradually approaches the Fermi level with the addition of Ru, the spin polarization is increased going from a low Ru content towards $x = 1.0$. Based on our calculations, we estimate the spin polarization of reallistic

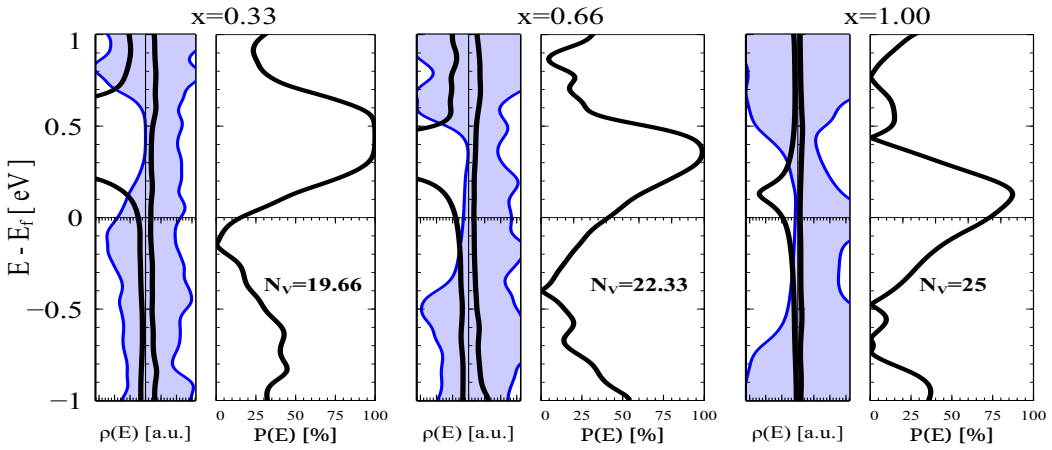

FIGURE 3.6: Transport properties for three stoichiometric Mn_2Ru_xGa compositions, x = 0.33, 0.66 and 1.0. Fermi level position, E_f, is determined by the valence electron number, N_V. *Left panel* - Spin resolved resistivity (black) and the corresponding DOS (blue). The data corespondingto the spin down channel is represented with negative values. *Right panel* - Calculated transport spin polarization.

MRG thin films to range from 30 % to 40 %, close to the compensation point, but not higher. For the experimentally characterised MRG thin films the spin polarization is expected to be in excess of 54 % [32]. We therefore conclude that our DFT calculations can not explain quantitatively the experimentally measured spin polarization.

Summary - In the previous discussion we have looked closely into the properties of stoichiometric MRG compounds. We have determined the ground state crystal structures and identified the ferrimagnetic ordering of the symmetry inequivalent Mn atoms as the ground-state spin configuration. The calculated ground state crystal lattice parameters were found to be smaller than the experimentally reported values, with a discrepancy, which can not be attributed to the standard DFT error. We have studied the magnetization compensation mechanism in detail to find that the theoretically calculated magnetization of MRG compounds strongly disagrees with the experimental measurements. Most notably, we do not observe the magnetization compensation. We find only a non-linear variation of the magnetization with the increasing Ru concentration. We have shown that this deviation from the generalized Slater-Pauling rule is due to the absence of half-metallicity. The latter was also confirmed by the transport calculations. The calculated transport spin polarization was found to be lower than the experimental value, by a factor of 1.5. From our result we can conclude that the DFT description of MRG films lacks vital ingredients necessary for an accurate description of MRG thin films. We have found that the strain alone can not account for these discrepancies, nor explain why the magnetization compensation occurs near $x < 0.6$ in the films. This leaves the site and/or the substitutional disorder as the main suspects for the aforementioned disagreement. The origin, the nature and the role of defects in MRG compounds

will be studied in the next section.

3.3 Relative Stability of the Competing Mn-Ru-Ga Heusler Phases: a High-Throughput Approach

In the previous section we have shown that the properties of real MRG thin films can be only partially explained by the properties of stoichiometric compounds. We are led to the conclusion that the films must be affected by a disorder of some sort. We have shown that the stoichiometric compounds lack both key features of a compensated ferrimagnetic half-metal (CFHM), namely the half-metallicity and the magnetization compensation are absent. The presence of disorder thus seems to be crucial for the emergence of these properties. Based on our results, disorder is expected to be stronger for compounds with a low Ru fraction, as the discrepancy between the experiment and the theory increases with the decreasing Ru concentration (cf. figure 3.4-b). For the compounds close to the Mn_2RuGa composition we anticipate that the films have the Mn:Ru ratio different from 2:1, going towards 1:1. This is inferred from the neutron diffraction study of Hori et al. [177]. They have found that the stoichiometric Mn_2RuGa has a magnetization of $\approx 1\,\mu_B\, f.u.^{-1}$, in good agreement with our DFT results. We attribute the formation of a meta-stable, nearly cubic, structure of the nominal Mn_2RuGa films to the substrate strain. We find that the properties of the films with a low Ru fraction can not be rationalized based on the DFT results presented so far. In this section we present the results of a high-throughput search for competing Mn-Ru-Ga phases having a (half-)Heusler crystal structure. We identify the energetically most favourable phases and then simulate the corresponding defects using supercells. The results allow us to qualitatively describe the effects of the disorder on the properties of realistic MRG alloys.

Limited computational resources prevent an *ab initio* description of the defect formation process in real materials. However, *ab initio* DFT calculations offer accurate total energies, which can in turn be used to compare the relative stability of compounds with different stoichiometry. We calculate the enthalpy of formation, ΔH, which is defined as a difference between the total energy of a compound and the sum of the energies of the most stable elemental, i.e. single element, phases of the constituents. For example, the enthalpy of formation of a hypothetical binary A_mB_n compound is calculated as

$$\Delta H = \frac{1}{N}\left[E\left(A_mB_n\right) - m \cdot E\left(A\right) - n \cdot \left(B\right)\right], \tag{3.3}$$

where $N = m + n$ is the number atoms in the formula unit for which the energy was calculated. The enthalpy of formation defined using the equation (3.3) measures the

#	4*a*	4*c*	4*b*	4*d*	ΔH [eV]	M [μ_B]	c/a	Vol [Å^3]
1	Mn	Ru	Ga	Ru	-0.28	2.18	1.0	56.33
2	Ga	Mn	Ga	Ru	-0.25	2.93	1.0	55.83
3	Mn	Mn	Ga	Ru	-0.24	0.07	1.2	55.88
4	Mn	Mn	Ga	Ru	-0.13	4.66	1.0	55.01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	Ga	Mn	Ga	–	0.01	3.14	1.0	52.01
12	Ru	Mn	Ga	–	0.09	0.19	1.0	44.70
13	Ru	Mn	Ga	Ru	0.10	4.44	1.0	59.20
14	–	Mn	Ga	Ru	0.17	4.50	1.0	49.60
15	Mn	Mn	Ga	–	0.18	0.47	1.0	46.54
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

TABLE 3.2: Calculated enthalpies of formation, ΔH, for the most stable competing Heusler phases of the 1221 structural and magnetic Mn-Ru-Ga cells investigated. The configurations investigated are limited to the primitive 4-atom Heusler cell (3-atom for half-Heuslers).

stability of a compound with respect to the decomposition into its constituents. In order to evaluate the most stable structure of a given stoichiometry one needs to minimize ΔH with respect to all possible decomposition paths. This makes the full stability analysis a formidable task. Here we are interested only in the relative stability, therefore it is sufficient to know ΔH for each compound considered. We have calculated the ΔH for 1221 structural and magnetic Mn-Ru-Ga Heusler cells [31]. The phases which are excluded by the experimental analysis, e.g. compounds having Ga-Ga and Ru-Ru nearest neighbors [32], were not included in the final stability analysis. We have also ignored the compositions where more than one atom per formula unit is replaced, e.g. we do not include the compounds like Ga_2Ru in the stability analysis. This would correspond to formation of multiple defects, which we believe is less likely to occur. We will later show that this approximation is sufficient for the understanding of real MRG films.

Our results are summarized in table 3.2. The most stable stoichiometry is Ru_2MnGa, followed by Ga_2MnRu and Mn_2RuGa. These results are in agreement with our hypothesis on the formation of an antiferromagnetic Ru_2MnGa phase in Ru rich MRG films. The formation of Ga_2MnRu is not likely due to the difference in stoichiometry. We note that Ru_2MnGa reported in table 3.2 is ferromagnetic. The apparent discrepancy comes from the fact that in the high-throughput calculations only primitive unit cells are considered. Thus Ru_2MnGa is ferromagnetically ordered by construction. DFT yields a positive enthalpy of formation for all half-Heusler Mn-Ru-Ga phases, i.e it predicts them as unstable against decomposition. We therefore conclude that the corresponding MRG films should be meta-stable structures. We observe that the addition of Ru, in general, tends to increase the stability of the compounds. The most stable half-Heusler stoichiometry is Ga_2Mn, followed by the RuMnGa. The Mn_2Ga composition, which we

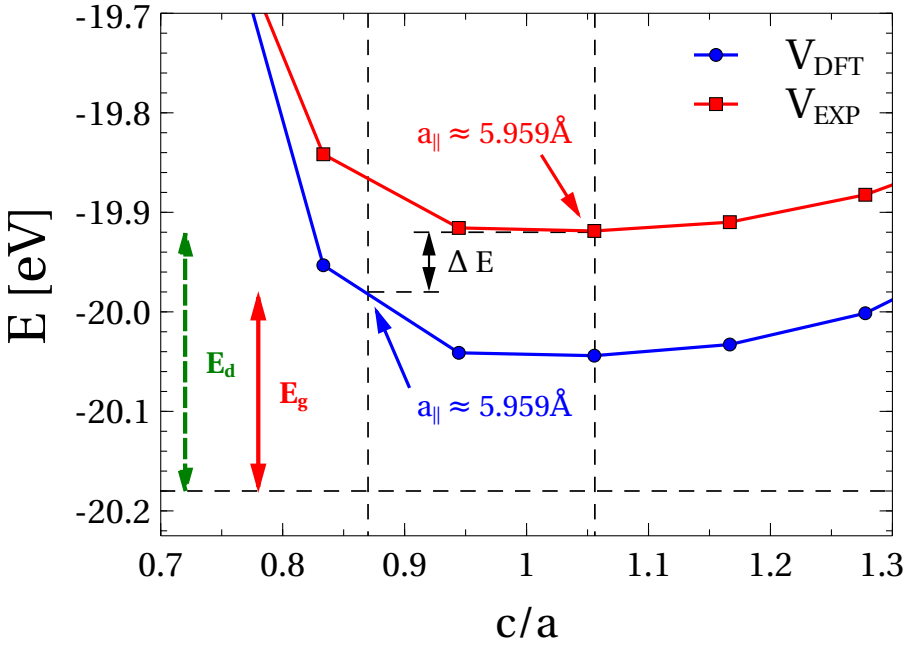

FIGURE 3.7: Isovolumetric energy profiles for stoichiometric Mn_2Ga calculated at the observed experimental and the relaxed DFT primitive unit cell volume, $V_{EXP} \approx 52.9$ Å and $V_{DFT} \approx 46.4$ Å, respectively. The theoretically expected c/a ratios are denoted by the vertical dashed lines. ΔE is the energy difference between the experimental and the relaxed DFT structure. E_d is the energy gain due to the formation of Ga defects. The net energy gain by forming the defects and taking the experimental structure is denoted by E_g.

have considered as a representative of the nominal Mn_2Ga films is predicted to be much higher in energy than the aforementioned, Ga and Ru rich, phases. It is therefore very likely that such defects occur in the real MRG films.

The high-throughput analysis of the relative phase stability of Mn-Ru-Ga compounds reveals the Mn-Ga and the Mn-Ru substitutions as the most likely defects to occur in the MRG films. In order to understand if the occurrence of these defects is more stable than the formation of the stoichiometric MRG phase, we construct supercells for the two MRG compositions, x =0 and 1/3, substitute a single Mn atom at the crystallographic positions 4a (or 4c) with Ga or Ru, and then compare the corresponding enthalpies of formation.

We find that the Mn-Ga substitution at the 4a site is energetically most favourable, for both compositions. For $x = 0.0$ we find $\Delta H = -0.26\,\text{eV}\,f.u.^{-1}$ and for $x \approx 0.3$ the $\Delta H = -0.15\,\text{eV}\,f.u.^{-1}$. In both cases the most favorable Ru substitution does not improve the stability of the compound with respect to the stoichiometric composition. Increasing Ru content in the supercell promotes the Mn-Ru defect stability, which makes the Mn-Ga substitution favourable only for the compounds with a low Ru fraction. In figure 3.7 we illustrate the results obtained for the Mn_2Ga supercell. In the absence of

Ga defects the lowest energy structure has the theoretically predicted volume, $V_{DFT} \approx 46.4\,\text{Å}$, and $c/a \approx 0.87$. The in-plane lattice constant of the MRG films is pinned by the MgO substrate to 5.959 Å, which leads to the tetragonal distortion. Even after taking into account the substrate strain, the theoretical structure is more stable than the experimentally observed one, by $\Delta E \approx 60\,\text{meV}\,f.u.^{-1}$. Substituting 1/3 of the Mn 4a sites by Ga and taking the experimental structure results in a net energy gain of $200\,\text{meV}\,f.u.^{-1}$, compared to the theoretically predicted unit cell. The formation of defects is therefore energetically more favourable than having a stoichiometric MRG phase.

3.4 The Effects of the Mn-Ga Substitution on the Properties of MRG Films

The high-throughput analysis presented in the previous section has shown that defect formation can be expected in real MRG films. We have found the antiferromagnetic Ru_2MnGa phase to be likely to form in Ru-rich MRG compounds. The existence of this phase explains the observed discrepancy in the magnetization between the Ru-rich MRG films and the theoretically predicted Mn_2RuGa phase. The DFT description of Mn_2RuGa, presented in Section 3.2, can thus be considered satisfactory. In Ru-poor films we find that the formation of Ga defects, i.e. the occurrence of Mn-Ga substitutions, plays an important role. These defects are intimately related to the magnetic, the structural, and the transport properties of the material. In this section we will investigate how these defects affect the properties of MRG films.

A direct consequence of the defect formation, i.e. the substitution of a magnetic 4a Mn atom by a non-magnetic Ga, is the change in magnetization. The substitution lowers the net magnetic moment of the supercell. From DFT calculations we find that the overall change is in the range of $1.9\,\mu_B$ to $2.6\,\mu_B$ per substitution. The exact value depends on the amount of Ru ions in the supercell. The magnetization decrease is almost entirely due to the magnetic moment of a Mn 4c atom, which is left uncompensated after the substitution. For compounds with a low Ru content, increasing the concentration of Ga defects will make the average moment of the 4c site greater than that of the 4a site, which is necessary to facilitate the compensation. We recall that in the stoichiometric MRG compounds the magnetization never changes its direction, i.e. it is determined by the Mn 4a spin sublattice. We take this as the "positive" direction. We have previously shown that the addition of Ru leads to an increase of the net magnetization by reducing the magnetic moment on the 4c site. Achieving the compensation solely by adding Ru is thus impossible. The occurrence of Mn-Ga substitution has just the opposite effect,

namely it reduces the effective magnetic moment of the 4a site. We therefore find it necessary to have a sufficiently high concentration of Ga defects in Ru-poor films, so that the net magnetization is determined by the 4c sublattice. The compensation can then be achieved by adding the correct amount of Ru. This implies that the compensation point, x_c, is determined by the defect concentration and not by the valence electron count, as proposed by Kurt et al. [32]. It follows that the position of the compensation point can be engineered by modifying the film thickness, i.e. by controlling the strain in the film.

In addition to the magnetization change, we find that the Mn-Ga substitution introduces electron doping, which has a direct consequence on the structural and the transport properties of the MRG films. Removing one Mn removes 5 valence electrons and 5 occupied d states. At the same time Ga introduces 1 p electron and 0 d states. This results in electronic doping of 1 e per substitution. It is known that the main group element in Heusler alloys reduces the effective number of d electrons [25], which are found to fill the low lying p states. Since Ga contributes 1 p electron and does not introduce new d states, the effective number of d electrons is increased by 1. We expect the doping to result in an excess electronic pressure, which will in turn elongate the tetragonal axis of the crystal. The in-plane lattice constant will remain unchanged, since it is pinned by the substrate. The transport properties are also expected to change. The added electrons will shift the Fermi level towards the spin gap and, consequently, the spin polarization should increase.

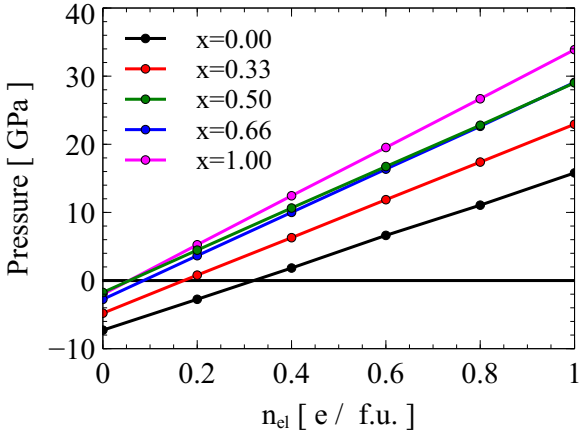

FIGURE 3.8: Calculated pressure for cubic MRG compounds at experimental (in-plane) lattice constant, $a = 5.97$ Å, as a function of electronic doping, n_{el}.

In order to understand how the electron doping affects the crystal structure we perform DFT calculations with excess electrons in the unit cell. The additional charge is compensated by the homogeneous positive background. The calculated pressure on the DFT unit cell, as a function of the excess electron concentration, n_{el}, is shown in figure 3.8.

From the figure we see that the pressure in a MRG unit cell of a fixed volume increases linearly with the number of excess electrons. More notably, in the absence of doping the pressure in the Ru-poor MRG unit cells is negative, i.e. the pressure is compressive. We can understand this if we recall that for these compositions DFT predicts equilibrium volumes, which are smaller than the experimental ones (cf. table 3.1). Electron doping thus provides the excess pressure needed for the materials to take on the experimental structure. We can also see that the Mn-Ga substitutions will occur in a larger proportion in Ru-poor films. For the compounds with $x \geq 0.5$ the amount of defects becomes substantially smaller, with $n_{\mathrm{el}} \approx 0.1\, e\, f.u.^{-1}$. This is consistent with the observed discrepancies in the magnetization between theory and experiment, further supporting the conclusion that the two phenomena have a common origin. Both phenomena are due to the Mn-Ga substitutions.

The pressure dependence on the electron doping allows us to estimate the number of defects. We note here that the electron doping, n_{el}, and the defect concentration may be equated, since we have already established that a Mn-Ga substitution introduces $1\, e$. The mechanical stability criteria then allows us to estimate n_{el}. The stability implies that the net pressure on the unit cell is zero. MRG compounds can be made stable at the experimental structure by introducing an additional pressure term, which we attribute entirely to the excess electrons. In other words, we assume that the difference between the DFT and the experimental volume (V_{DFT} and V_{EXP}, respectively) is due to the excess electron pressure, P_{el}. Following the DFT results shown in figure 3.8, we can express P_{el} as

$$P_{\mathrm{el}}(n_{\mathrm{el}}) = S_0 \cdot n_{\mathrm{el}} + P_{\mathrm{el}}(0), \tag{3.4}$$

where S_0 is the rate of change of the excess pressure with electron doping. The change in the pressure, $\Delta P = P_{\mathrm{el}}(n_{\mathrm{el}}) - P_{\mathrm{el}}(0)$, can then be related to the change in the volume, $\Delta V = V_{\mathrm{EXP}} - V_{\mathrm{DFT}}$, via the definition of the bulk modulus

$$B_0 = -V_0 \left(\frac{\Delta P}{\Delta V}\right)_{V=V_0}. \tag{3.5}$$

We estimate B_0 for each compound at the relaxed theoretical structure, i.e. at V_{DFT}, by fitting the Murnaghan equation of state [197, 198]. This approximation takes the compound volume to be a linear function of the electron doping. In doing so the effects of the chemical substitutions on the bulk modulus are neglected. Finally, we can express the electron doping as

$$n_{\mathrm{el}} = \frac{B_0}{S_0}\left[\left(\frac{c}{a}\right)_{\mathrm{exp}} \cdot \left(\frac{a_{\mathrm{exp}}}{a_0}\right)^3 - 1\right] \quad , \tag{3.6}$$

x	a_0 [Å]	B_0 [kbar]	S_0 [kbar e^{-1}]	P [kbar]	n_{el} [e]
0.00	5.69	596	231	99	0.43
0.33	5.88	2760	277	157	0.57
0.50	5.93	2140	307	65	0.21
0.66	5.95	3741	319	75	0.24
1.00	5.97	1496	359	15	0.04

TABLE 3.3: The electron doping estimate obtained using equation (3.6). The in-plane lattice constant, $a_{exp} = 5.97$ Å, and $c/a = 1.02$ are taken as parameters.

where a_{exp} and a_0 correspond to the experimental in-plane lattice constant and the relaxed theoretical lattice constant, respectively.

Using equation 3.6 and the experimental crystal structures we can estimate the defect concentration for each MRG compound. The results are summarised in table 3.3. We find that the Mn:Ga ratio of Ru-poor MRG compounds should be in between 1.6:1 and 2:1. The departure from the nominal stoichiometry (2:1) increases as we approach the Mn_2Ga stoichiometry. The strain-pressure model, discussed above, suggests that this departure is due the strain exerted on the compound by the substrate. This conclusion is further supported by the dependence of the experimental c/a ratio on the film thickness [185, Sup.Mat.]

The estimate of the electron doping, $n_{el}(x)$, allows us to quantify the magnetization change of MRG compounds due to the presence of the doping. We construct a simple model where the DFT magnetic moment is "corrected" by an amount proportional to the defect concentration. We can express the expected (experimental) magnetic moment, $M_{EXP}(x)$, to be

$$M_{EXP}(x) = M_{DFT}(x) - 2 \cdot n_{el}(x)\,, \tag{3.7}$$

where M_{DFT} is the theoretically calculated magnetic moment for an electron doped stoichiometric Mn_2Ru_xGa compound. The effective slope, $dM_{EXP}/dn_{el} = 2\,\mu_B\,e^{-1}$, is deduced from the same DFT calculations we have previously used to determine the relative stability of the Mn-Ga defects (see section 3.3). We now use the electron doping estimates obtained from the mechanical equilibrium criteria, shown in table 3.3 to evaluate $M_{EXP}(x)$. The results are shown in figure 3.9. We find that the corrections, obtained using equation (3.7) match the experimental values very well. We therefore conclude that the discrepancies between the DFT data and the experiments can be reconciled if the formation of Ga defects is taken into account. The magnetization corrections, which follow from the strain-pressure model then serve as a confirmation that the origin and the role of the defects in nominal MRG films has been properly understood.

Finally, we look at the effect the excess electrons have on the transport properties of MRG compounds. We calculate the Fermi level spin polarization, via equation (3.2), for

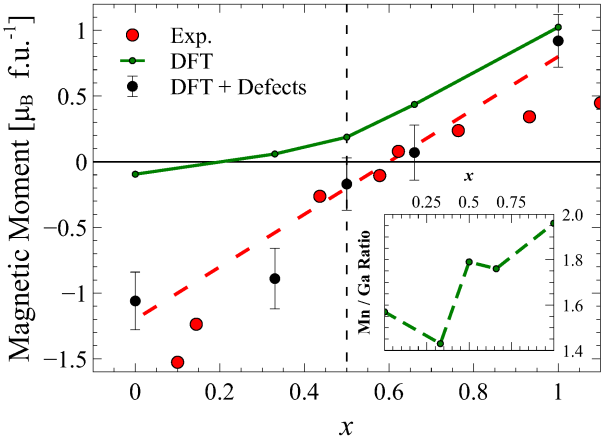

FIGURE 3.9: Magnetic moment of Mn_2Ru_xGa compounds. The DFT values calculated for stoichiometric compounds are shown by the green line. Correction to the DFT magnetization due to the formation of Ga defects is shown by the black points. The corresponding estimate of the Mn/Ga ratio is shown in the inset. The dashed line, equivalent to doping of $2\,e$ per Ru, is shown to guide the eye.

stoichiometric MRG compounds with excess electrons included in the DFT simulation. This allows us to estimate the number of excess electrons needed to achieve the half-metallicity and to estimate the spin polarization due to electron doping concentration derived from the strain-pressure model. The results, however, need to be interpreted with caution. The method employed is poorly suited to describe the transport in disordered materials. The change in the density of states due to Ga defects, which affects the Fermi level DOS the most, is unaccounted for. We find that the presence of excess Ga tends to close the spin gap by reducing the exchange splitting of the 4c Mn ions. In addition, the method is unable to correctly describe the conduction properties of electronic states close to the Fermi level. In small supercells the defects unavoidably form a crystal structure, rather than being distributed randomly. This in turn opens artificial conduction paths and, in general, increases the mobility of electrons occupying bands associated with the defects. In small supercells, both effects tend to increase the Fermi level conductivity and decrease the spin polarization. In reality, the two effects compete with each other, since the disorder decreases the mobility. Therefore, to properly address the transport properties one would need to construct a very large supercell and ensure that the spatial distribution of the defects is "realistic". This makes the approach impractical. In contrast, the magnetism in disordered MRG compounds can be described in a satisfactory manner using relatively small supercells. The difference is that the magnetic moment of a transition metal is localized around the atom and thus insensitive to the details of translational symmetry breaking introduced by the defects. The electronic transport on the other hand, which depends entirely on the nature of the Bloch states, is very sensitive to such details.

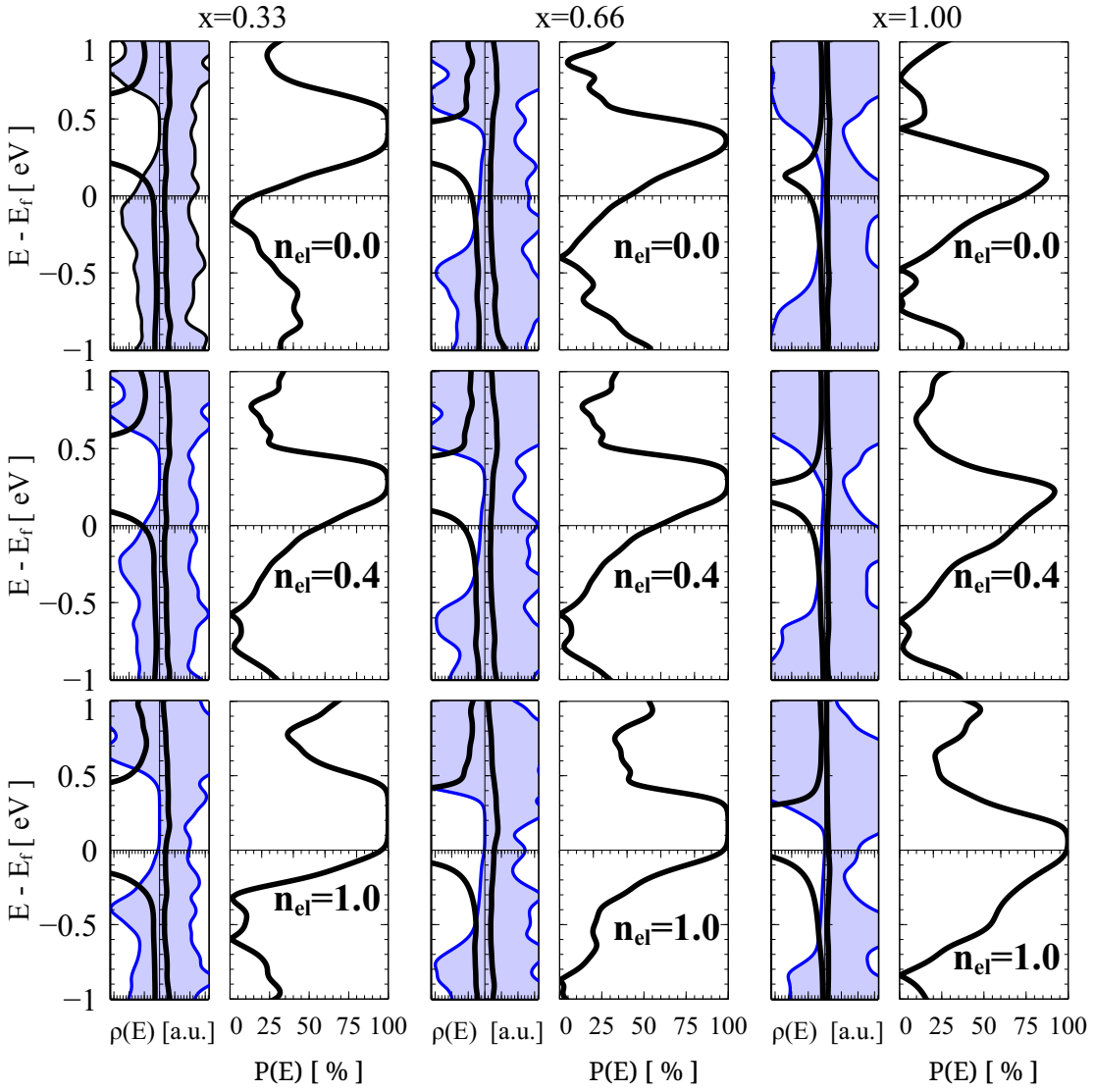

FIGURE 3.10: The spin resolved resistivity, $\rho(E)$, and the corresponding spin polarization, $P(E)$, as a function of energy. A scaled DOS is shown as a shaded background. The position of the spin gap is marked the divergence of the resistivity in the majority spin channel.

In figure 3.10 we show the spin resolved DOS and the transport spin polarization for three MRG compositions. In pristine compounds we find that the spin gap location, with respect to the Fermi level, varies with the Ru content. It starts at 0.4 eV above the Fermi level for x =1/3 and ends up at 0.15 eV for x =1. This trend is reflected in the spin polarization, which increases from 20 % to 60 %, respectively. For each MRG composition the electron doping gradually shifts the Fermi level towards the spin gap. However, regardless of the initial position of the spin gap, the half-metallicity is realized only when $n_{\mathrm{el}} = 1\, e\, f.u.^{-1}$, for all compositions. In figure 3.11 we show the spin polarization as a function of the Ru fraction, x, in greater detail. We see that the spin polarization is significantly smaller than the experimental values in the

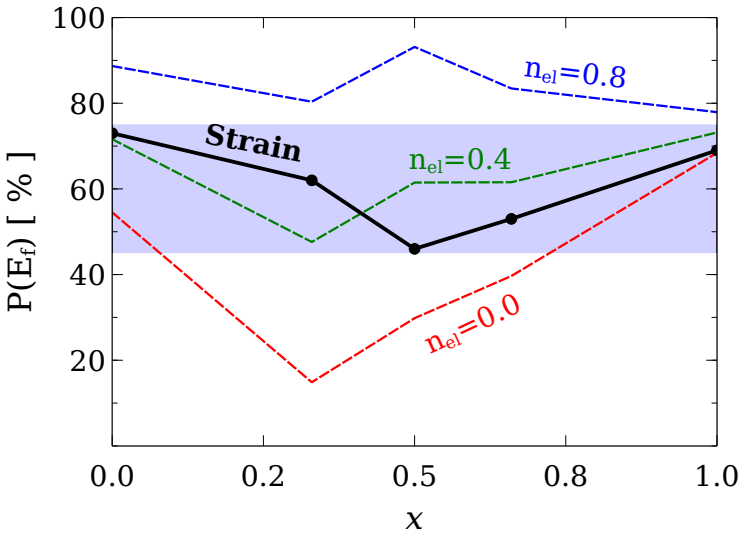

FIGURE 3.11: Calculated Mn_2Ru_xGa spin polarization, for different levels of electron doping, n_{el}. The strain induced polarization is shown by the thick black line. The experimentally reported range of values is shown by a shaded background.

absence of doping, $n_{\mathrm{el}} = 0\,e\,f.u.^{-1}$. A spin polarization curve corresponding to the doping concentrations derived from the strain-pressure model (see table 3.3) is shown by a thick black line. The estimated spin polarization falls right within the experimental range of values. We note that this can only be taken as a weak proof of the validity of the proposed strain-pressure model, since a fixed excess electron doping curve, namely $n_{\mathrm{el}} = 0.4\,e\,f.u.^{-1}$, exhibits an almost identical trend. Based only on the transport analysis these two models have to be regarded as equally valid. However, the Ga defect model is able to explain the structural and the magnetic properties, in addition to the transport, which makes it a more acceptable candidate. We therefore conclude that the inclusion of Ga defects leads to an overall improvement of the description of transport properties of realistic MRG compounds.

Our results suggest that a complete spin polarization can be achieved only with a doping of $1\,e\,f.u.^{-1}$, regardless of the Ru fraction. However, in thinner films the defect concentration should be larger than in thick films, since a greater fraction of the film is affected by the substrate strain. It is therefore a realistic possibility that the transport spin polarization in MRG thin films could be further increased by engineering the film thickness. In view of the previously discussed shortcomings of the employed method, namely the improper treatment of the disorder, we need to take these results with reserve. The shift of the Fermi level towards the spin gap is not the only effect that can increase the transport spin polarization, as defined by the equation (3.2). In stoichiometric MRG the only states at the Fermi level come from the Mn-Mn hybridization. The introduction of Ga defects, which break this bond, should decrease the mobility of electrons in both

spin bands. However, the already low density of states at the Fermi level is expected to result in poor conductivity of this spin channel. This would further enhance the spin polarization of the system. The competing effect is the gap closing due to the increased coordination of Mn by the non-magnetic Ga. This mechanism will provide extra DOS at the Fermi level in the spin down channel, increasing its mobility and reducing the overall spin polarization. To give an answer to the question, which of the two effects ultimately prevails, a large supercell study would be need. Demanding computational requirements prevent us to explore this subject further. Our study has shown that in thinner films the defect concentration should be larger than in thick films, since a greater fraction of the film is affected by the substrate strain. It is therefore a realistic possibility that the transport spin polarization in MRG thin films could be further increased by engineering the film thickness.

3.5 Conclusions

In this Chapter we have performed a thorough theoretical characterization of bulk MRG compounds with the hope of understanding if the MRG films can be used to realize a compensated ferrimagnetic half-metal. The absence of a demagnetizing field and a large spin polarization, comparable to values measured for known half-metallic materials [170, 199], make MRG films highly desirable for applications in antiferromagnetic spintronics. Previous theoretical work has dismissed the possibility of having half-metallicity in these materials on account of disorder [183]. However, the work has left much to be desired in terms of understanding the main features of these films.

In this work we have started off with a DFT analysis of stoichiometric MRG compounds. The structural, the magnetic, and the transport properties have been determined. Based on these results we have explained the microscopic mechanism of the magnetization compensation. The hybridization of the Ru with the Mn atoms at the 4c site was found to be driving the entire magnetization change. The magnetization compensation was missing for a wide range of volumes, including the experimental values. The magnetization showed appreciable volume dependence. Most notably, a non-linear variation with the Ru concentration was found for all volumes. We have traced the origin of this phenomenon to the absence of half-metallicity in stoichiometric compounds. By exploring the influence of the substrate strain on the compounds we have shown that the properties of the nominal Mn_2RuGa films can be understood in terms of the DFT results. In these films we have found that the strain most likely drives the material to take on the experimentally observed, nearly cubic, structure. The DFT ground state is almost degenerate in energy but yields a significantly larger c/a ratio. The discrepancy in the

magnetization has been attributed to the formation of an antiferromagnetic Ru_2MnGa phase. At the same time, the properties of the films with a Ru concentration $x \leq 0.5$ could not be explained by the results obtained for stoichiometric compounds. We have pointed at a number of discrepancies between the theoretically predicted structural, magnetic, and transport properties and the nominal MRG films. This has led us to the conclusion that stoichiometric composition can not be a good model for real films, in particular for the compounds with a low Ru content.

We have extended our study by performing a high-throughput search for competing MRG phases [31]. By combining experimental X-ray data and the relative stability analysis for these compounds, we have been able to single out the most energetically favourable competing phases. This information was used to construct a number of MRG supercells incorporating defects found by the high-througput analysis. The relative stability was once again used to find the Mn-Ga substitution at the 4c site as the energetically most favourable defect. The electronic doping introduced by the defects was related to the defect concentration by using the strain-pressure model, derived from the mechanical equilibrium criteria (cf. Section 3.4). We have found that the excess electrons introduce pressure, which in turn stabilizes the experimental structure. The pressure was then used to estimate the defect concentration. A simple model for the effective magnetic moment of MRG compounds, which establishes its dependence on the electron doping, was constructed to account for the presence of the defects. The parameters for the model were taken from the DFT calculations and the electron doping was estimated from the aforementioned strain-pressure model. The proposed effective magnetic moment model brings into agreement DFT and the experimentally measured values for the entire range of MRG compositions. More importantly, the model correctly captures the magnetic moment reduction on the 4a site, due to the defects, which in turn makes the compensation possible. The effective magnetic moment predicted by the model becomes compensated near $x = 0.6$, in good agreement with the experiments. We have thus shown that all the differences between DFT and the experiments have a common origin and can be reconciled by taking into account the Mn-Ga substitutions. The defects are nucleated by the substrate strain, which has its origin in the lattice mismatch between the relaxed MRG structure and the underlying MgO layer. The theoretical work presented here has predicted that MRG films should have a Mn:Ga ratio between 1.6 to 2.0, with the lower values corresponding to thinner, more strained, films. Both predictions were later confirmed by experimental data [185].

Finally, the transport properties of stoichiometric compounds in the presence of excess electrons have been calculated. The results show that electron doping is necessary to achieve full spin polarization, however this occurs only at $1\,e\,f.u.^{-1}$. These levels are much higher than what is expected to occur in nominal MRG films. Therefore, the

DFT results suggest that the MRG films remain normal metals even in the presence of the defects. The transport spin polarization, estimated for the doping levels extracted from the strain-pressure model, is found to reproduce the experimentally reported values (cf. figure 3.11). We have discussed the shortcomings of the method used to estimate the transport properties and concluded that an inadequate treatment of the spatial disorder of the defects is likely to yield a lower spin polarization. This is also expected on the basis of the experimental results, where we note that the surface spin polarization is measured. The latter is usually lower than the corresponding bulk values, which we calculate. The results presented here are thus likely to underestimate the actual experimental values. Calculations, which could provide us with a more reliable theoretical prediction are computationally too demanding, and remain to be made.

In conclusion, this study provides a comprehensive insight into the composition and properties of realistic MRG films. The high-throughput approach was utilized to extract information regarding the defect formation process, which has proved to be crucial for capturing the key properties of the films. We have shown that disorder facilitates the magnetization compensation and also provides means to engineer the properties of these materials. Disorder is expected to close the spin gap, however, the large spin polarization observed experimentally can be rationalized only if a transport spin gap is formed at the Fermi level due to the disorder. We therefore expect that transport half-metallicity can be achieved near $x = 0.7$ by further reducing the Mn:Ga ratio. To maximize the spin polarization, the position of the Fermi level can be finely tuned by varying the thickness of the films.

Chapter 4

Magnetism of Iron in Heusler Alloys: a Machine Learning Approach

4.1 Introduction

The structure-property relationship lies at the heart of every material science problem. From a macroscopic perspective, the material properties emerge from the interplay of the intrinsic features and the microstructure. The latter plays a crucial role in the engineering of technologically useful materials, but it is typically out of the scope of the *ab initio* investigations. The intrinsic properties are completely determined by the structure, i.e. by the chemical composition and the atomic positions. Here we restrict our attention to solids. The structure-property connection is implicit. It is given by the electronic structure of the material, which is nowadays routinely obtained using *ab initio* methods, such as density functional theory. From there one can, in principle, extract all the material properties, both for the ground state and in response to an external perturbation. Despite the tremendous progress made in the last 50 years, the *ab initio* studies are, for the most part, limited to small and highly symmetric structures (rarely containing more than 100 atoms). Studying realistic materials, where the microstructure plays an important role, or just a large number of small, idealized, structures, is still a time consuming and expensive process. However, developments in computer technology, both at the hardware and software level, are continuously reducing this barrier at an exponentially fast pace.

The futuristic research procedure is envisioned to be a loop where the experimental data are feed into the theoretical simulations and the predictions are then used to set up new

experiments [21]. The existing research methods are not well suited for this workflow, and new, data-oriented, tools and methods need to be developed. The theoretical data generation is addressed by the high-throughput computational approach [5], which today allows one a systematic screening of large portions of the chemical space. A number of databases, containing a range of material properties, are already available to be used as a starting point, e.g. for finding new materials tailored to a particular application, or for understanding the properties of entire classes of materials. The sheer amount of data precludes the classical, case study, approach to the analysis of the results. This has led to the introduction of the data-mining and the machine learning (ML) techniques into material science [4, 21]. Using the ML approach it is possible to classify materials, predict the value of a particular property, or to explore the hidden structure between the individual properties [23]. The ML algorithms are purposely built for this task and enable an objective and efficient analysis of multi-dimensional data. It is thus natural to seek ways to incorporate these methods into the standard material science toolbox.

The advantage of the ML approach is twofold. Firstly, ML algorithms are general and flexible, namely the same algorithm can be used to solve completely different problems. This eliminates the need for the development of computer codes specialized for a particular task. In addition, combining different types of data is possible, e.g. experimental and theoretical data, without the need to explicitly develop an internally self-consistent theory. Second, the ML approach is data-efficient. The data is never discarded, instead, it is continuously integrated and reused for solving new problems. However, this does not make the ML a "perfect" scientific tool. The ML approach relies on the assumption that the information of interest, e.g. the information about a particular material property, is already contained in a dataset, which is "rich enough". ML methods can then be used to perform the data analysis in order to extract the answer. ML algorithms are usually less accurate than *ab initio* methods and their statistical nature makes them less favourable if one is interested only in the properties of a particular material. The effort needed to obtain the training data and build a ML model also needs to be considered. It may happen that a direct calculation, although expensive, is still more affordable. The ML approach should, therefore, be understood as a complement to *ab initio* methods, rather than their replacement.

With the rising availability of materials data, two questions emerge: "*How can we use the availabe data?*" and "*Can we learn from the data?!*". ML techniques provide means to address these questions. They have already been successfully applied to the problems in material science [116–119, 200], demonstrating their potential. However, magnetic systems have not received much attention [35, 36]. Here we wish to demonstrate that ML methods can be successfully applied to problems in magnetism. Even more so, we wish to show that by using these methods one can study large datasets and gain a

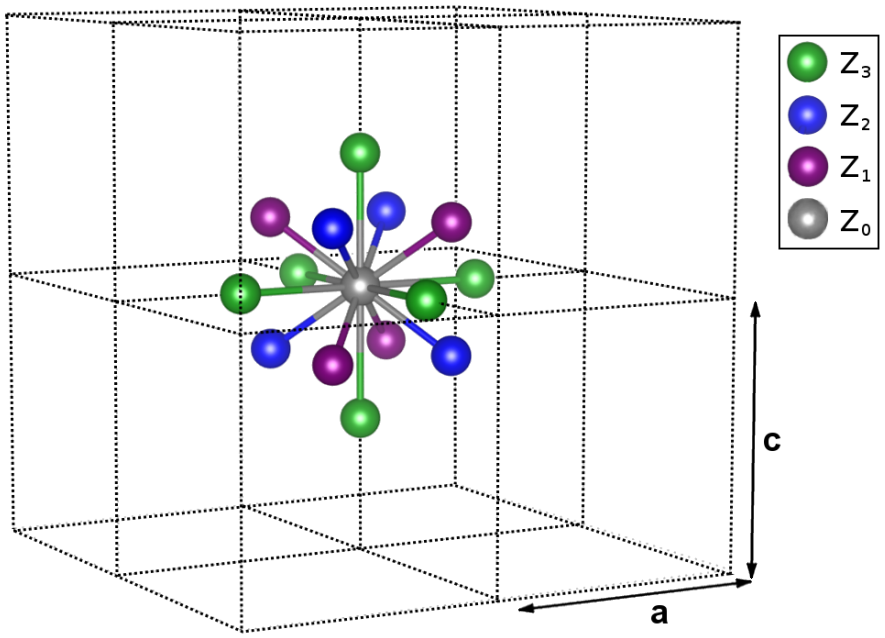

FIGURE 4.1: The local coordination of atomic sites in Heulser alloys. The neighbours of the central atom form two shells of different symmetry. Atoms belonging to the nearest neighbour shell, shown in blue and magenta, coordinate the central atom tetrahedrally. The next nearest neighbour shell is made out of 6 (green) atoms and has octahedral symmetry.

deeper understanding of the underlying physics. Starting from an in-house database of Heusler alloys, calculated at the DFT level, we build a ML model capable of predicting the magnetic moment of an atom in a given chemical environment. We study metallic Heusler alloys, where the magnetic moment originates from the valence d electrons. Therefore, the magnetism is of itinerant type. Nonetheless, the localized d orbitals and the charge neutrality make the $3d$ transition metal elements have a well defined (average) magnetic moment. The moment is modified by the presence of the neighbouring atoms. This can be due to, e.g. the crystal field, hybridization effects, or the electro-negativity of the neighbours. Therefore, the ML model is supposed to capture all these effects, in order to provide an accurate description of the magnetic moment. For practical applications the atoms from the middle of the $3d$ transition series are the most interesting. These have a large and stable magnetic moment, which is reflected in a high critical ordering temperature of their alloys. Of particular interest are Fe and Mn, which are both abundant and inexpensive. Mn offers the highest magnetic moment of all the transition metal atoms, usually $\approx 4\,\mu_{\mathrm{B}}$/atom. In practice, one expects the Mn moment to remain roughly constant, regardless of the choice of neighbours. For example, the stability of the Mn moment in the Heusler alloys has been know for a long time [192]. This makes Mn less interesting for the ML study we wish to perform. We are thus left with Fe as the only candidate.

In this work the environment of an atom is represented by a cluster, formed by the first two coordination shells, as shown in figure 4.1. The ML algorithm is trained to predict the magnetic moment of the central atom based on the list of the neighbours in the cluster. The ML model allows us to show that the information contained in the cluster is sufficient to define the magnetic moment. We then focus on the Fe-centered clusters, with the nearest neighbour shell composed of identical atoms. In other words, we consider the X_2FeZ Heusler alloy prototype as parent materials of the cluster. The model is then used to establish the relevant trends governing the magnetic moment. Once these are identified, we are able to select a representative set of materials to investigate their microscopic origin. We find that an interplay of an on-site Coulomb repulsion and hybridization with the nearest neighbours determines the magnetic moment. The thermodynamic stability analysis is performed to find that only the main group elements make viable candidates for the second coordination shell. Based on the established knowledge we propose the ideal prototype for an Fe-based Heusler alloy. We then use the ML model to characterize the magnetization for all the materials belonging to the proposed prototype. We thus demonstrate the potential of the ML approach for developing a fast, high-throughput, method for investigating magnetic materials.

4.2 Method

Here we develop and test various ML regression models for an investigation of magnetism in Heusler alloys. First we give an outline of the ML modelling process (section 4.2.1), we describe the general model building procedure, starting from the problem definition, and discuss all the steps involved in the process. In section 4.2.2 the discussion is specialized to ML models for magnetism. ML models for the volume and for the magnetic moment of Fe are introduced. We then discuss the specifics of the training procedure for each model, namely, the data selection, the scope of applicability, and the accuracy.

4.2.1 General Considerations

In the following discussion we will make a distinction between a ML model and a ML algorithm. We thus spend a moment to discuss the difference between the two. A ML algorithm can be viewed as a generic tool, which can be applied to an arbitrary data-based problem. In fact, as the name suggests, it is just an algorithm, which can be trained to make predictions for a given set of data. The algorithm *per se* is not specific to any application. In contrast, a ML model is obtained when a particular vector of input features is defined and paired with a ML algorithm. Here we assume that all free

parameters of the ML algorithm have been fixed in the training process. This gives us a "black box" model for the target property. The ML model can then be viewed as a function, which maps the input onto the output.

For the application it is crucial to estimate the accuracy of the model. This allows us to use the ML model judiciously. In practice two quality measures have to be distinguished. The first is the one used by the algorithm, while the second one is user defined. The algorithm uses an internal measure of the accuracy, the score, in order to facilitate the training. The definition of the score and its range of allowed values are algorithm dependent. This is typically a dimensionless quantity, which can not be used to compare different algorithms directly, but it is useful to fine-tune a particular algorithm. For example, the performance of the model can be estimated by using this number, e.g. the precision as a function of the training set size. For practical applications, we are primarily interested in a measure, which has a dimension and hence can be interpreted. For example for the regression problem one can measure the root mean square (RMS) error. This value is then reported as the precision of the ML model and can be used to compare different algorithms. In this work we will use the RMS error as user defined quality measure.

Building a ML model is a process, which involves a number of steps and considerations. The main benefit of the supervised learning approach is that a generic procedure can be defined. The main steps are shown in figure 4.2. Understanding the entire process can help us in deciding if the ML approach is the best way to solve the problem at hand and, if so, which algorithms and data should be used. In the following we expose and discuss the details involved in each step.

4.2.1.1 Defining the problem

The first step in building a ML model is to clearly define the problem that needs to be solved, namely to identify the target property and its nature. This is the starting point for all further considerations. If the property can be evaluated in some objective way, e.g. if it can be calculated or measured, then we can use supervised learning methods. As discussed previously, it is advantageous if we can formulate the problem in this way. Otherwise, the unsupervised methods need to be used. Here we will focus on the former. The range of the output variable also needs to be considered. If the target property takes a discrete set of values, or it can be reduced to that case, then one needs to solve a classification problem. For example, sometimes it is sufficient to know if a variable has a low or a high value. It is implicitly assumed here that it is possible to determine the class to which the variable belongs. In practice it may be too expensive to

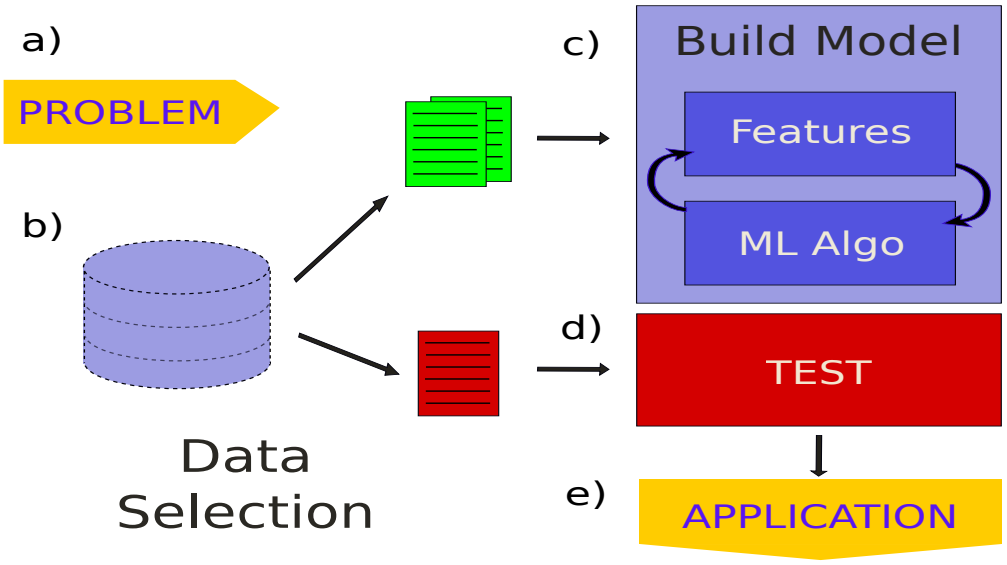

FIGURE 4.2: Supervised learning. The ML approach consists of 5 steps: **a)** defining the problem, **b)** selecting and pre-processing the data, **c)** building the ML model, **d)** testing the model, and, finally, **e)** applying the ML model to the problem of interest.

label all the samples, which would in turn preclude the supervised approach. Similarly, if we need to predict the value of a continuous property, e.g. the magnetic moment, then a regression problem needs to be solved. Therefore, the formulation of the problem decides the type of the algorithm, which will be used and the kind of data that needs to be collected in order to train the model. It may happen that the most natural approach is not the most practical one. One then needs to redefine the problem, or replace it with another, to make the computational task feasible. The statistical character of the ML approach also needs to be kept in mind. We can use the ML method only if the possibility of a large error, which may happen occasionally, is not critical for our objective. Otherwise a different approach must be used. We reiterate here that sometimes a direct calculation may be more preferable than the ML approach. This may be due to the required precision or just the computational cost involved. There is usually a significant overhead associated with constructing a ML model. All these aspects should be kept in mind. Carefully thinking about the ML problem can help us in deciding what is the best strategy to achieve our final objective.

4.2.1.2 Data Selection

The first step, the problem definition, is used to define what kind of data is needed. The data then needs to be obtained and put into a form, which is useful for building the ML model. This step, although conceptually straightforward, usually takes time, sometimes even more than building the model. The difficulty is that the data needs to be obtained from various sources, not all of which are easy to access and/or to process.

The data collection starts by identifying the sources, which contain useful data and deciding on the best strategy to extract it. Manual data collection is a slow and time consuming process. Automated computational techniques for performing this task are being actively developed, and a new field, commonly known as "data scraping" [201, 202], is starting to emerge. After the data has been extracted it is likely to be in an inhomogeneous format. The data then needs to be curated and brought into a standardized form. It is very fortunate if all the data is already found in a database. In this case only the standardization step might be needed. For example, consider the task of collecting experimental data. The data will be found scattered across different journals, arbitrarily formatted tables or graphs, etc. The extracted values will have been produced using different experimental techniques and expressed in various measurement units. The data curation then ensures that one can interpret what each data point means and where it comes from. This step already involves some data standardization, but only to the extent to make the data interpretable. The standardization implies performing a data transformation, which would allow us to directly compare different data points. In general, this may even be impossible. Such data may still be used for the ML, but the input to the algorithm will exhibit a strong "noise". Consequently, the precision of the final model will be lower. In practice one needs to find a balance between these two requirements.

The data preparation is tightly related to the model building procedure. As indicated in figure 4.2, the data from the database is passed to the next step in a structured, tabular form. The definition of the problem, shown in figure 4.2-a, sets the target property that needs to be evaluated by the ML model. The list of material properties in the database that can potentially be correlated with the target property is usually not large. The list can then be chosen *a priori*, using our knowledge of the underlying physics, or by performing a separate data exploration step. This may involve unsupervised learning methods, especially if our understanding of the problem is insufficient to make the decision. This list is later transformed and narrowed down to a useful set of features, in the model building step. The optimal transformation is unknown in advance and depends implicitly on the ML model. If a satisfactory model can not be obtained using the existing data one may need to come back to this step. Therefore, the data selection can also be viewed as a part of the model building procedure.

The dataset is then split into two parts, the training and the test set, indicated by the green and the red table in figure 4.2-b. The latter is crucial to establish the precision of the model in an objective way. The training set is typically larger, containing more than 50 % of the data. Ideally, the test set should be as large as possible. This is not difficult to achieve if a large amount of data is available. However, in practice the dataset sizes are often small and the test set size is then a compromise between being able to train

a model and being able to ascertain its precision. The subtle point regarding the data splitting step is that it needs to preserve the underlying data distribution. The purpose of the split is to ensure an objective test of the model. This implies that the data used for the training originates from the same "distribution", i.e. from the same generating process, as the data in the test set. In other words, the test data has to be representative. The split is performed with respect to the target property, for instance the value of the magnetic moment. This step is conceptually simple but it is best performed using a standard method, provided by the ML software package. The reason for this is that these methods have been thoroughly tested. In this way one avoids potential coding errors, which can lead to improperly split dataset. Such biased training set would likely yield a ML model, which performs poorly on the test data. Since the error is very unpredictable, it is very hard to detect and to correct it. Therefore, if possible, it is best to avoid such problems. After the split, the test dataset is stored in a separate file and used only after the ML model has been built.

4.2.1.3 Building the Model

The model building entails the feature selection and the algorithm training. These are two distinct steps but intimately connected in a self-consistency loop. The outcome of the training depends on the choice of the input features and in turn, the input is chosen such that the training score is maximized. Since each algorithm follows a different strategy to map the input data to the output, the optimal choice of the input features is algorithm dependent. The self-consistency loop starts by defining the input vector. This involves finding a suitable transformation of the available training data. Once the input vector is defined a ML algorithm is trained. For the training purposes a second data split, into the train and the validation set, is performed. This is needed since the precision of the algorithm has to be determined during the training. The difference between the validation and the test set lies in the fact that the information about the validation set is implicitly built into the ML model, while in the latter case it is not. The information feedback is mediated by human intervention, by using the validation score to improve the model. Therefore, the validation set can not be used to obtain an unbiased measure of the performance. This is why it is so important to keep the test data aside. Since the information about the validation set is already implicitly contained in the model it makes sense to use the entire set for the training. This is especially true when the dataset is small to begin with. When the train-validation set is small the training process may become sensitive to the details of the split. A single random split can be highly biased, since it may break the underlying distribution of the target property, by pure chance. Making a number of different splits reduces the probability of accidentally

selecting a biased dataset. The k-fold cross-validation is commonly used to avoid this shortcoming (cf. section 1.2.3). In this way a more robust ML model can be obtained and the dataset is used more efficiently.

The feature selection is the key step in building a ML model. However, a unique or standard way to perform such selection does not exist. Consequently, it is difficult to interpret the significance of the input features, which the ML model finds important, especially when the dimension of the input vector is large. Since in the material science we often want to gain a deeper understanding of the problem we are investigating, the difficulty to interpret the ML model may seem repulsing. However, this shortcoming may be circumvented if one takes a constructive approach to model building. One can start from a full physical description of the system and identify the key elements and the processes that define the quantity of interest. The data, which best represents these, is then included in the input vector. For each trial input the ML model is trained and the precision, evaluated on the test set, is measured. The model is then iteratively improved until a satisfactory precision is achieved. For example, all material properties are, in principle, defined by the structure and the composition of the material. This information should then be considered sufficient to describe any problem in the material science. However, it would be naive to expect that an arbitrary property can be described in this way. The core of the problem lies in the discrete nature of the chemical space. Very often, the internal task of a ML algorithm is to find a non-linear transformation, which makes the problem linear in some abstract hyperspace. The available data, e.g. the atomic numbers and the positions, are in general insufficient to define the necessary transformations. The ML algorithm will in general require less data to learn the rule the smoother the dependency between the input vector and the output is [122]. The input features should be chosen having this in mind. For example, the magnetic moment arises from the interaction of electronic degrees of freedom, which are not explicitly specified by the material structure. Therefore, one needs to resort to using auxiliary variables, which introduce more information about the underlying problem. Ideally, one would like to have linear dependencies between the input and the output. The feature selection is thus a slow, trial and error procedure, but it is often much faster than the typical throughput of the *ab initio* approach. The advantage is that one works with all the materials at once and the ML model obtained in this way should, ideally, include all the main trends that can be found in the dataset.

For the training one usually considers a whole range of ML algorithms, aimed at a specific ML task, e.g. the regression. A subset of algorithms, which perform best in the training are selected for testing. It is important to note that the quality of the ML algorithm is not only given by the score. If a ML algorithm has too many "internal degrees of freedom", the training scores will usually be high, however, they are false.

This is what is commonly known as the over-fitting. All ML algorithms come with a number of free parameters, which can be used to tune their behaviour. These are called the hyper-parameters. The purpose of the ML training step is to find an optimal set of parameters for a given algorithm. These parameters can be of different nature, but their common feature is that they define the capacity of the ML algorithm, i.e. its ability to represent different functions. This is directly related to the size of the space of admissible functions. If this space is too large the ML model can represent almost any function. Therefore, the task of training a ML algorithm is the one of maximizing the precision, while keeping the capacity of the algorithm minimal. This is known as the structural risk minimization [124]. A common technique that is used to ensure that the over-fitting does not occur is to plot the learning curve (see section 1.2.3). By plotting this curve we can directly see if the learning process is converging and to which accuracy. It allows us estimate if we have enough data and how much more would be needed to obtain improvement. This is of vital importance when one needs to estimate the "cost" of proceeding with the ML approach. If generating new data is expensive, it may be more appropriate to take a less precise but faster converging algorithm, i.e. the one that can be properly trained using a smaller dataset. The final decision depends on the specific requirements (precision, throughput, fidelity, etc.).

4.2.1.4 Testing the Model

The final step in the model building procedure is the verification. This step has already been mentioned a number of times and for a good reason. The test set is used to obtain an unbiased estimate of the ML precision. Since this set of data was never used for training one can estimate if the algorithm will generalize well. A direct practical use of this step is to detect the over-fitting. The latter is often difficult to avoid, especially if the training set is small. However, this does not mean that the ML model can not be used. It only means that the performance will be somewhat lower then that obtained during the training. The precision estimated using the test data will then give a better estimate of what can be expected.

It is important to note that the utility of the test step is not to detect a particular deficiency in a ML algorithm. Its primary purpose is to eliminate human error, i.e. human bias. During training the user gains understanding of the underlying data and tunes the ML model to make it more accurate. Therefore, the ML model building process is a loop, as indicated in figure 4.2-c. However, it is important to realize that the loop is, ultimately, closed by the human. For this reason it is impossible to obtain an objective measure of the quality of the ML model using only the training set. In the test step, the human is removed from the process and completely new data is used.

Any bias, implicit or explicit, that occurs during the training will manifest itself in a decreased test accuracy. Since this step is meant to correct for human error, it is crucial to strictly follow the train-validate-test protocol.

4.2.1.5 Application

Provided that the test results are satisfactory one can directly apply the ML model to the problem of interest. Compared to the *ab initio* approach, the desired property can now be evaluated at a significantly lower computational cost. The confidence level for the prediction is the one established through the testing phase. The only input that needs to be provided is the vector of features. The entire process of material characterization then becomes scalable and fast. However, there are two points to consider here. First, one should understand what is the scope of applicability of the model and, second, to which extent the results can be interpreted. The first point was reiterated throughout the entire model building discussion and remains valid for the application. Namely, the precision of the ML model strongly depends on the generating process, which connects the input features to the output. Therefore, a ML model can be safely used to characterize only the materials, which share common traits with those used for training. The users of a ML model need to understand for which systems the method is expected to be valid and for which it is not. It is somewhat paradoxical that although the ML model is practically a "black box" it should not be used as such. If the underlying physics of the materials under investigation changes significantly with respect to the training data, the ML model is likely to give a bad prediction. The reason is that the new data does not have the same underlying structure as the training data. For example, a ML model trained exclusively on insulators can not be expected, in general, to predict the properties of metals well. Another serious limitation is given by the crystal structure of the system. It was tacitly assumed so far that a common representation of the crystal structure can be used for all systems. However, this is often not possible and represents a serious challenge for building truly universal ML models [118, 119, 203]. This restricts the use of the model to only those systems for which the appropriate representation can be constructed. The second point concerns the possibility of generating new knowledge, using the ML approach. In material science one often aims at gaining a deeper understanding of the data, i.e. of the underlying physical mechanisms. A trained ML model, whose precision is high enough, should be able to distinguish between different materials and capture the main features of the data. The trade-off is that some spurious trends might also be introduced. The analysis of the ML model then needs to be performed cautiously, by comparing its prediction against the data. The advantage is that the ML model is a compact representation of the dataset and can be viewed as a function. If the model

uses a reasonably small number of input features one can study the relation between them, e.g. one can look at how the ML prediction varies with individual input features. The data can then be used to distinguish the real and the false trends predicted by the model. A data series following the principal trends can be identified and selected for further study, using *ab initio* methods, for example. In this way relevant features hidden in large datasets can be detected and analysed. Therefore, new knowledge can be extracted from the data. However, a thorough understanding of the ML model behavior is required. Finally, we stress that a ML approach becomes useful only if a compact enough representation of the data can be obtained. In practice, one first needs to go through the entire procedure of building a ML model in order to evaluate its practical utility. Following this approach will make sense only if the potential gain is much higher than the "cost" of building a model. In this sense, one of the most laborious steps is to gather and organize the input data. This is why the availability of "free" materials data [30, 204–207] represents a great potential for further developments in this field.

4.2.2 Machine Learning Model for Magnetism

Magnetism has been known and used for centuries and, despite the fact that we know almost everything about its origin, it is still a hard problem to address when it comes down to material design. The difficulty stems mainly from two aspects, the small energy scales involved, which are of the order of meV, and the complexity of the energy landscape. The requirement for high precision calculations is nowadays not critical, as many electronic structure codes can give accurate enough energies for this kind of problem. However, the complexity of the energy landscape is still an open issue. The usual approach is to map the calculated electronic structure onto a simpler model, e.g. onto the Heisenberg model, which can then be solved by using either Monte Carlo methods or by integrating the equation of motion in the time domain. In general, the electronic degrees of freedom are integrated out and only the spin is kept. The accuracy of the entire process then depends on how accurately the mapping is performed and how well the spin model represents the underlying physics. From an application point of view, there are two types of theoretical approach to magnetism, which are commonly used. The first is to obtain an accurate electronic structure using *ab initio* methods for a given spin configuration. This is useful for understanding the physics related to the electronic degrees of freedom and has a wide range of applications. For example, to study the steady state transport properties, effects of disorder and doping, etc. The second approach is to use the aforementioned models, describing the spin degrees of freedom, to study the dynamics, phase transitions, critical temperatures, and in general finite-temperature properties. To the best of our knowledge the ML models have not

been used so far to address magnetism in complex materials, like in Heusler alloys. The full problem described above is certainly a far too ambitious task to tackle at the moment, however, we believe that certain practical aspects, e.g. determining the magnetic moments of atoms in a given chemical environment, can be addressed. The reason for this is that the magnetic moment of an individual transition metal atom is often independent of the global spin configuration; otherwise the magnetic model approach would not work as well. In addition, the magnetic moment can be considered a local property, determined mainly by the atom and modified by the environment. It therefore makes sense to look at an atom and its few nearest neighbours as an effective model for the magnetic moment of the central atom. The problem can certainly be recast into a set of physical equations describing the "cluster" but, here we wish to use the ML approach to infer the effective model directly from the *ab initio* data. The cluster, which we take to describe the moment of an atom in Heusler alloys, is shown in figure 4.1. The advantage is that we work with the entire class of materials at once and we do not need to worry how to parametrize the model. The accuracy and the range of validity are determined automatically as a part of the model building process. The problem definition then implies that a regression algorithm is used for this task. Practical considerations, on the other hand, restrict us to use only the information about the structure and the atomic properties as the input. In this way there is no need to perform higher level calculations to feed the input of the ML model after the training is completed. The target property needs to be obtained from *ab initio* data.

In order to facilitate the training we need to obtain the magnetic moment data, which can then be associated to a particular cluster. We do not perform simulations of isolated clusters but instead take a Heusler alloy calculation and construct the cluster from there. In this way we implicitly take into account that the cluster is a part of the alloy. We note that this also makes the method less general, since the validation is performed only for the Heusler alloy data. Each calculation allows us to construct 4 clusters, one for each site in the Heusler formula unit. This is done by translating the origin of the cell to the central atom of the cluster. Assuming that the crystal is cubic, each site forms a fcc lattice and the local coordination of the site is always as shown in figure 4.1. In the case of direct Heusler alloys ($X_2\mathrm{Y}Z$), one can construct only 3 inequivalent clusters, since the two X atoms occupy two symmetry equivalent sites. The target property that we use is the site projected magnetic moment, namely its absolute value. In this way any dependence of the magnetic moment on the spin configuration becomes noise in the data. In other words, if the database contains multiple calculations for the same structure, where only the spin configuration is changed, the cluster model will have to map the same input vector onto (potentially) different magnetic moments. These differences in

the output will be seen as a noise by the model. This will ultimately limit the precision of the ML model.

The *ab initio* data was obtained from an in-house database of Heusler alloys, named "Materials Mine" [31, 133]. The calculations there were performed with the PBE local density functional approximation [71], using the PAW [81] pseudopotential method implemented in VASP [61, 147–150]. The GGA is commonly employed to describe the properties of metals. The lattice constants and the magnetic moments usually agree with the experimental results very well. However, the on-site correlations are described at a mean-field level, which may be important in some cases, e.g. in the case of non-collinear magnetism. At the moment the LDA/GGA is the only method suitable for a systematic large scale study. The trade-off is the introduction of a systematic error in the data. Ultimately, this is not critical, because this approximation makes a starting point for a number of higher-level, many-body, correction schemes.

In general, we are interested in ferromagnetic alloys with a high magnetic moment. Here we focus on regular Heusler alloys. Therefore, the crystal structure is the same for all the alloys considered. This removes the need to devise a universal scheme to encode the structure. Only the site occupancy and a lattice constant need to be specified in order to distinguish between different clusters. The lattice constant can be obtained from a DFT calculation but, for a practical application of the ML model one would like to avoid this. Otherwise the magnetic moment could be obtained using the same calculation. Therefore, in order to make the ML model practically useful, we must build a ML model which is able to predict the lattice constant of the Heusler alloy, given only on the information about the composition. This is done in Section 4.2.2.1. In Section 4.2.2.2 we then use this model to provide input for the magnetic moment model. This is an example of chaining procedure, where one ML model is used to provide input for another. Such strategy can be useful when the necessary data is missing or hard to obtain. The input thus becomes noisy, but this is not detrimental to the ML approach, it will only affect the final accuracy of the model.

4.2.2.1 Heusler Alloy Volume Regression Model

The inter-atomic distance plays a vital role in the physics of materials, since both the potential and the kinetic energy of the system depend on the atomic positions. It is therefore crucial to be able to estimate this quantity accurately. The general problem is hard to solve. The configuration where all of the atoms are in equilibrium needs to be found. In practice one is often interested in a particular crystal structure, where the atomic positions are known in advance (the crystallographic positions). In particular,

for a cubic crystal the structure is determined by a single parameter. This can be either the lattice constant or the volume. The Heusler alloys in our dataset in general have a tetragonal symmetry. This means that an additional parameter, e.g. the ratio of the in-plane and the perpendicular lattice constant, c/a, is needed to describe the structure. Predicting the c/a ratio is a non-trivial task since it arises from the interplay of the magnetic and the kinetic energy. For example, a structure having a high symmetry, like a cubic crystal, can give rise to degenerate energy levels close to the Fermi level, which in turn leads to a large kinetic energy. One way to reduce the total energy is to lower the symmetry of the system by distorting the crystal. This is known as the Jahn-Teller effect [208]. Alternatively, the degeneracy can be reduced by forming a magnetic moment, i.e. by the occurrence of the Stoner split of the density of states (cf. section 1.1.4) . If we assume that the volume of the alloy is fixed, the c/a ratio will be such that the total energy is minimal. Therefore, from a ML perspective, knowing the c/a ratio would probably require that the magnetic moment is supplied as input. However, predicting the latter is our final objective. For this reason we omit the c/a parameter from the model and choose to predict the volume of the unit cell. We will show later that this does not reduce the final accuracy of the model. We note in passing that the volume usually remains conserved upon the tetragonal distortion, i.e. the in-plane lattice usually contracts proportionately to the increase of the c/a ratio. Therefore, one can anticipate that the volume will be a smoother function to predict, compared to the in-plane lattice constant, for example. This makes the volume a better candidate as target property, since less training points can be used. The problem is then formulated as a ML regression task and the target quantity needs to be obtained from the *ab initio* data.

The data that is required for the training can be found in the Materials Mine database. However, it needs to be chosen carefully. For a given chemical composition, there is only one structure that has the minimal energy. Therefore the dataset needs to be filtered to select only these structures. Note that here we assume the material to be a regular Heusler alloy and that both the energy and the magnetic configuration have been relaxed. For this purpose we have reused the MCA dataset, discussed in Chapter 2, which contains 271 tetragonal Heusler structures. As previously discussed, selecting only tetragonal structures is not expected to affect the volume of the alloy significantly. We keep 30 % of the data for the test and the rest is used for training.

The ML model was built using the ridge regression algorithm, as implemented in scikit-learn package [159]. The input vector, $\boldsymbol{v}_{in}$, was constructed out of the atomic numbers, Z_i, the atomic volumes, V_i, and the atomic radii, R_i, for the three non-equivalent atoms, namely

$$\boldsymbol{v}_{in} = (\{Z_i\}, \{V_i\}, \{R_i\}) \ , \tag{4.1}$$

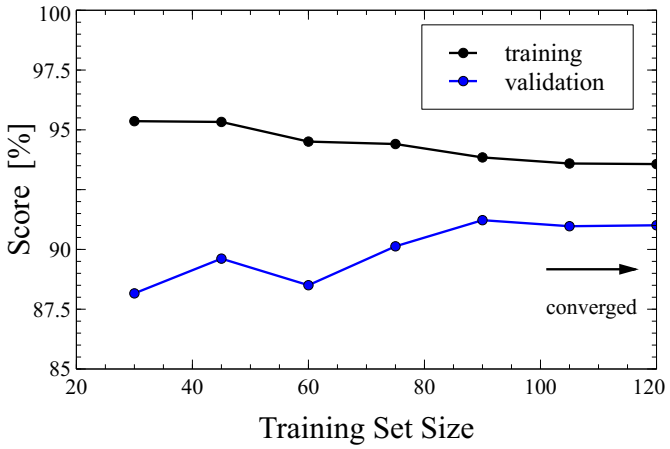

FIGURE 4.3: The learning curve for the volume regressor. The score is plotted as a function of the training set size. For each case the training and the validation score are shown. The feature space is spanned by a 9-dimensional input vector of Eq. 4.1.

where $i \in \{X, Y, Z\}$. The volumes were obtained from Mentel (2014) [209] and then the radius was calculated for each element. For this calculation we assume that the atoms are like homogeneous solid spheres. The learning curve for the model is shown in figure 4.3. The training and the validation score are plotted as a function of the training set size. Here the score represents the internal accuracy measure of the algorithm. We see that the training score decreases as the size of the set increases. At the same time the validation score improves. After a certain number of training points both curves flatten and approach a finite value. This behavior is expected in general as it was thoroughly discussed in section 1.2.3.

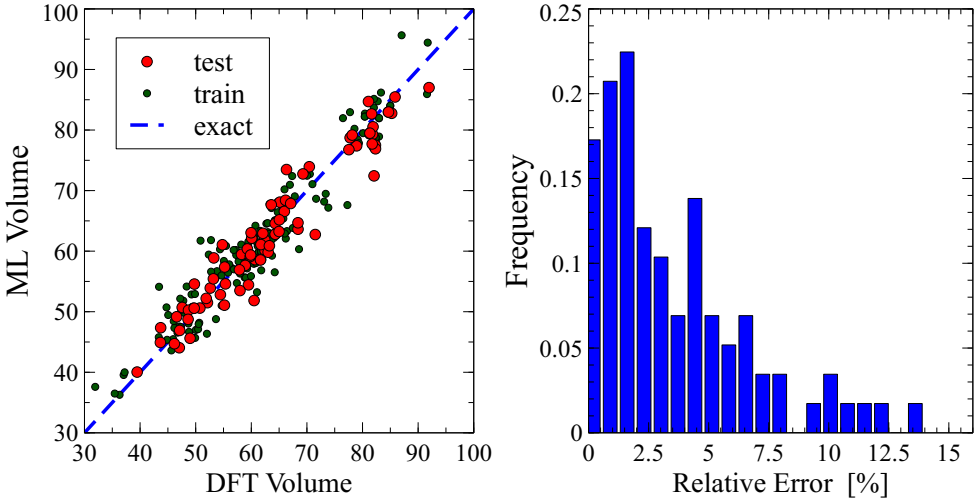

FIGURE 4.4: Accuracy of the volume regressor. *Left Panel:* The ML estimate is plotted against the reference DFT data. The points which belong to the training set (189 points) are shown smaller (green). The test data (82 points) are shown in red. The line of the exact agreement with the reference data is shown in blue. *Right Panel:* The frequency at which the regression error occurs, as a function of its relative value (ΔV / V_{DFT}). The distribution was estimated from the test data.

The root-mean-square error (RMSE) was calculated on the test dataset, finding RMSE to be $\approx 3.2\,\text{Å}^3$. The accuracy of the ML model is visualized in figure 4.4. In the left panel the ML prediction is plotted against the reference DFT data. The perfect agreement line is also shown (dashed). The scattering around this line represents the error of the ML model. We can see that the scattering is qualitatively the same for both the training and the validation data set. Similarly, the regression is equally accurate for the entire range of volumes. In the right panel, the probability of finding a relative error of a given size, i.e. its empirical counterpart, the frequency, is shown. We see that the probability of finding an error larger than 10 % is very small. The calculated RMSE can then be safely associated to the spread of the data around the exact (DFT) result. The attained accuracy is comparable to the systematic DFT error, when compared to the experimental data, ≈5 % for the volume. This makes the overall result rather satisfactory. We note that the ML model presented here does not require any special data as input, only the list of atoms, which appear in the Heusler formula unit. It is thus safe to use the model for a fast volume characterization. In particular, we can feed the data to the input of other ML algorithms, without resorting to additional DFT input. This chaining process allows us to build more sophisticated and scalable ML models.

4.2.2.2 Magnetic Moment Regression Model

In the introduction we have discussed the general outline of the ML problem, which we aim to solve. The main idea is to describe the environment of a magnetic atom by a cluster formed by its nearest neighbours. The goal is then to predict its magnetic moment based on the information about the local environment. The magnetic moments obtained from DFT are used for training the algorithm. Of particular interest to us are the Fe-containing Heusler alloys. We will specialize the ML model to this task, but the approach is completely general. Here we discuss the technical details related to building the magnetic moment model.

The data for the model is obtained from the Materials Mine database. For this task the required DFT information are the site projected magnetic moments. This data is later used only as a target property and hence the final model will be independent of the *ab initio* input. In this case we are not restricted only to the lowest energy structure of a given composition. Instead, we can consider all the crystal structures and the spin configurations associated to a particular Heusler formula unit. This allows us to have a much larger training set. From the available DFT data we keep only the alloys, which contain Fe. Then we construct 4 clusters for each calculation. We do not discard the duplicate clusters when they appear. We find that this does not change the final result in any way. The cluster representing the environment of the central atom

is built according to figure 4.1. The central atom is, in general, surrounded by the two inequivalent nearest neighbour sites, each containing 4 atoms. These sites coordinate the central site tetrahedrally. The second coordination shell is composed of 8 equivalent atoms, forming an octahedral coordination for the central atom. We note that Fe is not the central atom in all the clusters. Thus, in principle, the ML model can be used to predict the magnetic moment of all atoms that appear in Fe containing alloys. We construct in total 18 268 clusters, of which ≈ 7000 were used for the test.

For the regression we use the "Random Forest Regression" algorithm, implemented in scikit-learn [159], which was found to give the best results. The input vector, $\boldsymbol{v}_{in}$, was constructed as

$$\boldsymbol{v}_{in} = (\{Z_i\}, R_0, a_{\mathrm{lat}}, \{r_{0i}\}, \{N_i\}, S_0)\,, \tag{4.2}$$

where Z_i and R_i are the atomic number and the atomic radius of the atom i, respectively, and N_i is the valence. The atomic positions are labeled as in figure 4.1. Here $r_{0\mathrm{i}}$ is the distance between atoms 0 and i, scaled by the sum of their atomic radii. The "effective" cubic lattice constant, a_{lat}, is calculated from the volume of the parent Heusler structure. The Stoner parameter, S_0, of the central atom was obtained from Janak (1977) [158]. We find that this piece of information is crucial for obtaining a reasonable model. The lattice constant was obtained from the volume estimated by the regression model (Section 4.2.2.1). We chose to neglect the c/a parameter in our model for two reasons. The first is that it does not improve the overall accuracy in a significant way and the second is that it is difficult to estimate without resorting to the DFT information. Therefore, all the clusters in our model correspond to cubic crystal structures, although the target property was obtained from a tetragonal structure. This introduces a certain level of noise in the input data, since the model is unable to distinguish calculations with a different c/a ratio. Another source of noise comes from neglecting the spin configuration for which the DFT calculation has been made. This has a hidden benefit, namely the ML model can be used for all possible spin configurations. The expected precision is always the one established through the testing. We note that since there are many parameters, which can influence the precision of the ML model, there necessary exists a trade-off between the simplicity of the model and its accuracy.

In order to determine the accuracy of the model we have calculated the root-mean-square error. The RMS error measured on the test set was $0.4\,\mu_{\mathrm{B}}$. We have found that the error could not be improved by further specializing the ML model to clusters with Fe as the central atom. We believe there are two reasons for this. The first one is that the reduced size of the dataset does not leave enough training data. We find that ≈ 7000 clusters is barely sufficient to converge the learning curve. The second is the that the larger dataset also contains a large number of non-magnetic central atoms. The RMS error in both

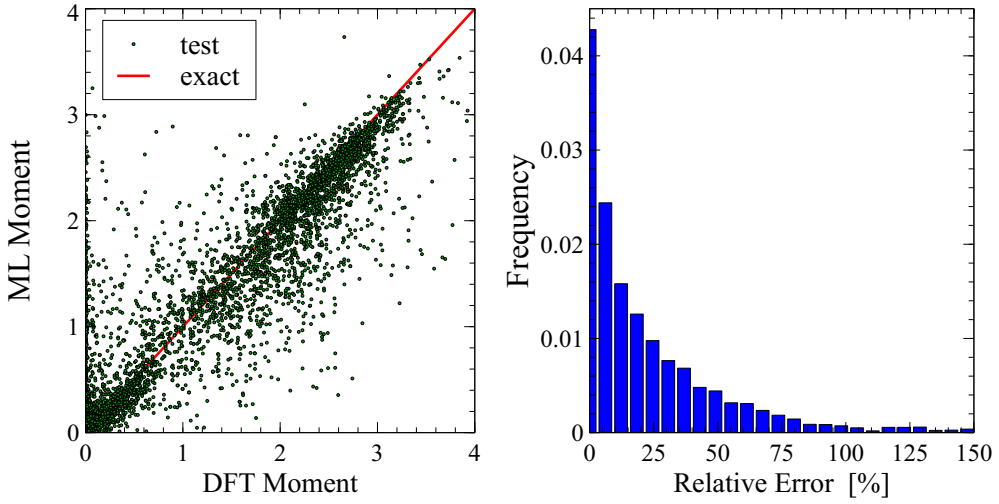

FIGURE 4.5: Accuracy of the magnetic moment regressor. *Left Panel:* The ML estimate is plotted against the reference DFT data. The line of the exact agreement data is shown in red. *Right Panel:* The frequency at which the regression error occurs, as a function of its relative value (ΔM / M_{DFT}). The distribution was estimated from the test data.

cases was found to be practically identical. This implies that the overall error is dictated mainly by the magnetic atoms. The accuracy of the ML model is shown in figure 4.5. In the left panel the ML estimate is plotted against the DFT magnetic moments for the Fe-containing alloys. The scattering is significant, however, the majority of the points lies near the exact agreement line. In the right panel the error density is plotted as a function of the relative error size. We see that even the very large errors have a non-vanishing frequency of occurrence, however the probability of having a large error decreases exponentially. The calculated RMSE of $\approx 0.5\,\mu_{\mathrm{B}}$ is thus reasonable. Similar results to the ones shown in the figure 4.5 are obtained even for the model specialized for the Fe-centered clusters. We note that the magnetic moment obtained from DFT has already an intrinsically large spread. In particular, the magnetic moment of Fe usually does not vary as much as the data would suggest. This may be partly related to problems in the convergence of the DFT calculations. For example, if two Fe atoms are close together they would tend to have an antiferromagnetic configuration, however, a DFT calculation may not find this ground state and the resulting magnetic moment may (erroneously) be zero. This can not be viewed as a completely wrong result, but instead as the best possible solution for a constrained spin configuration. We are thus forced to treat the data as having an intrinsic error component, i.e. the data exhibits a certain degree of noise. At present we have no systematic way to detect and eliminate at the outset this kind of errors from the dataset.

4.3 Results

In the previous section we have built a ML model that can be used to predict the magnetic moment of individual atoms in Fe-containing regular Heusler alloys, regardless of the actual spin configuration of the system. The model uses 4 parameters to estimate the magnetic moment of the central atom shown in figure 4.1. These are, namely, the list of 3 neighbour atom species and the volume of the parent alloy to which the cluster belongs. Our goal is to analyse the main trends of the Fe moment predicted by the ML model. In particular, we are interested in choosing the optimal Fe neighbours, such that the magnetic moment is maximized. We perform a proof-of-principle study, which demonstrates that the ML approach can be used to conduct data-driven research, where the relevant knowledge is extracted directly from the data. In this approach one starts from the data, determines the principal trends using the ML algorithms and then proceeds with a focused *ab inito* study of the established trends, aimed at understanding the underlying physics. Here we will limit our attention to Fe-centered Heusler alloys, where the first coordination shell is occupied by a single atomic species, X. This means that we can directly extend our conclusions to the $X_2\mathrm{Fe}Z$ regular Heusler alloys. The aforementioned restrictions leave us with only 3 structural parameters, whose influence on the magnetic moment need to be explored. We start our investigation by determining the role of the neighbours in the formation of the magnetic moment of Fe (Section 4.3.1) and then proceed with the application of the obtained knowledge. In section 4.3.2 we combine the ML models to determine the magnetization for all alloys belonging to the optimal prototype of the $X_2\mathrm{Fe}Z$ Heusler family. Based on our findings we discuss, which materials are the most suitable for technological applications.

4.3.1 Role of the Coordination Shells

The ML model can be viewed as a compact representation of the materials data that was selected for study. A well trained model should thus contain all the information about the physical trends in the dataset. However, this strictly applies only to a model, which has a high level of accuracy. In the previous section we have shown that the model for the magnetic moment exhibits strong noise. It is therefore possible that certain false trends might be introduced by the algorithm. For this reason we simultaneously look into different functional dependencies of the magnetic moment of Fe, given by the model, and its correspondence to the representative selection of DFT data. This allows us to evaluate the significance of the trends predicted by the ML model. Since the ML model is effectively a function of three variables it is hard to directly visualize the problem without making it incomprehensible. For this reason we focus on the individual

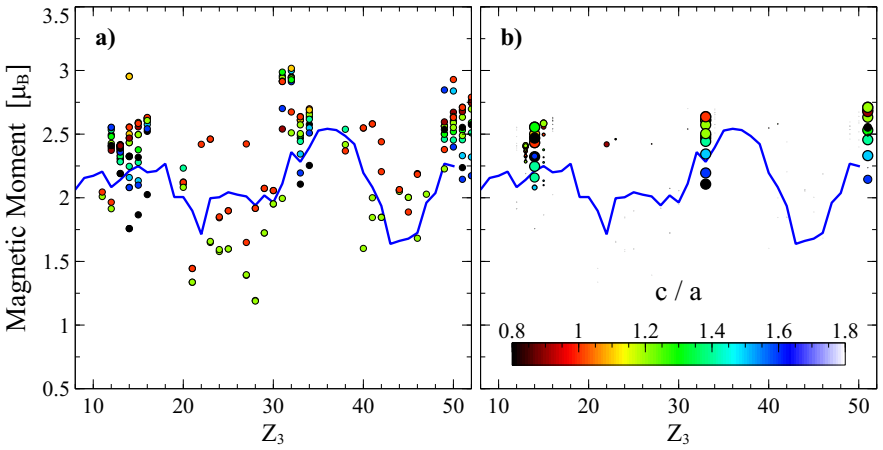

FIGURE 4.6: a) Magnetic moment of the Fe atom as a function of the next-nearest-neighbour atomic number, Z_3, for a Wigner-Seitz radius of 2.7 a_0. The c/a ratio of the parent Heusler alloy is color coded. A machine learning estimate is shown by a solid blue line. b) The same data as previously, only now the radius of the points is proportional to the calculated enthalpy of formation, ΔH [210]. Large circles correspond to more stable alloys.

functional dependencies first and then look for ways to simplify the analysis. We first investigate the role of the outer coordination shell for a fixed nearest neighbour. The outer shell is expected to have a weaker influence on the moment compared to the nearest neighbours. We show that the volume plays only a minor role for the physically relevant range of values. This allows us to study a single volume. The thermodynamic stability criteria and the ML model are then used to show that the next-nearest neighbour ion can be chosen from the main group without losing generality. This further reduces the complexity of the problem and in turn allows us to identify the best candidates for the nearest neighbour shell. With this knowledge we are then able to postulate the ideal prototype for a large-moment Fe-based Heusler alloy.

4.3.1.1 Next-Nearest Neighbour

We start the investigation by looking at the effect of the next-nearest neighbour (NNN) on the magnetic moment of the central atom, namely Fe. For this purpose we fix the 8 nearest neighbour (NN) atoms to Fe. The DFT data is sampled at a constant volume, defined by the Wigner-Seitz radius R_{WS} =2.7 a_0. This volume corresponds to a Heusler lattice constant of 5.8 Å. We find that the Wigner-Seitz radius provides a more intuitive measure of the interatomic distances. In figure 4.6-a we show the magnetic moment of Fe as a function of the NNN atomic number, Z_3. The corresponding ML trend is plotted together with the DFT data. The DFT data is color coded according to the c/a ratio of the crystal cell. We can see that there is no obvious correlation between the magnetic moment trend and this variable. We note that a more apparent correlation emerges at

smaller volumes. This means that the crystal field is not strong enough to determine the magnetic moment at physically relevant volumes. The ML pattern exhibits a periodic modulation as a function of Z_3. An inspection of the DFT data reveals that the main group elements and the transition metal (TM) atoms make two distinct classes of NNN. The DFT points belonging to the former group cluster around the maxima of the ML trend, indicating that the magnetic moment does not depend strongly on the choice of the main group element. Similarly the TM atoms increase the magnetic moment proportional to their valence. The ML trend is found to follow the main trend observed in the DFT data. We note that this is not the case in general. At lower volumes the ML trend becomes different while the DFT data distribution remains pretty much the same. However, the error in the magnetic moment never exceeds its expected value of $0.5\,\mu_{\mathrm{B}}$. It is not clear if the ML trend is simply erroneous or the data can be sampled in another way, which would explain the difference. Therefore, in order to get a complete picture one needs to combine both the DFT data and the ML model.

In figure 4.6-b we show the same data, where now the size of the data point is proportional to the thermodynamic stability of the parent Heusler compound. Here the stability is estimated by using the enthalpy of formation, which is calculated with respect to the decomposition into the most stable elemental phases (cf. equation 3.3). This allows us to understand which compounds are likely to exist and hence to narrow down the parameter space of the investigation. We can see that only the main group elements are likely to be stable. This is in agreement with the known chemistry of Heusler alloys. In the cluster configuration under investigation the NNN corresponds to the octahedral site in the Heulser unit cell, which is always occupied by a main group element [153]. In conclusion, the magnetic moment of Fe does not depend strongly on the choice of the NNN, as long as it is a main group element. This can be further confirmed by using the ML model. We can fix the volume in the model and vary both the NN and the NNN choice. We choose a TM for the former and a main group element for the latter. We find that, for a given NN, the magnetic moment remains roughly the same when the NNN is varied. With this choice the moment is maximized and the alloy is likely to be stable.

4.3.1.2 Nearest Neighbour

We now turn our attention to investigate the dependence of the magnetic moment on the choice of the nearest neighbour. Since we have found that the magnetic moment does not depend significantly on the choice of the NNN (Z_3), we allow a range of main group elements to occupy the outer coordination shell. In figure 4.7 we show the magnetic

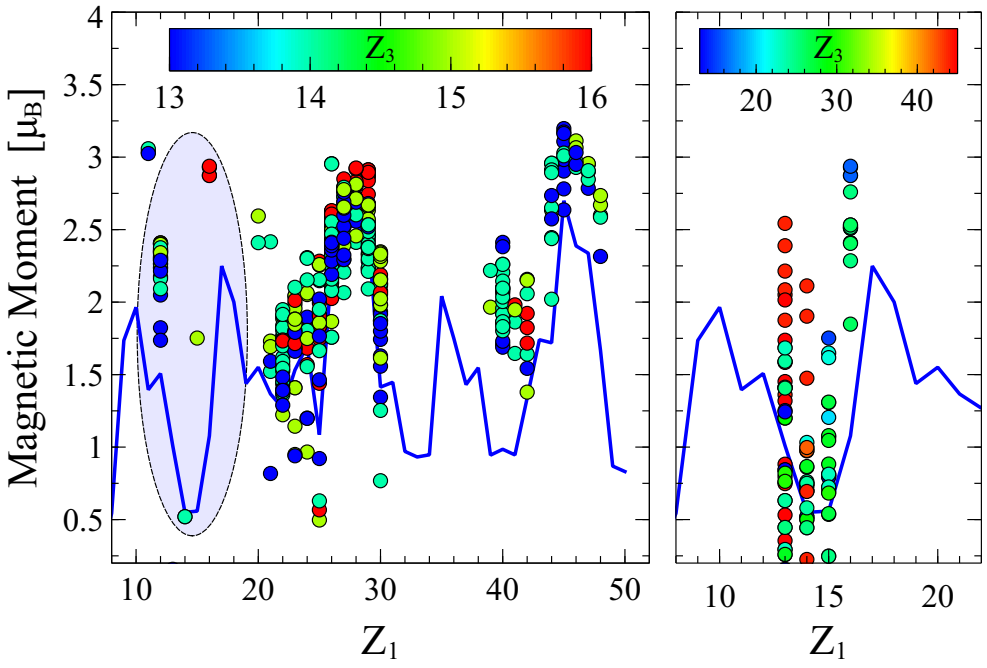

FIGURE 4.7: *Left Panel:* Magnetic moment of Fe for a wide range of nearest neighbours, at a constant Wigner-Seitz volume of $\approx 12\,\text{Å}^3$. The atomic number of the next nearest neighbour, Z_3, is color coded. We can notice a linear increase of the magnetic moment in the transition metal series which does not depend on Z_3. The corresponding machine learning trend is shown as a blue line. *Right Panel:* a data sample containing a wider range of main group elements. The data elucidates the origin of the oscillation in the machine learning trend throughout the main group series in the left panel, indicated by a shaded ellipse.

moment of Fe as a function of the NN atomic number, Z_1. The element Z_3 is now colour-coded and ranges from 13 to 16 (Al to S). The ML estimate is shown as a blue line. We see that the ML model captures the data distribution correctly. However, it exhibits large oscillations in the region where there is no data available. It is instructive to study this observation more closely. In the right panel we show a less stringent selection of the DFT data, namely Z_3 is allowed to vary in the range from 13 to 40. The new data explains the origin of the ML trend oscillations. This clearly demonstrates that the ML is able to provide a reasonable interpolation even when the data is missing. This is an important point. During the data exploration the researcher may select a biased dataset by accident, i.e. select the data which does not contain all the relevant information. This may in turn lead one to make wrong conclusions. The ML algorithm uses the entire dataset and thus offers an objective perspective. In contrast, one has to keep in mind that false ML trends may also occur. For this reason the ML predictions need to be carefully analysed and interpreted. The predicted trends should not be accepted nor rejected lightly. We now turn to the main magnetization trend. We observe that the magnetic moment increases linearly with the valence of the nearest neighbour ion in the TM series. As shown in figure 4.7, the same trend is observed for both the $3d$ and the

4*d* series. The late transition metals (LTMs): Co, Ni, Cu, Rh, Pd, and Ag, offer the largest moment and thus make the most desirable nearest neighbours of Fe. The moment becomes maximal when Ni or Pd constitute the first coordination shell. We note that this observation is valid only for a fixed volume. In real materials the magnetic moment will also depend on the equilibrium volume. The resulting magnetization trend will thus be more complex. The volume dependence changes with the choice of Z_1 and can not be qualified easily. Nevertheless, the ML model can be used to obtain a quantitative estimate for each particular case.

4.3.1.3 Physical Origin of the Magnetic Moment Trend

We have so far established that the magnetic moment of Fe is proportional to the valence of the nearest neighbour ion. Here we wish to understand the physical origin of this trend. For this reason we select a series of materials, which follow the established trend, namely: Sc_2FeSi, Ti_2FeAl, V_2FeSi, Cr_2FeSi, Mn_2FeS, Fe_3Al, Co_2FeAl, Ni_2FeSi, and Cu_2FeSi, all having a cubic $L2_1$ structure. We find that the trend does not depend on the c/a parameter and therefore it is sufficient to focus on the cubic structures. Note that the volume is unconstrained. The linear dependence of the magnetic moment on the valence is reminiscent of the Slater-Pauling curve [25]. Since only the magnetic moment of Fe increases the observed trend could be interpreted as a local version of this rule. However, this is misleading. In figure 4.8-a we show the total charge and the magnetization of Fe for the series above. We find that the total charge in the Fe *d* band is always 6, as required by the local charge-neutrality condition. In other words, there is no net charge transfer from the neighbours to the Fe atom. At the same time the magnetic moment increases proportionally to the NN atomic number, Z_1. The increase of the magnetic moment, therefore, originates from the on-site charge redistribution between the *d* states. We note that this excludes the Slater-Pauling mechanism, which would imply that the origin of the trend is *d* band filling. The orbital projected charge and the magnetization are shown in figure 4.8-b. It can be noticed that the t_{2g} charge decreases slightly, while the e_g increases as the atomic number of the NN TM gets larger. Similarly, the t_{2g} magnetization increases and the e_g one remains constant. We conclude that the charge transfer occurs from the t_{2g} spin down orbitals to both e_g spin orbitals, in equal proportion, so that the e_g moment remains unchanged. This is confirmed by figure 4.8-c, where the spin-resolved orbital charge is shown. If we plot the Fe projected density of states (PDOS), we can understand why the charge redistribution occurs. In figure 4.9 we show the PDOS of Fe for two different nearest neighbours. The PDOS was obtained directly from the raw database data. We can see that for the lighter NN ($Z_1 = 21$) the t_{2g} DOS is narrow and strongly spin-split. Note that both the spin up

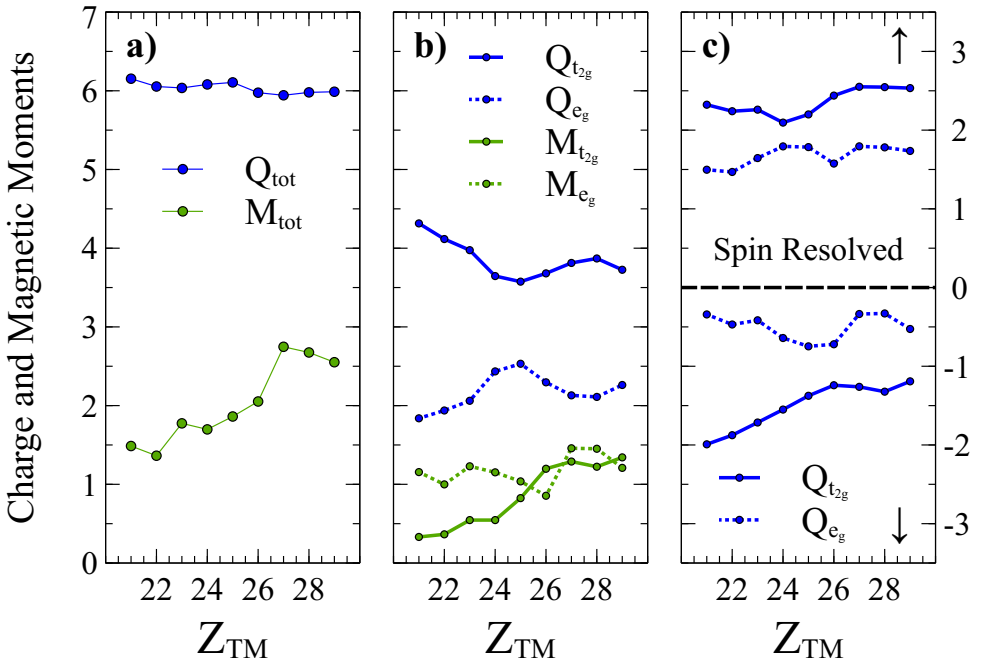

FIGURE 4.8: The integrated charge (blue) and the magnetization (green) of Fe in the TM_2FeZ clusters, as a function of the transition metal (TM) atomic number, Z_{TM}. **a)** the total charge and the magnetization. **b)** the orbital resolved charge and the magnetization. The t_{2g} and the e_g orbitals are shown as full and doted line, respectively. **c)** The orbital resolved charge for the spin up and the spin down channel represented with positive and negative values, respectively.

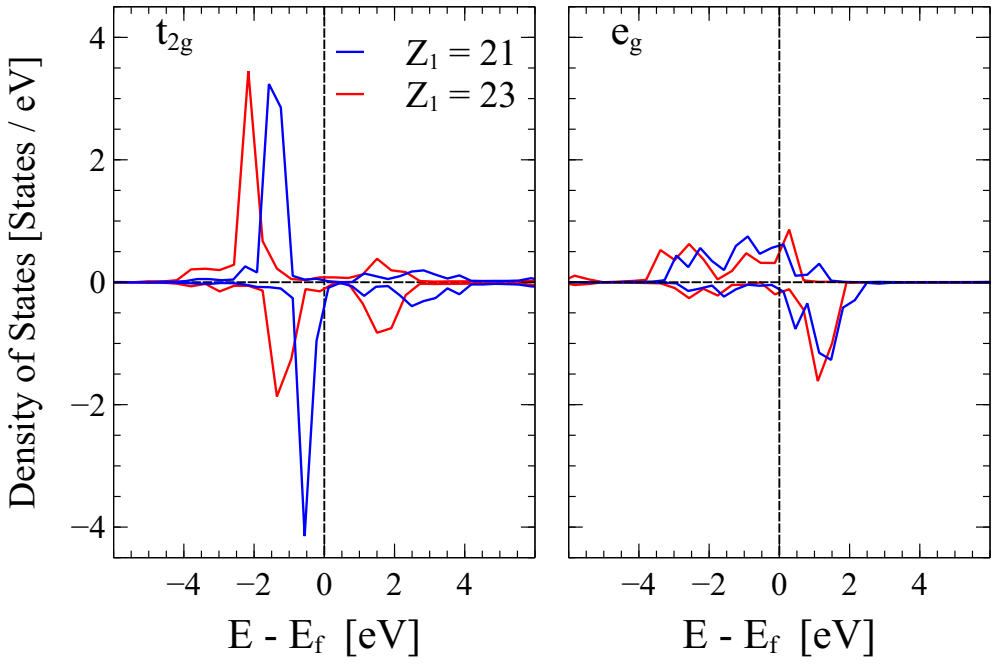

FIGURE 4.9: Projected density of states of Fe in TM_2FeZ clusters, where the selected TMs are the Sc ($Z_1 = 21$) and the V ($Z_1 = 23$). The t_{2g} band is shown in the left panel and the e_g in the right. The spin up and down channels are shown as the positive and the negative values, respectively.

and the spin down t_{2g} peaks are large and have no overlap. Moving down along the TM series, e.g. for $Z_1 = 23$, these states are broadened by the overlap with the NN. The two spin bands now overlap. This in turn gives rise to a repulsive Coulomb interaction and the charge from the t_{2g} spin down states is transferred to the e_g band. This is evident from the size of the t_{2g} spin down peak, which becomes smaller. From figure 4.9 we can also notice that the e_g states are broader than the t_{2g} ones. The former are spread over $\approx 4\,\text{eV}$, while the latter are confined to less than $1\,\text{eV}$. This indicates that the e_g states are strongly hybridized and that the charge is more delocalized. Since the Coulomb energy is proportional to the electron density the inter-orbital charge redistribution is energetically more favourable. We note in passing that this mechanism is an example of the strong electron-electron correlation problem, which is described within the DFT at the mean-field level. Despite the fact that DFT usually gives reasonable magnetic moments, these results suggest that strong correlations may play an important role in the magnetism of Heusler alloys.

4.3.2 Application

The knowledge obtained in the previous section allows us to propose LTM_2FeMG as an optimal prototype for large-moment Fe-based Heusler alloy. Here LTM and MG denote the late transition metal and the main group element, respectively. We note that our analysis applies only to regular Heusler alloys, since all the data was sampled from this class. However, this does not diminish the significance of our result. The LTMs in the proposed prototype are expected to occupy the tetrahedral sites, i.e. the alloy is expected to form the required $L2_1$ phase. This follows from a well established rule regarding the site ordering preference in Heusler alloys [153]. A number of Heusler alloys that belong to this prototype have already been synthesized. For example, Co_2FeSi is a well known ferromagnetic half-metal, with a critical temperature of 1100 K [211] . Other examples of related Heusler alloys include: Co_2FeAl, Cu_2FeAl, Ni_2FeAl, Ni_2FeGa, and Rh_2FeSn [212–216]. This means that there exists a good chance to grow thermodynamically stable alloys with this prototype. We now wish to use the ML model to calculate the magnetization for all the alloys belonging to this family.

The ML models presented here can be used to predict the volume of the alloy and the magnetic moment of Fe. However, this is not enough to evaluate the magnetization of the alloy. Two problems need to be addressed first, namely the estimation of the LTM moment and the magnetic ordering. For both tasks we resort to the DFT data that was used for the ML training. We calculate the average magnetic moments for each element, using the available information about the Fe-containing alloys. The bias in the moment correction, arising from the fact that the alloys under investigation are also part

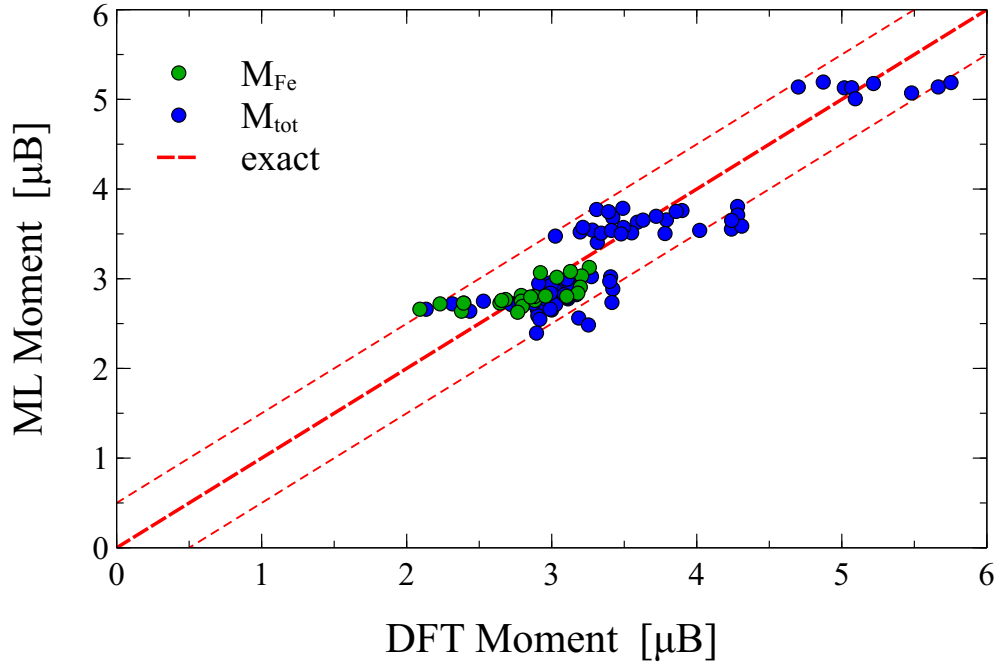

FIGURE 4.10: The magnetic moments in the LTM_2FeMG Heusler alloys. The machine learning regression results are compared to the corresponding DFT values. The magnetic moment of Fe is predicted directly by using the regression (green dots). The total magnetic moment of the alloy (blue dots) is estimated by using the empirical correction scheme described in the text. The exact agreement line (red) and the $\pm$ 0.5 μ_B error bars are shown to guide the eye.

of the dataset, is eliminated by using the average LTM moment. The latter can always be estimated from any representative (random) sample of the true distribution, making our approach completely general. In practice, one would generate the training data and estimate the aforementioned quantities from there. In this case, the training data would represent a sample of the entire chemical space. For the LTM_2FeMG prototype, only: Co, Ni, Rh, and Ir are found to have appreciable magnetic moments. Their respective values were found to be: 1.19 μ_B/atom, 0.37 μ_B/atom, 0.34 μ_B/atom and 0.35 μ_B/atom. The problem of the magnetic ordering is then circumvented by the fact that all the LTM atomic moments are oriented in the same direction as the Fe moment. Therefore, the total magnetic moment of the alloy can be obtained by adding the average values estimated above to the ML estimate for Fe. For a general application the magnetic ordering would need to be determined via some other means.

We proceed by constructing all possible alloys belonging to the proposed prototype. The main group elements were chosen among: Al, Si, P, Ga, Ge, As, In, Sn, and Sb, and the late transition metals among: Co, Ni, Cu, Rh, Pd, Ag, Ir, Pt, and Au. The volume and the magnetic moment of Fe were estimated using the respective ML models. The Fe moment was then "corrected" using the scheme discussed above. In this way the magnetization is estimated using only information about the composition of the unit cell.

The results are shown in figure 4.10. Both the Fe and the total magnetic moment are found to be in good agreement with the corresponding DFT values. The error was found to be $\lesssim 0.5\,\mu_B\,f.u.^{-1}$, for all alloys considered. It is interesting to observe that the total moment, which was obtained by using an approximate scheme, has roughly the same dispersion as the Fe moment. This is mainly due to the robustness of the LTM magnetic moments. In terms of maximizing the moment of Fe, the best performing alloys were based on Rh and Ir, reaching $3\,\mu_B$/atom. This is directly related to the volume of the alloy. As discussed in section 4.3.1.2, for a fixed volume, Ni and Pd nearest neighbours yield the maximal moment. The deviation from this order can thus be only due to the volume. Since these alloys have a relatively large volume the magnetization is lower than in the compounds based on the $3d$ LTMs. We find that a maximal magnetization of 1.2 T is achieved in Co_2-based magnets. The Ni_2- and the Cu_2-based based magnets follow, having a magnetization of 0.83 T and 0.65 T, respectively.

4.4 Conclusions

In this Chapter we have used the ML approach to investigate the relationship between structure and magnetism in Heusler alloys. We had two main objectives. The first one was to demonstrate that this approach can be used to estimate material properties, which are highly sensitive to the changes in structure and composition. In particular, we have focused on the Fe magnetic moment, since Fe is one of the most desirable components for making a practically useful magnet. Nevertheless, the approach presented here is completely general and can be straightforwardly applied to any other element. We have successfully built ML models, which have allowed us to predict the magnetization of an entire family of Heusler alloys using only a modest computational effort. The quantitative description provided by the ML models was crucial for achieving this objective. If the ML models were not accurate enough, a meaningful characterization of the materials would not have been possible. The second objective was to demonstrate that a relevant insight into the properties of a broad class of materials can be obtained by conducting a data-driven investigation. A database of Heusler alloy data was utilized as a source of information about the properties of these system. The ML model was then used to define a minimal parameter space needed to describe the distribution of the material properties. This is a crucial step, where a distribution of ≈ 5000 materials is reduced to a set of 5 parameters. The entire approach relies on this ability of the ML algorithms to make a compact representation of a large amount of materials data. We note that the physical intuition also plays an important role in this process. Namely, it has allowed us to formulate a sensible representation for the data, i.e. to convert the DFT data into a physically meaningful information. This is required if we wish to build

a ML model, which is amenable to interpretation. The relationship of the magnetic moment with respect to each parameter was then investigated. By imposing restrictions, e.g. by focusing only on the thermodynamically stable alloys, we were able to narrow down the parameter space to a single variable, namely the choice of the nearest neighbour. We could then study the main trend in more detail using *ab initio* data for only 10 to 20 materials. This was sufficient to establish that the origin of the trend lies in the on-site charge redistribution, induced by the overlap with the nearest neighbour. This conclusion can then be extended to the entire dataset, since a representative sample was chosen beforehand.

The knowledge of the behaviour of the magnetic moment and the information about the thermodynamic stability of alloys has allowed us to postulate LTM_2FeMG as the ideal prototype for an Fe-based Heusler alloy. We have used the ML models to estimate the magnetization for the entire family using only the information about the composition, i.e. the chemical formula of the alloy. The materials were then ranked according to the maximal attainable magnetization. We find that the maximal magnetization of $\approx 1.2\,\mathrm{T}$ can be achieved in Co_2-based alloys. At the top of the list we find the Co_2FeSi and the Co_2FeAl, which are well known high-performance Heusler alloys. Next on the list there are the Ni_2-alloys, which are related to a well known Heusler shape-memory alloy Ni_2MnGa [217]. However, if the cost of the materials is taken into account the Cu_2-based alloys become the best choice. This consideration becomes crucial for the production of "bulk" magnets, e.g. for applications in the automotive industry. To put things in perspective, the price of Co is approximately 8 times higher than the cost of Cu [218]. The performance gain offered by the Co_2-based alloys is thus economically justifiable only for high-tech applications, like the production of the thin film magnets for memory devices. We note that all of the best performing magnets identified by the ML screening were already synthesized, and many are used for technological applications. Nevertheless, the study presented here is still relevant. There are two reasons for this. The first is that we have performed a systematic investigation and therefore, it is possible to claim that probably all interesting Fe-based Heusler materials have already been discovered. The progress can thus only be made by alloying the aforementioned materials or by engineering the microstructure. These topics, however, are out of the scope of the present study. The second reason is the novelty of the methodological approach, which has allowed us to discover all the relevant Fe-based Heusler material families in a single study, within a time-frame of a few months. In comparison, the Heusler alloys have been actively studied for decades. In this sense the data-driven research has much to offer for an accelerated material discovery. Our task then is to adopt and further specialize the ML methods for the type of investigations that are commonly performed in condensed matter physics. We reiterate that these methods are complementary to the *ab initio*

approach. As we have demonstrated here, data-driven research can help us to select a statistically representative sample of materials, which can then be studied in more detail using the *ab initio* techniques. By establishing the set of relevant property trends beforehand, every specific study can be placed into the broader context, i.e. the relevance and the completeness of each particular study, in terms of explaining the overall property distribution, becomes transparent. In this way the research is done faster and both the computational and the human resources are used more efficiently.

Conclusions

In this Thesis we have applied the high-throughput (HT) and the machine learning (ML) methods to the problem of finding new magnetic materials and engineering existing ones. These are difficult and multifaceted problems, which we address from several aspects. With the exception of Chapter 3, where we had provided a full theoretical description for a material, namely MRG, which has a potential to make a significant impact on spintronics, the work presented here has mostly a methodological significance. The focus was to explore novel, data-oriented, methods and approaches for accelerating the material discovery process.

In Chapter 2 we have designed and performed a HT screening procedure for the discovery of permanent magnets. We were able to demonstrate that the HT screening procedure is feasible and can be employed to search efficiently for new magnetic materials. For the screening we have combined the data-mining approach and *ab initio* calculations. The former was used to identify a list of candidates, which meet the physical requirements for a permanent magnet, namely, to identify potentially stable magnets having a tetragonal ground state crystal structure. The *ab inito* methods were then used to accurately characterize the Curie temperature, T_{C}, and the magneto-crystalline anisotropy, MCA, for the 229 Heusler alloys, selected in the initial screening step. We find 8 candidates, in total, which meet all of the requirements for permanent magnet applications. By a direct inspection, for most of these we were able to identify a number of possible decomposition paths into binary alloys, which are not included in the stability analysis performed here. A full thermodynamic stability analysis should be performed to further refine this estimate, however, we do not do this here. Even so, a sample of 8 viable materials can easily be synthesized in the laboratory. We should keep in mind that the sole purpose of the HT study is to reduce the number of potential candidates to a handful, which can then be analysed in more detail. We find that one of the selected materials actually exists, and was previously found to grow in the inverse Heusler phase, in contrast with the theoretical predictions. Following the empirical site ordering rules in Heusler alloys [153], a similar outcome could be expected for a number of other candidates. We note that this kind of error in the theoretical characterization was

already reported in literature [152]. The HT data has allowed us to estimate that the high-T_C materials ($T_C \geq 1000\,K$) make roughly 20 % of the 229 material dataset, while the materials having MCA $\geq 0.8\,MJ\,m^{-3}$ make 50 %. However, the materials which meet all of the requirements for the permanent magnet application are found at a rate lower than 1 in 10 000, with respect to the overall population of the regular Heusler alloys. Permanent magnets thus seem to be intrinsically rare materials, in good agreement with the experimental observations. The analysis of the HT data has allowed us to identify the ideal composition for a hard magnet in the family of the Heusler alloys as the one where the X site is occupied by the $4d$ or the $5d$ transition metal atom, and the Y site is, ideally, occupied by Fe or Mn. The X site thus plays an important role in the formation of the MCA in full Heusler alloys. Unfortunately, this also limits the maximal magnetic moment to $\approx 4\,\mu_B\,f.u.^{-1}$, which is not sufficient for a technologically useful permanent magnet. Finally, we have constructed a ML classifier, which takes as the input the composition of the Heusler alloy. Based on this information it is capable of identifying soft magnets without any misclassification error. In this way we were able to improve the efficiency of the HT screening procedure by 30 %, without sacrificing the accuracy of the DFT approach. In addition, the analysis of the classifier predictions has further confirmed the established ideal site occupation for the hard magnetic materials. The conclusions of the ML investigation and the results obtained by direct calculation are thus equivalent. Therefore, the ML techniques could prove to be very useful when dealing with very large datasets.

In Chapter 3 we have combined the HT stability analysis and the *ab initio* investigations to resolve the discrepancy between the theoretical (DFT) description of the Mn_2Ru_xGa (MRG) thin films and the corresponding experimental results. We have shown that the DFT description of the ideal MRG films can not account for the experimental observations, with respect to structural, magnetic and transport properties. Using the HT data we have determined the relative stability the competing Mn-Ru-Ga phases and identified the potential defects that could form under the influence of the substrate strain. The defects were then studied using a supercell approach and the substitution of Mn by Ga, occurring on the 4a site, was identified as the energetically most favourable. We have then constructed models, which account for all of the key properties of the nominal MRG films. Namely, we show that strain leads to the formation of defects, which in turn modify the crystal structure through an electron doping mechanism and lower the overall magnetic moment. This is what makes possible the realization of the compensated ferrimagetic state through Ru doping. The magnetic moment correction, taking the defects into account, is found to be in good agreement with experimental results. At the same time, electron doping is found to drive the Fermi level towards the spin gap, thus systematically increasing the spin polarization. The theory predicts

the onset of the half-metallicity at an experimentally unattainable electron doping level, $\approx 1\,e\,f.u.^{-1}$. However, the effects of disorder were not included in this estimate. Taking into account the experimental evidence, we believe that the transport half-metallicity may be realized at much lower doping level, $\sim 0.4\,e\,f.u.^{-1}$, in films thinner than 30 nm, thanks to Mn-Ga substitution disorder. Our study thus paves the way for the realization of the first compensated ferrimagnetic half-metal.

In Chapter 4 we perform a fully data-driven ML investigation of the magnetism of Fe in regular Heulser alloys. We construct clusters describing the local chemical environment of Fe and build ML models capable of predicting its magnetic moment. We also introduce a ML model for the prediction of the volume of the regular Heusler alloys. The ML models were then used to study the trends in the Fe magnetic moment. We have shown that late transition metals, like Ni and Pd, maximize the Fe moment, for a fixed volume. The magnetic moment in the transition metal series was found to increase proportionally to the valence of the nearest neighbour. An inspection of the Fe projected charge densities and the density of states for a series of alloys following this trend has led us to the conclusion that the moment increases due to the on-site charge redistribution, between the t_{2g} and the e_g orbitals. The Fe atom always remains charge neutral. The charge redistribution was found to be driven by the e-e interactions in the d band, i.e. by the strong on-site correlation. The latter is rarely appreciated in the description of the magnetism of metallic systems. We note that no DFT calculations have been performed for this study. Instead, we have directly analysed the DFT data used to construct the database. Finally, we were able to postulate the LTM_2FeMG as the ideal prototype for an Fe-based Heusler alloy, where the LTM and the MG are the late transition metal and the main group element, respectively. We have used the aforementioned ML algorithms to characterize 81 alloys belonging to this family. The results were found to be in good agreement with the DFT data, and the error of the magnetic moment was found to be $\lesssim 0.5\,\mu_B$, for all the alloys. We have found Rh and Ir neighbours to yield the maximal moment of Fe, $\sim 3\,\mu_B$. We show that this is an effect of the volume. Due to the large volume these alloys are not the ones with the maximal magnetization. The maximal magetization of 1.2 T was found in Co_2FeZ alloys. Namely, we find the well known $Co_2Fe\{Si, Al\}$ alloys at the top of our list. The Ni_2- and the Cu_2- alloys follow. All these families have already found technological applications. The Ni_2-family is known for the shape memory applications, and the Cu_2-family represents a more economical alternative to the Co_2FeZ alloys. We have thus demonstrated that a purely data-driven investigation can lead to a discovery of technologically relevant materials. Although we do not discover any new material, by adopting the ML approach we were able to shorten the time of the discovery process, from decades to months. This clearly demonstrates the power of the machine learning and the data-driven approach in general.

In summary, we have utilized the HT data and the ML techniques to accelerate the discovery process of technologically relevant magnetic materials. We have shown that new magnets can be discovered by the HT screening and that the understanding of the complex magnetic materials can be improved using the HT data. Finally, we have shown that even pure data-oriented studies can be highly valuable, allowing not only a fast and scalable material characterisation, but also providing an access to a deeper understanding of the underlying physics of a broad class of materials. We thus clearly demonstrate the potential that these new methods have to offer for tackling the complexity of the structure-property relationship which underpins the design of new materials.

Appendix A

Mean-Field Curie Temeperature

Here we derive the expression for the mean-field (MF) Curie temperature for systems with multiple magnetic sublattices.

The MF approximation assumes that the average magnetic moment $M_i \equiv \langle M_i^z \rangle$, for the site i, depends only on the effective magnetic field, H^{eff}, arising from the neighbouring magnetic sites, M_j. In the classical approximation, which considers the thermal oscillations of a magnetic moment, represented by a vector of fixed length $|\boldsymbol{M_i}| \equiv \bar{M}_i$, the average magnetic moment is given by

$$M_i = \bar{M}_i \, \mathscr{B}_\infty(y_i) \ , \tag{A.1}$$

where $\mathscr{B}_\infty(y)$ is a limiting case of the Brillouin function, $\mathscr{B}_J(y)$ [65, p. 655]. The former is also known as the Langevin function. Here y_i is given by

$$y_i = \beta \bar{M}_i H^{\text{eff}} = \tag{A.2}$$

$$= \beta \bar{M}_i \left[H + \sum_j J_{ij} M_j \right] \ , \tag{A.3}$$

where $\beta^{-1} = k_B T$ and H is the (external) applied magnetic field. The J_{ij}'s are the magnetic exchange constants. For large temperatures we can use the leading term in the expansion of the Langevin function, $\mathscr{B}_\infty(y) \approx \frac{1}{3} y$, to obtain

$$M_i = \frac{\bar{M}_i}{3 k_B T} \left[\bar{M}_i H + \bar{M}_i \sum_j J_{ij} M_j \right] = \tag{A.4}$$

$$T x_i = \frac{\bar{M}_i}{3 k_B} H + \sum_j \Theta_{ij} x_j \ ,$$

where $x_i = M_i/\bar{M}_i$ and $\Theta_{ij} = (\bar{M}_i J_{ij} \bar{M}_j)/3k_B$. By introducing a matrix notation, where $\boldsymbol{x}$ and $\bar{\boldsymbol{M}}$ are vectors, such that $[\boldsymbol{x}]_i = x_i$ and $[\bar{\boldsymbol{M}}]_i = \bar{M}_i$, we can rewrite Eq. (A.4) as

$$T\boldsymbol{x} = \frac{1}{3k_B}\bar{\boldsymbol{M}}H + \boldsymbol{\Theta}\boldsymbol{x} = \tag{A.5}$$

$$(T - \boldsymbol{\Theta})\,\boldsymbol{x} = \frac{1}{3k_B}\bar{\boldsymbol{M}}H$$

and solve it for $\boldsymbol{M}$, to obtain

$$M_i = \frac{1}{3k_B}\bar{M}_i \left[(T - \boldsymbol{\Theta})^{-1}\,\bar{\boldsymbol{M}}\right]_i H \ , \tag{A.6}$$

where the total magnetic moment is given by $M = \sum_i M_i$. The magnetic moment is thus a linear function of the applied field H.

Using Eq. (A.6), the magnetic susceptibility, $\chi = \frac{M}{H}$, can be written as

$$\chi = \frac{1}{3k_B}\bar{\boldsymbol{M}}^T (T - \boldsymbol{\Theta})^{-1}\,\bar{\boldsymbol{M}} \ . \tag{A.7}$$

Furthermore, by using the similarity transformation, which diagonalizes $\boldsymbol{\Theta}$ ($\boldsymbol{\Theta}_D = \boldsymbol{S}\boldsymbol{\Theta}\boldsymbol{S}^{-1} = \boldsymbol{S}\boldsymbol{\Theta}\boldsymbol{S}^T$, since $\boldsymbol{\Theta}$ is a real symmetric matrix), we can easily show that Eq. (A.7) becomes

$$\chi = \frac{1}{3k_B}(\boldsymbol{S}\bar{\boldsymbol{M}})^T (T - \boldsymbol{\Theta}_D)^{-1} (\boldsymbol{S}\bar{\boldsymbol{M}}) \ , \tag{A.8}$$

where $\boldsymbol{\Theta}_D$ is a diagonal matrix. Therefore, the susceptibility diverges when T becomes equal to the eigenvalue of $\boldsymbol{\Theta}$, and the Curie temperature, T_C, is given by the largest eigenvalue [99]. Since a high temperature expansion was used to obtain Eq. (A.4), this is also known as the paramagnetic Curie temperature.

We will now make a connection between the equation (A.4) and a system with multiple magnetic sublattices. In this case the index i in Eq. (A.4) can be understood as a collective index $(\boldsymbol{R}_i, \alpha)$. Here $\boldsymbol{R}_i$ denotes the unit cell and α is the site index. The equation (A.4) then becomes

$$M_{\boldsymbol{R}_i,\alpha} = \frac{\bar{M}_{\boldsymbol{R}_i\alpha}}{3k_B T}\left[\bar{M}_{\boldsymbol{R}_i\alpha}H + \bar{M}_{\boldsymbol{R}_i\alpha}\sum_{\boldsymbol{R}_j\beta} J_{\boldsymbol{R}_i\alpha,\boldsymbol{R}_j\beta} M_{\boldsymbol{R}_j\beta}\right] \ . \tag{A.9}$$

By exploiting translational invariance, namely, $\bar{M}_{\boldsymbol{R}_i\alpha} = \bar{M}_{0\alpha}$ and $J_{\boldsymbol{R}_i\alpha,\boldsymbol{R}_j\beta} = J_{0\alpha,\boldsymbol{R}_j\beta}$, we can rewrite Eq. (A.9) as

$$M_{0\alpha} = \frac{\bar{M}_{0\alpha}}{3k_BT}\left[\bar{M}_{0\alpha}H + \bar{M}_{0\alpha}\sum_{\beta}\left(\sum_{\boldsymbol{R}_j} J_{0\alpha\boldsymbol{R}_j\beta}\right)M_{0\beta}\right] = \tag{A.10}$$

$$M_{\alpha} = \frac{\bar{M}_{\alpha}}{3k_BT}\left[\bar{M}_{\alpha}H + \bar{M}_{\alpha}\sum_{\beta} J_{\alpha\beta}M_{\beta}\right] \ ,$$

where we have dropped the unit cell indices for clarity, and defined a new exchange coupling matrix

$$J_{\alpha\beta} = \sum_{\boldsymbol{R}_j} J_{0\alpha,\boldsymbol{R}_j\beta} \ . \tag{A.11}$$

We note that Eq. (A.10) is mathematically identical to Eq. (A.4), therefore, all the previously obtained results remain valid. In particular, the T_{C} is obtained by diagonalizing the $\boldsymbol{\Theta}_{\alpha\beta}$ matrix, and the J_{ij}'s ($J_{0\alpha,\boldsymbol{R}_j\beta}$) are obtained from *ab initio* calculations.

Appendix B

List of Publications

- M. Žic, K. Rode, N. Thiyagarajah, Y.-C. Lau, D. Betto, J. M. D. Coey, S. Sanvito, K. J. O'Shea, C. A. Ferguson, D. A. MacLaren, and T. Archer. "Designing a fully compensated half-metallic ferrimagnet." Phys. Rev. B, **93**, 140202, (2016).

- D. Betto, K. Rode, N. Thiyagarajah, Y.-C. Lau, K. Borisov, G. Atcheson, M. Žic, T. Archer, P. Stamenov, and J. M. D. Coey. "The zero-moment half metal: How could it change spin electronics?" AIP Adv., **6**, 055601, (2016).

- S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Žic, T. Archer, P. Tozman, M. Venkatesan, J. Coey, and S. Curtarolo. "Accelerated discovery of new magnets in the Heusler alloy family." Sci. Adv., **3**, e1602241, (2017).

- M. Žic, T. Archer, S. Sanvito. "Designing magnetism in Fe-based Heusler alloys: a machine learning approach" arXiv:1706.01840, (2017). Submitted.

Appendix C

Glossary

Machine Learning Terminology

bayesian classifier minimizes the misclassification error by assigning a label of the most likely class to the output, for a given input vector (see Ref. [219, p. 37]).

bias error introduced by approximating the true function by a simpler one (see Ref. [219, p. 34]).

bias-variance trade-off describes the inability to simultaneously minimize the bias and the variance of a model due to the dependence of the learning process on the choice of the training data (see section 1.2.2).

classification a category of learning tasks, where the goal is to determine the class of an object described by the feature vector.

curse of dimensionality implies that any practically feasible dataset is "small", i.e. the sampling is coarse, in high-dimensional spaces (see section 1.2.1).

data distribution refers to the dependence of the target property on the independent variables, with respect to the entire sample population. The shape of a scatter plot or a histogram, for example.

data-mining a process of extracting knowledge from the data, e.g. using machine learning methods.

extra-sample error (generalization error) an average error a machine learning model would make on a new dataset, i.e. on data that was not previously used for training.

true/false positive an outcome of a binary classification task (with two possible outcomes: *true* and *false*) is labeled as a true positive if the class was correctly predicted as *true*. Otherwise the outcome is labeled as a false positive (see Refs. [23, 219]).

feature vector also known as the "predictor" or the "input vector", refers to an array of data passed as an input to the machine learning algorithm. Describes the target property.

high-throughput approach automated synthesis and characterization of materials, for an arbitrarily large number of crystal structures and chemical compositions, using e.g. computational or experimental methods (see Refs. [4, 5]).

hyper-parameters free parameters of a machine learning algorithm, which are used to tune the algorithm performance (e.g. model complexity).

in-sample error an average error of a machine learning model estimated using the training data.

k-fold cross-validation a procedure for estimating the extra-sample error of a learning model, using the training data (see section 1.2.3).

learning curve a graphical method used to analyse the performance of a machine learning model (see section 1.2.3).

loss function a function which measures the "distance" between two points belonging to the space on which the target property is defined. Used to measure the error of a machine learning prediction. For example, in the case of a regression task the loss function is typically taken to be a square (or an absolute value) of a difference between two real numbers (see section 1.2.1).

machine learning a discipline concerned with the development and application of computer algorithms capable of learning from the data, namely the algorithms which can utilize the training data to make future predictions.

model complexity describes the ability of a machine learning model to represent different mathematical functions, namely the functions which are used to fit the training data (see section 1.2.1).

model selection a process of selecting the optimal model for a given learning task (e.g. regression) and training data.

optimal model minimizes the risk of making a bad prediction.

over-fitting occurs when the model has to many degrees of freedom for given training data. Results in a high sensitivity of the learned model to the choice of the training data (high variance).

receiver operating curve a graphical method used to analyse the performance of a classifier (see Refs. [161, 162]).

regression a task of predicting the value of a continuous target property, e.g. predicting the value of a magnetic moment (see section 1.2.1).

risk functional defined as the expectation value of the loss function. Measures the difference between the correct response and the machine learning prediction, averaged over the entire data distribution.

screening a coarse selection procedure, where the accuracy is sacrificed in favour of high throughput. Typically, the materials to be characterized in detail, e.g. using *ab initio* methods, are selected from a large pool of candidates in this way.

separating hyperplane in a classification task, refers to a multidimensional plane separating two classes of objects (see section 1.2.1).

statistical approach an approach where one is concerned with describing (or understanding) the properties of an entire population of objects, rather than focusing on the properties of individual objects.

statistical learning a branch of statistics which is concerned with the problem of learning from the data.

structural risk minimization a strategy for choosing an optimal machine learning model, valid for any dataset (see section 1.2.1).

supervised learning a class of machine learning methods. The correct prediction can be determined by means of a "supervisor", e.g. a classification problem where the correct label can always be assigned (by a human, for example).

supervisor a (hypothetical) "agent", which can provide a correct response for any admissible input vector. For example, in this work the DFT calculations play the role of a supervisor, in a sense.

target property the property that one wishes to predict.

test error an average error of a machine learning model estimated using the test data. Directly estimates the generalization error.

unsupervised learning a class of machine learning methods. The "correct prediction" can not be defined or determined in a practical way. For example, in clustering problems the algorithm automatically decides which elements are "similar" and belong to the same class.

validation error an average error of a machine learning model estimated from the training data by withholding a part of the data (validation set) during the training. In contrast to the test error, the information about the validation error is used in the model selection process (see k-fold cross-validation, section 1.2.3).

variance error arising from the dependence of the interpolating function on the choice of the training data (see Ref. [219, p. 34])

Solid State Physics Terminology

band a distinct group of electronic states of an infinite periodic system (exhibits a discrete translational symmetry). A conduction band, for example.

bandgap the energy difference between the bottom of the lowest unoccupied band (conduction band) and the top of the highest occupied band (valence band).

Born-Oppenheimer approximation decouples the electronic and the ionic degrees of freedom (see section 1.1.1).

chemical space a space spanned by materials of all possible structures and chemical compositions.

coordination shell a sphere of constant radius, centred on an atom, on which at least one of the neighbours of the central atom lies.

correlation energy takes into account the fact that particles in a many-body system move in a correlated manner (see section 1.1.1).

crystal defects imperfections in the crystal structure, which destroy the ideal (discrete) translational symmetry. For example, if two ions swap places in a given unit cell, we have site disorder.

Curie temperature critical temperature of a ferromagnet. Above this temperature the magnetic order is lost and the ferromagnet becomes a paramagnet.

density of states the density of electronic states as a function of energy.

elemental phase a material phase, consisting of only one atomic species.

enthalpy of formation a measure of thermodynamic stability of a material (see section 3.3)

exchange energy a quantum mechanical contribution to the electron-electron interaction energy, due to the fact that electrons must satisfy the Pauli exclusion principle (see section 1.1.1).

Fermi level the highest occupied energy level at zero-Kelvin temperature. It is well defined only for metals.

ferrimagnet material which has at least two symmetry inequivalent magnetic sites in the unit cell.

fully-compensated ferrimagnet ferrimagnet where different magnetic sublattices compensate each other to achieve zero net magnetization, for a given temperature.

half-metal behaves like a normal metal for one spin band and like a semiconductor (or an insulator) for the other.

KKR method electronic structure method based on the multiple-scattering theory (see Ref. [139])

magnetic anisotropy implies the existence of a preferential axis (not necessarily unique) along which the magnetization lies in the ground state. An uniform global rotation of the magnetization, away from this axis, takes the system into an excited state.

magneto-crystalline anisotropy an intrinsic contribution to the magnetic anisotropy, arising from the spin-orbit coupling.

magneto-elastic coupling a change in the crystal structure of a material strongly affects its magnetization, and vice versa (see Ref. [91]).

many-body effects refers to the effects that arise due to the correlated motion of a large number of particles, e.g. electrons (see Chapter 1).

mean-field approximation replaces the complex interaction of a particle with the other particles by an effective single-particle interaction, namely the interaction with a "mean" field.

mobility gap absence of electrically conductive states in the energy window around the Fermi level.

pseudopotential an effective ionic potential, that acts only on the valence band electrons (see section 1.1.2.5)

quasi-particle a particle-like excitation (typically fermionic) of a many-body ground state (see Ref. [65, Ch. 17]).

saturation magnetization maximally attainable magnetization, for a given magnet.

self-interaction error error arising from an approximate treatment of electron-electron interactions, where electrons end up exhibiting self-interaction (see section 1.1.1).

soft/hard magnet magnets which are easily magnetized by an external magnetic field are called soft. The opposite is valid for the hard magnets (see Ref. [27]).

spin band/channel the band structure of electronic states labelled by a particular spin projection, e.g. spin up states. The term "spin channel" is typically used when discussing transport properties, e.g. electrical conductivity.

spin gap a gap in the electronic density of states, which occurs for only one spin band.

spin polarization refers to an unequal number of spin up and spin down states at the Fermi level (see Chapter 3).

spin-current (pure) a flux of the magnetic moment, originating from electron spins (see Chapter 3).

spin-orbit coupling coupling between the orbital and spin degrees of freedom arising in relativistic quantum mechanics (see Ref. [45]).

spin-polarized current electric current where the number of spin up and spin down charge carriers is not equal. In addition to the flux of the electric charge there exists an associated flux of the magnetic moment (see Chapter 3).

spin-spiral method method used to calculate the spin-wave excitation energies (see section 1.1.4.1).

stray field the magnetic field produced by a magnet outside of its volume (see Ref. [27]).

structure-property relationship a relationship between the structure and chemical composition of a material, and its properties.

uniaxial anisotropy a type of anisotropy, exhibited by a system which has a principal axis of symmetry, e.g. in tetragonal crystals.

Wigner-Seitz radius radius of a sphere whose volume is given by $V_{\mathrm{WS}} = V/N$, where V is the unit cell volume and N is the number of atoms in the unit cell.

Bibliography

[1] B. Dieny, V. S. Speriosu, S. Metin, S. S. P. Parkin, B. A. Gurney, P. Baumgart, and D. R. Wilhoit. "Magnetotransport properties of magnetically soft spinvalve structures (invited)." *J. Appl. Phys.*, **69**, 4774, (1991).

[2] T. Miyazaki and N. Tezuka. "Spin polarized tunneling in ferromagnet/insulator/-ferromagnet junctions." *J. Magn. Magn. Mater.*, **151**, 403, (1995).

[3] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey. "Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions." *Phys. Rev. Lett.*, **74**, 3273, (1995).

[4] W. F. Maier, K. Stöwe, and S. Sieg. "Combinatorial and High-Throughput Materials Science." *Angew. Chem. Int. Ed.*, **46**, 6016, (2007).

[5] S. Curtarolo, G. L. W. Hart, M. Buongiorno-Nardelli, N. Mingo, S. Sanvito, and O. Levy. "The high-throughput highway to computational materials design." *Nat. Mater.*, **12**, 191, (2013).

[6] A. Jain, G. Hautier, C. J. Moore, S. Ping Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder. "A high-throughput infrastructure for density functional theory calculations." *Comp. Mater. Sci.*, **50**, 2295, (2011).

[7] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder. "Predicting Crystal Structures with Data Mining of Quantum Calculations." *Phys. Rev. Lett.*, **91**, 135503, (2003).

[8] S. Curtarolo, D. Morgan, and G. Ceder. "Accuracy of ab initio methods in predicting the crystal structures of metals: A review of 80 binary alloys." *Calphad*, **29**, 163, (2005).

[9] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder. "Predicting crystal structure by merging data mining with quantum mechanics." *Nat. Mater.*, **5**, 641, (2006).

[10] L. Yu and A. Zunger. "Identification of Potential Photovoltaic Absorbers Based on First-Principles Spectroscopic Screening of Materials." *Phys. Rev. Lett.*, **108**, 068701, (2012).

[11] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen. "Computational screening of perovskite metal oxides for optimal solar light capture." *Energy & Environmental Science*, **5**, 5814, (2012).

[12] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, and J. K. Nørskov. "Computational high-throughput screening of electrocatalytic materials for hydrogen evolution." *Nat. Mater.*, **5**, 909, (2006).

[13] J. Greeley and J. K. Nørskov. "Electrochemical dissolution of surface alloys in acids: Thermodynamic trends from first-principles calculations." *Electrochim. Acta*, **52**, 5829, (2007).

[14] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, and B. Smit. "In silico screening of carbon-capture materials." *Nat. Mater.*, **11**, 633, (2012).

[15] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr. "Large-scale screening of hypothetical metal-organic frameworks." *Nat. Chem.*, **4**, 83, (2011).

[16] S. V. Alapati, J. K. Johnson, and D. S. Sholl. "Large-Scale Screening of Metal Hydride Mixtures for High-Capacity Hydrogen Storage from First-Principles Calculations." *J. Phys. Chem. C*, **112**, 5258, (2008).

[17] W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson, and S. Curtarolo. "High-Throughput Combinatorial Database of Electronic Band Structures for Inorganic Scintillator Materials." *ACS. Comb. Sci.*, **13**, 382, (2011).

[18] R. Armiento, B. Kozinsky, M. Fornari, and G. Ceder. "Screening for high-performance piezoelectrics using high-throughput density functional theory." *Phys. Rev. B*, **84**, 014103, (2011).

[19] A. Roy, J. W. Bennett, K. M. Rabe, and D. Vanderbilt. "Half-Heusler Semiconductors as Piezoelectrics." *Phys. Rev. Lett.*, **109**, 037602, (2012).

[20] J. Yang, H. Li, T. Wu, W. Zhang, L. Chen, and J. Yang. "Evaluation of Half-Heusler Compounds as Thermoelectric Materials Based on the Calculated Electrical Transport Properties." *Adv. Funct. Mater*, **18**, 2880, (2008).

[21] B. G. Sumpter, R. K. Vasudevan, T. Potok, and S. V. Kalinin. "A bridge for accelerating materials by design." *npj Computational Materials*, **1**, 15008, (2015).

[22] E. Alpaydin. *Introduction to machine learning.* MIT Press, (2010).

[23] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning.* Springer, (2008).

[24] A. R. Williams, V. L. Moruzzi, A. P. Malozemoff, and K. Terakura. "Generalized Slater-Pauling curve for transition-metal magnets." *IEEE Trans. Magn.*, **19**, 1983, (1983).

[25] I. Galanakis, P. Dederichs, and N. Papanikolaou. "Slater-Pauling behavior and origin of the half-metallicity of the full-Heusler alloys." *Phys. Rev. B*, **66**, 174429, (2002).

[26] G. H. Fecher, H. C. Kandpal, S. Wurmehl, C. Felser, and G. Schönhense. "Slater-Pauling rule and Curie temperature of Co_2-based Heusler compounds." *J. Appl. Phys.*, **99**, 08J106, (2006).

[27] J. M. D. Coey. *Magnetism and Magnetic Materials.* Cambridge University Press, (2010).

[28] J. M. D. Coey. "Hard Magnetic Materials: A Perspective." *IEEE Trans. Magn.*, **47**, 4671, (2011).

[29] T. Graf, S. S. P. Parkin, and C. Felser. "Heusler Compounds: A Material Class With Exceptional Properties." *IEEE Trans. Magn.*, **47**, 367, (2011).

[30] "AFLOWLIB." http://aflowlib.org/.

[31] http://www.materials-mine.com.

[32] H. Kurt, K. Rode, P. Stamenov, M. Venkatesan, Y.-C. Lau, E. Fonda, and J. M. D. Coey. "Cubic Mn_2Ga Thin Films: Crossing the Spin Gap with Ruthenium." *Phys. Rev. Lett.*, **112**, 027201, (2014).

[33] H. Van Leuken and R. A. De Groot. "Half-metallic antiferromagnets." *Phys. Rev. Lett*, **74**, 1171, (1995).

[34] D. Betto, K. Rode, N. Thiyagarajah, Y.-C. Lau, K. Borisov, G. Atcheson, M. Žic, T. Archer, P. Stamenov, and J. M. D. Coey. "The zero-moment half metal: How could it change spin electronics?" *AIP Advances*, **6**, 055601, (2016).

[35] K. Takahashi and Y. Tanaka. "Material synthesis and design from first principle calculations and machine learning." *Comp. Mater. Sci.*, **112**, 364, (2016).

[36] A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig, and A. Mar. "High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds." *Chem. Mater.*, **28**, 7324, (2016).

[37] J. C. Slater. "The Theory of Complex Spectra." *Phys. Rev.*, **34**, 1293, (1929).

[38] D. R. Hartree. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods." *Math. Proc. Cambridge Philos. Soc.*, **24**, 89, (1928).

[39] D. R. Hartree. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion." *Math. Proc. Cambridge Philos. Soc.*, **24**, 111, (1928).

[40] V. Fock. "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems." *Z. Physik*, **61**, 126, (1930).

[41] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory.* Springer-Verlag, Berlin Heidelberg, (1990).

[42] I. Shavitt. "The history and evolution of configuration interaction." *Mol. Phys.*, **94**, 3, (1998).

[43] Čížek J. "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell–Type Expansion Using Quantum-Field Theoretical Methods." *J. Chem. Phys.*, **45**, 4256, (1966).

[44] R. P. Feynman. "Space-time approach to quantum electrodynamics." *Phys. Rev.*, **76**, 769, (1949).

[45] R. J. Elliott. "Theory of the effect of spin-orbit coupling on magnetic resonance in some semiconductors." *Phys. Rev.*, **96**, 266, (1954).

[46] C. Kittel. *Introduction to Solid State Physics.* John Wiley & Sons, Inc, (2005).

[47] C. Barreteau, D. Spanjaard, and M.-C. Desjonquères. "An efficient magnetic tight-binding method for transition metals and alloys." *Comptes Rendus Physique*, **17**, 406, (2016).

[48] J. K. L. MacDonald. "Successive approximations by the Rayleigh-Ritz variation method." *Phys. Rev.*, **43**, 830, (1933).

[49] C. Møller and M. S. Plesset. "Note on an approximation treatment for many-electron systems." *Phys. Rev.*, **46**, 618, (1934).

[50] L. Hedin. "New method for calculating the one-particle Green's function with application to the electron-gas problem." *Phys. Rev.*, **139**, A796, (1965).

[51] S. Blügel, N. Helbig, V. Meden, and W. D. (Eds.). *Computing Solids.* Forschungszentrum Jülich GmbH, (2014).

[52] R. M. Martin. *Electronic Structure.* Cambridge University Press, (2004).

[53] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas." *Phys. Rev.*, **136**, 864, (1964).

[54] N. D. Mermin. "Thermal Properties of the Inhomogeneous Electron Gas." *Phys. Rev.*, **137**, A1441, (1965).

[55] E. Runge and E. K. U. Gross. "Density-functional theory for time-dependent systems." *Phys. Rev. Lett.*, **52**, 997, (1984).

[56] U. von Barth and L. Hedin. "A local exchange-correlation potential for the spin polarized case." *J. Phys. C: Solid State Phys.*, **5**, 1629, (1972).

[57] A. K. Rajagopal. "Inhomogeneous relativistic electron gas." *J. Phys. C: Solid State Phys.*, **11**, L943, (1978).

[58] A. H. MacDonald and S. H. Vosko. "A relativistic density functional formalism." *J. Phys. C: Solid State Phys.*, **12**, 2977, (1979).

[59] M. Levy. "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem." *Proc. Natl. Acad. Sci.*, **76**, 6062, (1979).

[60] E. H. Lieb. "Density functionals for coulomb systems." *Int. J. Quantum Chem.*, **24**, 243, (1983).

[61] G. Kresse and J. Furthmüller. "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set." *Phys. Rev. B*, **54**, 11169, (1996).

[62] "Local spindensity functional theory of noncollinear magnetism (invited)." *J. Appl. Phys.*, **63**, 3482, (1988).

[63] F. G. Eich and E. K. U. Gross. "Transverse Spin-Gradient Functional for Noncollinear Spin-Density-Functional Theory." *Phys. Rev. Lett.*, **111**, (2013).

[64] P. Kurz, F. Förster, L. Nordström, G. Bihlmayer, and S. Blügel. "Ab initio treatment of noncollinear magnets with the full-potential linearized augmented plane wave method." *Phys. Rev. B*, **69**, 024415, (2004).

[65] N. W. Ashcroft and N. D. Mermin. *Solid State Physics.* Saunders College, (1976).

[66] J. F. Janak. "Proof that $\frac{dE}{dn_i} = \varepsilon_i$ in density-functional theory." *Phys. Rev. B*, **18**, 7165, (1978).

[67] P. Haas, F. Tran, and P. Blaha. "Calculation of the lattice constant of solids with semilocal functionals." *Phys. Rev. B*, **79**, (2009).

[68] L. J. Sham and M. Schlüter. "Density-functional theory of the band gap." *Phys. Rev. B*, **32**, 3883, (1985).

[69] M. K. Y. Chan and G. Ceder. "Efficient Band Gap Prediction for Solids." *Phys. Rev. Lett.*, **105**, 196403, (2010).

[70] B. Alder and D. Ceperly. "Ground State of the Electron Gas by a Stochastic Method." *Phys. Rev. Lett.*, **45**, 56, (1980).

[71] J. P. Perdew, K. Burke, and M. Ernzerhof. "Generalized Gradient Approximation Made Simple." *Phys. Rev. Lett.*, **78**, 1396, (1997).

[72] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka. "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits." *J. Chem. Phys.*, **123**, 062201, (2005).

[73] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. "Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids." *Phys. Rev. Lett.*, **91**, (2003).

[74] C. Adamo and V. Barone. "Toward reliable density functional methods without adjustable parameters: The PBE0 model." *J. Chem. Phys.*, **110**, 6158, (1999).

[75] J. Heyd, G. E. Scuseria, and M. Ernzerhof. "Hybrid functionals based on a screened Coulomb potential." *J. Chem. Phys.*, **118**, 8207, (2003).

[76] A. D. Becke. "Correlation energy of an inhomogeneous electron gas: A coordinatespace model." *J. Chem. Phys.*, **88**, 1053, (1988).

[77] J. Heyd and G. E. Scuseria. "Efficient hybrid density functional calculations in solids: Assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional." *J. Chem. Phys.*, **121**, 1187, (2004).

[78] T. M. Henderson, J. Paier, and G. E. Scuseria. "Accurate treatment of solids with the HSE screened hybrid." *physica status solidi (b)*, **248**, 767, (2011).

[79] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán. "Screened hybrid density functionals applied to solids." *J. Chem. Phys.*, **124**, 154709, (2006).

[80] G. Kresse and J. Hafner. "Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements." *J. Phys. Condens. Matter.*, **6**, 8245, (1994).

[81] P. E. Blöchl. "Projector augmented-wave method." *Phys. Rev. B*, **50**, 17953, (1994).

[82] K. Lejaeghere, G. Bihlmayer, T. Bjorkman, P. Blaha, S. Blugel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Duak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Granas, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Kucukbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordstrom, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunstrom, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, and S. Cottenier. "Reproducibility in density functional theory calculations of solids." *Science*, **351**, aad3000, 2016.

[83] R. A. de Groot, F. M. Mueller, P. G. v. Engen, and K. H. J. Buschow. "New Class of Materials: Half-Metallic Ferromagnets." *Phys. Rev. Lett.*, **50**, 2024, (1983).

[84] N. Yang, X. Xu, G. Zhang, and B. Li. "Thermal transport in nanostructures." *AIP Advances*, **2**, 041410, (2012).

[85] J. H. van Vleck. *Theory of Electric and Magnetic Susceptibilities*. Clarendon Press, (1932).

[86] P. Weiss. "L'hypothèse du champ molèculaire et la propriètè frromagnètique." *J. Phys.*, **6**, 661, (1907).

[87] S. R. *Simple Models of Magnetism*. Oxford University Press, (2008).

[88] J. B. Goodenough. *Magnetism and the Chemical Bond*. Wiley, New York, (1963).

[89] C. Zener. "Interaction between the d shells in the transition metals." *Phys. Rev.*, **81**, 440, (1951).

[90] P. Fazekas. *Lecture Notes on Electron Correlation and Magnetism*. World Scientific, (1999).

[91] E. R. Callen and H. B. Callen. "Static magnetoelastic coupling in cubic crystals." *Phys. Rev.*, **129**, 578, (1963).

[92] S. O. Demokritov, V. E. Demidov, O. Dzyapko, G. A. Melkov, A. A. Serga, B. Hillebrands, and A. N. Slavin. "Bose–Einstein condensation of quasi-equilibrium magnons at room temperature under pumping." *Nature*, **443**, 430, (2006).

[93] P. Buczek, A. Ernst, P. Bruno, and L. Sandratskii. "Energies and Lifetimes of Magnons in Complex Ferromagnets: A First-Principle Study of Heusler Alloys." *Phys. Rev. Lett.*, **102**, 247206, (2009).

[94] M. Ležaić, P. Mavropoulos, G. Bihlmayer, and S. Blügel. "Exchange interactions and local-moment fluctuation corrections in ferromagnets at finite temperatures based on noncollinear density-functional calculations." *Phys. Rev. B*, **88**, 134403, (2013).

[95] A. Jacobsson, B. Sanyal, M. Ležaić, and S. Blügel. "Exchange parameters and adiabatic magnon energies from spin-spiral calculations." *Phys. Rev. B*, **88**, 134427, (2013).

[96] M. Pajda, J. Kudrnovský, I. Turek, V. Drchal, and P. Bruno. "Ab initio calculations of exchange interactions, spin-wave stiffness constants, and Curie temperatures of Fe, Co, and Ni." *Phys. Rev. B*, **64**, 174402, 2001.

[97] P. Mavropoulos. *Multiscale Modelling of Magnetic Materials: From the Total Energy of the Homogeneous Electron Gas to the Curie Temperature of Ferromagnets.* Supercomputing Centre, Forschungszentrum Jülich, (2009).

[98] A. I. Liechtenstein, M. I. Katsnelson, V. P. Antropov, and V. A. Gubanov. "Local spin density functional approach to the theory of exchange interactions in ferromagnetic metals and alloys." *J. Magn. Magn. Mater.*, **67**, 65, (1987).

[99] E. Şaşoğlu, L. Sandratskii, and P. Bruno. "Pressure dependence of the Curie temperature in Ni_2MnSn Heusler alloy: A first-principles study." *Phys. Rev. B*, **71**, (2005).

[100] L. Nordstrom, M. S. S. Brooks, and B. Johansson. "Calculation of orbital magnetism and magnetocrystalline anisotropy energy in YCo_5." *J. Phys. Condens. Matter.*, **4**, 3261, (1992).

[101] G. H. O. Daalderop, P. J. Kelly, and M. F. H. Schuurmans. "Magnetocrystalline anisotropy of $RECo_5$ compounds." *J. Mag. Mag. Mater.*, **104-107**, 737, (1992).

[102] J. H. van Vleck. "On the anisotropy of cubic ferromagnetic crystals." *Phys. Rev.*, **52**, 1178, (1937).

[103] A. Crépieux and C. Lacroix. "Dzyaloshinsky–Moriya interactions induced by symmetry breaking at a surface." *J. Magn. Magn. Mater.*, **182**, 341, (1998).

[104] I. Dzyaloshinsky. "A thermodynamic theory of weak ferromagnetism of antiferromagnetics." *J. Phys. Chem. Solids*, **4**, 241, (1958).

[105] T. Moriya. "Anisotropic superexchange interaction and weak ferromagnetism." *Phys. Rev.*, **120**, 91, (1960).

[106] T. Moriya. "New mechanism of anisotropic superexchange interaction." *Phys. Rev. Lett.*, **4**, 228, (1960).

[107] P. Bruno. "Tight-binding approach to the orbital magnetic moment and magnetocrystalline anisotropy of transition-metal monolayers." *Phys. Rev. B*, **39**, 865, (1989).

[108] C. Andersson, B. Sanyal, O. Eriksson, L. Nordström, O. Karis, D. Arvanitis, T. Konishi, E. Holub-Krappe, and J. H. Dunn. "Influence of Ligand States on the Relationship between Orbital Moment and Magnetocrystalline Anisotropy." *Phys. Rev. Lett.*, **99**, 177207, (2007).

[109] Y.-S. Kim, K. Hummer, and G. Kresse. "Accurate band structures and effective masses for InP, InAs, and InSb using hybrid functionals." *Phys. Rev. B*, **80**, 035203, (2009).

[110] R. Wu and A. J. Freeman. "Spinorbit induced magnetic phenomena in bulk metals and their surfaces and interfaces." *J. Magn. Magn. Mater.*, **200**, 498, (1999).

[111] X. Wang, W. D.-S., R. Wu, and A. J. Freeman. "Validity of the force theorem for magnetocrystalline anisotropy." *J. Magn. Magn. Mater.*, **159**, 337, (1996).

[112] G. H. O. Daalderop, P. J. Kelly, and M. F. H. Schuurmans. "First-principles calculation of the magnetocrystalline anisotropy energy of iron, cobalt, and nickel." *Phys. Rev. B*, **41**, 11919, (1990).

[113] M. Methfessel and J. Kübler. "Bond analysis of heats of formation: application to some group VIII and IB hydrides." *J. Phys. F*, **12**, 141, (1982).

[114] D. Halliday, R. Resnick, and J. Walker. *Fundamentals of Physics.* Wiley, (2013).

[115] J. Schmidhuber. "Deep learning in neural networks: An overview." *Neural Networks*, **61**, 85, 2015.

[116] M. Attarian Shandiz and R. Gauvin. "Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries." *Comp. Mater. Sci.*, **117**, 270, (2016).

[117] E. Bélisle, Z. Huang, S. Le Digabel, and A. E. Gheribi. "Evaluation of machine learning interpolation techniques for prediction of physical properties." *Comp. Mater. Sci.*, **98**, 170, (2015).

[118] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross. "How to represent crystal structures for machine learning: Towards fast prediction of electronic properties." *Phys. Rev. B*, **89**, 205118, (2014).

[119] 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).

[120] J. Si, A. Barto, W. Powell, and D. Wunsch. *Handbook of Learning and Approximate Dynamic Programming.* IEEE Press, (2004).

[121] S. Shalev-Shwartz. "Online Learning and Online Convex Optimization." *Foundations and Trends in Machine Learning*, **4**, 107, (2012).

[122] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, (1999).

[123] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning.* Cambridge University Press, (2014).

[124] V. N. Vapnik. "An overview of statistical learning theory." *IEEE transactions on neural networks*, **10**, 988, (1999).

[125] B. Efron. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics*, **7**, 1, (1979).

[126] Y. Bengio and Y. Grandvalet. "No unbiased estimator of the variance of k-fold cross-validation." *J. Mach. Learn. Res.*, **5**, 1089, (2004).

[127] G. C. Cawley and N. L. C. Talbot. "On over-fitting in model selection and subsequent selection bias in performance evaluation." *J. Mach. Learn. Res.*, **11**, 2079, (2010).

[128] I. Zutic, J. Fabian, and S. D. Sarma. "Spintronics: Fundamentals and applications." *Rev. Mod. Phys.*, **76**, 323, (2004).

[129] P. Mcguiness, O. Akdogan, A. Asali, S. Bance, F. Bittner, J. M. D. Coey, N. M. Dempsey, J. Fidler, D. Givord, O. Gutfleisch, M. Katter, D. Le Roy, S. Sanvito, T. Schrefl, L. Schultz, C. Schwöbl, M. Soderžnik, S. Šturm, P. Tozman, K. Üstüner, M. Venkatesan, T. G. Woodcock, K. Žagar, and S. Kobe. "Replacement and Original Magnet Engineering Options (ROMEOs): A European Seventh Framework Project to Develop Advanced Permanent Magnets Without, or with Reduced Use of, Critical Raw Materials." *JOM*, **67**, 1306, (2015).

[130] P. Tozman, M. Venkatesan, G. A. Zickler, J. Fidler, and J. M. D. Coey. "Enhanced energy product in Y-Co-Fe magnets intermediate between Nd-Fe-B and ferrite." *Appl. Phys. Lett.*, **107**, 032405, (2015).

[131] E. C. Stoner and E. P. Wohlfarth. "A Mechanism of Magnetic Hysteresis in Heterogeneous Alloys." *Philos. Trans. R. Soc. London, Ser. A*, **240**, 599, (1948).

[132] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Žic, T. Archer, P. Tozman, M. Venkatesan, J. Coey, and S. Curtarolo. "Accelerated discovery of new magnets in the Heusler alloy family." *Sci. Adv.*, (2017). Accepted.

[133] I would like to thank Dr. T. Archer for providing me with a direct access to the Materials Mine database.

[134] N. Drebov, A. Martinez-Limia, L. Kunz, A. Gola, T. Shigematsu, T. Eckl, P. Gumbsch, and C. Elsässer. "*Ab initio* screening methodology applied to the search for new permanent magnetic materials." *New J. Phys.*, **15**, 125023, (2013).

[135] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy. "AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations." *Comp. Mater. Sci.*, **58**, 227, (2012).

[136] J. Groff and P. Weinberg. *SQL The Complete Reference, 3rd Edition.* McGraw-Hill, Inc., New York, NY, USA, 3 edition, (2010).

[137] J. Kübler, A. R. William, and C. B. Sommers. "Formation and coupling of magnetic moments in Heusler alloys." *Phys. Rev. B*, **28**, 1745, (1983).

[138] For the KKR Jij calcualtions $R_{cut} = 1.0\ a_{lat}$ and a $20 \times 20 \times 20$ kpoint mesh was used. The FLEUR calcuations were made using a plane-wave cutoff parameter $k_{max} = 4.0\ a.u.^{-1}$ and a $14 \times 14 \times 14$ kpoint mesh. The Jij's were calculated for the first 200 neighbour shells, using a $9 \times 9 \times 9$ qpoint mesh. An exception is the Co_2FeSi, where we have calculated Jij's for the first 50 shells.

[139] H. Ebert, D. Ködderitzsch, and J. Minár. "Calculating condensed matter properties using the KKR-Green's function methodrecent developments and applications." *Rep. Prog. Phys.*, **74**, 096501, (2011).

[140] H. Ebert et al. "The Munich SPR-KKR package, version 6.3." http://ebert.cup.uni-muenchen.de/SPRKKR.

[141] http://www.flapw.de.

[142] C. G. Broyden. "A Class of Methods for Solving Nonlinear Simultaneous Equations." *Math. Comp.*, **19**, 577, (1965).

[143] I would like to thank Dr. Ph. Mavropoulos for suggesting this approach to me.

[144] K. Tillmanns. *Magnetic phase transitions from electronic structure theory.* M.S. thesis, RWTH Aachen, (2013).

[145] L. Wollmann, S. Chadov, J. Kübler, and C. Felser. "Magnetism in cubic manganese-rich Heusler compounds." *Phys. Rev. B*, **90**, 214420, (2014).

[146] L. Wollmann, S. Chadov, J. Kübler, and C. Felser. "Magnetism in tetragonal manganese-rich Heusler compounds." *Phys. Rev. B*, **92**, 064417, (2015).

[147] G. Kresse and J. Hafner. "Ab initio molecular dynamics for liquid metals." *Phys. Rev. B*, **47**, 558, (1993).

[148] G. Kresse and J. Hafner. "Ab initio molecular-dynamics simulation of the liquid-metalamorphous-semiconductor transition in germanium." *Phys. Rev. B*, **49**, 14251, (1994).

[149] G. Kresse and J. Furthmüller. "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set." *Comp. Mater. Sci.*, **6**, 15, (1996).

[150] G. Kresse and D. Joubert. "From ultrasoft pseudopotentials to the projector augmented-wave method." *Phys. Rev. B*, **59**, 1758, (1999).

[151] K. Rode, N. Baadji, D. Betto, Y.-C. Lau, H. Kurt, M. Venkatesan, P. Stamenov, S. Sanvito, J. M. D. Coey, E. Fonda, E. Otero, F. Choueikani, P. Ohresser, F. Porcher, and G. André. "Site-specific order and magnetism in tetragonal Mn_3Ga thin films." *Phys. Rev. B*, **87**, 184429, (2013).

[152] T. Gasi, V. Ksenofontov, J. Kiss, S. Chadov, A. Nayak, M. Nicklas, J. Winterlik, M. Schwall, P. Klaer, P. Adler, and C. Felser. "Iron-based Heusler compounds Fe_2YZ: Comparison with theoretical predictions of the crystal structure and magnetic properties." *Phys. Rev. B*, **87**, 064411, (2013).

[153] T. Graf, F. Casper, J. Winterlik, B. Balke, G. H. Fecher, and C. Felser. "Crystal Structure of New Heusler Compounds." *Zeitschrift für anorganische und allgemeine Chemie*, **635**, 976, (2009).

[154] M. Gilleßen and R. Dronskowski. "A combinatorial study of inverse Heusler alloys by first-principles computational methods." *J. Comp. Chem.*, **31**, 612, (2009).

[155] D. R. Lide (Editor). *Handbook of CHEMISTRY and PHYSICS.* CRC Press, (2004).

[156] K. J. G. *Introducion to Aluminium Alloys and Tempers.* ASM International, (2000).

[157] http://echa.europa.eu.

[158] J. F. Janak. "Uniform susceptibilities of metallic elements." *Phys. Rev. B*, **16**, 255, (1977).

[159] 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, (2011).

[160] T. M. Dunn. "Spin-orbit coupling in the first and second transition series." *Transactions of the Faraday Society*, **57**, 1441, (1961).

[161] T. Fawcett. "An introduction to ROC analysis." *Pattern Recogn. Lett.*, **27**, 861, (2006).

[162] C. D. Brown and H. T. Davis. "Receiver operating characteristics curves and related decision measures: A tutorial." *Chemometrics and Intelligent Laboratory Systems*, **80**, 24, (2006).

[163] H. Adachi and S. Maekawa. *Thermal Effects in Spintronics: Physics and Applications*, page 1553. Springer Netherlands, Dordrecht, (2016).

[164] M. J. Stevens, A. L. Smirl, R. D. R. Bhat, A. Najmaie, J. E. Sipe, and H. M. van Driel. "Quantum Interference Control of Ballistic Pure Spin Currents in Semiconductors." *Phys. Rev. Lett.*, **90**, 136603, (2003).

[165] A. Fert and I. A. Campbell. "Two-current conduction in nickel." *Phys. Rev. Lett.*, **21**, 1190, (1968).

[166] K. E. H. M. Hanssen and P. E. Mijnarends. "Positron-annihilation study of the half-metallic ferromagnet NiMnSb: theory." *Phys. Rev. B*, **34**, 5009, (1986).

[167] K. E. H. M. Hanssen, P. E. Mijnarends, L. P. L. M. Rabou, and K. H. J. Buschow. "Positron-annihilation study of the half-metallic ferromagnet NiMnSb: Experiment." *Phys. Rev. B*, **42**, 1533, (1990).

[168] M. I. Katsnelson, V. Y. Irkhin, L. Chioncel, A. I. Lichtenstein, and R. A. de Groot. "Half-metallic ferromagnets: From band structure to many-body effects." *Rev. Mod. Phys.*, **80**, 315, (2008).

[169] P. A. Dowben. "Are half-metallic ferromagnets half metals? (invited)." *J. Appl Phys.*, **95**, 7453, (2004).

[170] J. M. D. Coey and S. Sanvito. "Magnetic semiconductors and half-metals." *J. Phys. D: Appl. Phys.*, **37**, 988, (2004).

[171] P. Wadley, B. Howells, J. Železný, C. Andrews, V. Hills, R. P. Campion, V. Novak, K. Olejnik, F. Maccherozzi, S. S. Dhesi, S. Y. Martin, T. Wagner, J. Wunderlich, F. Freimuth, Y. Mokrousov, J. Kune, J. S. Chauhan, M. J. Grzybowski, A. W. Rushforth, K. W. Edmonds, B. L. Gallagher, and T. Jungwirth. "Electrical switching of an antiferromagnet." *Science*, **351**, 587, (2016).

[172] J. Železný, H. Gao, K. Výborný, J. Zemen, J. Mašek, A. Manchon, J. Wunderlich, J. Sinova, and T. Jungwirth. "Relativistic Néel-Order Fields Induced by Electrical Current in Antiferromagnets." *Phys. Rev. Lett.*, **113**, 157201, (2014).

[173] T. Jungwirth, X. Marti, P. Wadley, and J. Wunderlich. "Antiferromagnetic spintronics." *Nat. Nanotechnol.*, **11**, 231, (2016).

[174] X. Hu. "Half-Metallic Antiferromagnet as a Prospective Material for Spintronics." *Adv. Mater.*, **24**, 294, (2012).

[175] M. Hakimi, M. Venkatesan, K. Rode, K. Ackland, and J. M. D. Coey. "The zero-magnetization Heusler ferrimagnet." *J. Appl. Phys.*, **113**, 17B101, (2013).

[176] L. Yang, B. Liu, F. Meng, H. Liu, H. Luo, E. Liu, W. Wang, and G. Wu. "Magnetic properties of Heusler alloy Mn_2RuGe and Mn_2RuGa ribbons." *J. Magn. Magn. Mater.*, **379**, 1, (2015).

[177] T. Hori, M. Akimitsu, H. Miki, K. Ohoyoama, and Y. Yamaguchi. "Magnetic properties of $(Mn_{1-x}Ru_x)_3Ga$ alloys." *Appl. Phys. A-Mater.*, **74**, s737, (2002).

[178] S. Wurmehl, H. C. Kandpal, G. H. Fecher, and C. Felser. "Valence electron rules for prediction of half-metallic compensated-ferrimagnetic behaviour of Heusler compounds with complete spin polarization." *J. Phys.-Condens. Mat.*, **18**, 6171, (2006).

[179] I. Galanakis, P. Mavropoulos, and P. H. Dederichs. "Electronic structure and Slater–Pauling behaviour in half-metallic Heusler alloys calculated from first principles." *J. Phys. D Appl. Phys.*, **39**, 765, (2006).

[180] E. Şaşıoğlu. "Nonzero macroscopic magnetization in half-metallic antiferromagnets at finite temperatures." *Phys. Rev. B*, **79**, 100406, (2009).

[181] I. Galanakis, E. Şaşoğlu, S. Blügel, and K. Özdoğan. "Voids-driven breakdown of the local-symmetry and Slater-Pauling rule in half-metallic Heusler compounds." *Phys. Rev. B*, **90**, 064408, (2014).

[182] N. Thiyagarajah, Y.-C. Lau, D. Betto, K. Borisov, J. M. D. Coey, P. Stamenov, and K. Rode. "Giant spontaneous Hall effect in zero-moment Mn_2Ru_xGa." *Appl. Phys. Lett.*, **106**, 122402, (2015).

[183] I. Galanakis, K. Özdoğan, E. Şaşioğlu, and S. Blügel. "Effect of disorder on the magnetic properties of cubic Mn_2Ru_xGa compounds: A first-principles study." *J. Appl. Phys.*, **116**, 033903, (2014).

[184] D. Betto, N. Thiyagarajah, Y.-C. Lau, C. Piamonteze, M.-A. Arrio, P. Stamenov, J. M. D. Coey, and K. Rode. "Site-specific magnetism of half-metallic Mn_2Ru_xGa thin films determined by x-ray absorption spectroscopy." *Phys. Rev. B*, **91**, 094410, (2015).

[185] M. Žic, K. Rode, N. Thiyagarajah, Y.-C. Lau, D. Betto, J. M. D. Coey, S. Sanvito, K. J. O'Shea, C. A. Ferguson, D. A. MacLaren, and T. Archer. "Designing a fully compensated half-metallic ferrimagnet." *Phys. Rev. B*, **93**, 140202, (2016).

[186] M. Ležaić, P. Mavropoulos, and S. Blügel. "First-principles prediction of high Curie temperature for ferromagnetic bcc-Co and bcc-FeCo alloys and its relevance to tunneling magnetoresistance." *Appl. Phys. Lett.*, **90**, 082504, (2007).

[187] V. P. Antropov, M. I. Katsnelson, B. N. Harmon, M. van Schilfgaarde, and D. Kusnezov. "Spin dynamics in magnets: Equation of motion and finite temperature effects." *Phys. Rev. B*, **54**, 1019, (1996).

[188] M. Fähnle, R. Drautz, R. Singer, D. Steiauf, and D. V. Berkov. "A fast ab initio approach to the simulation of spin dynamics." *Comp. Mater. Sci.*, **32**, 118, (2005).

[189] B. Skubic, J. Hellsvik, L. Nordström, and O. Eriksson. "A method for atomistic spin dynamics simulations: implementation and examples." *J. Phys. Condens. Matter.*, **20**, 315203, (2008).

[190] G. Madsen and D. Singh. "BoltzTraP. A code for calculating bandstructure dependent quantities." *Comput. Phys. Commun.*, **175**, 67, (2006).

[191] W. W. Schulz, P. B. Allen, and N. Trivedi. "Hall coefficient of cubic metals." *Phys. Rev. B*, **45**, 10886, (1992).

[192] J. Kübler, A. R. William, and C. B. Sommers. "Formation and coupling of magnetic moments in Heusler alloys." *Phys. Rev. B*, **28**, 1745, (1983).

[193] H. Z. Luo, H. W. Zhang, Z. Y. Zhu, L. Ma, S. F. Xu, G. H. Wu, X. X. Zhu, C. B. Jiang, and H. B. Xu. "Half-metallic properties for the Mn_2FeZ (Z=Al, Ga, Si, Ge, Sb) Heusler alloys: A first-principles study." *J. Appl. Phys.*, **103**, 083908, (2008).

[194] N. Xing, H. Li, J. Dong, R. Long, and C. Zhang. "First-principle prediction of half-metallic ferrimagnetism of the Heusler alloys Mn_2CoZ (Z=Al, Ga, Si, Ge) with a high-ordered structure." *Comp. Mater. Sci.*, **42**, 600, (2008).

[195] S. Chadov, J. Kiss, and C. Felser. "Improving Spin-Transport by Disorder." *Adv. Funct. Mater*, **23**, 832, (2013).

[196] J. Chen, H. Luo, P. Jia, F. Meng, G. Liu, E. Liu, W. Wang, and G. Wu. "Site preference and electronic structure of Mn_2RuSn: A theoretical study." *J. Mag. Mag. Mater.*, **365**, 132, (2014).

[197] C.-L. Fu and K.-M. Ho. "First-principles calculation of the equilibrium ground-state properties of transition metals: Applications to Nb and Mo." *Phys. Rev. B*, **28**, 5480, (1983).

[198] F. D. Murnaghan. "The Compressibility of Media under Extreme Pressures." *Proc. Natl. Acad. Sci. USA*, **30**, 244, (1944).

[199] L. Makinistian, M. M. Faiz, R. P. Panguluri, B. Balke, S. Wurmehl, C. Felser, E. A. Albanesi, A. G. Petukhov, and B. Nadgorny. "On the half-metallicity of Co_2FeSi Heusler alloy: Point-contact Andreev reflection spectroscopy and *ab initio* study." *Phys. Rev. B*, **87**, (2013).

[200] O. Isayev, D. Fourches, E. N. Muratov, C. Oses, K. Rasch, A. Tropsha, and S. Curtarolo. "Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints." *Chem. Mater.*, **27**, 735, (2015).

[201] https://en.wikipedia.org/wiki/Data_scraping. Accessed: 2017-02-27.

[202] J. Hirschey. "Symbiotic relationships: Pragmatic acceptance of data scraping." *Berkeley Tech. LJ*, **29**, 897, (2014).

[203] 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, (2016).

[204] "The Materials Project." http://www.materialsproject.org/.

[205] "The Harvard Clean Energy Project Database." http://cepdb.molecularspace.org/.

[206] "OQMD: An Open Quantum Materials Database." http://oqmd.org/.

[207] "COMPUTATIONAL MATERIALS REPOSITORY." http://cmr.fysik.dtu.dk/.

[208] H. A. Jahn and E. Teller. "Stability of Polyatomic Molecules in Degenerate Electronic States. I. Orbital Degeneracy." *Proc. R. Soc. London, Ser. A*, **161**, 220, (1937).

[209] L. M. Mentel. "mendeleev." https://bitbucket.org/lukaszmentel/mendeleev., 2014.

[210] The radius of the point is calculated as $r = \exp\left(-\frac{\Delta H}{0.2\,\mathrm{eV}}\right)$, where ΔH is the enthalpy of formation.

[211] S. Wurmehl, G. H. Fecher, H. C. Kandpal, V. Ksenofontov, C. Felser, and H.-J. Lin. "Investigation of Co_2FeSi: The Heusler compound with highest Curie temperature and magnetic moment." *Appl. Phys. Lett.*, **88**, 032503, (2006).

[212] K. Kobayashi, R. Y. Umetsu, R. Kainuma, K. Ishida, T. Oyamada, A. Fujita, and K. Fukamichi. "Phase separation and magnetic properties of half-metal-type $Co_2Cr_{1-x}Fe_xAl$ alloys." *Appl. Phys. Lett.*, **85**, 4684, (2004).

[213] M. Zhang, Y. Cui, Z. Liu, G. Liu, J. Chen, G. Wu, Y. Sui, Y. Liu, Z. Qian, E. Brück, and F. R. de Boer. "A new Heusler compound Cu_2FeAl: electronic structure, magnetism and transport properties." *physica status solidi (a)*, **201**, 1570, (2004).

[214] W. Zhang, Z. Qian, J. Tang, L. Zhao, Y. Sui, H. Wang, Y. Li, W. Su, M. Zhang, Z. Liu, G. Liu, and G. Wu. "Superparamagnetic behaviour in melt-spun Ni_2FeAl ribbons." *J. Phys. Condens. Matter.*, **19**, 096214, (2007).

[215] Z. H. Liu, M. Zhang, Y. T. Cui, Y. Q. Zhou, W. H. Wang, G. H. Wu, X. X. Zhang, and G. Xiao. "Martensitic transformation and shape memory effect in ferromagnetic Heusler alloy Ni_2FeGa." *Appl. Phys. Lett.*, **82**, 424, (2003).

[216] J. C. Suits. "Structural Instability in new magnetic Heusler compounds." *Solid State Commun.*, **18**, 423, (1976).

[217] P. Entel, V. D. Buchelnikov, V. V. Khovailo, A. T. Zayak, W. A. Adeagbo, M. E. Gruner, H. C. Herper, and E. F. Wassermann. "Modelling the phase diagram of magnetic shape memory Heusler alloys." *J. Phys. D: Appl. Phys.*, **39**, 865, (2006).

[218] http://www.infomine.com/investment/metal-prices/. Accessed: 2017-02-27.

[219] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning.* Springer, (2013).