

ELECTRONIC STRUCTURE THEORY: APPLICATIONS AND GEOMETRICAL ASPECTS

BY SINISA COH

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Physics and Astronomy

Written under the direction of
David Vanderbilt
and approved by

New Brunswick, New Jersey

October, 2011

ABSTRACT OF THE DISSERTATION

Electronic structure theory: Applications and geometrical aspects

by Sinisa Coh

Dissertation Director: David Vanderbilt

This thesis contains several applications of the first-principles electronic-structure theory with special emphasis in parts of the thesis on the geometrical aspects of the theory. We start by reviewing the basics of the first-principles electronic-structure methods which are then used throughout the thesis. The first application of these methods is on the analysis of the stability and lattice dynamics of α - and β -cristobalite phases of SiO_2 . We also map the complete low-energy landscape connecting these two structures and give implications on the phase transition in this compound. Next we study a family of $Pbnm$ perovskites that are promising candidates for silicon-compatible high- K dielectrics. We calculate their structure and dielectric response, and compare with experimental results where available. The third application of these methods is to the large isosymmetric reorientation of oxygen octahedra rotation axes in epitaxially strained perovskites. We explain the origin of the peculiar energy landscape topology as

a function of epitaxial strain. In the part of the thesis devoted to the geometrical aspects of electronic structure theory, we begin by extending the concept of electronic polarization to a Chern insulators. These insulators are characterized by a non-zero off-diagonal σ_{xy} conductivity tensor component, quantized in units of e^2/h . Finally we discuss another geometrical quantity, the Chern-Simons orbital magnetoelectric coupling. We present a first-principles based calculation of this quantity in several compounds, and motivated by recent developments in the theory of topological insulators, we speculate about the existence of “large- θ materials,” in which this kind of coupling could be unusually large.

Acknowledgements

I had amazing time as a graduate student at Rutgers University working towards this thesis and here would like to thank everyone who helped make that happen.

I am grateful to fellow graduate students and postdocs at Rutgers for their support, companionship and all the good time we spent together. Thank you Aatish, Adina, Alexey, Aline, Andrei, Andrew, Anil, Anindya, Anthony, Chioun, Chuck, Daniel, Darakhshan, Deepak, Eamonn, George, Grant, James, John, Jun-hee, Kasturi, Kshitij, Lucia, Manjul, Marietta, Maryam, Michael, Oscar, Patrick, Rebecca, Roberto, Ryan, Simon, Sushmita, Sylvie, Tahir, Takeshi, Vijay, Yuval.

I have learned a lot from discussions with Darrell Schlom, Donald Hamann, Ivo Souza, James Rondinelli, Karin Rabe, Morrel Cohen and from classes by David Vanderbilt, Gregory Moore, Kristjan Haule, Premala Chandra and Sang-Wook Cheong.

I would especially like to thank my wife Senia, my parents Ćiril and Branka, siblings Smiljana, Dobriša and Tomislav for their unconditional love and support.

Finally, my thanks goes to David Vanderbilt for all inspiring discussions we had, challenging projects, guidance, humor and for being an overall amazing adviser.

Dedication

To everyone who suffered during the Yugoslav wars.

Table of Contents

Abstract	ii
Acknowledgements	iv
Dedication	v
List of Tables	xi
List of Figures	xii
1. Introduction	1
2. Methods	4
2.1. Parameter free method to compute properties of solids	4
2.1.1. Density-functional theory	5
2.1.2. Kohn-Sham reformulation	7
Collinear spin density	11
Noncollinear spin density	13
2.1.3. Periodic external potential	14
2.1.4. Pseudopotential theory	17
Norm-conserving pseudopotentials	19
Ultrasoft pseudopotentials	23
Projector augmented waves	25
2.1.5. Relativistic effects in solids	28

2.1.6.	Density-functional perturbation theory	32
2.2.	Use of the geometry of the wavefunctions	34
2.2.1.	Band insulators	35
2.2.2.	Gauge freedom	35
2.2.3.	Geometric quantities	36
2.2.4.	Geometric bulk physical quantities	38
2.2.5.	Topological insulators	42

I A few applications of total-energy electronic structure methods **44**

3.	Structural stability and lattice dynamics of SiO₂ cristobalite .	45
3.1.	Introduction	46
3.2.	Preliminaries	49
3.2.1.	Cristobalite structures	49
3.2.2.	Computational methods	51
3.3.	Results	51
3.3.1.	Structural properties of $\tilde{\alpha}$ and $\tilde{\beta}$ structures	51
	Comparison with rigid-unit geometry	54
3.3.2.	Phonons	56
	Phonons at Γ in $\tilde{\alpha}$ -cristobalite	56
	Phonons at Γ and M in $\tilde{\beta}$ -cristobalite	58
	Relation to unstable phonons in the cubic phase	60
	Comparison with experiment for α -cristobalite	61
3.3.3.	SiO ₂ cristobalite stability analysis	64
	Rigid unit mode analysis	66

Energy landscape inside 3D manifolds	68
Cell volume at minima and saddle point	74
Domain walls	75
3.4. Discussion	75
3.5. Summary	81
4. Si-compatible candidates for high-K dielectrics with the Pbnm perovskite structure	82
4.1. Introduction	83
4.1.1. Compounds under consideration	85
4.2. Preliminaries	86
4.2.1. Structure of Pbnm perovskites	86
4.2.2. Computational methods	86
4.3. Results and discussion	91
4.3.1. Structural properties	91
Rare-earth scandates	91
Rare-earth yttrates	92
CaZrO ₃ , SrZrO ₃ , and SrHfO ₃	92
La ₂ BB'O ₆ compounds	92
Rare-earth rare-earth perovskites	93
4.3.2. Comparison with model of perfectly rigid octahedra	99
4.3.3. Dielectric properties	100
Rare-earth scandates	101
Rare-earth yttrates	102
CaZrO ₃ , SrZrO ₃ , and SrHfO ₃	102
La ₂ BB'O ₆ compounds	103
Rare-earth rare-earth perovskites	103

4.3.4.	Decomposition of the ionic contribution to the dielectric tensor	109
4.3.5.	Compounds with $R\bar{3}c$ symmetry	111
4.3.6.	Correlation between structural and dielectric properties . .	112
4.3.7.	Antisite substitutions	113
	Analysis of antisite defects in LaLuO_3	114
	Discussion	116
4.4.	Summary	116
5.	Large isosymmetric reorientation of oxygen octahedra rotation axes in epitaxially strained perovskites	119
5.1.	Introduction	119
5.2.	Calculation details and notation	121
5.3.	Strain-stabilized structures	123
	5.3.1. Microscopic structure evolution	124
5.4.	Origin of the isosymmetric transition	125
5.5.	Accessing and applications of the isosymmetric transition	129
5.6.	Conclusion	130
II	Geometrical aspects of electronic structure theory	131
6.	Electric polarization in a Chern insulator	132
6.1.	Introduction and motivation	132
6.2.	Results	134
6.3.	Conclusion	143
7.	Chern-Simons orbital magnetoelectric coupling in generic insulators	144

7.1.	Introduction	144
7.2.	Background and motivation	147
7.2.1.	Units and conventions	147
7.2.2.	Theory of orbital magnetoelectric coupling	148
7.2.3.	Relation between bulk and surface properties	149
	Electric polarization and surface charge	150
	OMP and surface anomalous Hall conductivity	151
7.2.4.	Motivation and relationship to strong \mathbb{Z}_2 topological insulators	154
	Time-reversal symmetry constraints on θ	154
	Prospects for large- θ materials	157
7.2.5.	General symmetry considerations	159
7.3.	Methods	160
7.3.1.	Review of Berry formalism	160
7.3.2.	Numerical evaluation of θ	163
7.3.3.	Computational details	167
7.4.	Results and discussion	168
7.4.1.	Conventional magnetoelectrics	168
	Calculation of θ in Cr_2O_3	168
	Other conventional magnetoelectrics	171
7.4.2.	Strong \mathbb{Z}_2 topological insulators	172
7.4.3.	\mathbb{Z}_2 -derived nontopological insulators with broken symmetries	174
7.5.	Summary and outlook	176
8.	Conclusion	180
	References	182

List of Tables

3.1. Lattice constants (in Å) and Wyckoff structural parameters for $\tilde{\alpha}$ ($P4_12_12$) and $\tilde{\beta}$ ($I\bar{4}2d$) cristobalite SiO_2	53
3.2. Infrared-active phonon modes at Γ in $\tilde{\alpha}$ -cristobalite ($P4_12_12$). . .	57
3.3. Raman-only phonon modes at Γ in $\tilde{\alpha}$ -cristobalite ($P4_12_12$).	58
3.4. Γ phonons in $\tilde{\beta}$ -cristobalite ($I\bar{4}2d$).	59
3.5. M phonons in $\tilde{\beta}$ -cristobalite ($I\bar{4}2d$).	59
3.6. Relations between unstable phonons in “ideal structure” and phonons in $\tilde{\alpha}$ and $\tilde{\beta}$ structures.	60
3.7. Computed mode frequencies for α -cristobalite and tentative assignments to measured mode frequencies.	62
4.1. Structural parameters of the $Pbnm$ perovskites we considered. . .	94
4.2. Structural parameters of the $\text{La}_2\text{BB}'\text{O}_6$ perovskites we considered.	97
4.3. Dielectric parameters of the perovskites with space group $Pbnm$ that we considered.	105
4.4. Dielectric parameters of the $\text{La}_2\text{BB}'\text{O}_6$ perovskites we considered.	107
7.1. Magnetic point groups for which a generic non-zero CSOMP is allowed by symmetry.	160

List of Figures

3.1. Projection on x - y plane of the $\tilde{\alpha}$ and $\tilde{\beta}$ structures.	50
3.2. Ground state energy vs. volume for $\tilde{\alpha}$, $\tilde{\beta}$, and cubic cristobalite structures of SiO_2	52
3.3. Structural parameters of the $\tilde{\alpha}$ structure vs. volume per formula unit.	55
3.4. Structural parameters of the $\tilde{\beta}$ structure vs. volume per formula unit.	56
3.5. Five linearized RUMs in ideal cristobalite $Z=4$ structure.	66
3.6. Energy in plane defined by the $\tilde{\alpha}$ structure, $\tilde{\beta}$ structure, and saddle point.	69
3.7. Energy landscape in the plane defined by ϕ_{α_1} and ϕ_{β_1} rotation angles.	71
3.8. Energy landscape in the plane defined by ϕ_{α_1} and $\phi_{\alpha'_1}$ rotation angles.	73
3.9. Sketch of important states in one of the three-dimensional rigid-unit subspaces discussed in the text.	77
4.1. Photographs of single crystal rare-earth scandates grown by the Czochralski method.	84
4.2. 20-atom primitive cell of a $Pbnm$ -distorted ABO_3 perovskite.	87
4.3. Projection on the a - c plane of the structure with $\theta_R > 0$ or with $\theta_M > 0$	87
4.4. Unit cell volumes of all rock-salt rare-earth nitrides.	89

4.5.	Structural information for all systems we considered.	98
4.6.	Dielectric information for all systems we considered.	108
4.7.	Eigenvalues of a force constant matrix $\Phi_{i\alpha,j\beta}$	110
4.8.	Correlation between the dielectric tensor z -anisotropy and the mismatch in the oxygen octahedra rotation angle.	114
5.1.	Evolution of the total energy as a function of epitaxial strain. . .	122
5.2.	GaO ₆ rotation angles, the octahedral distortion parameter Δ , and the La displacements.	123
5.3.	Energy landscape evolution as a function of strain.	127
6.1.	Sketch of $\theta_1(k_2)$ in a Chern insulator and computed $P_1(\alpha)$ and $\mathcal{P}_1(\alpha)$ for the modified Haldane model.	136
6.2.	Sketch of a band structure of a finite ribbon of a Chern insulator.	139
6.3.	Hybrid Wannier function centers.	142
7.1.	Identical samples cut from a strong \mathbb{Z}_2 topological insulator, but with two different surface preparations.	155
7.2.	Schematic view of the allowable values of θ in different parts of the two-parameter space of some unspecified model Hamiltonian. . . .	158
7.3.	Graphical interpretation of Eqs. (7.20), (7.21) and (7.22).	162
7.4.	Rhombohedral unit cell of Cr ₂ O ₃ and schematic of hexagonal unit cell of Bi ₂ Se ₃ with imposed local Zeeman field on Bi atoms.	169
7.5.	Calculated value of θ in Cr ₂ O ₃ for varying densities of k -space grids.	170
7.6.	Calculated θ in Cr ₂ O ₃ as a function of spin-orbit coupling strength.	172
7.7.	Calculated value of θ in Bi ₂ Se ₃ for varying densities of k -space grids.	174
7.8.	Calculated value of θ and induced magnetic moment on the Bi atom for Bi ₂ Se ₃ with artificially applied staggered Zeeman field on Bi atoms.	175

Chapter 1

Introduction

Confused. To shake it out of its state of complacency. I'm afraid I'm not personally qualified to confuse cats, but I can recommend an extremely good service. Here is their card.

Episode 5 of Monty Python's Flying Circus

Electronic structure theory based on the first-principles pseudopotential density-functional theory lies at the center of the modern theoretical understanding of materials. By allowing for a direct solution of the fundamental quantum mechanical equations on a computer, these methods have made possible, using no adjustable parameters, the calculation of properties of the materials even before they are created in the laboratory. This thesis contains several applications of these electronic-structure methods.

Special attention is given in parts of this thesis to the geometrical aspects of electronic structure theory. It has been realized only recently that there are some specific properties of periodic solids that can be calculated from the geometric (i.e. topological) properties of the electron wave functions themselves. These are the electronic polarization (in 1993), the anomalous Hall conductivity (in 1999), and the Chern-Simons orbital magnetoelectric coupling (in 2008).

Chapter 2 introduces some basic methods used throughout the thesis, such as

the density-functional theory, various pseudopotential methods, and the density-functional perturbation theory. There is also a discussion of the Berry potential, Berry curvature, and other geometrical quantities.

The main results of this thesis are presented in Chapters 3 through 7. Specifically, in Chapter 3 we present a study of the structural stability and lattice dynamics of SiO_2 in its α - and β -cristobalite phases. We characterize the low-lying energy landscape connecting these phases and give possible implications for the phase-transition mechanism.

In Chapter 4 we compute from first principles the dielectric tensor components and crystal structures of five classes of perovskites with the $Pbnm$ space group. This group of materials is believed to be stable in contact with silicon and is therefore a good candidate for technologically important high- K dielectrics.

The third application of first-principles methods is presented in Chapter 5. Here we show how bi-axial epitaxial strain can induce anomalously large structural changes in the perovskites. We present a detailed analysis of the origin of such structural changes.

In the second part of the thesis we focus on geometrical aspects of electronic structure theory. In Chapter 6 we discuss the extension of the Berry-phase concept of polarization to a special kind of insulators, namely, to the so called Chern insulators or the anomalous quantum Hall insulators. This generalization requires special care because of the partial occupation of chiral edge states.

Finally, in Chapter 7 we present a first-principles based method to calculate the Chern-Simons orbital magnetoelectric coupling in various materials. Motivated by the recent developments in the connection with the theory of strong \mathbb{Z}_2 topological insulators, we suggest that, in special cases, this coupling could be as large or larger than the total magnetoelectric coupling in known magnetoelectrics like Cr_2O_3 .

The work in Chapters 3 through 7 is based on Refs. [1, 2, 3, 4, 5] respectively.
We conclude this thesis in Chapter 8.

Chapter 2

Methods

In this chapter we briefly review some of the concepts used throughout this thesis. In the first part we give an overview of density-functional theory followed by a discussion of various pseudopotential techniques, including norm-conserving pseudopotentials, ultrasoft pseudopotentials, and the projector augmented waves method. Next we discuss the treatment of relativistic effects in solids, and mention the density-functional perturbation theory. In the second part of this overview we discuss the geometrical quantities in the theory of periodic solids, and give some material properties which can be expressed in terms of these quantities.

2.1 Parameter free method to compute properties of solids

The quantum mechanical description of interacting electrons and nuclei is given by the Hamiltonian

$$\begin{aligned}
 H = & -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - \frac{\hbar^2}{2M_I} \sum_I \nabla_I^2 \\
 & + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}
 \end{aligned} \tag{2.1}$$

where m_e and M_I are electron and ion mass respectively, Z_I is the atomic number of I -th ion, e is the electron charge, and \hbar is the reduced Planck constant. Since ionic masses are much larger than the mass of electron ($M_I \gg m_e$) we can treat the ions as if they are static (Born-Oppenheimer approximation). Therefore we can discard the ion kinetic energy from Eq. (2.1) and treat the ionic positions \mathbf{R}_I as external parameters in the electronic Hamiltonian

$$H = -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 + \sum_i V_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + E_{\text{ion}}. \quad (2.2)$$

Here

$$V_{\text{ext}}(\mathbf{r}) = \sum_I \frac{Z_I e^2}{|\mathbf{r} - \mathbf{R}_I|} \quad (2.3)$$

is the external potential created by the ions, and E_{ion} is the constant term arising from the ion-ion interaction.

2.1.1 Density-functional theory

Density-functional theory provides us an approximate way to solve the many-body electron problem defined by the Hamiltonian Eq. (2.2). This theory is based on two theorems proved by Hohenberg and Kohn in Ref. [6] and in a reformulation of the theory by Kohn and Sham in Ref. [7]. For simplicity here we deal with the case that the ground state of Hamiltonian Eq. (2.2) is nondegenerate, but the generalization to the degenerate case exists as well Ref. [8].

We begin by stating first the Hohenberg-Kohn theorem.

Theorem A: There exists a one-to-one map between the external potential $V_{\text{ext}}(\mathbf{r})$ and the ground state electron density $n(\mathbf{r})$.

Proof. For a given external potential $V_{\text{ext}}(\mathbf{r})$ there exists a unique ground-state wavefunction Ψ and therefore a unique density $n(\mathbf{r})$. Now let us prove that the ground state density $n(\mathbf{r})$ uniquely determines the external potential V_{ext} .

Let us assume that there exist two external potentials V_{ext} and V'_{ext} with the same ground-state density $n(\mathbf{r})$. Unless V_{ext} and V'_{ext} differ by a constant, the corresponding ground-state wavefunctions Ψ and Ψ' will be different since they correspond to different Hamiltonians H and H' .

By the variational principle we have

$$\langle \Psi | H | \Psi \rangle < \langle \Psi' | H | \Psi' \rangle \quad (2.4)$$

since Ψ is a true ground state of H . Therefore we conclude that

$$E < E' + \langle \Psi' | V - V' | \Psi' \rangle, \quad (2.5)$$

$$E < E' + \int [V(\mathbf{r}) - V'(\mathbf{r})] n(\mathbf{r}) d\mathbf{r}. \quad (2.6)$$

If we started with the variational principle for Ψ' instead of Ψ , we would have concluded

$$E' < E + \int [V'(\mathbf{r}) - V(\mathbf{r})] n(\mathbf{r}) d\mathbf{r}. \quad (2.7)$$

This leads to the contradiction $E + E' < E + E'$, and therefore it is impossible that two external potentials result in the same ground-state density.

Next let us state second theorem by Hohenberg and Kohn.

Theorem B: There exists a universal functional $F[n]$ independent of V_{ext} such that the functional $E[n] = F[n] + \int V_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{\text{ion}}$ has the following properties: in the space of all densities n such that $\int n(\mathbf{r})d\mathbf{r} = N$, the functional $E[n]$ has a global minimum at the exact ground state density, and the minimal energy

is equal to the exact ground-state energy.

Proof. For simplicity we will work with V-representable electron densities, defined as a density corresponding to the ground-state density of some one-particle Hamiltonian with some external potential V_{ext} .

Let us define a functional $F[n]$ in the following way. By virtue of previous theorem, to a given V-representable density $n(\mathbf{r})$ one can assign a ground-state wavefunction $\Psi[n]$ corresponding to some external potential V_{ext} . Now we define $F[n]$ as the expectation of the kinetic energy and the electron-electron interaction energy for that wavefunction $\Psi[n]$:

$$F[n] = -\frac{\hbar^2}{2m_e} \sum_i \langle \Psi[n] | \nabla_i^2 | \Psi[n] \rangle + \frac{e^2}{2} \sum_{i \neq j} \langle \Psi[n] | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \Psi[n] \rangle. \quad (2.8)$$

From Eq. (2.2) it is clear that ground-state energy corresponding to density $n(\mathbf{r})$ is

$$E[n] = F[n] + \int V_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{\text{ion}} \quad (2.9)$$

and that this functional is minimized by the exact ground-state density by virtue of the variational principle.

2.1.2 Kohn-Sham reformulation

Theorem B from the previous section stated that in order to find the exact ground-state energy and density, one needs to find a minimum of the corresponding functional given in Eq. (2.9) under appropriate constraints.

Since it is otherwise hard to find useful approximations for the universal functional $F[n]$, one often works with the Kohn-Sham formulation of the variational problem following theorem B. Let us first consider an auxiliary noninteracting

N -electron system described by

$$H_{\text{aux}} = -\frac{\hbar^2}{2m_e}\nabla^2 + V_{\text{KS}}(\mathbf{r}). \quad (2.10)$$

The single-particle solutions $\phi_i(\mathbf{r})$ of this system satisfy

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V_{\text{KS}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}), \quad (2.11)$$

and if one solves for the N lowest-energy solutions, then the corresponding ground-state electron density is given simply by

$$n(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2. \quad (2.12)$$

Even though no rigorous proof exists [9], we will assume that the exact ground-density can be expressed in terms of Eq. (2.12), i.e., in terms of the density of some auxiliary non-interacting electron problem.

The key idea of the Kohn-Sham reformulation is to express the density in Eq. (2.9) in terms of solutions ϕ_i of some non-interacting system, and then minimize the energy functional in Eq. (2.9) with respect to the ϕ_i . As we are about to show, the reason for doing this is that this reformulation allows for practical approximations to the universal functional $F[n]$.

Let us first define the kinetic (T_{KS}) and Hartree (E_{H}) energy of our auxiliary non-interacting system described by H_{aux} as

$$T_{\text{KS}}[n] = -\frac{\hbar^2}{2m_e} \sum_i \langle \phi_i[n] | \nabla^2 | \phi_i[n] \rangle, \quad (2.13)$$

$$E_{\text{H}}[n] = \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \quad (2.14)$$

Here T_{KS} is a functional of density since theorem A applied to the auxiliary

Hamiltonian H_{aux} guarantees that V_{KS} is uniquely given by the ground state density $n(\mathbf{r})$, and therefore the solutions ϕ_i are also given uniquely by the density $n(\mathbf{r})$.

By adding and subtracting T_{KS} and E_{H} to the energy functional in Eq. (2.9), we can rewrite it as

$$\begin{aligned} E[n] &= (T_{\text{KS}}[n] - T_{\text{KS}}[n] + E_{\text{H}}[n] - E_{\text{H}}[n]) + F[n] + \int V_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{\text{ion}} \\ &= T_{\text{KS}}[n] + E_{\text{H}}[n] + \int V_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{\text{ion}} + E_{\text{xc}}[n] \end{aligned} \quad (2.15)$$

where we have defined the *exchange-correlation* functional as

$$E_{\text{xc}}[n] = F[n] - T_{\text{KS}}[n] - E_{\text{H}}[n]. \quad (2.16)$$

Therefore E_{xc} is defined as the difference between the exact expectation value of the kinetic energy and electron-electron interaction and the expectation value of the kinetic energy T_{KS} and Hartree energy E_{H} of the auxiliary system.

By removing the auxiliary kinetic energy T_{KS} and the long-ranged Hartree energy E_{H} from the functional $F[n]$, we have arrived at a functional E_{xc} which can be much better approximated than F itself. This is the main reason for the Kohn-Sham reformulation.

With the help of Eq. (2.12), we have rewritten the density in terms of independent-electron orbitals $\phi_i(\mathbf{r})$. Therefore, in searching for the minimal value of the energy functional Eq. (2.15) (or Eq. (2.9)), we can vary over the orbitals $\phi_i(\mathbf{r})$ instead of the density $n(\mathbf{r})$ itself. The condition that $E[n]$ does not change under variation of the orbitals ϕ_i , or formally that

$$\frac{\delta E}{\delta \phi_i(\mathbf{r})} = 0 \quad (2.17)$$

under constraint $\langle \phi_i | \phi_j \rangle = \delta_{ij}$, leads to the requirement that the orbitals ϕ_i need to satisfy Eq. (2.10) with potential V_{KS} equal to

$$V_{\text{KS}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}). \quad (2.18)$$

Here we have defined V_{H} and V_{xc} as

$$V_{\text{H}}(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (2.19)$$

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}}{\delta n(\mathbf{r})}. \quad (2.20)$$

In summary, in the Kohn-Sham formulation of the density functional theory, instead of directly minimizing the energy functional $E[n]$, one has to find the N lowest-energy solutions of the Kohn-Sham equation

$$H_{\text{KS}}\phi_i(\mathbf{r}) = \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{xc}}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}). \quad (2.21)$$

The density $n(\mathbf{r})$ corresponding to these orbitals $\phi_i(\mathbf{r})$ (given by Eq. (2.12)) is then the exact ground state density, and $E[n]$ is the exact ground state energy. The only unknown functional in Eq. (2.21) is the exchange-correlation potential V_{xc} , and in practical calculations one needs to make an appropriate approximation for V_{xc} .

In local density approximation (LDA) to the density functional theory one assumes that E_{xc} is given by

$$E_{\text{xc}}[n] = \int n(\mathbf{r}) \epsilon_{\text{xc}}^{\text{hom}}[n(\mathbf{r})] d\mathbf{r}. \quad (2.22)$$

Here $\epsilon_{\text{xc}}^{\text{hom}}(n)$ is the exchange-correlation energy density of a homogeneous electron

gas with the density n . For function ϵ_{xc} one often uses various parametrization of the results of a very precise Monte-Carlo calculation given in Ref. [10].

It is important to note that eigenfunction and eigenvalues of the Kohn-Sham Hamiltonian in Eq. (2.21) can not be interpreted as real electron wavefunctions and eigenenergies. Kohn-Sham orbitals were introduced in Eq. (2.11) only as a convenient way to parametrize the total electron density and introduce practical approximation to the exchange-correlation functional $E_{\text{xc}}[n]$. Consistent with inability to interpret Kohn-Sham eigenenergies as real electron energies, total electronic energy is not simply given by the sum of occupied Kohn-Sham eigenenergies ϵ_i from Eq. (2.21), instead one has to evaluate explicitly energy functional $E[n]$ from Eq. (2.15) for the density n given by Eq. (2.12) through the Kohn-Sham eigenfunctions. In more convenient form for the numerical calculations, total electronic energy E can be calculated from the Kohn-Sham eigenenergies *and* the ground state density $n(\mathbf{r})$ as

$$E = \sum_i \epsilon_i - \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{\text{xc}} - \int V_{\text{xc}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}, \quad (2.23)$$

which can easily be shown by multiplying Eq. (2.21) with $\phi_i^*(\mathbf{r})$ from the left and integrating over the whole space.

Collinear spin density

Density-functional theory as presented so far has been formulated in terms of the electron density $n(\mathbf{r})$ alone. In some cases it is useful to formulate the theory in terms of the spin polarized densities $n_{\uparrow}(\mathbf{r})$ and $n_{\downarrow}(\mathbf{r})$, defined as the densities of the up and down components of spin respectively.

In that case one arrives at the Kohn-Sham equation separately for up and

down Kohn-Sham orbitals $\phi_{i\sigma}(\mathbf{r})$ with σ denoting the spin index,

$$\left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{xc}}^\sigma(\mathbf{r}) \right] \phi_{i\sigma}(\mathbf{r}) = \epsilon_{i\sigma} \phi_{i\sigma}(\mathbf{r}). \quad (2.24)$$

Since the theory is now formulated in terms of both $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$, the exchange-correlation potential $V_{\text{xc}}^\sigma(\mathbf{r}) = \delta E_{\text{xc}} / \delta n^\sigma(\mathbf{r})$ becomes spin-dependent.

As for the LDA, one can make a local spin density approximation (LSDA) to the exchange-correlation energy functional

$$E_{\text{xc}}[n_\uparrow, n_\downarrow] = \int n(\mathbf{r}) \epsilon_{\text{xc}}^{\text{hom}} [n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})] d\mathbf{r} \quad (2.25)$$

where $\epsilon_{\text{xc}}^{\text{hom}} [n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})]$ is the energy density of the homogeneous electron gas with constant spin up and spin down densities n_\uparrow and n_\downarrow .

Instead of using spin densities $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$ as variables, one often uses total electron density $n(\mathbf{r}) = n_\uparrow(\mathbf{r}) + n_\downarrow(\mathbf{r})$ and local spin magnetization $m(\mathbf{r}) = \mu_e [n_\uparrow(\mathbf{r}) - n_\downarrow(\mathbf{r})]$, where μ_e is the magnetic moment of an electron. Now, instead of using $V_{\text{xc}}^\sigma(\mathbf{r}) = \delta E_{\text{xc}} / \delta n^\sigma(\mathbf{r})$ we can use average exchange-correlation potential $\bar{V}_{\text{xc}}(\mathbf{r}) = \delta E_{\text{xc}} / \delta n(\mathbf{r})$ and effective local magnetic field $B_{\text{xc}}(\mathbf{r}) = -\delta E_{\text{xc}} / \delta m(\mathbf{r})$. With these definitions, Kohn-Sham equations from Eq. (2.24) become,

$$\begin{aligned} \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \bar{V}_{\text{xc}}(\mathbf{r}) - \mu_e B_{\text{xc}}(\mathbf{r}) \right] \phi_{i\uparrow}(\mathbf{r}) &= \epsilon_{i\uparrow} \phi_{i\uparrow}(\mathbf{r}), \\ \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \bar{V}_{\text{xc}}(\mathbf{r}) + \mu_e B_{\text{xc}}(\mathbf{r}) \right] \phi_{i\downarrow}(\mathbf{r}) &= \epsilon_{i\downarrow} \phi_{i\downarrow}(\mathbf{r}). \end{aligned} \quad (2.26)$$

Total electronic energy E can now be computed analogously as in the spin unpolarized case,

$$E = \sum_i \epsilon_i - \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{\text{xc}} - \int \bar{V}_{\text{xc}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \int B_{\text{xc}}(\mathbf{r})m(\mathbf{r})d\mathbf{r}. \quad (2.27)$$

Noncollinear spin density

More generally, one can formulate density-functional theory in terms of the full spin-density matrix $n_{\sigma\sigma'}(\mathbf{r})$. The Kohn-Sham equation in this case becomes

$$\left\{ \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right] \delta_{\sigma\sigma'} + V_{\text{xc}}^{\sigma\sigma'}(\mathbf{r}) \right\} \phi_{i\sigma'}(\mathbf{r}) = \epsilon_i \phi_{i\sigma}(\mathbf{r}) \quad (2.28)$$

where the exchange-correlation potential $V_{\text{xc}}^{\sigma\sigma'}(\mathbf{r}) = \delta E_{\text{xc}} / \delta n^{\sigma\sigma'}(\mathbf{r})$ can have non-diagonal components as well and sum over σ' index is implicit. This functional is usually constructed by diagonalizing locally at each point \mathbf{r} a density matrix $n_{\sigma\sigma'}(\mathbf{r})$, and then using the same homogeneous electron gas kernel $\epsilon_{\text{xc}}^{\text{hom}}[n_{\uparrow}, n_{\downarrow}]$ as in Eq. (2.25) for these two diagonal components.

Analogously as in the case of the collinear spin, we can perform a change of variables from $n_{\sigma\sigma'}(\mathbf{r})$ to $n(\mathbf{r})$ and $\mathbf{m}(\mathbf{r})$ defined as

$$n(\mathbf{r}) = \text{Tr} [n_{\sigma\sigma'}(\mathbf{r})] \quad \text{and} \quad \mathbf{m}(\mathbf{r}) = \mu_e \sum_{\sigma_1\sigma_2} \boldsymbol{\sigma}_{\sigma_1\sigma_2} n_{\sigma_1\sigma_2}(\mathbf{r}), \quad (2.29)$$

where $\boldsymbol{\sigma}$ is a vector of three Pauli matrices. Again, we define average exchange-correlation potential $\bar{V}_{\text{xc}}(\mathbf{r}) = \delta E_{\text{xc}} / \delta n(\mathbf{r})$ and effective local magnetic field vector

$\mathbf{B}_{\text{xc}}(\mathbf{r}) = -\delta E_{\text{xc}}/\delta \mathbf{m}(\mathbf{r})$ and now Kohn-Sham equation Eq. (2.28) reads

$$\left\{ \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \bar{V}_{\text{xc}}(\mathbf{r}) \right] \delta_{\sigma\sigma'} - \mu_e \mathbf{B}_{\text{xc}}(\mathbf{r}) \cdot \boldsymbol{\sigma}_{\sigma\sigma'} \right\} \phi_{i\sigma'}(\mathbf{r}) = \epsilon_i \phi_{i\sigma}(\mathbf{r}). \quad (2.30)$$

Total energy in the case of noncollinear calculation can be calculated similarly as in the collinear case through

$$E = \sum_i \epsilon_i - \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{\text{xc}} - \int \bar{V}_{\text{xc}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \int \mathbf{B}_{\text{xc}}(\mathbf{r}) \cdot \mathbf{m}(\mathbf{r})d\mathbf{r}. \quad (2.31)$$

2.1.3 Periodic external potential

We will often be interested in solving the Kohn-Sham Eq. (2.21) with a periodic external potential V_{ext} corresponding to some periodic arrangement of ions \mathbf{R}_I . In that case the total potential V_{KS} acting on the Kohn-Sham orbitals will be periodic as well. For simplicity, in this section we will refer to V_{KS} simply as $V(\mathbf{r})$ since this analysis will pertain to any periodic external potential.

Here we will review the formulation of Bloch's theorem, which applies in such a case.

The periodicity of $V(\mathbf{r})$ in three dimensions is expressed as

$$V(\mathbf{r} + \mathbf{a}_i) = V(\mathbf{r}) \quad (2.32)$$

for some triplet of vectors \mathbf{a}_i not in the same plane. (For now we consider only

three-dimensional crystals; the generalization to d dimensions is trivial.) Therefore there exists an entire lattice of translation vectors

$$\mathbf{R}_m = m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3 \quad (2.33)$$

labeled by a triplet of integer indices $m = \{m_1, m_2, m_3\}$ which keep V invariant as in Eq. (2.32). Equivalently, we can say that the translation operators T_m corresponding to vector \mathbf{R}_m commute with Hamiltonian H .

Now let us first formally introduce the concept of the reciprocal vector \mathbf{G} as a vector satisfying

$$e^{i\mathbf{G}\cdot\mathbf{R}_m} = 1 \quad (2.34)$$

for all translation vectors \mathbf{R}_m . One can show that the set of all such vectors \mathbf{G} again form a lattice and can all be expressed in the terms of triplet of vectors \mathbf{b}_i

$$\mathbf{G}_l = l_1 \mathbf{b}_1 + l_2 \mathbf{b}_2 + l_3 \mathbf{b}_3, \quad (2.35)$$

and integers l_i . Using Eq. (2.34) one can show that vectors \mathbf{b}_i can be defined in terms of vectors \mathbf{a}_i through the nine relations $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$.

Since the set of translation operators T_m corresponding to translation vectors \mathbf{R}_m commute with each other and with the Hamiltonian H , there exists a basis of eigenstates $\phi_i(\mathbf{r})$ which simultaneously diagonalizes all translation operators T_m and the Hamiltonian H . Furthermore, it can be shown that these basis functions can be put in the form

$$\phi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \quad (2.36)$$

where $u_{n\mathbf{k}}$ is a function satisfying the *cell-periodicity* condition

$$u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}_m) = u_{n\mathbf{k}}(\mathbf{r}). \quad (2.37)$$

Here \mathbf{k} is an arbitrary 3-vector and n is an integer labeling states with the same \mathbf{k} . This is the well known Bloch theorem.

Since the planewave function $\exp(i\mathbf{G} \cdot \mathbf{r})$ for any reciprocal vector \mathbf{G} is itself a cell-periodic function, both $\phi_{n\mathbf{k}}(\mathbf{r})$ and $\phi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r})$ can be written in the same form of $e^{i\mathbf{k} \cdot \mathbf{r}}$ times the cell-periodic function. Therefore we can impose a more strict form of ansatz Eq. (2.36)

$$\phi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \phi_{n\mathbf{k}}(\mathbf{r}). \quad (2.38)$$

which gives rise to the following condition

$$u_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}) = e^{-i\mathbf{G} \cdot \mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}). \quad (2.39)$$

Therefore given a state labeled by $\{n, \mathbf{k}\}$, the states at $\{n, \mathbf{k} + \mathbf{G}\}$ are redundant for any reciprocal vector $\mathbf{G} \neq \mathbf{0}$.

Using Eq. (2.36) as an ansatz to the solution of Hamiltonian

$$\left[-\frac{\hbar^2}{2m_e} \nabla^2 + V(\mathbf{r}) \right] \phi_{n\mathbf{k}}(\mathbf{r}) = \epsilon_{n\mathbf{k}} \phi_{n\mathbf{k}}(\mathbf{r}) \quad (2.40)$$

we arrive at the equation

$$\left[-\frac{\hbar^2}{2m_e} (\nabla + i\mathbf{k})^2 + V(\mathbf{r}) \right] u_{n\mathbf{k}}(\mathbf{r}) = \epsilon_{n\mathbf{k}} u_{n\mathbf{k}}(\mathbf{r}) \quad (2.41)$$

for $u_{n\mathbf{k}}$ itself. Therefore the cell-periodic functions $u_{n\mathbf{k}}(\mathbf{r})$ are not eigenvectors of

H , but of $H(\mathbf{k})$ defined as

$$H(\mathbf{k}) = -\frac{\hbar^2}{2m_e} (\nabla + i\mathbf{k})^2 + V(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (2.42)$$

We can think of the set of solutions $u_{n\mathbf{k}}$ for some chosen range of n as eigenvectors of a family of Hamiltonians $H_{\mathbf{k}}$ labeled by the continuous variable \mathbf{k} . If we further note that we need to solve these Hamiltonians in the space of periodic functions satisfying Eq. (2.37), our problem is completely defined.

Since by Eq. (2.39) the eigenfunction $u_{n\mathbf{k}}$ uniquely determines the eigenfunction $u_{n,\mathbf{k}+\mathbf{G}}$, it is enough to consider only one of them in further analysis. By convention we will only consider the family of Hamiltonians $H_{\mathbf{k}}$ with \mathbf{k} restricted to the parallelepiped spanned by vectors \mathbf{b}_l . More precisely, we will consider \mathbf{k} vectors from the set

$$\mathcal{B} = \{k_1\mathbf{b}_1 + k_2\mathbf{b}_2 + k_3\mathbf{b}_3 \mid k_i \in [0, 1]\}. \quad (2.43)$$

Such set of \mathbf{k} vectors we refer to as the *Brillouin zone*.¹

We can consider \mathcal{B} to have the topology of a 3-torus since the states at $k_i = 0$ and $k_i = 1$ are related by Eq. (2.39). This viewpoint will be important later when we discuss geometrical aspects of the electronic structure theory in Sec. 2.2.

2.1.4 Pseudopotential theory

In this section we will introduce the concept of pseudopotentials. The natural setting to develop this formalism is that of an isolated atom with spherical symmetry.

First let us consider a Schrodinger equation for some spherical potential $V(r)$

¹Strictly speaking the *Brillouin zone* is defined as a Wigner-Seitz cell in the reciprocal space.

written in atomic units as

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V(r) - \epsilon \right] \phi_l(r) = 0. \quad (2.44)$$

Here $\phi_l(r)$ is the radial part of the wavefunction related to the full wavefunction $\psi(\mathbf{r})$ via $\psi(\mathbf{r}) = (1/r)\phi_l(r)Y_{lm}(\theta, \phi)$ where $Y_{lm}(\theta, \phi)$ is a spherical harmonic. In the theory of scattering one can show that the wavefunction far away from the scattering region is given by

$$\lim_{|\mathbf{r}| \rightarrow \infty} \psi(\mathbf{r}, \epsilon) = \exp(i\mathbf{q} \cdot \mathbf{r}) + i \frac{\exp(iqr)}{qr} \sum_l (2l+1) \exp(i\eta_l) \sin(\eta_l) P_l[\cos(\theta)] \quad (2.45)$$

where the first term corresponds to an incoming planewave with momentum \mathbf{q} and energy $\epsilon = q^2/2$ (in atomic units) while the second term describes the scattered wave. All of the dependence on the scattering potential V itself is encoded in the energy-dependent *phase shift* $\eta_l(\epsilon)$ for angular channel l . The phase shifts $\eta_l(\epsilon)$ can be calculated from matching the wavefunction and its first derivative inside and outside the scattering region.

From Eq. (2.45) it is obvious that the scattering properties at a given energy ϵ and angular momentum l do not change if the phase shift $\delta_l(\epsilon)$ is changed by an integer multiple of 2π . Therefore if we focus only on a certain energy window $[\epsilon_1, \epsilon_2]$ it is possible to have qualitatively very different scattering potentials V giving rise to the same (up to 2π) phase shift $\delta_l(\epsilon)$ and therefore having the same scattering properties. This gives us some freedom to tailor the potential V to our specific needs.

The ionic potential acting on the electrons in Eq. (2.3) consists of a series of Coulomb potentials which diverge at the ion position as $\sim r^{-1}$ when $r \rightarrow 0$. The theory of pseudopotentials, based on the observations above about $\delta_l(\epsilon)$, allows

one to replace the potential from Eq. (2.3) with a much softer potential \hat{V}_{PS} that can be treated more easily in the numerical calculations.

In what follows we will describe three schemes to replace the Coulomb potential Ze/r of the ion having charge Z with a softer potential having similar scattering properties in the relevant energy window. In all three cases this pseudopotential can be written in the form

$$\hat{V}_{\text{PS}} = V_{\text{loc}}(r) + \sum_i D_{ii'} |\beta_i\rangle \langle \beta_{i'}| \quad (2.46)$$

with an appropriately chosen potential V_{loc} and orbitals β_i that vanish outside the core region. Even though the operator \hat{V}_{PS} is no longer simply a “potential”, it is conventional to refer to it as a pseudopotential.

Norm-conserving pseudopotentials

We start by discussing norm-conserving pseudopotentials as first introduced in Ref. [11]. Norm-conserving pseudopotentials are chosen to satisfy following conditions.

1. For a given atomic configuration, the valence eigenvalues of the true all-electron potential and the pseudopotential are the same.
2. Outside a chosen core radius r_c , the valence eigenfunctions of the true all-electron potential and the pseudopotential are the same.
3. The total charge of the true all-electron and pseudo eigenfunctions inside core radius r_c are the same.
4. The logarithmic derivative and the energy derivative of the logarithmic derivative of the true all-electron and pseudo eigenfunctions agree for $r > r_c$.

The first condition guarantees that for a single atomic configuration, the true all-electron potential and the pseudopotential will give the same eigenenergies ϵ_i . The second condition guarantees that in the bonding region, away from the core, the wavefunctions match as well.

Furthermore, the equality of energy derivatives given in the fourth condition requires that the scattering of the pseudopotential and the true all-electron potential match not only at ϵ_i but also at energies close to ϵ_i . Since bonding with neighboring atoms will change the eigenenergies away from the atomic energies ϵ_i , pseudopotentials satisfying the fourth requirement should give a good description of the bonding properties.

The third requirement guarantees that the electrostatic potential of the true all-electron and pseudo wavefunction are the same. Furthermore, the third requirement is equivalent to the equality of energy derivatives given in the fourth condition due, to the identity

$$\frac{d}{d\epsilon} \frac{d}{dr} \ln \phi_l(R, \epsilon) = -\frac{1}{\phi_l^2(R, \epsilon)} \int_0^R \phi_l^2(r, \epsilon) dr \quad (2.47)$$

which is valid for any solution ϕ_l of Eq. (2.44) for arbitrary $V(r)$. Therefore, if the total charge within some radius r_c is the same for two different radial potentials $V(r)$, then the energy derivatives of the logarithmic derivatives of their eigenfunctions will be the same.

Now we will describe a procedure for constructing a pseudopotential satisfying these four constraints.

The first step is to solve the Kohn-Sham equation (2.21) for external potential V_{ext} given by a potential of a single ion Ze/r in some reference atomic configuration. Due to the spherical symmetry, we can write the Kohn-Sham equation for

each l -component $\phi_l(r)$ of a wavefunction $\psi_l(\mathbf{r})$ as

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + \frac{Ze}{r} + V_H(r) + V_{xc}(r) \right] \phi_l(r) = \epsilon_l \phi_l(r). \quad (2.48)$$

Next one has to divide the set of eigenfunctions ϕ_l into *core* and *valence* functions. Our goal is now to construct a pseudopotential in form of Eq. (2.46) that will reproduce the properties of the chosen set of *valence* functions corresponding to the four criteria given above.

First, we will construct an operator \hat{V}^{PS} such that the pseudofunctions ϕ_l^{PS} satisfying the four criteria above are given as solutions to

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + \hat{V}^{\text{PS}} + V_H^{\text{PS}}(r) + V_{xc}^{\text{PS}}(r) \right] \phi_l^{\text{PS}}(r) = \epsilon_l \phi_l^{\text{PS}}(r). \quad (2.49)$$

Here potentials $V_H^{\text{PS}}(r)$ and $V_{xc}^{\text{PS}}(r)$ are ordinary Hartree and exchange-correlation potentials, but evaluated using the density of the pseudo wavefunctions $\psi_l^{\text{PS}}(r)$.

We will require the pseudopotential operator \hat{V}^{PS} to be of the form of Eq. (2.46), or more specifically, we will require that

$$\hat{V}^{\text{PS}} = V_{\text{loc}}(r) + \sum_{lm} B_l |\chi_{lm}\rangle \langle \chi_{lm}|, \quad (2.50)$$

where l and m are angular momentum indices.

There are various approaches to choosing coefficients B_l and functions χ_{lm} . For example one approach given in Ref. [12] is first to explicitly construct the solutions $\phi_l^{\text{PS}}(r)$ and local potential $V_{\text{loc}}(r)$ to satisfy the desired four properties. For $r > r_c$ we have to choose pseudofunction $\phi_l^{\text{PS}}(r)$ to be equal to the true all-electron wavefunction and V_{loc} has to equal the true all-electron ionic potential. For $r < r_c$ we are then free to choose some smooth functions which satisfy the necessary criteria.

Once functions $\phi_l^{\text{PS}}(r)$ and $V_{\text{loc}}(r)$ have been chosen, we can construct the pseudopotential from Eq. (2.50) by defining

$$\chi_{lm}(\mathbf{r}) = \left\{ \epsilon_l - \left[-\frac{1}{2}\nabla^2 + V_{\text{loc}}(r) + V_{\text{H}}^{\text{PS}}(r) + V_{\text{xc}}^{\text{PS}}(r) \right] \right\} \psi_{lm}^{\text{PS}}(\mathbf{r}) \quad (2.51)$$

and

$$B_l = \frac{1}{\langle \chi_{lm} | \psi_{lm}^{\text{PS}} \rangle}. \quad (2.52)$$

It is clear from this definition that ψ_{lm}^{PS} is indeed an eigenvector of the Hamiltonian in Eq. (2.49).

As shown in Ref. [12], more accurate norm-conserving pseudopotentials can be achieved by generalizing the ansatz of Eq. (2.50) to include multiple projector functions $\beta_{lm\tau}$, indexed by (typically two) τ values for a single value of l and m :

$$\hat{V}^{\text{PS}} = V_{\text{loc}}(r) + \sum_{lm} \left[\sum_{\tau\tau'} B_{l\tau\tau'} |\beta_{lm\tau}\rangle \langle \beta_{lm\tau'}| \right] \quad (2.53)$$

From here we can proceed in the analogous way as before by constructing pseudofunctions $\phi_{l\tau}^{\text{PS}}$ not only to satisfy the four conditions given above, but also to satisfy the more general norm-conserving property

$$\int_0^{r_c} [\phi_{l\tau'}(r)]^* \phi_{l\tau}(r) dr = \int_0^{r_c} [\phi_{l\tau'}^{\text{PS}}(r)]^* \phi_{l\tau}^{\text{PS}}(r) dr. \quad (2.54)$$

While it is not obvious, it turns out that this property will guarantee that the energy derivative will be well reproduced at all energies $\epsilon_{l\tau}$ for all τ . One can

show that by choosing

$$B_{l\tau\tau'} = \langle \phi_{l\tau}^{\text{PS}} | \chi_{lm\tau'} \rangle, \quad (2.55)$$

$$|\beta_{lm\tau}\rangle = \sum_{\tau'} [B_l^{-1}]_{\tau'\tau} |\chi_{lm\tau'}\rangle, \quad (2.56)$$

we have again constructed a pseudopotential operator \hat{V}^{PS} such that pseudofunctions $\phi_{l\tau}^{\text{PS}}$ are eigenvectors of Hamiltonian in Eq. (2.49).

Ultrasoft pseudopotentials

Now we will describe the construction of ultrasoft pseudopotentials following the presentation of Ref. [12].

Norm-conserving pseudopotentials were defined by Eq. (2.49), which we will generalize here to include an overlap operator \hat{S} that appears on the right side of the secular equation

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + \hat{V}^{\text{PS}} \right] \phi_{l\tau}^{\text{PS}}(r) = \epsilon_{l\tau} \hat{S} \phi_{l\tau}^{\text{PS}}(r), \quad (2.57)$$

turning this equation into a *generalized* eigenvalue problem. The index τ here distinguishes between multiple solutions for the same value of l . Furthermore, we require the generalized eigenfunctions $\phi_{l\tau}^{\text{PS}}$ to satisfy the generalized orthonormalization condition $\langle \phi_{l\tau}^{\text{PS}} | \hat{S} | \phi_{l'\tau'}^{\text{PS}} \rangle = \delta_{ll'} \delta_{\tau\tau'}$.

Denoting indices l, m, τ by the single index i , we will again express the pseudopotential operator as

$$\hat{V}_{\text{PS}} = V_{\text{loc}}(r) + \sum_i D_{ij} |\beta_i\rangle \langle \beta_j|. \quad (2.58)$$

Additionally in Eq. (2.57) we have absorbed for a moment the Hartree and exchange-correlation potentials into $V_{\text{loc}}(r)$.

Now we will define the operator \hat{S} in such a way that the energy derivative of the logarithmic derivative of the pseudo wavefunction will be reproduced for an arbitrary choice of the pseudo wavefunctions ϕ_i^{PS} . This is in contrast to the norm-conserving scheme in which one has to impose constraints on the pseudo wavefunction (as in Eq. (2.54)) in order to reproduce correctly the energy derivative. For this reason, in the ultrasoft scheme there is more freedom in choosing the pseudofunctions ϕ_i^{PS} and therefore they can be chosen to be more smooth than the norm-conserving ones.

First let us define the matrix Q_{ij} to be the difference between the left and right hand sides of Eq. (2.54). Now we can define the pseudopotential \hat{V}^{PS} by

$$D_{ij} = B_{ij} + \epsilon_j Q_{ij}. \quad (2.59)$$

and operator \hat{S} as

$$\hat{S} = 1 + \sum_{ij} Q_{ij} |\beta_i\rangle \langle \beta_j|. \quad (2.60)$$

Here the functions β_i and the matrix B_{ij} are defined as in the case of the norm-conserving pseudopotentials. With this choice of operators \hat{S} and \hat{V}^{PS} , the only imposed requirement on the pseudofunctions ϕ_i^{PS} is that they match the true all-electron wavefunctions beyond the cutoff radius.

Additional terms arising from the introduction of operator \hat{S} makes ultrasoft pseudopotential method more complicated than the norm-conserving pseudopotential method. Nevertheless, this reduces the total computational cost. For further details on this method, please see Ref. [12].

Projector augmented waves

The projector augmented waves method developed in Ref. [13] is a third way we will describe which allows one to treat the fast oscillations of valence electrons near the electron cores.

We start by considering the Hilbert space of all wavefunctions orthogonal to the core electron wavefunctions. The functions of interest in this space have rapid oscillations near the core which make them hard to treat numerically. Therefore we consider another Hilbert space, also referred to as the pseudo Hilbert space, which contains numerically more easily treated functions without these rapid oscillations. Additionally we will consider a linear operator \mathcal{T} that maps between these two spaces. With the help of this operator we can do all of our calculations in the pseudo Hilbert space. If we are interested in, for example, the expectation of some operator \mathcal{O} , one can simply transform the pseudo functions $|\tilde{\psi}\rangle$ to the real Hilbert space $\mathcal{T}|\tilde{\psi}\rangle$ and calculate $\langle\tilde{\psi}|\mathcal{T}^\dagger\mathcal{O}\mathcal{T}|\tilde{\psi}\rangle$ instead of dealing with $\langle\psi|\mathcal{O}|\psi\rangle$ directly. We will write the operator \mathcal{T} in the form

$$\mathcal{T} = \mathbf{1} + \mathcal{T}_0 \tag{2.61}$$

with \mathcal{T}_0 acting only in some spherical (augmentation) region around the atom. Here we will again work in the context of a free isolated atom.

Now let us choose a set of functions $|\phi_i\rangle$ and $|\tilde{\phi}_i\rangle$ which we will refer to as the all-electron and pseudo partial waves respectively. Let us assume that the $|\phi_i\rangle$ functions are chosen to be orthogonal to the atomic core states, and that both the $|\phi_i\rangle$ and the $|\tilde{\phi}_i\rangle$ are complete within the augmentation region – in other words, that they span the entire (pseudo) Hilbert space. With this assumption, we can

define the operator \mathcal{T} by imposing that

$$|\phi_i\rangle = (\mathbf{1} + \mathcal{T}_0) |\tilde{\phi}_i\rangle. \quad (2.62)$$

An obvious choice for the all-electron partial waves $|\phi_i\rangle$ is to take the solutions of the Schrodinger equation for a single atom. In that case the index i stands for the radial wavefunction indices $\{n, l, m\}$. From the definition in Eq. (2.61) it is clear that the pseudo partial waves $|\tilde{\phi}_i\rangle$ will then have to equal the all-electron partial waves outside the augmentation region, since \mathcal{T}_0 acts only inside the augmentation region.

Since, set of pseudo partial waves $|\tilde{\phi}_i\rangle$ is complete in the augmentation region, every pseudo wavefunction $\tilde{\Psi}$ can be expanded in terms of the pseudo partial waves as

$$|\tilde{\Psi}\rangle = \sum_i c_i |\tilde{\phi}_i\rangle. \quad (2.63)$$

This pseudo wavefunction is then mapped by \mathcal{T} into

$$\mathcal{T}|\tilde{\Psi}\rangle = |\Psi\rangle = \sum_i c_i |\phi_i\rangle \quad (2.64)$$

where $|\phi_i\rangle$ are all-electron partial waves. Therefore we can write the all-electron wavefunction as

$$|\Psi\rangle = |\tilde{\Psi}\rangle + \sum_i \left[|\phi_i\rangle - |\tilde{\phi}_i\rangle \right] c_i. \quad (2.65)$$

If we require operator \mathcal{T} to be linear, then from Eq. (2.65) it is clear that the coefficients c_i themselves must be linear functionals of $|\tilde{\Psi}\rangle$. The most general form of a linear functional is a scalar product with some constant function. We will

denote this function with $\langle \tilde{p}_i |$ and refer to it as a projector function. Therefore we have

$$c_i = \langle \tilde{p}_i | \tilde{\Psi} \rangle. \quad (2.66)$$

Since the pseudo partial waves form a complete basis, and expansion in Eq. (2.63) is unique, it follows that

$$\langle \tilde{p}_i | \tilde{\phi}_j \rangle = \delta_{ij}. \quad (2.67)$$

The projection operator \mathcal{T} now can be compactly written as

$$\mathcal{T} = 1 + \sum_i \left[|\phi_i\rangle - |\tilde{\phi}_i\rangle \right] \langle \tilde{p}_i|. \quad (2.68)$$

The expectation $\langle \psi | \mathcal{O} | \psi \rangle$ of any local operator \mathcal{O} can be written in terms of the pseudofunctions $\langle \tilde{\psi} | \tilde{\mathcal{O}} | \tilde{\psi} \rangle$ where the pseudo operator $\tilde{\mathcal{O}}$ is defined as

$$\begin{aligned} \tilde{\mathcal{O}} &= \mathcal{T}^\dagger \mathcal{O} \mathcal{T} \\ &= \mathcal{O} + \sum_{ij} |\tilde{p}_i\rangle \left[\langle \phi_i | \mathcal{O} | \phi_j \rangle - \langle \tilde{\phi}_i | \mathcal{O} | \tilde{\phi}_j \rangle \right] \langle \tilde{p}_j|. \end{aligned} \quad (2.69)$$

Additionally we can choose an arbitrary operator \mathcal{V} acting within the core region and add

$$\mathcal{V} - \sum_{ij} |\tilde{p}_i\rangle \langle \tilde{\phi}_i | \mathcal{V} | \tilde{\phi}_j \rangle \langle \tilde{p}_j| \quad (2.70)$$

to Eq. (2.69) without changing its expectation value for any pseudo wavefunction. One can use this freedom in order to remove singularities in the core region for any operator \mathcal{O} . This can, for example, be used to remove the singularity in the

Coulomb potential of an ion.

In numerical applications, one can now deal with smooth pseudo wavefunctions $|\tilde{\psi}\rangle$ which can be expanded in a relatively small set of planewaves. All the integrations involving rapidly varying partial waves can be done on a radial grid, since these functions can be written as a radial function times a spherical harmonic.

2.1.5 Relativistic effects in solids

Especially in the case of heavier atoms, relativistic effects in solids can be important. One example of such an effect is the spin-orbit interaction. Luckily, these effects originate mostly from the deep core-level electrons. For this reason, as we are about to show, it is enough to properly include the relativistic effects only in the generation of the pseudopotentials.

The generalization of the quantum mechanical description of electrons in external fields to a relativistically invariant theory leads to the Dirac equation [14, 15]

$$\left\{ c\boldsymbol{\alpha} \cdot \left[\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{r}) \right] + V(\mathbf{r}) + \beta m_e c^2 \right\} \boldsymbol{\Psi} = i\hbar \frac{\partial}{\partial t} \boldsymbol{\Psi}. \quad (2.71)$$

Here $\boldsymbol{\Psi}$ is a 4-component wavefunction, \mathbf{A} and V are external vector and scalar potentials acting on the electron, $\mathbf{p} = -i\hbar\nabla$, while $\boldsymbol{\alpha}$ and β are the 4×4 matrices

$$\alpha_i = \begin{bmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{bmatrix} \quad \beta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (2.72)$$

where σ_i are the Pauli matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2.73)$$

The proper description of the electron-electron interaction in a relativistic framework is given by the theory of quantum electrodynamics.

Starting from this relativistic theory including electron-electron interactions, it was shown in Ref. [16] how one can derive the relativistic version of the density functional theory formalism and the corresponding relativistic Kohn-Sham equations. To do so, one has to proceed in a similar manner as for the nonrelativistic theory, but we will not present that derivation here. Here we will just give the final form of the relativistic Kohn-Sham equations [16]

$$\left\{ c\boldsymbol{\alpha} \cdot \left[\mathbf{p} - \frac{e}{c} \mathbf{A}_{\text{DKS}}(\mathbf{r}) \right] + V_{\text{DKS}}(\mathbf{r}) + \beta m_e c^2 \right\} \Psi_i(\mathbf{r}) = E_i \Psi_i(\mathbf{r}) \quad (2.74)$$

where the effective potentials \mathbf{A}_{DKS} and V_{DKS} are defined as

$$V_{\text{DKS}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \frac{e^2}{c} \int \frac{j^0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + ec \frac{\delta E_{\text{xc}}[j^\mu]}{\delta j^0(\mathbf{r})}, \quad (2.75)$$

$$\mathbf{A}_{\text{DKS}}(\mathbf{r}) = \mathbf{A}_{\text{ext}}(\mathbf{r}) + \frac{e}{c} \int \frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + c \frac{\delta E_{\text{xc}}[j^\mu]}{\delta \mathbf{j}(\mathbf{r})}. \quad (2.76)$$

The structure of these equations closely resembles that of the Dirac equation given in Eq. (2.71) for an electron in an effective scalar and vector potential. The difference with respect to the nonrelativistic theory is that we are not working here with functionals of density alone, but with functionals of the entire current density 4-vector $j^\mu = \{j^0, \mathbf{j}\}$.

Now let us assume that there is no external magnetic field ($\mathbf{A}_{\text{ext}} = \mathbf{0}$), and let us also neglect the retardation part of the Coulomb interaction (Gaunt term) since it scales as c^{-2} . With these assumptions, we can neglect the first two terms in $\mathbf{A}_{\text{DKS}}(\mathbf{r})$. Furthermore, it is often enough to replace the relativistic exchange-correlation energy $E_{\text{xc}}[j^\mu]$ with its nonrelativistic form.

With these simplifications, we arrive at the following relativistic generalization

of the Kohn-Sham equation:

$$\left[c\boldsymbol{\alpha} \cdot \mathbf{p} + V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{xc}}(\mathbf{r}) + \beta m_e c^2 \right] \boldsymbol{\Psi}_i(\mathbf{r}) = E_i \boldsymbol{\Psi}_i(\mathbf{r}). \quad (2.77)$$

In order to make this equation look more similar to the nonrelativistic version in Eq. (2.21), we have defined the electron density in terms of the 4-vector Kohn-Sham orbitals simply as $n(\mathbf{r}) = \sum_i \boldsymbol{\Psi}_i^\dagger(\mathbf{r}) \boldsymbol{\Psi}_i(\mathbf{r})$. The difference with Eq. (2.21) is that here the kinetic energy operator is different, the wavefunctions have four components, and one is supposed to solve only for positive-energy solutions E_i .

For our purposes it will be enough to solve the relativistic equation (2.21) only for isolated atoms in the process of pseudopotential generation. Therefore, let us consider the special case of Eq. (2.21) with $V_{\text{ext}}(\mathbf{r}) = Ze^2/|\mathbf{r}|$ and let us denote simply by V the sum of V_{ext} , the Hartree potential, and the exchange-correlation potential in Eq. (2.77). Therefore one has to solve

$$(c\boldsymbol{\alpha} \cdot \mathbf{p} + V + \beta m_e c^2) \boldsymbol{\Psi}_i(\mathbf{r}) = E_i \boldsymbol{\Psi}_i(\mathbf{r}). \quad (2.78)$$

In the case of a single ion, the total potential V is spherical, $V = V(r)$. Therefore we can reduce Eq. (2.78) to a radial equation for each value of the total angular momentum j and projection m .

We start by choosing an ansatz for $\boldsymbol{\Psi}_i(\mathbf{r})$ of the form [17]

$$\boldsymbol{\Psi}_{njml}(\mathbf{r}) = \begin{pmatrix} \frac{1}{r} G_{njl} \phi_{jml} \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{r}}{r^2} F_{njl} \phi_{jml} \end{pmatrix} \quad (2.79)$$

where G_{njl} and F_{njl} are corresponding large and small radial functions and ϕ_{jml} are the angular wavefunctions with total angular momentum j , projection m and angular momentum l . With the help of this ansatz we can cast Eq. (2.78) into

an equation for radial functions only:

$$\frac{1}{\hbar c} (\epsilon_{njl} - V) G_{njl}(r) = -\frac{dF_{nlj}(r)}{dr} + \frac{\kappa}{r} F_{nlj}(r), \quad (2.80)$$

$$\frac{1}{\hbar c} (\epsilon_{njl} + 2m_e c^2 - V) F_{njl}(r) = +\frac{dG_{nlj}(r)}{dr} + \frac{\kappa}{r} G_{nlj}(r). \quad (2.81)$$

Here $\epsilon_{njl} = E_{njl} - m_e c^2$ and if $j = l + 1/2$, then $\kappa = -(l + 1) = -(j + 1/2)$, while if $j = l - 1/2$, then $\kappa = l = j + 1/2$.

Let us now assume that we have solved the set of coupled equations (2.80) and (2.81) selfconsistently for some atom. This means that we have calculated the potential V and the radial parts of the wavefunction G_{njl} and F_{njl} .

For valence electrons the eigenenergy ϵ_{njl} will be much smaller than $m_e c^2$ (which is $\sim 5 \cdot 10^5$ eV). For this reason, and because of the additional $2m_e c^2$ term in Eq. (2.81), we can conclude that for valence electrons the $|G_{njl}(r)|$ component will be much larger than $|F_{njl}(r)|$. This is not necessarily the case for core electrons of very heavy ions. Additionally, outside the core region, the potential $V(r)$ will also be small. Therefore, for the valence electrons outside the core region, we can use Eq. (2.81) to approximate $F_{njl}(r)$ by [18]

$$F_{njl}(r) = \frac{\hbar}{2mc} \left[\frac{dG_{nlj}(r)}{dr} + \frac{\kappa}{r} G_{nlj}(r) \right] + \mathcal{O}(c^{-2}), \quad (2.82)$$

which is approximately correct up to and including order c^{-1} . Inserting this into Eq. (2.80) we obtain

$$-\frac{\hbar^2}{2m} \left[\frac{d^2}{dr^2} - \frac{\kappa(\kappa + 1)}{r^2} \right] G_{nlj}(r) = (\epsilon_{njl} - V) G_{njl}(r) + \mathcal{O}(c^{-2}), \quad (2.83)$$

which looks just like an ordinary radial Schrodinger equation but with the function $G_{nlj}(r)$ replacing the Schrodinger wavefunction.

Therefore, valence electrons outside the core region are well described up to

order c^{-1} by the Schrodinger equation (2.83), while for the core electrons we have to use the fully relativistic equations (2.80) and (2.81). This does not mean that relativistic effects are not included in equation for the valence electrons, since the potential V is determined selfconsistently from both the core and the valence electron densities.

As suggested in Ref. [18], one can now repeat the pseudopotential construction of Sec. 2.1.4, but for the relativistic potential V and the radial solutions $G_{njl}(r)$. Pseudopotentials constructed in this way will include the relativistic effects of the core electrons, but they will not require solution of relativistic equations for the valence (pseudo) electrons.

2.1.6 Density-functional perturbation theory

By solving the Hamiltonian in the Born Oppenheimer approximation given in Eq. (2.2), one can obtain the total energy of the solid $E(\{\mathbf{R}_I\})$ as a function of the set of ionic positions $\{\mathbf{R}_I\}$. With this information we can also study the classical dynamics of the ions as determined by the Hamiltonian

$$H_I = \sum_I \frac{1}{2} M_I \left(\frac{\partial \mathbf{R}_I}{\partial t} \right)^2 + E(\{\mathbf{R}_I\}). \quad (2.84)$$

After expanding $E(\{\mathbf{R}_I\})$ around a stationary point and looking for oscillating solutions, one can arrive at a secular equation

$$\det \left| \frac{1}{\sqrt{M_I M_J}} \frac{\partial^2 E(\{\mathbf{R}_I\})}{\partial \mathbf{R}_I \partial \mathbf{R}_J} - \omega^2 \right| = 0 \quad (2.85)$$

where the ω are the oscillating frequencies.

Among other things, density-functional perturbation theory allows for the efficient calculation of the second derivatives of the energy appearing in Eq. (2.85),

and therefore allows one to calculate the ionic lattice (phonon) vibration frequencies ω . An extensive review of density functional perturbation theory can be found in Ref. [19] and in the original work in Refs. [20, 21, 22].

The energy $E(\{\mathbf{R}_I\})$ can be calculated as a ground-state expectation value of the electronic Hamiltonian given in Eq. (2.2), $E(\{\mathbf{R}_I\}) = \langle \Psi | H | \Psi \rangle$. By explicitly calculating the second derivative of $E(\{\mathbf{R}_K\})$ with respect to the ionic position \mathbf{R}_I and \mathbf{R}_J , we get

$$\frac{\partial^2 E(\{\mathbf{R}_K\})}{\partial \mathbf{R}_I \partial \mathbf{R}_J} = \int \frac{\partial n(\mathbf{r})}{\partial \mathbf{R}_J} \frac{\partial V_{\text{ext}}(\mathbf{r})}{\partial \mathbf{R}_I} d\mathbf{r} + \int n(\mathbf{r}) \frac{\partial^2 V_{\text{ext}}(\mathbf{r})}{\partial \mathbf{R}_I \partial \mathbf{R}_J} d\mathbf{r} + \frac{\partial^2 E_{\text{ion}}}{\partial \mathbf{R}_I \partial \mathbf{R}_J}. \quad (2.86)$$

Here $V_{\text{ext}}(\mathbf{r})$ is just the ionic potential given by Eq. (2.3), and repeated here for completeness,

$$V_{\text{ext}}(\mathbf{r}) = \sum_I \frac{Z_I e^2}{|\mathbf{r} - \mathbf{R}_I|}. \quad (2.87)$$

The only unknown in Eq. (2.86) is the derivative of the ground-state electron density $n(\mathbf{r})$ with respect to the ionic position \mathbf{R}_I .

In the Kohn-Sham formulation of the density-functional theory, the ground-state electron density is given as a sum over the density of the occupied Kohn-Sham orbitals $\phi_i(\mathbf{r})$. Therefore its derivative with respect to the ionic position \mathbf{R}_I is simply

$$\partial_{\mathbf{R}_I} n(\mathbf{r}) = 2\text{Re} \sum_i \phi_i^*(\mathbf{r}) \partial_{\mathbf{R}_I} \phi_i(\mathbf{r}). \quad (2.88)$$

Here we have introduced the notation $\partial_{\mathbf{R}_I}$ to stand for $\partial/\partial \mathbf{R}_I$.

The variation $\partial_{\mathbf{R}_I} |\phi_i\rangle$ of the Kohn-Sham orbitals $|\phi_i\rangle$ can be obtained by

standard perturbation theory as

$$(H_{\text{KS}} - \epsilon_i) \partial_{\mathbf{R}_I} |\phi_i\rangle = -(\partial_{\mathbf{R}_I} V_{\text{KS}} - \partial_{\mathbf{R}_I} \epsilon_i) |\phi_i\rangle. \quad (2.89)$$

Here H_{KS} is the Kohn-Sham Hamiltonian defined in Eq. (2.21) and V_{KS} is the Kohn-Sham effective Hamiltonian from Eq. (2.18).

The first order change in the Kohn-Sham orbital energy is also given by a standard result of the perturbation theory,

$$\partial_{\mathbf{R}_I} \epsilon_i = \langle \phi_i | \partial_{\mathbf{R}_I} V_{\text{KS}} | \phi_i \rangle, \quad (2.90)$$

while the first-order change in V_{KS} can be calculated directly from its definition as

$$\partial_{\mathbf{R}_I} V_{\text{KS}}(\mathbf{r}) = \partial_{\mathbf{R}_I} V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{\partial_{\mathbf{R}_I} n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{dV_{\text{xc}}(n)}{dn} \partial_{\mathbf{R}_I} n(\mathbf{r}). \quad (2.91)$$

Equations (2.88), (2.89), (2.90) and (2.91) have to be solved selfconsistently in a similar manner as the Kohn-Sham equations. From the solution of these equations one then gets the first-order change in the electron density $\partial_{\mathbf{R}_I} n(\mathbf{r})$, from which the second derivative matrix of the total energy $E(\{\mathbf{R}_I\})$ can be calculated through Eq. (2.86).

2.2 Use of the geometry of the wavefunctions

In this section we will introduce some geometric quantities that can be calculated from the cell-periodic part $u_{n\mathbf{k}}$ of the wavefunction of an electron in a periodic solid.

2.2.1 Band insulators

For a given value of \mathbf{k} , we will assume that the indices $n \geq 1$ of the eigenvectors $u_{n\mathbf{k}}$ and the eigenvalues $\epsilon_{n\mathbf{k}}$ are assigned in such a way that $\epsilon_{n\mathbf{k}} \leq \epsilon_{n+1,\mathbf{k}}$. In the case of degenerate energies the ambiguity in the assignment of index n will be of no importance.

We define a *band insulator* to be a system in which eigenstates $u_{n\mathbf{k}}$ are occupied for $n \leq N$ and empty for $n > N$ for some integer N . Let us further also require that $\epsilon_{N+1,\mathbf{k}}$ is strictly larger than $\epsilon_{N\mathbf{k}}$. (The conventional definition of band insulators also assumes that there is an energy E , independent of \mathbf{k} , such that $\epsilon_{N\mathbf{k}} < E$ and $\epsilon_{N+1,\mathbf{k}} > E$ for all \mathbf{k} . This distinction will not be important unless specified otherwise.)

2.2.2 Gauge freedom

The set of occupied states $u_{n \leq N\mathbf{k}}$ in a band insulator spans a complex N -dimensional vector space for each point \mathbf{k} in the Brillouin zone. We will refer to this vector space as $\mathcal{L}_{\mathbf{k}}$ and it is defined as

$$\mathcal{L}_{\mathbf{k}} = \text{Span}(u_{1\mathbf{k}}, \dots, u_{N\mathbf{k}}). \quad (2.92)$$

Therefore we can think of $\mathcal{L}_{\mathbf{k}}$ as a family of vector spaces labelled by \mathbf{k} . Each of these spaces is a subspace of the Hilbert space of periodic functions satisfying the cell-periodicity criterion given in Eq. (2.37).

If we had considered the space spanned by both the occupied and all the unoccupied states, then this vector space would have been equal to the entire Hilbert space at each \mathbf{k} point. Therefore this family of vector spaces would have trivial topology of a direct product between the Hilbert space of cell-periodic functions and Brillouin zone.

In what follows we will mostly be interested in the vector space $\mathcal{L}_{\mathbf{k}}$ and not in the individual eigenstates of the Hamiltonian themselves. Therefore we do not need to choose the basis for $\mathcal{L}_{\mathbf{k}}$ to be eigenstates $u_{n\mathbf{k}}$ themselves, but are free to choose arbitrarily rotated set of states

$$\tilde{u}_{n\mathbf{k}} = \sum_m U_{mn}^{\mathbf{k}} u_{m\mathbf{k}} \quad (2.93)$$

with some unitary \mathbf{k} -dependent matrix $U^{\mathbf{k}}$. This freedom of choice of basis we will refer to as a *gauge choice*. Often we will refer to the basis vectors in arbitrary gauge simply as $u_{n\mathbf{k}}$, without the tilde symbol.

Matrices $U^{\mathbf{k}}$ have to conserve our convention from Eq. (2.39) and therefore we need to have

$$U^{\mathbf{k}+\mathbf{G}} = U^{\mathbf{k}} \quad (2.94)$$

for any vector \mathbf{k} and any reciprocal vector \mathbf{G} .

2.2.3 Geometric quantities

Now we will define some geometric quantities that we will use later. For a set of functions $u_{n\mathbf{k}}$ in an arbitrary gauge, we define the Berry connection, an $N \times N$ matrix of Cartesian 3-vectors, as [23]

$$\mathcal{A}_{mn,\mathbf{k}j} = i \langle u_{m\mathbf{k}} | \partial_j | u_{n\mathbf{k}} \rangle \quad (2.95)$$

where ∂_j is the derivative with respect to the j -th Cartesian component of \mathbf{k} . Next we define the Berry curvature, an $N \times N$ matrix of antisymmetric Cartesian

3×3 tensors, as [23]

$$\mathcal{F}_{mn,\mathbf{k}ij} = -\partial_i \mathcal{A}_{mn,\mathbf{k}j} + \partial_j \mathcal{A}_{mn,\mathbf{k}i} + i[\mathcal{A}_{\mathbf{k}i}, \mathcal{A}_{\mathbf{k}j}]_{mn}, \quad (2.96)$$

where the commutator of $N \times N$ matrices is taken. The trace of the Berry curvature matrix is given by

$$f_{\mathbf{k}ij} = \text{tr} \mathcal{F}_{mn,\mathbf{k}ij} = 2\text{Im} \sum_n \langle \partial_i u_{n\mathbf{k}} | \partial_j u_{n\mathbf{k}} \rangle. \quad (2.97)$$

Here tr denotes a trace over the occupied states. Finally we define the quantum metric, a 3×3 symmetric Cartesian tensor, as [23]

$$g_{\mathbf{k}ij} = \text{Re} \sum_n \langle \partial_i u_{n\mathbf{k}} | Q_{\mathbf{k}} | \partial_j u_{n\mathbf{k}} \rangle, \quad (2.98)$$

where the sum is performed over the set of occupied states n and $Q_{\mathbf{k}} = 1 - P_{\mathbf{k}}$, where $P_{\mathbf{k}}$ is the projection operator onto the set of occupied states at point \mathbf{k} , i.e.,

$$P_{\mathbf{k}} = \sum_m |u_{m\mathbf{k}}\rangle \langle u_{m\mathbf{k}}|. \quad (2.99)$$

The quantum metric tensor and the trace of the Berry curvature can be expressed as the real and twice the imaginary parts of the same object

$$\sum_n \langle \partial_i u_{n\mathbf{k}} | Q_{\mathbf{k}} | \partial_j u_{n\mathbf{k}} \rangle \quad (2.100)$$

due to Eq. 2.97 and 2.98.

The Berry connection $\mathcal{A}_{mn,\mathbf{k}j}$, the curvature $\mathcal{F}_{mn,\mathbf{k}ij}$, its trace $f_{\mathbf{k}ij}$, and the quantum metric tensor $g_{\mathbf{k}ij}$ are quantities which depend only on the ‘‘geometry’’ of the occupied states. For example, $\mathcal{A}_{mn,\mathbf{k}j}$ is just an $N \times N$ matrix which

measures the relative phase between wavefunctions at two neighboring points in the Brillouin zone. A similar interpretation can be given to the remaining objects.

2.2.4 Geometric bulk physical quantities

Several important material properties can be evaluated knowing only the geometrical properties of the band structure. These include the electric polarization (Ref. [24]), the intrinsic anomalous Hall conductivity (Ref. [25, 26, 27]), the Chern-Simons orbital magnetoelectric coupling (Ref. [28, 29]), the invariant part of the spread functional (Ref. [30]), and the optical conductivity sum rule (Ref. [31]).

The electric polarization \mathcal{P}_{el} already appears in dimension $d = 1$, and it can be evaluated as an integral of the Berry connection over the one-dimensional Brillouin zone (BZ) as [24]

$$\mathcal{P}_{\text{el}} = -\frac{e}{2\pi} \int_{\text{BZ}} dk \text{tr} \mathcal{A}_k, \quad (2.101)$$

where the trace is performed over the band indices of the Berry connection. The integrand is also referred to as the Chern-Simons 1-form, and its integral over the BZ is well known to be defined only modulo 2π . Any periodic adiabatic evolution of the Hamiltonian $\mathcal{H}(\lambda)$ whose first Chern number in (k, λ) space is non-zero will change the integral above by a multiple of 2π [24].

Polarization as defined in Eq. 2.101 includes only the contribution of the electron charges to the polarization. In order to arrive at the total polarization of a crystal (modulo a quantum), one needs to add to Eq. 2.101 the contribution from the ionic charges as well. Since ions are usually treated as classical objects, their polarization can simply be calculated via, $\mathcal{P}_{\text{ion}} = e \sum_I Z_I R_I$. Here eZ_I is the charge of the nucleus and R_I is its classical position. Total polarization \mathcal{P} is

then given by the sum of these two contributions,

$$\mathcal{P} = \mathcal{P}_{\text{el}} + \mathcal{P}_{\text{ion}}. \quad (2.102)$$

Unlike one-dimensional systems, crystals in $d = 2$ can have an anomalous Hall conductivity. Since this conductivity is strictly off-diagonal, one direction is needed for the applied electric field and another for the measured current. For a metal, the intrinsic contribution from a band crossing the Fermi level can be evaluated as a line integral [25, 26, 27]

$$\sigma^{\text{AH}} = \frac{e^2}{h} \frac{1}{2\pi} \oint_{\text{FL}} d\mathbf{k} \cdot \mathcal{A}_{\mathbf{k}} \quad (2.103)$$

over the Fermi loop (the boundary of the Fermi sea). Fully-filled deeper bands can also make a quantized contribution given by a similar integral, but around the entire BZ; this is the only contribution in the case of a quantum anomalous Hall insulator [32]. In both cases, the gauge choice on the boundary of the region should be consistent with a continuous, but not necessarily k -periodic, gauge in its interior; alternatively, each expression can be converted to an area integral of a Berry curvature to resolve any uncertainty about branch choice, i.e.,

$$\sigma_{ij}^{\text{AH}} = \frac{e^2}{h} \frac{1}{4\pi^2} \int_{\text{BZ}} d^2k \, \text{tr} \mathcal{F}_{\mathbf{k}ij}, \quad (2.104)$$

where the trace is performed over the occupied set of states at each \mathbf{k} .

The magnetoelectric coupling tensor α_{ij} is defined as

$$\alpha_{ij} = \left(\frac{\partial \mathcal{P}_i}{\partial B_j} \right)_{\mathcal{E}} = \left(\frac{\partial M_j}{\partial \mathcal{E}_i} \right)_B, \quad (2.105)$$

where \mathcal{P}_i is the electric polarization induced by the magnetic field B_j , or equivalently, M_j is the magnetization induced by the electric field \mathcal{E}_i . In the special case that the induced response (\mathcal{P} or \mathbf{M}) remains parallel to the applied field (\mathbf{B} or \mathcal{E}), the tensor α is purely diagonal with equal diagonal elements, and its strength can be measured by a dimensionless scalar parameter θ defined via

$$\alpha_{ij}^{\text{iso}} = \frac{\theta e^2}{2\pi h} \delta_{ij}. \quad (2.106)$$

Unlike one- or two-dimensional systems, three-dimensional systems can have an isotropic magnetoelectric coupling. Two directions are needed for the application of the magnetic field, but then a third direction is needed because the induced polarization has to be in the direction of the applied magnetic field. The Chern-Simons orbital magnetoelectric coupling (CSOMP) can be evaluated in $d = 3$ as a BZ integration of a quantity involving the Berry connection, namely

$$\theta = -\frac{1}{4\pi} \int_{\text{BZ}} d^3k \epsilon_{ijk} \text{tr} \left[\mathcal{A}_i \partial_j \mathcal{A}_k - \frac{2i}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k \right]. \quad (2.107)$$

The integrand in this expression is known as the Chern-Simons 3-form, and its integral over the entire BZ is again ill-defined modulo 2π , since any periodic adiabatic evolution of the Hamiltonian $\mathcal{H}(\lambda)$ whose second Chern number in (\mathbf{k}, λ) space is non-zero will change θ by an integer multiple of 2π [28, 29].

The CSOMP is only one part of the magnetoelectric response. The remaining part of the response involves non-geometric terms which explicitly depend on the band energy or the Hamiltonian (see Ref. [33, 34] for details).

In four dimensions ($d = 4$) there exists a generalization of a two-dimensional quantum anomalous Hall insulator [35, 28]. The response of such an insulator [28] is characterized by a current density j induced by the electromagnetic potential

A , given by

$$j_\mu = \frac{C_2}{8\pi^2} \epsilon_{\mu\nu\rho\sigma\tau} \partial_\nu A_\rho \partial_\sigma A_\tau. \quad (2.108)$$

This means that, for example, the field configuration

$$A_x = 0 \quad A_y = B_z x \quad A_z = -E_z t \quad A_w = A_t = 0, \quad (2.109)$$

corresponding to magnetic and electric fields in direction z will induce a current

$$j_w = \frac{C_2}{4\pi^2} B_z E_z \quad (2.110)$$

along the remaining fourth spatial direction. The integer C_2 can be calculated as an integral over the Brillouin zone via

$$C_2 = \frac{1}{32\pi^2} \int_{\text{BZ}} d^4 k \epsilon_{ijkl} \text{tr} [\mathcal{F}_{ij} \mathcal{F}_{kl}]. \quad (2.111)$$

The same quantity could be extended to the $d = 4$ metallic case, in which case the integration would be done only over the occupied part of the Brillouin zone. The same caveats as in the case of anomalous Hall conductivity would apply.

Now we will just briefly mention two remaining quantities which can be calculated from a knowledge of the quantum metric tensor $g_{\mathbf{k}ij}$. One of these quantities is the invariant part of the spread functional Ω_{I} , which for a three-dimensional insulator can be calculated as [30]

$$\Omega_{\text{I}} = \frac{V}{8\pi^3} \int_{\text{BZ}} d^3 k \text{Tr} g_{\mathbf{k}}. \quad (2.112)$$

Here V is the unit cell volume and the trace Tr is performed over the Cartesian indices of the metric tensor. Finally, the integral of the quantum metric tensor is

related to the frequency integral of the optical conductivity $\sigma_{ij}(\omega)$ as [31]

$$\frac{\hbar}{\pi e^2} \int_0^\infty \frac{d\omega}{\omega} \text{Re}\sigma_{ij}(\omega) = \frac{1}{8\pi^3} \int_{\text{BZ}} d^3k g_{\mathbf{k}ij}. \quad (2.113)$$

2.2.5 Topological insulators

Each of the four quantities Eq. (2.101), (2.103), (2.107) and (2.111) gives rise to one of four kinds of topological insulators. A comprehensive review of topological insulators is given in Ref. [36], and their complete classification can be found in Refs. [37, 38, 39, 40].

The dipole moment of a finite $d = 1$ crystal changes sign under space inversion. However, the bulk quantity \mathcal{P} defined in Eq. 2.102, which is related to the total dipole moment of a crystal, is well defined only modulo a quantum of charge e . Therefore, for a system which is inversion symmetric, one needs to have

$$\mathcal{P} = -\mathcal{P} \pmod{e}. \quad (2.114)$$

However, there are two distinct values of \mathcal{P} which satisfy this equation, namely

$$\mathcal{P} = 0 \quad \text{or} \quad \mathcal{P} = \frac{e}{2}. \quad (2.115)$$

Of course, each value still can be changed by a quantum e under a gauge transformation.²

An inversion-symmetric insulator having $\mathcal{P} = 0$ cannot be smoothly deformed into an insulator having $\mathcal{P} = e/2$, and vice versa. In this sense, this is the simplest example of a *topological insulator*.

The group associated to this kind of insulator is \mathbb{Z}_2 , because the expression

²Further classification of inversion symmetric insulators is given in Ref. [41].

for \mathcal{P} is additive over bands and it can either be trivial (0) or non-trivial ($e/2$).

Another example of a topological insulator appears in $d = 2$ and does not require any additional symmetry constraints. The integration of the Berry curvature in Eq. (2.104) has to result in an integer. Therefore this integer is a topological invariant of that insulator, since it cannot change under a smooth evolution of a Hamiltonian. The group associated to this insulator is \mathbb{Z} because Eq. (2.104) is again additive under addition of bands to the integration. A model example of this kind of topological insulator was first given in Ref. [32].

The Chern-Simon orbital magnetoelectric coupling, which exists in a $d = 3$ insulator, will change sign under either time-reversal or space inversion symmetry. Again, as in the case of polarization, the CSOMP can be changed under a gauge transformation by 2π . Therefore a system containing any or both of these symmetries needs to have

$$\theta = -\theta \pmod{2\pi}. \quad (2.116)$$

which again leads to two possibilities

$$\theta = 0 \quad \text{or} \quad \theta = \pi. \quad (2.117)$$

The group associated to this third kind of topological insulator is \mathbb{Z}_2 .

This kind of topological insulator was first predicted theoretically in Refs. [42, 43, 44] and then also confirmed experimentally in Refs. [45, 46, 47] on various materials.

Finally, in four dimensions there is another kind of topological insulator associated with the integer C_2 as defined in Eq. (2.111). This again leads to a group \mathbb{Z} associated with this kind of insulator. Unfortunately, four-dimensional crystals do not occur in nature.

Part I

A few applications of total-energy electronic structure methods

Chapter 3

Structural stability and lattice dynamics of SiO₂ cristobalite

Among the phases of SiO₂ are α - and β -cristobalites, which have a long and somewhat controversial history of proposed structural assignments and phase-transition mechanisms. Recently, Zhang and Scott found new indications that the higher-temperature β phase has space group $I\bar{4}2d$ and, by assuming a group-subgroup relationship between phases, they argued that the lower-temperature α phase should have lower symmetry than that of the widely-accepted $P4_12_12$ space group. With this motivation, we use first-principles calculations to investigate the energy, structure, and local stability of $P4_12_12$ and $I\bar{4}2d$ structures. We also compute the frequencies of the zone-center phonon modes in both structures, as well as certain zone-boundary modes in the $I\bar{4}2d$ structure, and compare with experiment. We then argue that the various $P4_12_12$ and $I\bar{4}2d$ enantiomorphs can be grouped into three clusters, each of which is identified with a three-dimensional manifold of structures of $P2_12_12_1$ symmetry in which the $P4_12_12$ and $I\bar{4}2d$ appear as higher-symmetry special cases. We find that there are relatively high energy barriers between manifolds, but low barriers within a manifold. Exploring the energy landscape within one of these manifolds, we find a minimal-energy path connecting $P4_12_12$ and $I\bar{4}2d$ structures with a surprisingly low barrier of ~ 5 meV

per formula unit. Possible implications for the phase-transition mechanism are discussed.

3.1 Introduction

The fact that SiO_2 can exist in numerous crystalline and amorphous forms, and its status as one of the most prevalent compounds on earth, has stimulated a long history of experimental and theoretical investigation. Here we focus on the α (“low”) and β (“high”) cristobalite phases, which are stable near the melting temperature and are metastable at room temperature.

The structure of the higher-temperature β phase has a history of controversy. Early indications of a cubic structure with 180° bond angles (space group $Fd\bar{3}m$) [48] were challenged by others [49, 50] who hypothesized that the true β -phase structure has lower symmetry and that the apparent cubic structure arises from averaging over spatial domains or dynamical fluctuations. In particular, Wright and Leadbetter [50] argued for a tetragonal structure belonging to space group $I\bar{4}2d$ (D_{2d}^{12}). While some subsequent works have provided support for this identification [51, 52, 53], other authors maintain that it is better to describe the β phase as a dynamically disordered one having overall $Fd\bar{3}m$ symmetry but with a large population of rigid-unit-mode (RUM) fluctuations [54, 55]. To some degree, the argument may be semantic; if the fluctuations have strong short-range correlations in space and time, it is difficult to distinguish this picture from one of dynamic domains of a lower-symmetry structure [56]. Thus, for example, either picture may be able to explain the fact that there are two more first-order lines in the Raman and infrared spectra than would be expected from $Fd\bar{3}m$ symmetry [53], and the question of which description is “correct” might depend on the time and length scales of the experimental probes in question.

In contrast, the assignment of the α -cristobalite phase to the tetragonal $P4_12_12$ (D_4^4) space group [57] has until recently been noncontroversial. However, based on a reexamination of Raman and infrared vibrational spectroscopies, Zhang and Scott [53] have recently raised new questions about the identity of the α phase. By using Raman spectroscopy to study small single crystals of β -cristobalite, these authors argued that the β structure must be D_{2d} , not cubic, and assuming a group-subgroup relationship for the β -to- α transition, concluded that the α phase should have some lower symmetry such as D_2 instead of D_4 . The apparent D_4 symmetry of α -cristobalite could result from spatial or dynamic averaging over D_2 domains, in analogy to what had been proposed for the β phase. To support their assumption that a group-subgroup relationship should hold, Zhang and Scott also pointed to the temperature dependence of the optical phonon frequencies near the phase transition as being inconsistent with a reconstructive phase transition [58] and as suggesting a nearly second-order behavior, although arguing in the opposite direction are the facts that the latent heat and volume change at the transition are quite substantial [59].

In their paper, Zhang and Scott [53] reexamined earlier Raman and infrared spectroscopic measurements not only on the α - and β -cristobalite SiO_2 [60], but also on α and β AlPO_4 (Ref. [61]) and α BPO_4 (Ref. [62]) cristobalites. Note that the replacement of Si atoms by Al and P (or B and P) atoms immediately reduces the symmetry according to $P4_12_12$ (D_4^4) \rightarrow $C222_1$ (D_2^5) for the α phase ¹ and $I\bar{4}2d$ (D_{2d}^{12}) \rightarrow $I\bar{4}$ (S_4^2) for the β phase. Also of possible relevance is the pressure-induced phase transition from α - SiO_2 to a high-pressure monoclinic cristobalite phase [63]. The relationship of these other cristobalites to the α and β phases of SiO_2 is an interesting avenue for future exploration, but falls outside the scope of

¹Note that this is a different D_2 space group than the $P2_12_12_1$ (D_2^4) one discussed elsewhere in this chapter.

the present work.

First-principles calculations of the structural and lattice dynamical properties of SiO_2 have a long and productive history [64, 65, 52, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77]. While quite a few of these works specifically address the α -cristobalite structure [64, 65, 71, 73, 70, 66, 77], questions about its stability and about possible pathways from the α to the β phase have not been fully explored.

In the present work, we have carried out first-principles calculations for candidate α and β cristobalite structures in the framework of density-functional theory (DFT) in order to check the stability of both phases and to explore the energy landscape connecting them. We have also calculated phonon frequencies and infrared activities for both α and β phases, and explored how the phonon modes in the different phases are related to each other and to those of the high-symmetry cubic phase. Our calculations are effectively zero-temperature ones, and thus cannot properly treat the temperature-induced α - β cristobalite phase transition. Nevertheless we hope that the information obtained from these calculations can eventually be built into a realistic statistical-mechanical theory that correctly takes the RUM fluctuations into account in its description of the α and β phases at experimentally relevant temperatures.

This chapter is organized as follows. In Sec. 3.2 we give a brief review of α and β cristobalite structures and describe the methods used in the calculations. Then, in Sec. 3.3, we present the results of our calculations of structural and lattice vibrational properties of the two phases and of the energy landscape connecting them. We discuss those results in Sec. 3.4. Finally, we summarize the work in Section 3.5.

3.2 Preliminaries

3.2.1 Cristobalite structures

In order to describe the structures of the SiO_2 α and β cristobalite phases, it is easiest to start by considering the “ideal cristobalite” structure, which is constructed by placing Si atoms in a diamond structure with oxygen atoms located midway between each pair of nearest-neighbor Si atoms. This structure has the space group $Fd\bar{3}m$ (O_h^7) and has two formula units per primitive unit cell. Each Si atom with its four surrounding O atoms forms a tetrahedron, so the whole structure can be visualized as a network of tetrahedra connected at their apices.

The generally accepted structure of α -cristobalite is arrived at by starting from the ideal structure and making nearly rigid rotations of the tetrahedra about $[100]$ and $[010]$ axes (in the original diamond cubic frame), leading to a tetragonal structure with its axis along \hat{z} . This is illustrated in Fig. 3.1(a), but in the conventional tetragonal frame, related to the original cubic frame by a 45° rotation about \hat{z} . The tetrahedral rotations are also accompanied by small strains and tetrahedral translations needed to keep the apices coincident, as would be expected from enforcement of the rigid-unit constraints. The space group of the structure is $P4_12_12$ (D_4^4), and since the four rotations shown in Fig. 3.1(a) are all different, the number of formula units per primitive unit cell is now increased to four.

As mentioned earlier, diffraction experiments on the β -cristobalite phase tend to give inconclusive results because of spatial and dynamical averaging. Nevertheless, based on the comparison of structure factors predicted by various disorder models and the ones obtained in their x-ray diffraction experiments, Wright and Leadbetter [50] concluded that the β -cristobalite has local $I\bar{4}2d$ (D_{2d}^{12}) space-group

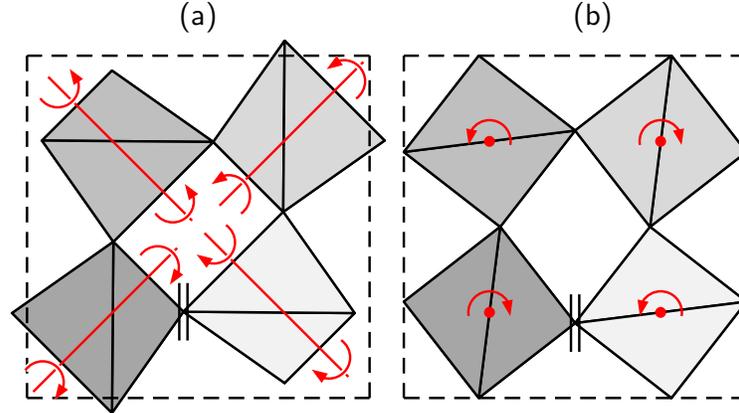


Figure 3.1: Projection on x - y plane of the (a) $\tilde{\alpha}$ and (b) $\tilde{\beta}$ structures, proposed as candidates for α and β cristobalite phases respectively. Darker shading is used to represent more distant tetrahedra so that the spiral structure of the connected tetrahedrons becomes evident; double vertical lines indicate that the adjoining tetrahedra are actually disconnected because they are separated in the z -direction.

symmetry. Their proposed structure can also be constructed from the ideal structure, but this time by rotating all the tetrahedra around the \hat{z} axis, yielding the structure shown in 3.1(b). The number of formula units per primitive unit cell remains at two as in the ideal structure (although it can alternatively be described, as in Fig. 3.1(b), by a doubled conventional cell containing four formula units). Again, the structure is highly consistent with the rigid-unit constraints.

Because we do not want to presuppose an identification of a particular experimentally observed *phase* with a particular *crystal structure*, we henceforth adopt a notation in which the phases are identified by labels “ α ” and “ β ” without tildes, whereas the putative crystal structures shown in 3.1(a) and (b) will be referred to as “ $\tilde{\alpha}$ ” and “ $\tilde{\beta}$ ” structures, respectively. Our working hypothesis is that the α and β phases have microscopic crystal structures of type $\tilde{\alpha}$ and $\tilde{\beta}$ respectively, but we adhere to a distinction in the notation in order to discriminate clearly between the specified structures used in our calculations and the hypothetical identification of these with experimental phases.

3.2.2 Computational methods

The calculations were carried out using the ABINIT implementation [78] of density-functional theory with Perdew-Burke-Ernzerhof [79] version of the generalized gradient approximation (GGA) for electron exchange and correlation. Since it is the smallest unit cell that contains both $\tilde{\alpha}$ and $\tilde{\beta}$ structures, all calculations were performed on the four-formula-unit computational cell shown in Fig. 3.1, even though the primitive cell is smaller in the $\tilde{\beta}$ structure. The Brillouin zone was sampled by a $4 \times 4 \times 4$ Monkhorst-Pack grid [80]. Structural properties were computed using projector augmented-wave [13] potentials converted from ultrasoft pseudopotentials [12] with a plane-wave cutoff of 22 Ha unless otherwise specified, while phonon frequencies, eigenvectors, and Born charges were computed [81] using norm-conserving Trouiller-Martins pseudopotentials [82] at an energy cutoff of 50 Ha (after repeating the structural relaxation using these potentials). The acoustic sum rule was imposed on the force constants, and charge neutrality was imposed on the Born charges. Throughout the chapter, the symmetry analysis associated with crystal space groups has been carried out using the Bilbao package [83, 84].

3.3 Results

3.3.1 Structural properties of $\tilde{\alpha}$ and $\tilde{\beta}$ structures

We started our calculations by considering the ideal cubic structure and relaxing its volume, obtaining $a_c=7.444 \text{ \AA}$ for the lattice constant of its eight-formula-unit cubic cell. Then, working in the frame of the four-formula-unit tetragonal cell, we found the phonon frequencies at the Γ point of its Brillouin zone, corresponding to phonons at the Γ point and one X point [namely $(2\pi/a_c)(001)$ or equivalently $(2\pi/a_c)(110)$ in the cubic frame] of the primitive two-formula-unit fcc cell. For

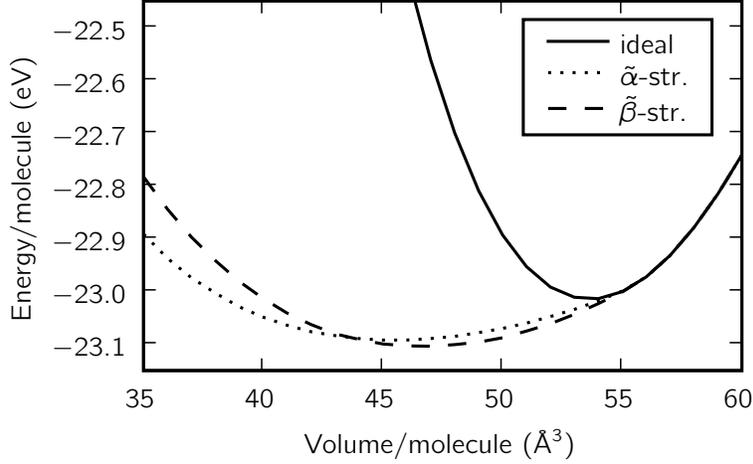


Figure 3.2: Ground state energy per formula unit (eV) vs. volume per formula unit (\AA^3) for $\tilde{\alpha}$, $\tilde{\beta}$, and cubic cristobalite structures of SiO_2 .

the “ideal structure” of space group $Fd\bar{3}m$, the symmetry decomposition of these phonons into irreducible representations is

$$\Gamma(\text{ideal}) = 1A_{2u} \oplus 1E_u \oplus 2T_{1u} \oplus 1T_{2u} \oplus 1T_{2g}, \quad (3.1)$$

$$X(\text{ideal}) = 3X_1 \oplus 1X_2 \oplus 2X_3 \oplus 3X_4. \quad (3.2)$$

(The translational T_{1u} mode has been omitted.) The E_u mode and all X modes are doubly degenerate; the T_{1u} , T_{2u} and T_{2g} modes are triply degenerate; and A_{2u} is non-degenerate.

We found that the triply-degenerate T_{2u} mode at Γ is unstable with an imaginary frequency of $i83 \text{ cm}^{-1}$. All other optical phonons have real frequencies, the lowest being at 250 cm^{-1} . Furthermore, one of the doubly-degenerate (X_4) modes is unstable with a frequency of $i53 \text{ cm}^{-1}$. We thus conclude that the ideal cristobalite structure is unstable with respect to these distortions.

Next we imposed distortions corresponding to these unstable modes and did a full relaxation of the structure subject to the symmetry constraints of the resulting space group. The unstable ($i53 \text{ cm}^{-1}$) mode at X leads to the space group $P4_12_12$

Table 3.1: Lattice constants (in Å) and Wyckoff structural parameters for $\tilde{\alpha}$ ($P4_12_12$) and $\tilde{\beta}$ ($I\bar{4}2d$) cristobalite SiO_2 .

	Present theory	Previous theory ¹	Expt. ²
$\tilde{\alpha}$ -cristobalite			
a	5.0730	5.1190	4.9570
c	7.0852	7.1683	6.8903
Si(u)	0.3001	0.2869	0.3047
O(x)	0.2384	0.2439	0.2381
O(y)	0.1081	0.0777	0.1109
O(z)	0.1819	0.1657	0.1826
$\tilde{\beta}$ -cristobalite			
a	7.1050	7.226	7.131 ³
c	7.4061	7.331	7.131 ³
O(x)	0.1051	0.0896	0.079

(or $P4_32_12$) which corresponds to $\tilde{\alpha}$ -cristobalite, while the ($i83\text{ cm}^{-1}$) mode at Γ takes us to the space group $I\bar{4}2d$ of $\tilde{\beta}$ -cristobalite. The energy of the relaxed ground state as a function of volume per formula unit is shown for both cases in Fig. 3.2, with the energy of the cubic phase also shown for reference. The corresponding structural parameters at the energy minimum are given in Table 3.1.

From Fig. 3.2 it is clear that the $\tilde{\alpha}$ and $\tilde{\beta}$ structures indeed have lower energies than the ideal cristobalite when the volume becomes smaller than some critical volume $V_0 \sim 55 \text{ \AA}^3$. (Above this volume, the imposed distortions disappear during relaxation and the structure returns to the ideal one.) We find that both the $\tilde{\alpha}$ and $\tilde{\beta}$ structures have a quite similar dependence of energy on volume. According to our calculation, the relaxed $\tilde{\beta}$ structure has a slightly lower energy than that of the $\tilde{\alpha}$ structure (12 meV per formula unit). This appears to be in conflict with the experimental situation, since the α phase is experimentally more stable at lower temperatures. However, when we repeated our calculations using a local-density approximation (LDA) exchange-correlation functional [85], the $\tilde{\beta}$

structure was found to be lower by 1 meV per formula unit. We thus conclude that the small energy difference between the two nearly-degenerate structures is a quantity that is too delicate to be reliably obtained by our DFT calculations. A similar discrepancy between the results from LDA and GGA functionals was found in Ref. [70].

Comparison with rigid-unit geometry

We also analyzed what happens to the bond lengths and angles in the $\tilde{\alpha}$ and $\tilde{\beta}$ structures as a function of volume V . For $V < V_0$, the O–Si–O bond angles and Si–O bond lengths inside the tetrahedra are found to remain almost constant, while the Si–O–Si bond angles change by $\sim 35^\circ$. The details are shown to be very close to the predictions of a picture of tilting of perfectly rigid tetrahedra. The fact that the three phases are indistinguishable for $V > V_0$ is also easily explained, since the tilts of rigid tetrahedra can only decrease the volume of the ideal structure. Thus, for $V > V_0$ the tetrahedra cannot remain rigid and the Si–O bond length must increase, and only when V becomes smaller than some volume V_0 will one of the RUMs condense in the structure in order to maintain the preferred Si–O bond lengths.

In a picture in which the rigid-unit constraints are perfectly enforced, it turns out that the structures of $\tilde{\alpha}$ and $\tilde{\beta}$ symmetries are completely determined by a single parameter, which can be taken to be the volume V per formula unit relative to the corresponding value V_0 in the ideal cubic structure. (That is, V_0 is the volume below which rigid distortions start to appear, as explained in Sec. 3.3.1.) Now we check to see how closely our structures, as optimized from the first-principles calculations, match with this elementary model.

The solid curves in Figs. 3.3 and 3.4 show the mathematical predictions of this elementary model, obtained by applying rotations of types $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ (see

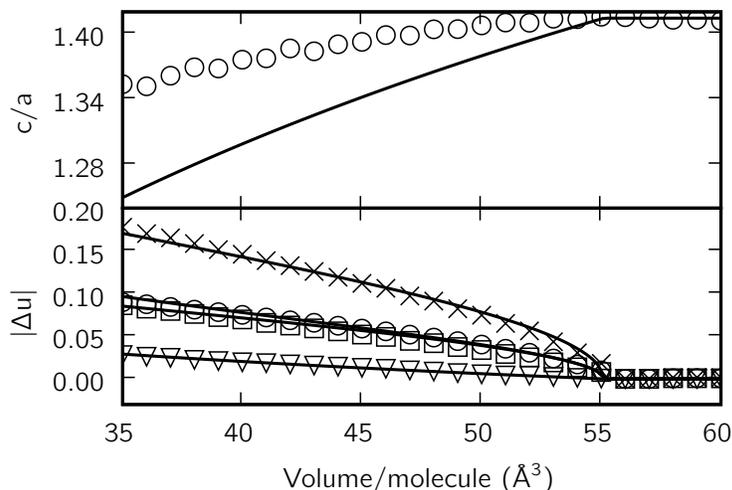


Figure 3.3: Structural parameters of the $\tilde{\alpha}$ structure vs. volume per formula unit. Top panel: c/a ratio. Bottom panel: absolute values of deviations of internal parameters $u(\text{Si})$ (squares), $x(\text{O})$ (triangles), $y(\text{O})$ (crosses), and $z(\text{O})$ (circles) from ideal-cubic values. Symbols represent first-principles calculations; lines are fits to an ideal rigid-unit geometry.

Fig. 3.5) in such a way as to keep the tetrahedra perfectly rigid. (For $V > V_0$, the elementary model cannot be satisfied, and the ideal cubic parameters are plotted instead.) The symbols shown in Figs. 3.3 and 3.4 denote the results of our first-principles calculations where, for each specified value of V , the volume was treated as a constraint while all other structural parameters were relaxed. The fit was optimized by choosing a common $V_0 = 55.1 \text{ \AA}^3$ for both $\tilde{\alpha}$ and $\tilde{\beta}$ structures. For reference, the first-principles equilibrium volumes are 45.7 and 46.7 \AA^3 for the $\tilde{\alpha}$ and $\tilde{\beta}$ structures, respectively.

We find that the agreement is extraordinarily good for all of the internal parameters, but that there are some significant discrepancies in the c/a ratios. At first sight this may seem contradictory: why are the c/a ratios off by many percent, while the Si-O bond lengths agree within $\sim 0.05\%$? The answer is connected with the presence of volume-preserving tetragonal distortions of low energy cost. In such a distortion, each tetrahedron is stretched slightly along c and compressed in a (or vice versa), and it happens that the tetrahedral angle of $\arccos(1/\sqrt{3})$ is

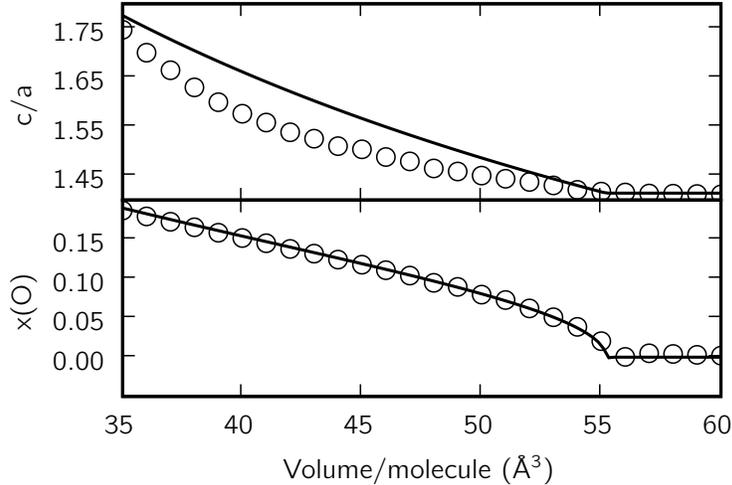


Figure 3.4: Structural parameters of the $\tilde{\beta}$ structure vs. volume per formula unit. Top panel: c/a ratio. Bottom panel: $x(O)$. Symbols represent first-principles calculations; lines are fits to an ideal rigid-unit geometry.

precisely the one at which Si-O bond lengths are preserved to first order in the distortion amplitude. While the O-Si-O bond angles do change at first order, this may entail a smaller energy cost than for bond-length changes. As expected from this analysis, we find that our first-principles O-Si-O bond angles differ from the ideal ones by $\sim 4\%$. In short, it appears that it is energetically more important to preserve bond lengths than bond angles, and that for geometrical reasons this translates into an enhanced freedom for the c/a ratio to deviate from the ideal rigid-unit geometry.

3.3.2 Phonons

Phonons at Γ in $\tilde{\alpha}$ -cristobalite

We next repeated the calculation of the phonon frequencies for the fully relaxed $\tilde{\alpha}$ and $\tilde{\beta}$ cristobalite structures. We did this in order to compare with experimental measurements and also to check the stability of the structures and to investigate, at least at harmonic order, the nature of the energy landscape around these structures. An analysis extending beyond the harmonic approximation will be

Table 3.2: Infrared-active phonon modes at Γ in $\tilde{\alpha}$ -cristobalite ($P4_12_12$). (E modes are also Raman-active.) For A_2 modes, ω_{LO} refers to a phonon with $\hat{q} = \hat{z}$, while for E modes ω_{LO} refers to \hat{q} lying in x-y plane.

Irrep	ω_{TO} (cm ⁻¹)	\tilde{Z}_λ	ω_{LO} (cm ⁻¹)
E	127	0.05	128
E	259	0.04	260
A ₂	285	0.20	293
E	357	0.18	360
E	440	0.74	507
A ₂	462	0.67	515
E	584	0.23	591
A ₂	751	0.52	764
E	752	0.02	752
A ₂	1050	1.52	1201
E	1170	0.17	1165
E	1048	1.55	1208

presented in Sec. 3.3.3.

The decomposition of the optical Γ phonons into irreducible representations for the $\tilde{\alpha}$ structure in space group $P4_12_12$ is

$$\Gamma(\tilde{\alpha}) = 4A_1 \oplus 4A_2 \oplus 5B_1 \oplus 4B_2 \oplus 8E. \quad (3.3)$$

(The translational A_2 and E zero modes have been omitted.) Only the E modes are doubly degenerate; all others are non-degenerate.

Tables 3.2 and 3.3 present the phonon frequencies at the Γ point for the fully relaxed $\tilde{\alpha}$ -cristobalite structure. All phonon frequencies are positive, although some appear to be rather low in frequency. For the infrared (IR) active modes shown in Table 3.2, the transverse mode frequencies were computed initially, and their mode dynamical charges were also computed using

$$\tilde{Z}_{\lambda,\alpha}^* = \sum_{i\beta} \frac{1}{\sqrt{M_i}} \xi_{i,\lambda\beta} Z_{i,\alpha\beta}^* \quad (3.4)$$

Table 3.3: Raman-only phonon modes at Γ in $\tilde{\alpha}$ -cristobalite ($P4_12_12$).

Irrep	ω (cm ⁻¹)	Irrep	ω (cm ⁻¹)	Irrep	ω (cm ⁻¹)
B ₁	29	B ₁	358	A ₁	1046
B ₁	103	A ₁	378	B ₁	1049
A ₁	197	B ₂	410	B ₂	1109
B ₂	275	B ₁	745		
A ₁	350	B ₂	750		

where $\xi_{i,\lambda\beta}$ is an eigenvector of the dynamical matrix, $Z_{i,\alpha\beta}^*$ is the Born atomic charge tensor, and M_i is the mass of the i -th atom in amu. The norms of the mode-charge vectors $\tilde{Z}_\lambda^* = [\sum_\alpha (\tilde{Z}_{\lambda,\alpha}^*)^2]^{1/2}$ are also given in the Table. The longitudinal dynamical matrix was then constructed and diagonalized using standard methods [81], and the resulting LO mode frequencies are presented in the last column of Table 3.2. It can be seen that there are wide variations in the mode dynamical charges, and consequently, large variations in the LO–TO splittings.

Phonons at Γ and M in $\tilde{\beta}$ -cristobalite

Similar calculations of phonon frequencies were also carried out for the $\tilde{\beta}$ structure proposed by Wright and Leadbetter [50] for β -cristobalite. Since the primitive cell of the $\tilde{\beta}$ structure contains only two formula units while the $\tilde{\alpha}$ structure contains four, it should be kept in mind that the Γ point of the $\tilde{\alpha}$ structure maps not only into the Γ point of the $\tilde{\beta}$ structure, but also into a second point that would be denoted as $X = (2\pi/a_c)(110)$ in the original fcc frame, or $(2\pi/a)(100)$ (where $a \simeq a_c/\sqrt{2}$) in the rotated frame of Fig. 3.1; we shall refer to this as the M point in accordance with the conventional labeling of the body-centered-tetragonal (bct) primitive cell in the latter frame. The decompositions of the Γ and M phonons

Table 3.4: Γ phonons in $\tilde{\beta}$ -cristobalite ($I\bar{4}2d$). In italics we show for each phonon in $\tilde{\beta}$ -cristobalite a closest phonon in $\tilde{\alpha}$ -cristobalite ($P4_12_12$).

Phonon in $\tilde{\beta}$ structure			Closest in $\tilde{\alpha}$ structure	
Irrep	ω (cm ⁻¹)	\tilde{Z}_λ	Irrep	ω (cm ⁻¹)
Infrared and Raman				
E	126	0.06	<i>E</i>	<i>127</i>
B ₂	425	0.79	<i>A₂</i>	<i>462</i>
E	444	0.72	<i>E</i>	<i>440</i>
E	748	0.51	<i>E</i>	<i>752</i>
B ₂	1038	1.56	<i>A₂</i>	<i>1050</i>
E	1047	1.52	<i>E</i>	<i>1048</i>
Raman only				
A ₁	289		<i>B₁</i>	<i>29</i>
B ₁	406		<i>A₁</i>	<i>350</i>
B ₁	737		<i>B₁</i>	<i>745</i>
Inactive				
A ₂	357		<i>B₂</i>	<i>410</i>
A ₂	1097		<i>B₂</i>	<i>1109</i>

Table 3.5: M phonons in $\tilde{\beta}$ -cristobalite ($I\bar{4}2d$). In italics we show for each phonon in $\tilde{\beta}$ -cristobalite a closest phonons in $\tilde{\alpha}$ -cristobalite ($P4_12_12$).

Phonon in $\tilde{\beta}$ structure		Closest in $\tilde{\alpha}$ structure			
Irrep	ω (cm ⁻¹)	Irrep	ω (cm ⁻¹)	Irrep	ω (cm ⁻¹)
M ₃ M ₄	35	<i>B₁</i>	<i>103</i>	<i>A₁</i>	<i>197</i>
M ₅	281	<i>E</i>	<i>259</i>	<i>E</i>	<i>357</i>
M ₁ M ₂	316	<i>B₂</i>	<i>275</i>	<i>A₂</i>	<i>285</i>
M ₃ M ₄	336	<i>B₁</i>	<i>358</i>	<i>A₁</i>	<i>378</i>
M ₅	372	<i>E</i>	<i>259</i>	<i>E</i>	<i>357</i>
M ₅	586	<i>E</i>	<i>584</i>		
M ₁ M ₂	780	<i>B₂</i>	<i>750</i>	<i>A₂</i>	<i>751</i>
M ₃ M ₄	1045	<i>A₁</i>	<i>1046</i>	<i>B₁</i>	<i>1049</i>
M ₅	1162	<i>E</i>	<i>1170</i>		

Table 3.6: Relations between unstable phonons in “ideal structure” and phonons in $\tilde{\alpha}$ and $\tilde{\beta}$ structures.

	Ideal cm ⁻¹	$\tilde{\alpha}$ -cristobalite cm ⁻¹	Irrep	$\tilde{\beta}$ -cristobalite cm ⁻¹	Irrep
Γ	$i83$	29	B ₁	289	A ₁
Γ	$i83$	127	E	126	E
Γ	$i83$	127	E	126	E
X	$i53$	197	A ₁	35	M ₃ M ₄
X	$i53$	103	B ₁	35	M ₃ M ₄

into irreducible representations for the $\tilde{\beta}$ structure in space group $I\bar{4}2d$ are

$$\Gamma(\tilde{\beta}) = 1A_1 \oplus 2A_2 \oplus 2B_1 \oplus 2B_2 \oplus 4E, \quad (3.5)$$

$$M(\tilde{\beta}) = 2M_1M_2 \oplus 3M_3M_4 \oplus 4M_5. \quad (3.6)$$

(Translational B₂ and E zero modes have been omitted.) All M-point modes and Γ -point E modes are doubly degenerate, while other modes are non-degenerate.

In Tables 3.4 and 3.5 we present the our results for the Γ -point and M-point phonon modes, respectively, of $\tilde{\beta}$ -cristobalite. The frequencies given for the IR-active modes at Γ are the transverse ones only. The tables also show the correspondences between the Γ modes in the $\tilde{\alpha}$ -cristobalite structure and the Γ and M modes in the $\tilde{\beta}$ -cristobalite structure, as determined by comparing phonon eigenvectors.

Relation to unstable phonons in the cubic phase

The triply degenerate Γ -point mode of the cubic structure having imaginary frequency $i83 \text{ cm}^{-1}$, which condensed to form the $\tilde{\beta}$ structure, now has positive frequencies of 289 cm^{-1} for the non-degenerate A₁ mode and 126 cm^{-1} for the E doublet in the $\tilde{\beta}$ structure. This same triplet corresponds, in the $\tilde{\alpha}$ structure, to the lowest-frequency phonon of frequency 29 cm^{-1} , which has symmetry B₁, and

to an E doublet at 127 cm^{-1} having almost the same frequency as in the $\tilde{\beta}$ structure. The doubly-degenerate unstable mode of the cubic structure at $i53 \text{ cm}^{-1}$, which condensed to form the $\tilde{\alpha}$ structure, now appears in the $\tilde{\alpha}$ structure at frequencies 197 cm^{-1} and 103 cm^{-1} with symmetries A_1 and B_1 , respectively. In the $\tilde{\beta}$ structure, on the other hand, the same doublet appears as the lowest-frequency phonon in that structure, namely the doublet at 35 cm^{-1} with symmetry M_3M_4 . These relations between the unstable phonons in the “ideal structure” and the phonons in $\tilde{\alpha}$ and $\tilde{\beta}$ structures are shown in Table 3.6.

Comparison with experiment for α -cristobalite

In view of the recent questions that have been posed about the identity of the α -cristobalite phase [53], we have carried out a more detailed analysis of the phonons in the $\tilde{\alpha}$ structure. In particular, we have calculated the LO frequencies of the Γ -point phonons in $\tilde{\alpha}$ -cristobalite as a function of the angle at which the limit $\hat{q} \rightarrow 0$ is taken. It turns out that the labels A_2 and E are not well-defined at arbitrary \hat{q} because of mixing between modes of these symmetries. Moreover, it can happen that if one starts with an E mode at $\hat{q} \parallel \hat{z}$ and follows the branch as \hat{q} is rotated, one arrives at an A_2 mode when \hat{q} lies in the x - y plane, or vice versa. Experiments have typically been done on powder samples, so that one should in principle average the phonon spectrum over all possible directions for $q \rightarrow 0$. Moreover, some phonon modes with E symmetry have a very small LO-TO splitting, so they would most likely appear in experiment as a single line.

For all these reasons, a direct comparison of experimental data with our results as presented in Tables 3.2 and 3.3 is problematic. Nevertheless, we attempt such a comparison in Table 3.7. Despite the difficulties, the agreement with experimental data is rather good, with a few exceptions that will be discussed shortly. We generally underestimate the experimental frequencies by $\sim 20 \text{ cm}^{-1}$ for lower

Table 3.7: Left: Computed mode frequencies and irreps for α -cristobalite, with direction of dynamical polarization in parentheses for IR-active modes. The modes that are adiabatically connected as \hat{q} is rotated from \hat{x} to \hat{z} appear on the same line. All modes other than A_2 modes are Raman-active. Right: Tentative assignments to measured mode frequencies in powder samples.

Present theory					Experimental data			
$\hat{q} \parallel \hat{x}$		$\hat{q} \parallel \hat{z}$			IR ^g	Raman ^h		
cm ⁻¹	Irrep	cm ⁻¹	Irrep	Notes	cm ⁻¹	cm ⁻¹	Irrep ⁱ	Notes
1208	E(x)	1201	A ₂ (z)	<i>a,e</i>	1144		A ₂	<i>c</i>
1170	E(y)	1170	E(y)	<i>d</i>	1196	1193	E	<i>d</i>
1165	E(x)	1170	E(x)	<i>d</i>				
1109	B ₂	1109	B ₂	<i>d</i>		1188	B ₂	<i>d</i>
1050	A ₂ (z)	1048	E(x)	<i>e</i>				
1048	E(y)	1048	E(y)		1100	—	E	<i>c</i>
1049	B ₁	1049	B ₁	<i>d</i>		1086	A ₁ or B ₁	<i>d</i>
1046	A ₁	1046	A ₁	<i>d</i>		1076	A ₁ or B ₁	
751	A ₂ (z)	764	A ₂ (z)	<i>d</i>	798			
752	E(y)	752	E(y)	<i>b</i>				
752	E(x)	752	E(x)	<i>b</i>				
750	B ₂	750	B ₂	<i>d</i>		796		<i>d</i>
745	B ₁	745	B ₁			785	B ₁	
591	E(x)	584	E(x)	<i>d</i>				
584	E(y)	584	E(y)	<i>d</i>	625	—	E	<i>d</i>
507	E(x)	515	A ₂ (z)	<i>a,e</i>				
462	A ₂ (z)	440	E(x)	<i>e</i>	495		A ₂	<i>c</i>
440	E(y)	440	E(y)		480	485?	E	<i>c</i>
410	B ₂	410	B ₂	<i>d</i>		426	A ₁ or B ₂	<i>d</i>
378	A ₁	378	A ₁	<i>d</i>				
360	E(x)	357	E(x)	<i>d</i>				
357	E(y)	357	E(y)	<i>d</i>	380	380	E	<i>d</i>
358	B ₁	358	B ₁	<i>d</i>				
350	A ₁	350	A ₁			368	A ₁ or B ₁	
285	A ₂ (z)	293	A ₂ (z)	<i>d</i>	300		A ₂	<i>d</i>
275	B ₂	275	B ₂	<i>d</i>		286	B ₂	<i>d</i>
260	E(x)	259	E(x)	<i>d</i>				
259	E(y)	259	E(y)	<i>d</i>	276	275	E	<i>d</i>
197	A ₁	197	A ₁	<i>d,f</i>		233	A ₁	<i>d</i>
128	E(x)	127	E(x)	<i>f</i>				
127	E(y)	127	E(y)	<i>f</i>	147	—	E	
103	B ₁	103	B ₁	<i>d,f</i>		121	B ₁	<i>d</i>
29	B ₁	29	B ₁	<i>f</i>		50	B ₁	

^a Not pure LO at all \hat{q} . ^b LO-TO splitting is negligible. ^c Part of structured peak. ^d Inactive in β phase. ^e In $\tilde{\beta}$ structure the A₂ component also becomes Raman active. ^f Corresponds to RUM mode in ideal cristobalite. ^g Observation from [60]. ^h Observation of 50 cm⁻¹ mode is from [86], all others from [87]. ⁱ Empirical assignments from [60].

frequency phonons and by $\sim 35 \text{ cm}^{-1}$ for higher frequencies, but otherwise our results reproduce the experimental pattern of frequencies, and the irrep assignments are also consistent with those obtained from empirical models [60]. Furthermore, the identification of the modes that are not expected to be active in the β phase (fifth and ninth columns of Table 3.7) because they correspond to zone boundary modes in the β phase (see Table 3.5) or because they are inactive Γ -point modes (see Table 3.4) agrees well with the results reported in Ref. [60].

The first anomaly is related to the experimentally observed IR mode with a frequency of 798 cm^{-1} in the α phase that does not disappear upon transition to the β phase. Finnie *et al.* [88] explained this by suggesting that a two-phonon process in the β phase replaces the fundamental mode of the α phase. Our calculations identify two almost-degenerate IR modes that are close to this frequency, an A_2 mode at 751 cm^{-1} and an E mode 752 cm^{-1} . In the $\tilde{\alpha}$ structure both of these modes are IR active, but the Born charge of the E mode is only 0.02 while that of A_2 is 0.52, which means that the E mode in the $\tilde{\alpha}$ structure is almost invisible. In the $\tilde{\beta}$ structure, the A_2 mode disappears since it is no longer at Γ . On the other hand, the E mode remains at the Γ point and its Born charge is increased to 0.51. These results suggest a possible explanation for the ‘‘anomaly,’’ namely that there are two IR modes in the α phase; one of them is much weaker than the other, but upon the transition to the β phase, the stronger one disappears by symmetry while the weaker one greatly increases its IR activity. The reason why the 752 cm^{-1} E phonon in the $\tilde{\alpha}$ structure acquires a larger Born charge upon converting to the $\tilde{\beta}$ structure is that it gets some admixture of the 440 cm^{-1} E mode, which has a much larger Born charge (0.74) than that of the 752 cm^{-1} mode (0.02).

The second anomaly is related to the 1076 cm^{-1} mode that is still present upon the transition to the β phase in the form of a fairly broad feature (see Fig.

1 in Ref. [87]), whereas it would be expected to vanish by symmetry. Swainson *et al.* [60] attributed this mode to a possible higher-order process. We think that it could also be related to the fact that the 1050 cm^{-1} A_2 mode that is Raman inactive in the $\tilde{\alpha}$ -phase becomes Raman active in $\tilde{\beta}$ -phase.

We also predict several phonon modes that are not seen experimentally, such as the Raman-active modes at 127, 358, 378, 584, 752 and 1048 cm^{-1} . Since we have not computed Raman matrix elements, it is possible that the Raman intensities are small for these modes. We also find one weak IR-active mode at 752 cm^{-1} that is not seen in the experiments.

A very low-frequency phonon at 50 cm^{-1} has been reported in α cristobalite [89, 86, 90]. We believe this most likely corresponds to the B_1 phonon that we have calculated to appear at 29 cm^{-1} , corresponding closely to the RUM mode that takes the ideal cubic cristobalite structure into the $\tilde{\beta}$ structure. The same conclusion regarding the lowest-frequency B_1 phonon was reached in Ref. [60]. The minimal-energy path between α and β phases that is related to this low-frequency phonon is discussed in Sec. 3.3.3.

3.3.3 SiO₂ cristobalite stability analysis

As shown in Sec. 3.3.2, all calculated optical phonons in $\tilde{\alpha}$ and $\tilde{\beta}$ cristobalites have $\omega^2 > 0$, indicating that the relaxed structure is stable with respect to those modes. In view of the suggestion in Ref. [53] that the α phase might locally have D_2 rather than D_4 point-group symmetry, we checked carefully for instabilities leading from the $\tilde{\alpha}$ structure to D_2 structures, but found none. The possible subgroups of $P4_12_12$ (D_4^4) having D_2 symmetry (without reduced translational symmetry) are $C222_1$ (D_2^5) and $P2_12_12_1$ (D_2^4), and the phonon distortions leading to these symmetry-lowered structures are the ones of B_2 and B_1 symmetry respectively. The lowest-frequency mode of B_2 symmetry is at 275 cm^{-1} , so there is

certainly no sign of an instability there. On the other hand, the lowest-frequency B_1 phonon is nearly soft at 29 cm^{-1} , suggesting that the $\tilde{\alpha}$ structure is nearly unstable to a spontaneous transformation into the $P2_12_12_1$ structure. To check this possibility more carefully, we started from the relaxed $\tilde{\alpha}$ structure and followed the distortion corresponding to the 29 cm^{-1} B_1 phonon, and confirmed that the energy increases monotonically (no double-well structure). Moreover, starting from one of these structures having a small amount of the 29 cm^{-1} mode frozen in, a subsequent relaxation inside the resulting space group $P2_12_12_1$ lead to a recovery of the starting space group $P4_12_12$. We thus conclude, at least within our zero-temperature first-principles calculations, that the $\tilde{\alpha}$ structure is locally stable, i.e., does not spontaneously lower its symmetry to D_2 .

Nevertheless, the presence of several modes of quite low frequency in the $\tilde{\alpha}$ -cristobalite structure may be suggestive of low-energy pathways leading from the $\tilde{\alpha}$ to the $\tilde{\beta}$ structure or between domains of the $\tilde{\alpha}$ structure. For example, we have shown above that the lowest-frequency 29 cm^{-1} mode in the $\tilde{\alpha}$ -cristobalite structure corresponds to a phonon of the “ideal structure” that leads to the $\tilde{\beta}$ structure, and vice versa. This might suggest that there is a relatively low energy barrier in the configuration space that connects one structure to the other. Other phonons from the unstable triplet and doublet in the “ideal structure” have frequencies that are higher, but still low enough to suggest that there is a low energy barrier for creation of the domains. In order to clarify these issues, we shall explore the energy landscape around the $\tilde{\alpha}$ and $\tilde{\beta}$ structures in more detail in Sec. 3.3.3. First, however, we begin with a general discussion of RUMs in the cristobalite phases in the next subsection.

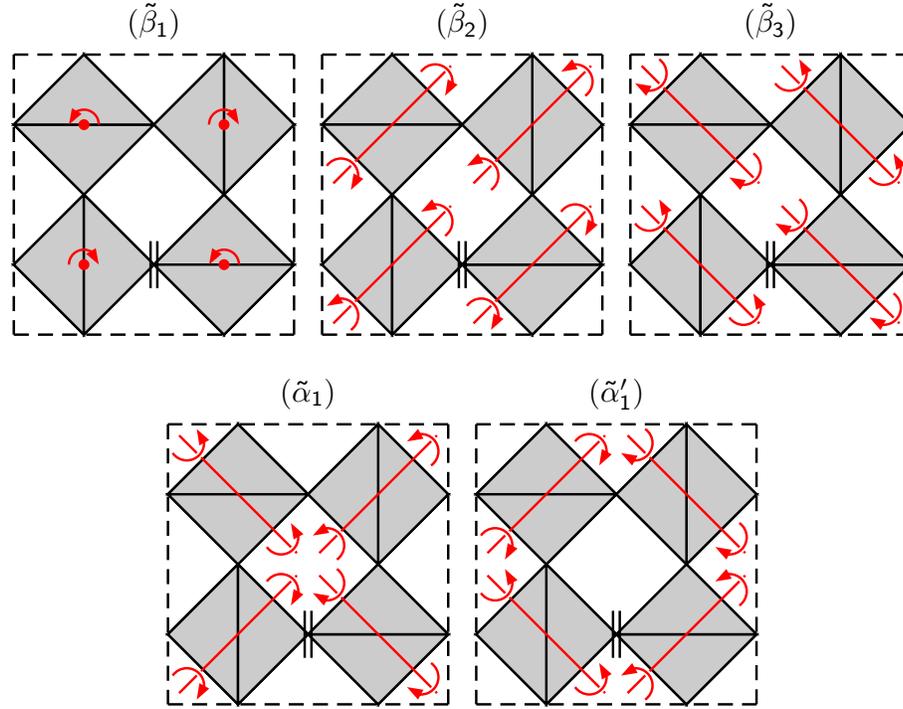


Figure 3.5: Five linearized RUMs in ideal cristobalite $Z=4$ structure.

Rigid unit mode analysis

Here we analyze the RUMs present in the high-symmetry cubic structure [91], but constrained to maintain the periodicity of the four-formula-unit ($Z=4$) cell of the $\tilde{\alpha}$ -cristobalite structure (i.e., containing two unit cells of the $\tilde{\beta}$ -cristobalite structure). When these constraints are imposed, we find that there are five linearly independent RUM distortions as shown in Fig. 3.5. The first three of these distortions, which we denote as $\tilde{\beta}_1$, $\tilde{\beta}_2$, and $\tilde{\beta}_3$, consist of tetrahedral rotations of alternating signs about a single Cartesian axis, and carry the system into the $I\bar{4}2d$ symmetry of the $\tilde{\beta}$ phase. The last two, which we denote as $\tilde{\alpha}_1$ and $\tilde{\alpha}'_1$, consist of a pattern of rotations around the $[110]$ and $[1\bar{1}0]$ axes in the frame of Fig. 3.1, together with small translations of the tetrahedra needed to keep them connected at their apices, and carry the system into the $P4_12_12$ (or, for $\tilde{\alpha}'_1$, into the enantiomorphic $P4_32_12$) symmetry of the $\tilde{\alpha}$ -cristobalite structure. Not shown

in Fig. 3.1 are RUM rotations $\tilde{\alpha}_2$ and $\tilde{\alpha}'_2$ associated with a second X point, and $\tilde{\alpha}_3$ and $\tilde{\alpha}'_3$ associated with a third X point.

Within the context of an ideal rigid-unit geometry (in which no additional relaxations are allowed), one can make the following mathematical analysis. The freezing in of the ideal $\tilde{\beta}_1$ RUM leads to a $\tilde{\beta}$ structure oriented as in Sec. 3.3.1, whereas the freezing in of the ideal $\tilde{\alpha}_1$ RUM leads to an $\tilde{\alpha}$ structure as in that section. In the ideal $\tilde{\beta}$ structure, all five of the modes shown in Fig. 3.5 remain as true RUMs – i.e., the tetrahedra can remain undistorted to first order in the mode amplitudes. Thus, all five modes are expected to have low frequencies in a more realistic description. However, modes $\tilde{\alpha}_2$, $\tilde{\alpha}'_2$, $\tilde{\alpha}_3$, and $\tilde{\alpha}'_3$ are no longer RUMs when a finite $\tilde{\beta}_1$ RUM is present.

In the $\tilde{\alpha}_1$ structure, only the $\tilde{\alpha}_1$, $\tilde{\alpha}'_1$, and $\tilde{\beta}_1$ distortions remain as true RUMs. However, the $\tilde{\beta}_2$ and $\tilde{\beta}_3$ modes at least share the same translational symmetry, and so may be expected to have somewhat low frequencies. The remaining $\tilde{\alpha}_2$, $\tilde{\alpha}'_2$, $\tilde{\alpha}_3$, and $\tilde{\alpha}'_3$ modes are incompatible both in the RUM sense and in their translational periodicity [92].

Not surprisingly, when we impose the translational periodicity consistent with the five modes shown in Fig. 3.1, we confirm that these five distortions correspond quite closely to the five unstable phonon modes that we found in the ideal structure. The unstable Γ modes correspond to $\tilde{\beta}_1$, $\tilde{\beta}_2$ and $\tilde{\beta}_3$, while the unstable X modes correspond to $\tilde{\alpha}_1$ and $\tilde{\alpha}'_1$. They also correspond closely to the low-frequency phonons in the $\tilde{\alpha}$ and $\tilde{\beta}$ structures as discussed in Sec. 3.3.2.

Extending the mathematical analysis of the compatibility of RUMs, it can be shown that there is an entire three-dimensional subspace of rigid-unit structures (i.e., with the tetrahedral rigidity condition satisfied exactly) in which finite rotations of type $(\alpha_1, \alpha'_1, \beta_1)$ are simultaneously present, and having the space group $P2_12_12_1$ that is induced if any two of them are present. In a similar way, there

are two additional three-dimensional (3D) manifolds $(\alpha_2, \alpha'_2, \beta_2)$ and $(\alpha_3, \alpha'_3, \beta_3)$ corresponding to different choices of the X point and thus having different $Z=4$ supercells. The three subspaces meet only at a single point (the cubic phase with all angles vanishing), and RUMs selected from different 3D manifolds are always incompatible with each other in the sense that the perfect tetrahedral rigidity cannot be preserved when imposing both. This picture has important consequences for our understanding of the possible paths connecting domains of the $\tilde{\alpha}$ and $\tilde{\beta}$ structures, as discussed below.

Energy landscape inside 3D manifolds

After we have explained the origin of the low-energy phonons in the $\tilde{\alpha}$ and $\tilde{\beta}$ structures by relating them to RUM modes, we would now like to explore the energy landscape around these structures. To do so, we begin by finding a configuration space containing both structures. Since there is no group-subgroup relation between the $\tilde{\alpha}$ and $\tilde{\beta}$ structures, we seek a maximal common subgroup of both structures. In the present case, this leads to the space group $P2_12_12_1$ (D_2^4).

In the $P2_12_12_1$ configuration space, the $\tilde{\alpha}$ and $\tilde{\beta}$ structures represent two special points, and we know that the energy has local minima at these points because all computed phonon frequencies were found to be positive there. But then we also expect that there must be at least one saddle point connecting these points. To search for this saddle point, we started from the midpoint between the $\tilde{\alpha}$ and $\tilde{\beta}$ structures in the 12-dimensional $P2_12_12_1$ configuration space (described by nine internal coordinates and three cell parameters), and identified the unit vector \hat{e} pointing between the two structures. We then applied a simple saddle-point search strategy in which component of the force vector parallel to \hat{e} was reversed in sign before executing the steepest-descent update. This algorithm can be expected to succeed if the saddle point is not too far from the midpoint

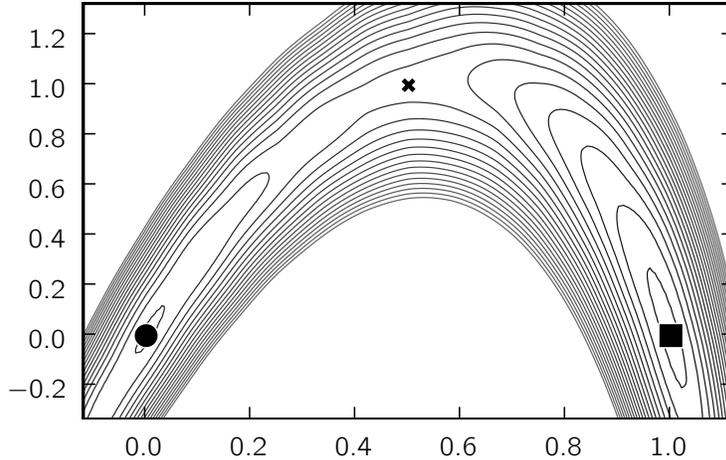


Figure 3.6: Energy in plane defined by the $\tilde{\alpha}$ structure (filled circle), $\tilde{\beta}$ structure (filled square), and saddle point (cross). Coordinates are chosen such that these structures occur at $(0, 0)$, $(1, 0)$, and $(0.5, 1)$, respectively. The energy difference separating contours is 3 meV per SiO_2 formula unit.

and if the principal axis of the negative Hessian eigenvalue at the saddle point is roughly parallel to \hat{e} . In the present case, it led to a rapid convergence on the desired saddle point. Surprisingly, we find that the saddle point has a very low energy, only 5 meV per formula unit above that of the $\tilde{\alpha}$ structure, or 17 meV above that of the $\tilde{\beta}$ structure.

The three points representing the $\tilde{\alpha}$ and $\tilde{\beta}$ structures and the saddle point determine a plane in the 12-dimensional configuration space. To confirm that the path running through the saddle point encounters only a single barrier, we have plotted the structural energy in this plane (without relaxation of other coordinates) in Fig. 3.6. We have somewhat arbitrarily carried out a linear transformation on the coordinates in such a way that the $\tilde{\alpha}$ and $\tilde{\beta}$ structures lie at $(0, 0)$ and $(1, 0)$ respectively, while the saddle point lies at $(0.5, 1)$, in Fig. 3.6. The results confirm the picture of a simple barrier of 5 meV encountered when going from the $\tilde{\alpha}$ to the $\tilde{\beta}$ structure.

Note that a transformation that would lead from the $\tilde{\alpha}$ to the $\tilde{\beta}$ structure along a straight line in configuration space would have an enormously higher barrier

of 195 meV per formula unit. This is because the straight-line path is a poor approximation to a RUM. If instead we follow a curved minimum-energy path from $\tilde{\alpha}$ through the saddle to $\tilde{\beta}$ and compute the relaxed Si-O bond lengths and O-Si-O bond angles along this path, we find that these remain almost constant. This strongly suggests that this minimum-energy path may be well approximated by some RUM-like distortion.

In the Sec. 3.3.3, we pointed out that within the framework of ideal rigid-unit rotations, there is an entire three-dimensional subspace of structures for which the tetrahedral rigidity conditions are satisfied exactly, in which finite rotations of all three types are present. We label an arbitrary configuration in this 3D manifold by $(\alpha_1, \alpha'_1, \beta_1)$, where by convention the order of operations is $\tilde{\alpha}_1$ followed by $\tilde{\alpha}'_1$ and then $\tilde{\beta}_1$. The space group at a generic point in this 3D manifold is $P2_12_12_1$, the same one we have just been discussing. It thus seems likely that the minimum-energy path in Fig. 3.6 may correspond approximately to a path from the point $(\alpha_1, 0, 0)$ to the point $(0, 0, \beta_1)$ and lying, at least approximately, in the two-dimensional (2D) subspace $(\alpha_1, 0, \beta_1)$.

To test this conjecture, we first created an ideal rigid-unit structure for each pair of angles (α_1, β_1) on a two-dimensional mesh. We then used our first-principles calculations to relax each structure subject to the constraint that these two angle variables should not change. Technically, we did this by carrying out the minimization of the energy in the ten-dimensional subspace orthogonal to the two-dimensional surface for each starting point (α_1, β_1) . We typically found that these relaxations were small, confirming the approximate validity of the RUM picture.

The energy surface determined in this way is plotted as a function of rotation angles α_1 and β_1 in Fig. 3.7. The minima corresponding to the $\tilde{\alpha}$ structure are immediately visible near the left and right sides of the figure, while those

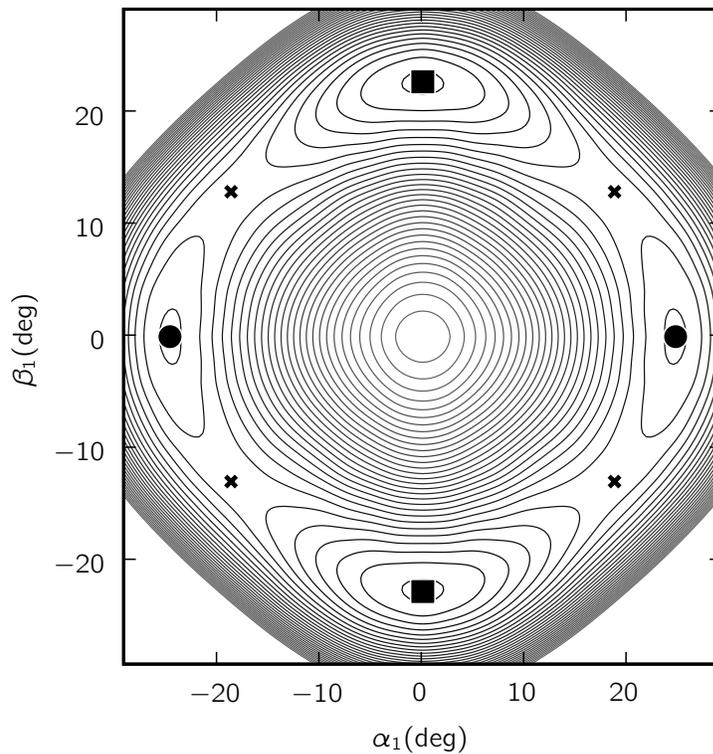


Figure 3.7: Energy as a function of rotation angles ϕ_{α_1} and ϕ_{β_1} , corresponding to rotations $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ shown in Fig. 3.5. The origin corresponds to “ideal” cristobalite. Filled squares at top and bottom denote $\tilde{\beta}$ minima, filled circles at left and right denote $\tilde{\alpha}_1$ minima, and crosses denote saddle points, as in Fig. 3.6. The energy difference separating contours is 3 meV per SiO_2 formula unit.

corresponding to the slightly lower-energy $\tilde{\beta}$ structure appear near the top and bottom. The minimum-energy path appears to be roughly circular on this plot, and four equivalent saddle points are apparent at $\alpha_1 \simeq \pm 19^\circ$ and $\beta_1 \simeq \pm 13^\circ$. These saddle points are equivalent to the one identified in Fig. 3.6, with a barrier height of 5 meV per formula unit relative to the $\tilde{\alpha}$ structure. We thus confirm the presence of a very low-energy barrier between these structures, and identify it as approximating a certain path in the space of rigid-unit rotations. A video animation showing the evolution of the structure along this path is provided in the supplementary material [93].

It is important to note that, according to the simplified model of Eq. (1) of Ref. [54], the energy would remain exactly zero on the entire (α_1, β_1) surface of Fig. 3.7 since the ideal rigid-unit structures satisfy the rigidity conditions analytically. The RUM framework envisages extensions to make the model more realistic; one way to do this is by adding an energy term that depends on the relative tilts of neighboring tetrahedra [94]. We tried this by introducing a simple double-well potential model that penalizes the departure of the Si-O-Si bond angles from a preferred bending angle. In this model the change of total energy per formula unit is

$$\Delta E = \frac{E_0}{N} \sum_i \left[-2 \left(\frac{\pi - \phi_i}{\pi - \phi_0} \right)^2 + \left(\frac{\pi - \phi_i}{\pi - \phi_0} \right)^4 \right], \quad (3.7)$$

where the sum runs over all N Si-O-Si bond angles ϕ_i in the unit cell. We found that we could obtain an optimal fit ² to the results of our first-principles calculations using parameters $E_0 = 83$ meV per formula unit and $\phi_0 = 145^\circ$. The energy landscape of the fitted model looks very similar to the results plotted in Fig. 3.7.

²The fitting was done to minimize the RMS error of four quantities, namely the equilibrium angles and the energy difference (relative to the symmetric structure) for the α and β structures.

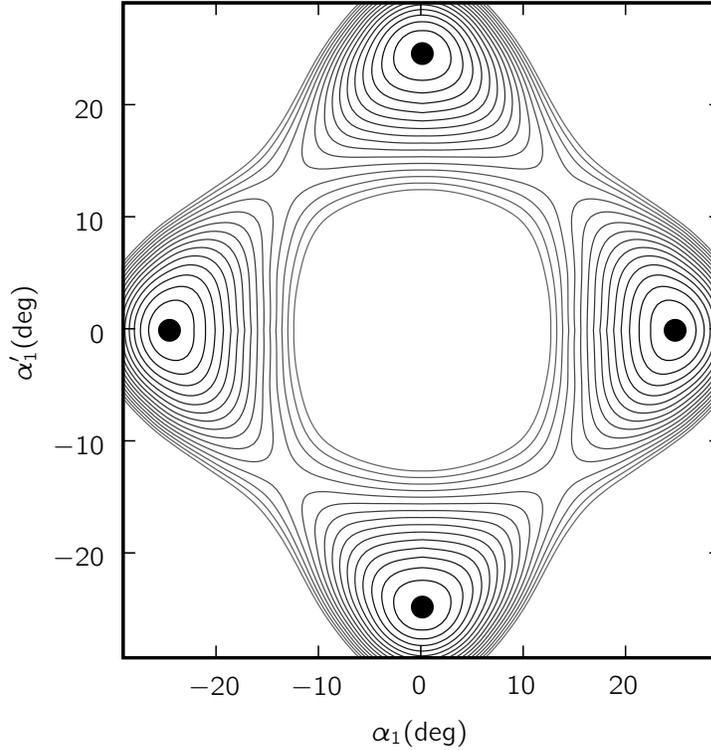


Figure 3.8: Energy as a function of rotation angles ϕ_{α_1} and $\phi_{\alpha'_1}$, corresponding to rotations $\tilde{\alpha}_1$ and $\tilde{\alpha}'_1$ shown in Fig. 3.5. The origin corresponds to “ideal” cristobalite. The four minima (filled circles) correspond to the $\tilde{\alpha}'_1$ structure (top and bottom) and to the $\tilde{\alpha}_1$ structure (left and right). The energy difference separating contours is 3 meV per SiO_2 formula unit.

In particular, the overall circular aspects of the energy landscape and minimal-energy path in Fig. 3.7 are reproduced. However, the model unfortunately assigns identical energies to the $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ structures, and moreover predicts a path connecting them on which the energy remains completely flat. This happens because, for any given pair of angles (α_1, β_1) on or near this path, one can find a small $\tilde{\alpha}'_1$ such that the rigid-unit structure $(\alpha_1, \alpha'_1, \beta_1)$ has all its Si-O-Si bond angles exactly equal to ϕ_0 . Therefore our simplified model of Eq. (3.7), or any other model that depends solely on the Si-O-Si angles, predicts a zero-barrier path between $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ structures. This behavior is reminiscent of an early model of Nieuwenkamp [95] for β cristobalite, in which the Si-O-Si bond was assumed to rotate freely on an annulus lying in the plane that is equidistant between Si atoms.

To further test the model of Eq. (3.7), we performed first-principles calculations on a mesh of (unrelaxed) structures in the 2D space ($\alpha_1, \alpha'_1, \beta_1=0$). The resulting energy landscape is shown in Fig. 3.8. The barrier between $\tilde{\alpha}_1$ and $\tilde{\alpha}'_1$ structures is now about 35 meV, substantially higher than for the path connecting $\tilde{\alpha}$ and $\tilde{\beta}$ structures. When we use the same fitting parameters obtained earlier, we again get very good overall agreement; the energy landscape obtained from our model has the same diamond-like appearance as in Fig. 3.8, and saddle points appear in very similar locations. Moreover, the barrier of 41 meV predicted by the model is in quite good agreement with the first-principles value of 35 meV. However, in this case the picture presented by Fig. 3.8 is somewhat misleading, because it turns out that the entire minimum-energy path lying in the $\beta_1=0$ plane is unstable, and falls to lower energy as $\tilde{\beta}_1$ is turned on. Thus, the apparent saddle points in Fig. 3.8 are actually stationary points with *two* negative eigenvalues in the 3D ($\alpha_1, \alpha'_1, \beta_1$) space. Within the model of Eq. (3.7), in fact, the lowest-energy path connecting the $\tilde{\alpha}_1$ and $\tilde{\alpha}'_1$ structures is actually completely flat, being composed of a segment connecting $\tilde{\alpha}_1$ to $\tilde{\beta}_1$ and then another connecting $\tilde{\beta}_1$ to $\tilde{\alpha}'_1$. This observation agrees with our first-principles calculations, since if we start from the purported saddle-point configuration and do a structural relaxation subject to the constraint that $\alpha_1 = \alpha'_1$, the structure is found to converge to the $\tilde{\beta}_1$ structure as expected.

Cell volume at minima and saddle point

Because we have found the unit-cell parameters to be very sensitive to details of the calculation, we increased the energy cutoff from 22 Ha to 30 Ha in order to obtain an accurate description of the volume changes along the minimum-energy path. We obtain a volume per formula unit of 45.7 and 46.7 Å³ for the $\tilde{\alpha}$ and $\tilde{\beta}$ local minima respectively, so that the volume is about 2.2% larger for

the latter. This is in qualitative agreement with experiments, which show that the β structure is about 5% larger [96], and implies that applied pressure would tend to favor the $\tilde{\alpha}$ phase and raise the α -to- β transition temperature. At the saddle point, we find that the volume per formula unit is 46.8 \AA^3 , which is just slightly larger than for either of the parent-phase structures. This finding may be of interest for future studies of the pressure-dependence of the phase-transition mechanism.

Domain walls

The barriers discussed in Sec. 3.3.3 refer to transformation pathways in which the crystal remains periodic and transforms homogeneously, and the energy barriers are given per unit cell. It would also be of interest to consider the energies of domain walls between various $\tilde{\alpha}$ and $\tilde{\beta}$ structures. This is beyond the scope of the present investigation, but the results for homogeneous transformations may give some hints as to what could be expected. For example, we speculate that domain walls connecting $\tilde{\alpha}$ and $\tilde{\beta}$ structures belonging to the same 3D rigid-unit manifold will probably have a rather low energy per unit area, while those connecting structures belonging to different 3D manifolds would be expected to have much higher energies.

3.4 Discussion

In this section we give a brief overview of several previously proposed models of α and β cristobalite phases, and discuss how the results of our calculations relate to those models.

The RUM model of Ref. [54] describes the β phase as an average cubic structure that has strong dynamical fluctuations occurring simultaneously into RUMs

in all allowed regions of the Brillouin zone. A simplified version of this picture would be one in which the tetrahedra are assumed to be completely free to pivot around their apices, as in Eq. (1) of Ref. [54]. In general, the simultaneous excitation of more than one RUM will have an associated energy cost because the tetrahedra typically cannot remain perfectly rigid while undergoing both kinds of distortion simultaneously. However, as an exception, we have identified 3D rigid-unit manifolds within which the geometrical constraints *can* simultaneously be satisfied. Within the model of Eq. (1) of Ref. [54], or the split-atom model of Ref. [97], the energy landscape within this special 3D manifold would be completely flat, and one would expect that freezing in of one RUM of type $\tilde{\alpha}_1$, $\tilde{\alpha}'_1$ or $\tilde{\beta}_1$ would have no consequence on the energy profile of one of these other RUM distortions.

However, once one goes beyond the simplest versions of the model and includes terms that depend on the Si-O-Si bond angles at the apices, our calculations indicate that the RUM distortions of type $\tilde{\alpha}_1$, $\tilde{\alpha}'_1$ and $\tilde{\beta}_1$ become coupled and have a rich energy landscape. As a step in this direction, the more sophisticated split-atom model having an additional energy term depending on Si-O-Si bond angles [94] should provide an improved description. However, we note that even this model, or any model based solely on Si-O-Si bond angles, still has a nonphysical behavior in that it would necessarily predict zero-energy barriers between the $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ structures, as discussed at the end of Sec. 3.3.3. Nevertheless, we believe that the split-atom and similar models can provide important complementary information to ours, since they are not restricted to periodic supercell structures as ours are.

Among the models of cristobalite phase transitions is also the model of Hatch and Ghose [98]. They argue that the β phase is dynamically and spatially fluctuating between the 12 different possible $\tilde{\alpha}$ domains having $P4_12_12$ space-group

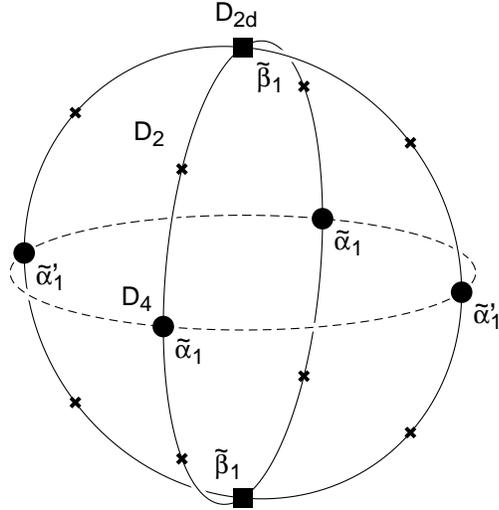


Figure 3.9: Sketch of important states in one of the three-dimensional rigid-unit subspaces discussed in the text. Local energy minima associated with $\tilde{\alpha}$ (D_4) and $\tilde{\beta}$ (D_{2d}) structures are indicated by filled circles and squares respectively. Remainder of space, including saddle points (crosses), has D_2 symmetry.

symmetry. The counting arises because there are three different X points; each exhibits a doublet of degenerate modes leading to enantiomorphic $\tilde{\alpha}_1$ and $\tilde{\alpha}'_1$ structures (see Fig. 3.5); and the tetrahedra can rotate by $\pm\phi$. The model is based on symmetry arguments and assumes that all of the barriers separating these 12 $\tilde{\alpha}$ structures are small. However, our work suggests that the barriers separating different types of $\tilde{\alpha}$ domains have very different barriers. Furthermore, their model does not take into account the fact that the $\tilde{\beta}$ structure is easily accessible with a very low barrier, suggesting that fluctuations into the $\tilde{\beta}$ structure may be more important than some of the other $\tilde{\alpha}$ structures.

Finally, O’Keeffe and Hyde [51] do discuss a path connecting $\tilde{\alpha}$ and $\tilde{\beta}$ structures, but it is of a different type than those discussed above since it connects $\tilde{\alpha}$ and $\tilde{\beta}$ structures belonging to different 3D rigid-unit manifolds. In our notation, their path would connect $\tilde{\alpha}_1$ or $\tilde{\alpha}'_1$ to $\tilde{\beta}_2$ or $\tilde{\beta}_3$, etc. Such a path would involve the simultaneous application of RUM rotations that are incompatible with each other, and as such would be expected to have a high energy barrier.

To clarify our view of the cristobalite phase transitions, we start by emphasizing once again the existence of three distinct 3D rigid-unit manifolds, as described above at the end of Sec. 3.3.3. To review, one of these is described by rotation angles $(\alpha_1, \alpha'_1, \beta_1)$ giving rise to structures of space group $P2_12_12_1$ whose translational periodicity is that corresponding to the X point $(2\pi/a_c)(001)$ or equivalently $(2\pi/a_c)(110)$. This manifold contains the $\tilde{\alpha}_1$, $\tilde{\alpha}'_1$, and $\tilde{\beta}_1$ structures, and their partners with reversed sense of rotation, as shown schematically in Fig. 3.9. The second and third 3D subspaces are described by rotations $(\alpha_2, \alpha'_2, \beta_2)$ and $(\alpha_3, \alpha'_3, \beta_3)$, with periodicities set by X points $(2\pi/a_c)(010) = (2\pi/a_c)(101)$ and $(2\pi/a_c)(100) = (2\pi/a_c)(011)$, respectively. We have found that these three subspaces are essentially incompatible, in the sense that it is not possible to combine rotations taken from any two of them into a combination that preserves the rigid-unit constraints. This occurs in part because these three 3D rigid-unit subspaces have incompatible translational symmetries, but also because of incompatibilities in the patterns of rotations.

The structure of the space sketched in Fig. 3.9 is intended to reflect a three-level hierarchy of energies and energy barriers as suggested by our analysis. In the model of Eq. (3.7), the energy is degenerate for all six of the structures shown in Fig. 3.9, as well as on the solid curves connecting them [93]. According to our first-principles results, this picture is modified so that the $\tilde{\alpha}$ and $\tilde{\beta}$ structures are local minima, with low-energy saddle points (~ 5 meV) between them (see Fig. 3.7). The low curvature of the energy surface along these curves is reflected in the presence in Table 3.6 of a very soft 29 cm^{-1} B_1 mode starting from the $\tilde{\alpha}$ structure, and a 35 cm^{-1} M_3M_4 doublet starting from the $\tilde{\beta}$ structure³. While our calculations have the $\tilde{\alpha}$ structures at a slightly higher energy than the $\tilde{\beta}$ ones,

³The B_1 phonon at 103 cm^{-1} in the $\tilde{\alpha}$ structure (see Tables 3.3 and 3.6) corresponds to moving along the dashed curve in Fig. 3.9.

this is presumably reversed in the true physical system.

The next energy scale in the hierarchy is that associated with the direct paths between $\tilde{\alpha}$ structures in the same 3D manifold, indicated by the dashed lines in Fig. 3.9. As shown in Fig. 3.8, this energy is on the order of ~ 35 meV, so that the true minimum-energy path between neighboring $\tilde{\alpha}$ structures goes instead through (or perhaps nearly through) the $\tilde{\beta}$ structures.

Finally, the highest energies are associated with the barriers separating any of the structures in Fig. 3.9 from any of the structures in the other two 3D subspaces. These barriers are on the order of 80 meV, the energy needed to pass through the undistorted cubic phase. While not enormously larger than the 35 meV mentioned above, this is high enough that we do not expect these barriers to be especially relevant for the phase transitions in this system.

We can now speculate on the nature of the phase transition between α and β cristobalites. We propose that in the lower-temperature α phase, the system is locally frozen onto one of the minima of type $\tilde{\alpha}$ in one of the 3D manifolds, but with substantial fluctuations along the low-energy paths leading to the two neighboring $\tilde{\beta}$ structures in the same manifold. Then, in the higher-temperature β phase, we speculate that the system instead shifts over and condenses locally onto one of these $\tilde{\beta}$ structures, but with substantial fluctuations along the low-energy paths leading to the four neighboring $\tilde{\alpha}$ structures, all in the same 3D manifold. The fact that there are four low-energy paths to fluctuate along, instead of two, is consistent with the fact that the β phase (being the higher-temperature phase) has higher entropy. If the system were truly to freeze onto a single $\tilde{\beta}$ structure, it would be globally tetragonal, with space group $I\bar{4}2d$. However, it is also possible that the system forms on some larger scale into spatiotemporal domains composed of $\tilde{\beta}$ structures from all three of the 3D manifolds, giving an overall average $Fd\bar{3}m$ structure in accord with the picture espoused in Refs. [54, 55].

Let us return for a moment to the recent work of Zhang and Scott [53], who argued that their Raman studies of single crystals of β -cristobalite were inconsistent with O_h symmetry. Assuming D_{2d} symmetry instead for the β phase, these authors then noted that D_4 is not a subgroup of D_{2d} , and thus that the existence of a group-subgroup relation for the phase transition would rule out the assignment of the α phase to the $D_4 \tilde{\alpha}$ structure. On this basis, they suggested that a lower symmetry, such as D_2 , should be considered for α -cristobalite. Our view, instead, is that a group-subgroup relation does not have to hold for the transition, since the transition is known to be of first order, and thus assignments of D_{2d} and D_4 for the α and β phases respectively are not inconsistent. As pointed out in the introduction, while certain spectroscopic signatures of the transition are indicative of a weakly first-order transition, the volume change and latent heat at the transition are substantial. The transition may perhaps be described as a reconstructive transition in the sense of Tolédano and Dmitriev [58], although in the present case the rearrangements of atoms can occur very gently, because of the existence of very low-barrier paths of D_2 symmetry connecting the D_4 ($\tilde{\alpha}$) and D_{2d} ($\tilde{\beta}$) structures. The situation may be somewhat analogous to the tetragonal-to-orthorhombic and orthorhombic-to-rhombohedral transitions in ferroelectric perovskites such as BaTiO_3 and KNbO_3 , where the presence of low-barrier paths of monoclinic symmetry is associated with the weakly first-order nature of the transitions [99].

Unfortunately our calculations are carried out at 0 K with crystal periodicity imposed. It is therefore difficult to draw any firm conclusions about the nature of the phase transitions between cristobalite phases, especially if fluctuations are as important as we think they are, and much of what we have said above must remain speculative. Nevertheless we hope that the results of our calculations

will be of use in developing improved models that may allow for realistic finite-temperature modeling of the phase transitions in this system, ultimately leading to a resolution of the controversies that have surrounded this system over the years.

3.5 Summary

Based on first-principles calculations, we have performed a detailed analysis of the $\tilde{\alpha}$ ($P4_12_12$) and $\tilde{\beta}$ ($I\bar{4}2d$) structures of cristobalite SiO_2 . In particular, we have confirmed that both structures are locally stable against all possible distortions associated with Γ -point modes of the four-formula-unit conventional cell. We have calculated phonon frequencies for the $\tilde{\alpha}$ and $\tilde{\beta}$ structures, compared these to the experimental values, and discussed how the phonons in these two structures are related to each other. We have also tried to resolve some experimental anomalies that were found in spectroscopic studies of the cristobalite phases. Finally, we have explored the energy landscape connecting the $\tilde{\alpha}$ and $\tilde{\beta}$ structures. We have emphasized the existence of three distinct 3D manifolds of structures, each of which contains both $\tilde{\alpha}$ and $\tilde{\beta}$ structures that can be connected to each other within the manifold by paths with a surprisingly small barrier of 5 meV per formula unit, while paths connecting different manifolds have a much higher barrier. While our calculations do not properly treat fluctuations, we nevertheless have speculated on the possible consequences of our findings for the understanding of the α - β phase transition in cristobalite SiO_2 .

Chapter 4

Si-compatible candidates for high- K dielectrics with the $Pbnm$ perovskite structure

We analyze theoretically from first-principles the dielectric tensor components and crystal structure of five classes of $Pbnm$ perovskites. All of these materials are believed to be stable on silicon and are therefore promising candidates for high- K dielectrics. We also analyze the structure of these materials with various simple models, decompose the lattice contribution to the dielectric tensor into force constant matrix eigenmode contributions, explore a peculiar correlation between structural and dielectric anisotropies in these compounds and give phonon frequencies and infrared activities of those modes that are infrared-active. We find that CaZrO_3 , SrZrO_3 , LaHoO_3 , and LaYO_3 are among the most promising candidates for high- K dielectrics among the compounds we considered. We also compare our calculations with available experimental data.

4.1 Introduction

As a result of the ongoing down-scaling of complementary metal-oxide-semiconductor (CMOS) integrated circuits, the SiO_2 gate oxide of field effect transistors is getting thinner and thinner in every new generation of devices [100]. Therefore the leakage current due to quantum-mechanical tunneling through the dielectric interface is increasing. One way to reduce this current is to replace SiO_2 with a material that has a higher dielectric constant. Such a high- K dielectric layer with the same effective dielectric thickness (i.e., providing the same capacitance) could be physically thicker and thus reduce the gate leakage.

In order for this replacement material to be useful in practical applications on silicon, it also needs to be stable in contact with silicon up to $\sim 1000^\circ\text{C}$, and among other things it must also have an appropriate band alignment with silicon [101, 102, 103]. Currently, a hafnia-based dielectric is used as a replacement to SiO_2 in advanced CMOS transistors in production [104, 105, 106]. There are, however, drawbacks to this material too, e.g., the limited K that it provides and undesirable threshold voltage shifts arising from highly mobile oxygen vacancies [107]. This brings up the natural question: *which other materials exist that would satisfy these requirements and would enable the scaling of MOSFETs to continue beyond today's hafnia-based dielectrics?*

The stability of single component oxides on silicon has been demonstrated both experimentally and from thermodynamic analysis [102], and a candidate list of multicomponent oxide materials has been compiled [101]. A promising group of these materials consists of perovskite oxides having a $Pbnm$ (or closely related $P2_1/c$) space group. These compounds are at the focus of the present work. Some of them have been studied in thin-film form [108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126], but the full dielectric tensor of

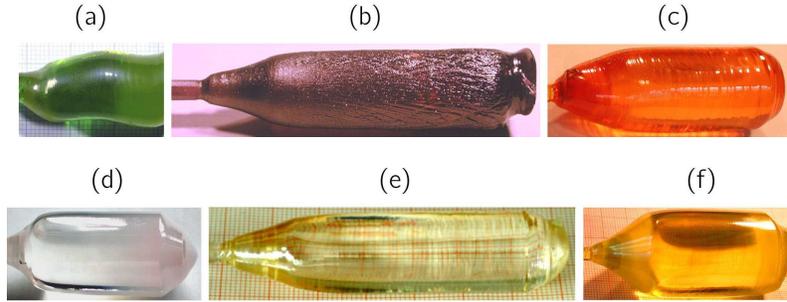


Figure 4.1: Photographs of single crystal rare-earth scandates grown along the $[110]$ direction by the Czochralski method. (a) PrScO_3 with a diameter of 12 mm, (b) NdScO_3 with a diameter of 16 mm, (c) SmScO_3 with a diameter of 18 mm, (d) GdScO_3 with a diameter of 32 mm, (e) TbScO_3 with a diameter of 18 mm, and (f) DyScO_3 with a diameter of 32 mm.

these materials has not yet been established, making the selection of materials best suited for high- K applications difficult. Some of these materials could also be of interest for microwave dielectric applications [127, 128]. Thus we decided to study the structural and dielectric properties of these compounds. The calculations are carried out using density-functional theory, and we compare the results with experimental data. To our knowledge, previous theoretical calculations have been carried out in only a few cases [129, 130, 131].

The chapter is organized as follows. Explanations of theoretical methods used in this work are given in Sec. 4.2. The main results on the structural and dielectric properties are given and discussed in Sec. 4.3. There we also discuss the correlations between the structural and dielectric properties of these perovskites, decompose the ionic contribution of the dielectric tensor into components arising from various force constant matrix eigenvectors, and discuss the effect of B_A antisite defects on the dielectric properties. We finish with a brief summary in Sec. 4.4.

4.1.1 Compounds under consideration

In this work, we consider the following five groups of perovskites having the $Pbnm$ space group.

The first group are rare-earth scandates having formula $A\text{ScO}_3$ where A is a rare-earth atom. In Sec. 4.3.3 we report full dielectric tensors for LaScO_3 , PrScO_3 , NdScO_3 , SmScO_3 , GdScO_3 , TbScO_3 and DyScO_3 . Note that HoScO_3 , ErScO_3 , TmScO_3 , YbScO_3 , LuScO_3 , and YScO_3 do not form single crystals with the perovskite structure from the melt at atmospheric pressure. Rather, they form solid solutions of $A_2\text{O}_3$ and Sc_2O_3 , i.e., $(A,\text{Sc})_2\text{O}_3$, with the bixbyite structure [132, 133, 134]. Nevertheless, LuScO_3 [126] and YbScO_3 [135] have been formed in perovskite form as thin films via epitaxial stabilization, and others might be made in the same way. To analyze trends within this group of compounds, we also did the calculations of dielectric tensors on LuScO_3 and YScO_3 in the $Pbnm$ perovskite structure; see Sec. 4.3.3 for the details.

The second group consists of rare-earth yttrates with formula $A\text{YO}_3$. Only one such compound, LaYO_3 , is known to form a perovskite [136], but to analyze trends the dielectric tensor of DyYO_3 in the perovskite structure was also calculated.

In the third group we consider CaZrO_3 , SrZrO_3 and SrHfO_3 perovskites.

The fourth group of compounds have the formula $\text{La}_2B\text{B}'\text{O}_6$ where the B atom is either Mg or Ca and B' is either Zr or Hf. Little is known experimentally about these compounds, and single crystals of these compounds have not been made [137].

The last group of compounds we considered have the formula $AA'\text{O}_3$, where both A and A' are rare-earth atoms. These include the 11 of such compounds that are known to form the perovskite structure with space group $Pbnm$ at atmospheric pressure: LaHoO_3 , LaErO_3 , LaTmO_3 , LaYbO_3 , LaLuO_3 , CeTmO_3 , CeYbO_3 ,

CeLuO₃, PrYbO₃, PrLuO₃, and NdLuO₃ [136, 138, 139]. We calculated the dielectric and structural properties of all of these compounds.

4.2 Preliminaries

4.2.1 Structure of Pbnm perovskites

The ideal cubic perovskite ABO_3 consists of a network of corner-shared octahedra, each with an oxygen on its vertices and a B atom at its center, and A ions that are 12-fold coordinated in the spaces between octahedra. It is well known that perovskites having sufficiently small A -site ions (i.e., a small Goldschmidt tolerance factor [140, 141]) often allow for a distorted perovskite structure that has a rotated framework of oxygen octahedra and displaced A -site ions. This lowers the space group symmetry from cubic $Pm\bar{3}m$ (O_h^1) to orthorhombic $Pbnm$ (D_{2h}^{16}), and the number of ABO_3 formula units per primitive cell increases from $Z = 1$ to $Z = 4$, as shown in Fig. 4.2. The rotations of the octahedra in the $Pbnm$ space group can be decomposed into two steps. The first step is the rotation around the $[110]$ direction of the original cubic frame (the cubic frame is rotated by 45° around the z axis with respect to the $Pbnm$ frame) by an angle θ_R as in Fig. 4.3(a), and the second step is a rotation around $[001]$ by θ_M as in Fig. 4.3(b). The rotations must be done in that order to prevent distortions of the octahedra. The pattern of neighboring octahedral rotations is denoted by $(a^- a^- c^+)$ in Glazer notation [142] (or see the directions of the arrows in Fig. 4.3). These rotations also allow for the displacement of A -site ions in the x - y plane without further lowering of the space-group symmetry.

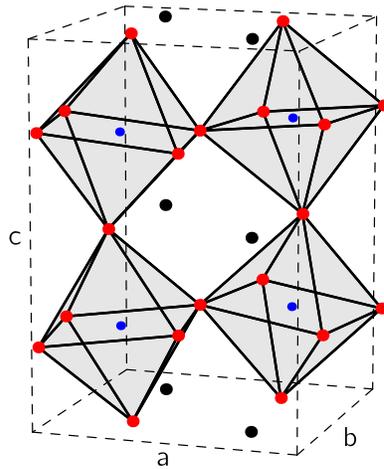


Figure 4.2: 20-atom primitive cell of a $Pbnm$ -distorted ABO_3 perovskite. A -site atoms are shown in black, B -site atoms in blue (at the centers of the octahedra) and oxygen atoms in red (at the vertices of the octahedra). Orthorhombic unit cell vectors (a , b , and c) are also indicated.

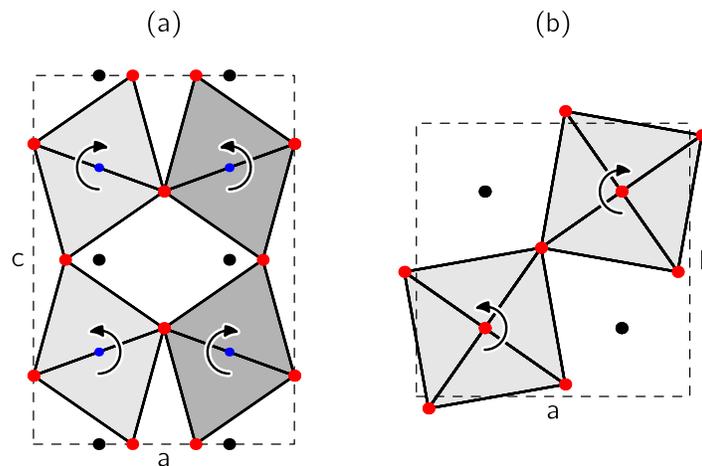


Figure 4.3: (a) Projection on the a - c plane of the structure with $\theta_R > 0$ and $\theta_M = 0$. (b) Projection on the a - b plane for $\theta_R = 0$ and $\theta_M > 0$. Color coding of atoms and positions of axes labels are the same as in Fig. 4.2. In (a), a darker shading is used to indicate the two octahedra that are further away along the y -coordinate; in (b), the two bottom octahedra are exactly below the two top octahedra.

4.2.2 Computational methods

The main computational method we are using is the density-functional theory as implemented in the Quantum-Espresso package [143]. The exchange-correlation functional was approximated using a generalized gradient approximation (GGA) of the Perdew-Burke-Ernzerhof type [79] and ultrasoft pseudopotentials were employed [12]. The electronic wavefunctions were expanded in a basis of plane waves with kinetic energy up to 40 Ry, while the charge density was expanded up to 300 Ry. The Brillouin zone was sampled using a $4 \times 4 \times 4$ Monkhorst-Pack grid [80].

A new set of ultrasoft pseudopotentials [12] for the lanthanoid series of rare earths, from La to Lu, were generated for the present project. In all cases the f -shell filling was chosen as appropriate for the $3+$ valence state: one f electron for Ce, two for Pr, etc. The f electrons were then considered to be in the core (and un-spin-polarized) for the purposes of generating the pseudopotentials. Thus, the f electrons are not explicitly included in the solid-state calculations. Such an approximation can be justified whenever the strong on-site Coulomb interactions of electrons in the f shell drive the occupied f states well below, and the unoccupied states well above, the energy range of interest for spd bonding in the crystal. Of course, this will not be a good approximation for some heavy-fermion or mixed-valent systems, and in any case our approach is obviously unable to describe phenomena involving magnetic ordering of f electrons at low temperature. Nevertheless, we believe that this approach is quite reasonable for the present purposes.

The artificial nature of the scattering in the f channel did, however, pose some problems in the pseudopotential construction. In particular, we found that the lattice constant of a perovskite containing the rare-earth atom in question could differ for two pseudopotentials having different scattering properties in the

f channel; this causes problems since the usual approach of matching to the all-electron f scattering is not appropriate in the present case. To ameliorate this problem, the f -channel parameters of these pseudopotentials were optimized so that resulting pseudopotentials would give the “correct” cell volumes for simple rare-earth compounds. Since the GGA typically overestimates crystal volumes by about 1-2% [144], the optimization was actually done in order to produce a corresponding overestimate in a consistent fashion.

For this procedure, our compounds of choice were the rare-earth nitrides with the simple rock-salt structure. The experimental [145] and calculated volumes of these nitrides are indicated in Fig. 4.4. Note that the volumes of CeN, PrN, and GdN show an anomalous behavior that is presumably due to strong correlation effects associated with the proximity to a mixed-valent regime ¹, and therefore they will not be correctly treated by our GGA calculation. To avoid this problem we first carried out a smoothed fit of the experimental volumes versus atomic number over the lanthanoide nitride series, but with CeN, PrN, and GdN omitted from the fit, as shown by the solid line in Fig. 4.4. We then used these fitted values to set the target volumes for the optimization of the pseudopotentials.

We used density-functional perturbation theory [19] to calculate the dielectric response. Both purely electronic ϵ^{el} and ionic ϵ^{ion} contributions were calculated [146]. The electronic part is defined as

$$\epsilon_{\alpha\beta}^{\text{el}} = \delta_{\alpha\beta} + 4\pi \left. \frac{\partial P_{\alpha}}{\partial \mathcal{E}_{\beta}} \right|_{u=0}, \quad (4.1)$$

where P_{α} is the polarization induced by the electric field \mathcal{E}_{β} while all ions are held fixed ($u = 0$). The remaining component of the dielectric response is by definition the ionic contribution ϵ^{ion} .

¹In the case of GdN the anomaly is more likely the result of the large spin splitting and strong ferromagnetism associated with the huge magnetic moment of the $4f^7$ configuration.

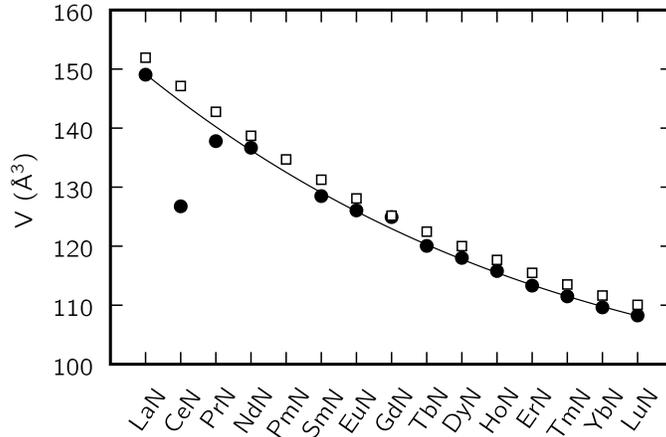


Figure 4.4: Unit cell volumes of all rock-salt rare-earth nitrides, in \AA^3 . Empty squares are the results obtained using our optimized ultrasoft pseudopotentials for rare-earth atoms; solid circles are experimental [145] results. The solid line is a fit to the experimental values excluding CeN, PrN, and GdN.

This ionic part can be calculated from the force-constant matrix $\Phi_{i\alpha,j\beta}$ and the Born effective charge matrix $Z_{i,\alpha\beta}$. The force-constant matrix is defined as

$$\Phi_{i\alpha,j\beta} = \frac{\partial^2 E}{\partial u_{i\alpha} \partial u_{j\beta}}, \quad (4.2)$$

where E is the total energy of the system and $u_{i\alpha}$ is the displacement of the i -th atom along the direction α . We will denote the n -th normalized eigenvector of this matrix as $\xi_{i\alpha}^n$ and its eigenvalue as μ_n . The Born effective charge matrix is defined as

$$Z_{i,\alpha\beta} = \frac{V}{e} \frac{\partial P_\alpha}{\partial u_{i\beta}}, \quad (4.3)$$

where P_α is the polarization induced in a crystal by the displacement of the i -th atom in the direction β . V is the volume of the unit cell and e is the electron

charge. Finally, the ionic part of the dielectric tensor can be written as

$$\epsilon_{\alpha\beta}^{\text{ion}} = \frac{4\pi e^2}{V} \sum_n \frac{1}{\mu_n} Q_\alpha^n Q_\beta^n, \quad (4.4)$$

where the charge Q_α^n of the n -th eigenmode is defined through the effective charge matrix as $Q_\alpha^n = \sum_{i\beta} Z_{i,\alpha\beta} \xi_{i\beta}^n$.

4.3 Results and discussion

4.3.1 Structural properties

We focus first on the structural properties of these systems. The structure of the *Pbnm* perovskites is described by three orthorhombic lattice constants plus two *A*-site and five oxygen Wyckoff parameters. Figure 4.5 shows graphically the most important structural parameters of these systems, while Table 4.1 gives detailed information on all the structural parameters. Rotational angles in Table 4.1 and in Fig. 4.5 were calculated by fitting the structural parameters to a model in which the octahedra are perfectly rigid (see Sec. 4.3.2 for the details of this model).

Overall we find good agreement with experimental values for the structural parameters. The Wyckoff coordinates in particular are in excellent agreement with experiments, with the average error being on the order of $2 \cdot 10^{-3}$. The volume of the unit cell, on the other hand, is consistently overestimated by 1-2%, as is usually expected from the GGA exchange-correlation functional, and as we would expect from our construction of the rare-earth pseudopotentials.

All structures show an angle θ_R that is about $\sqrt{2}$ times larger than θ_M . Therefore, consecutive rotations by θ_R and θ_M can be considered approximately as a single rotation around a $\{111\}$ axis in the cubic frame. That is, the actual $(a^- a^- c^+)$ pattern of rotations is very nearly $(a^- a^- a^+)$ in the Glazer notation [142]. See

Sec. 4.3.6 for a more detailed discussion.

Rare-earth scandates

The rare-earth scandates $A\text{ScO}_3$ show a decrease in volume by $\sim 9\%$ while going along the series from $A=\text{La}$ to $A=\text{Dy}$ (the calculated primitive unit cell volume is 271.40 \AA^3 for LaScO_3 and 249.81 \AA^3 for DyScO_3). On the other hand, the Sc-O distance remains nearly constant along the series (2.12 \AA for LaScO_3 and 2.11 \AA for DyScO_3), which means that the change in volume is almost entirely due to the larger octahedral rotation angles for DyScO_3 as compared to LaScO_3 . Our calculations also show that the same trend continues all the way to LuScO_3 .

Rare-earth yttrates

The rare-earth yttrates have a very similar behavior as the rare-earth scandates. The main quantitative structural difference between the two comes from the fact that yttrium is a larger ion than scandium. This leads to a larger volume for the yttrates, and also a larger rotation angle due to a smaller tolerance factor.

CaZrO_3 , SrZrO_3 , and SrHfO_3

SrZrO_3 and SrHfO_3 have quite similar structural properties. The main difference can be traced to the fact that Hf is a smaller ion than Zr. Therefore, the calculated average Hf-O distance is 2.07 \AA , while the average Zr-O distance is 2.11 \AA . Furthermore, their octahedral rotation angles are about 1.7 times smaller than in the rare-earth scandates.

In CaZrO_3 the average Zr-O distance is 2.10 \AA , which is very close to the corresponding distance in SrZrO_3 and SrHfO_3 . Thus, the main reason why CaZrO_3 has a smaller volume than SrZrO_3 is because of the larger rotation angles in CaZrO_3 .

La₂BB'O₆ compounds

We consider La₂BB'O₆ compounds with B =Mg or Ca and B' =Zr or Hf. These compounds are expected to exhibit rock-salt ordering of the B -site ions as a result of the difference in charge and ionic radius between the B and B' ions [147]. This ordering reduces the symmetry from the orthorhombic $Pbnm$ to the monoclinic $P2_1/c$ (C_{2h}^5) space group.

The structural properties for these systems are reported in Fig. 4.5 and in Table 4.2. The rotational angles are obtained by a fit to the rigid-octahedra model in which we have allowed for different sizes of B - and B' -centered octahedra. (See the end of Sec. 4.3.2 for details.)

The unit cell volume is larger by about 5 Å³ per primitive cell for the compounds containing Zr than for those containing Hf. On the other hand, compounds with Ca are larger by about 28 Å³ than those containing Mg. Similarly, the rotation angles are larger in compounds containing Ca than in those with Mg. The discrepancy between octahedral sizes is largest for La₂CaHfO₆ (12% linear increase) and smallest for La₂MgZrO₆ (0.4% linear increase).

Rare-earth rare-earth perovskites

We now briefly analyze the structural properties of $Pbnm$ perovskites of type $AA'O_3$ where both A and A' are rare-earth atoms. All eleven compounds we considered are known experimentally to form the perovskite structure in the $Pbnm$ space group [136, 138, 139].

Among these 11 compounds, the largest unit-cell volume of 311.58 Å³ is found in LaHoO₃, and the smallest of 292.32 Å³ is in NdLuO₃. Oxygen octahedral rotation angles are quite large in all of these compounds and show very little variation from one compound to another. The trends of the rotation angles are

as expected from a tolerance-factor analysis: perovskites with smaller A -site ions but the same B -site ions have larger oxygen octahedral rotation angles, and the opposite is true for the B -site ions.

Table 4.1: Structural parameters of the $Pbnm$ perovskites we considered. Our calculated values are indicated with the letter T, experimental values with E and other theoretical data with O. Fitted octahedra rotation angles θ_R and θ_M are given in degrees, see Sec. 4.3.2 for the details. A -site ions occupy the $4c$ site and in cubic configuration $x_1 = 0$ and $y_1 = 1/2$. B -site ions are on the $4a$ site. One type of oxygen sites are at $4c$ and in the cubic case $x_2 = 0$ and $y_2 = 0$. The remaining oxygens are at $8d$ and $x_3 = 1/4$, $y_3 = 1/4$ and $z_3 = 0$ in the cubic case.

		Unit cell parameters						Wyckoff coordinates						Model		
		a (Å)	b (Å)	c (Å)	V (Å ³)	b/a	c/a	x_1	y_1	x_2	y_2	x_3	y_3	z_3	θ_R (°)	θ_M (°)
LaScO ₃	T	5.7030	5.8414	8.1469	271.40	1.0243	1.4285	0.0123	0.5467	-0.0977	-0.0323	0.2059	0.2943	0.0523	14.9	9.9
	E ^a	5.6803	5.7907	8.0945	266.25	1.0194	1.4250	0.0100	0.5428	-0.0968	-0.0277	0.2073	0.2958	0.0521	14.6	9.6
PrScO ₃	T	5.6372	5.8367	8.0908	266.21	1.0354	1.4352	0.0148	0.5537	-0.1049	-0.0377	0.2014	0.2979	0.0557	16.4	11.0
	E ^j	5.608	5.780	8.025	260.1	1.0307	1.4310	0.0121	0.5507	-0.1052	-0.0395	0.1977	0.3008	0.0555	16.2	11.2
NdScO ₃	T	5.6077	5.8317	8.0667	263.80	1.0399	1.4385	0.0159	0.5562	-0.1083	-0.0401	0.1997	0.2992	0.0574	17.0	11.5
	E ^b	5.577	5.777	8.005	257.9	1.0359	1.4354	0.0133	0.5532	-0.1088	-0.0418	0.1953	0.3020	0.0571	16.8	11.8
SmScO ₃	T	5.5483	5.8067	8.0196	258.37	1.0466	1.4454	0.0176	0.5596	-0.1159	-0.0455	0.1964	0.3019	0.0610	18.2	12.5
	E ^b	5.531	5.758	7.975	254.0	1.0410	1.4419	0.0149	0.5566	-0.1163	-0.0468	0.1935	0.3037	0.0609	17.9	12.5
GdScO ₃	T	5.4987	5.7794	7.9861	253.79	1.0510	1.4524	0.0191	0.5617	-0.1222	-0.0502	0.1941	0.3036	0.0642	19.0	13.3
	E ^b	5.481	5.745	7.929	249.7	1.0482	1.4466	0.0163	0.5599	-0.1209	-0.0501	0.1912	0.3052	0.0628	18.7	13.2
TbScO ₃	T	5.4764	5.7646	7.9735	251.72	1.0526	1.4560	0.0198	0.5624	-0.1251	-0.0524	0.1932	0.3043	0.0656	19.4	13.7
	E ^k	5.4543	5.7233	7.9147	247.07	1.0493	1.4511	0.0167	0.5603	-0.1239	-0.0545	0.1900	0.3054	0.0643	19.1	13.7
DyScO ₃	T	5.4560	5.7501	7.9629	249.81	1.0539	1.4595	0.0203	0.5630	-0.1276	-0.0545	0.1923	0.3050	0.0669	19.7	14.1
	O ^f	5.449	5.739	7.929	248.0	1.0532	1.4551	0.019	0.562	-0.130	-0.057	0.190	0.307	0.068	19.9	14.1
	E ^b	5.443	5.717	7.901	245.9	1.0503	1.4516	0.0174	0.5616	-0.1262	-0.0561	0.1886	0.3063	0.0659	19.4	13.9
LaYO ₃	T	5.9035	6.1225	8.5810	310.16	1.0371	1.4535	0.0173	0.5506	-0.1284	-0.0581	0.1948	0.3035	0.0689	18.9	13.6

Continued on next page

		Unit cell parameters						Wyckoff coordinates						Model		
		a	b	c	V	b/a	c/a	x_1	y_1	x_2	y_2	x_3	y_3	z_3	θ_R	θ_M
		(Å)	(Å)	(Å)	(Å ³)										(°)	(°)
	E ^l	5.890	6.086	8.511	305.1	1.0333	1.4450									
CaZrO ₃	T	5.5974	5.7875	8.0416	260.51	1.0340	1.4367	0.0133	0.5506	-0.1078	-0.0413	0.1976	0.2999	0.0572	16.6	11.6
	E ^m	5.5831	5.7590	8.0070	257.45	1.0315	1.4341	0.0122	0.5495	-0.1044	-0.0401	0.1976	0.3000	0.0554	16.2	11.4
SrZrO ₃	T	5.8068	5.8602	8.2323	280.14	1.0092	1.4177	0.0070	0.5311	-0.0759	-0.0195	0.2140	0.2857	0.0401	11.6	7.7
	O ^g	5.652	5.664	7.995	255.9	1.0021	1.4145	0.007	0.534	-0.107	-0.036	0.199	0.301	0.056	14.7	10.0
	E ^c	5.7963	5.8171	8.2048	276.65	1.0036	1.4155	0.0040	0.5242	-0.0687	-0.0133	0.2154	0.2837	0.0363	10.4	7.2
SrHfO ₃	T	5.7552	5.7754	8.1365	270.45	1.0035	1.4138	0.0052	0.5230	-0.0660	-0.0128	0.2209	0.2792	0.0346	10.0	6.2
	O ^h	5.6887	5.7016	8.0455	260.95	1.0023	1.4143	0.006	0.528	-0.073	-0.016	0.2166	0.2834	0.0385	10.8	7.0
	E ^d	5.7516	5.7646	8.1344	269.70	1.0023	1.4143	0.0040	0.5160	-0.0630	-0.0140	0.2189	0.2789	0.0335	9.6	6.4
LaHoO ₃	T	5.9135	6.1367	8.5859	311.58	1.0377	1.4519	0.0170	0.5508	-0.1293	-0.0587	0.1936	0.3044	0.0692	19.0	13.6
	E ⁱ	5.884	6.094	8.508	305.1	1.0357	1.4460									
LaErO ₃	T	5.8971	6.1174	8.5509	308.48	1.0374	1.4500	0.0169	0.5509	-0.1272	-0.0567	0.1945	0.3036	0.0681	18.7	13.3
	E ⁱ	5.870	6.073	8.465	301.8	1.0346	1.4421									
LaTmO ₃	T	5.8829	6.1002	8.5190	305.73	1.0369	1.4481	0.0167	0.5509	-0.1252	-0.0549	0.1951	0.3030	0.0670	18.5	13.1
	E ⁱ	5.859	6.047	8.453	299.5	1.0321	1.4427									
LaYbO ₃	T	5.8692	6.0833	8.4890	303.09	1.0365	1.4464	0.0164	0.5509	-0.1233	-0.0532	0.1957	0.3024	0.0659	18.3	12.8
	E ⁱ	5.843	6.033	8.432	297.2	1.0325	1.4431									
LaLuO ₃	T	5.8579	6.0695	8.4646	300.95	1.0361	1.4450	0.0162	0.5508	-0.1217	-0.0518	0.1962	0.3020	0.0651	18.2	12.6
	E ^e	5.8259	6.0218	8.3804	294.00	1.0336	1.4385	0.0138	0.5507	-0.121	-0.056	0.193	0.307	0.063	17.9	12.7
CeTmO ₃	T	5.8520	6.0870	8.4984	302.72	1.0401	1.4522	0.0179	0.5528	-0.1283	-0.0575	0.1942	0.3036	0.0686	19.0	13.5
	E ⁱ	5.828	6.035	8.405	295.6	1.0355	1.4422									
CeYbO ₃	T	5.8381	6.0707	8.4671	300.09	1.0398	1.4503	0.0176	0.5529	-0.1264	-0.0558	0.1948	0.3031	0.0676	18.8	13.3

Continued on next page

	Unit cell parameters						Wyckoff coordinates						Model		
	a (Å)	b (Å)	c (Å)	V (Å ³)	b/a	c/a	x_1	y_1	x_2	y_2	x_3	y_3	z_3	θ_R (°)	θ_M (°)
	E ⁱ	5.806	6.009	8.373	292.1	1.0350	1.4421								
CeLuO ₃	T	5.8270	6.0578	8.4420	297.99	1.0396	1.4488	0.0174	0.5529	-0.1249	-0.0544	0.1952	0.3027	0.0667	18.7 13.1
	E ⁱ	5.793	5.997	8.344	289.9	1.0352	1.4404								
PrYbO ₃	T	5.8085	6.0544	8.4481	297.10	1.0423	1.4544	0.0185	0.5542	-0.1296	-0.0585	0.1936	0.3039	0.0692	19.3 13.7
	E ⁱ	5.776	5.995	8.368	289.8	1.0379	1.4488								
PrLuO ₃	T	5.7974	6.0424	8.4217	295.01	1.0423	1.4527	0.0183	0.5543	-0.1281	-0.0571	0.1936	0.3036	0.0683	19.1 13.5
	E ⁱ	5.768	5.991	8.340	288.2	1.0387	1.4459								
NdLuO ₃	T	5.7699	6.0270	8.4062	292.32	1.0446	1.4569	0.0193	0.5555	-0.1310	-0.0596	0.1932	0.3042	0.0699	19.5 14.0
	E ⁱ	5.737	5.974	8.311	284.8	1.0413	1.4487								

^aReference [148]. ^bReference [149]. ^cReference [150]. ^dReference [151]. ^eReference [138]. ^fReference [129]. ^gReference [130].

^hReference [131]. ⁱReference [136]. ^jReference [152]. ^kReference [153]. ^lReference [154]. ^mReference [155].

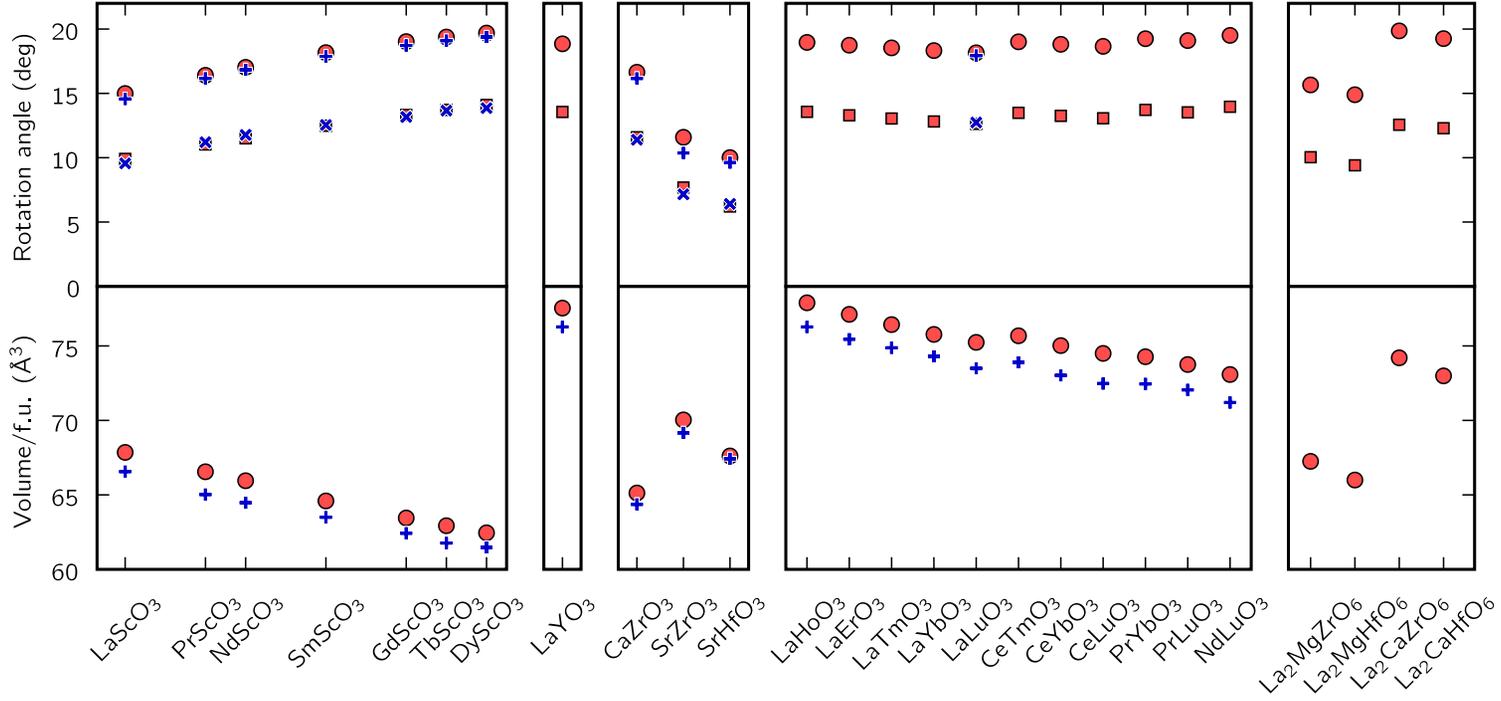


Figure 4.5: Structural information for all systems we considered. Bottom pane shows volume in \AA^3 per formula unit (f. u.) of ABO_3 . Calculated values are shown as red circles, and experimental values as blue plus symbols, if available. Top pane shows oxygen octahedra rotation angles in degrees. Theoretical values are shown with red circles and squares and experimental values with blue plus and cross symbols. θ_R angles are shown with circles and plus symbols while θ_M angles with squares and cross symbols. For $\text{La}_2BB'O_6$ systems the average of θ_M and θ'_M is given. For numerical values see Table 4.1 and 4.2.

4.3.2 Comparison with model of perfectly rigid octahedra

In $Pbnm$ perovskites, a rigid rotation of the oxygen octahedra by θ_R followed by another rigid rotation by θ_M (see Fig. 4.3) leads to Wyckoff parameters given by

$$x_2 = -\frac{1}{2\sqrt{2}} \tan \theta_R, \quad (4.5)$$

$$y_2 = -\frac{1}{2\sqrt{2}} \sin \theta_R \tan \theta_M, \quad (4.6)$$

$$x_3 = \frac{1}{4} \left(1 - \frac{\tan \theta_M}{\cos \theta_R} \right), \quad (4.7)$$

$$y_3 = \frac{1}{4} (1 + \cos \theta_R \tan \theta_M), \quad (4.8)$$

$$z_3 = \frac{1}{4\sqrt{2}} \tan \theta_R, \quad (4.9)$$

Here we have denoted the Wyckoff coordinates of the oxygen atoms at the $4c$ Wyckoff point with x_2 and y_2 , while those of the remaining oxygen atoms at the $8d$ point are denoted with x_3 , y_3 , and z_3 . The Wyckoff coordinates of the A -site ion at the $4c$ point are denoted by x_1 and y_1 , but these are left unspecified in our rigid-octahedra model. It also leads to orthorhombic lattice constants given by

$$a = \sqrt[3]{\frac{V_0}{\sqrt{2}}} \cos \theta_R \cos \theta_M, \quad (4.10)$$

$$b = \sqrt[3]{\frac{V_0}{\sqrt{2}}} \cos \theta_M, \quad (4.11)$$

$$c = \sqrt[3]{2V_0} \cos \theta_R \quad (4.12)$$

where V_0 is the volume the structure would have if the octahedra were rotated rigidly back to $\theta_R = \theta_M = 0$.

The Wyckoff parameters and unit-cell ratios from our calculations can be well fitted by Eqs. (4.5)-(4.12) (see Table 4.1 and Fig. 4.5 for the values of the fitted angles). By far the largest discrepancy is found for Wyckoff parameter y_2 . For

a typical system (e.g., LaScO_3) the discrepancy between calculated and fitted y_2 values is about 0.016, or 50% with respect to the difference from the cubic case. For the remaining oxygen Wyckoff coefficients, the discrepancy averages about 0.003, or $\sim 5\%$.

The rotation angles for the $\text{La}_2\text{BB}'\text{O}_6$ systems were obtained by fitting their structural parameters to a slightly more complicated model of rigid octahedra than the one given in Eqs. (4.5)-(4.12). In this model, we first change the relative sizes of B - and B' -centered octahedra. The ratio of their linear sizes is denoted by d/d' . We then proceed with the rotation by an angle θ_R around the $[110]$ axis in the cubic frame. Finally, we perform a rotation of the B -centered octahedra around $[001]$ by an angle θ_M , and of the B' -centered octahedra by an angle θ'_M around the same axis. The resulting fitted values of these parameters are given in Table 4.2.

4.3.3 Dielectric properties

In this section we discuss the dielectric properties of the materials included in our study. $Pbnm$ perovskites are orthorhombic and thus have diagonal dielectric tensors, with $\epsilon_{xx} \neq \epsilon_{yy} \neq \epsilon_{zz}$ in general. In addition to reporting these components, we also focus on analyzing the results in terms of the three linear combinations

$$\bar{\epsilon} = \frac{1}{3} (\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}), \quad (4.13)$$

$$\Delta\epsilon_{\parallel} = \epsilon_{xx} - \epsilon_{yy}, \quad (4.14)$$

$$\Delta\epsilon_{\perp} = \epsilon_{zz} - \frac{1}{2} (\epsilon_{xx} + \epsilon_{yy}), \quad (4.15)$$

representing the average dielectric tensor, a measure of the x - y anisotropy, and a measure of z anisotropy, respectively. This choice of parameters was made to simplify the analysis of trends of dielectric properties of these compounds. The

theoretical – and where available, experimental – results for the dielectric-tensor components are reported in Figure 4.6 and in Table 4.3. The theoretical values are further decomposed in Table 4.3 into purely electronic or frozen-ion contributions ϵ^{el} and lattice-mediated contributions ϵ^{ion} . We find that the electronic contribution is roughly five times smaller than the ionic one, is nearly isotropic, and does not show a dramatic variation from one perovskite to another. Thus, it is clear that the lattice-mediated ionic contributions play by far the dominant role in the observed dielectric tensors and their anisotropies.

We now consider each of our chosen classes of *Pbnm* perovskites in turn, orienting the presentation from the point of view of the theoretical calculations, but mentioning the comparison with experiment where appropriate.

Rare-earth scandates

All rare-earth scandates $A\text{ScO}_3$ have rather similar values for their isotropically-averaged dielectric constants, falling between about $\bar{\epsilon} = 26$ and $\bar{\epsilon} = 28$. The xx component for all these systems is larger than the yy component by about $\Delta\epsilon_{\parallel} = 4$. On the other hand, the zz component changes significantly from LaScO_3 to DyScO_3 . In LaScO_3 the average of the xx and yy components is almost as large as the zz component ($\Delta\epsilon_{\perp} = -1$), while in DyScO_3 , the zz component is larger by about $\Delta\epsilon_{\perp} = 9$ than the average of xx and yy components.

These results are in good agreement with experiment, especially for $\bar{\epsilon}$ and $\Delta\epsilon_{\parallel}$. On the other hand, $\Delta\epsilon_{\perp}$ is consistently larger in experiments by about 3-5, but the trend of increasing $\Delta\epsilon_{\perp}$ is present in both theory and experiment.

As was mentioned earlier, rare-earth atoms heavier than Dy (i.e., Ho-Lu) and Y itself do not form single-crystal scandates. Nevertheless, at least some (YbScO_3 [135] and LuScO_3 [126]) can form *Pbnm* perovskites in thin-film form. In order to establish the trends of the dielectric properties for these materials, we

calculated the dielectric tensors of LuScO_3 and YScO_3 . The dielectric tensor of LuScO_3 shows the continuation of the trend from LaScO_3 to DyScO_3 . Both xx and yy components are slightly smaller than for DyScO_3 , their numerical values being 23.5 and 21.4 respectively. On the other hand, the zz component (44.8) is larger than for DyScO_3 (32.6) and for LaScO_3 (27.4). YScO_3 has dielectric tensor components of 26.9, 23.0, and 37.7 for its xx , yy , and zz components, respectively.

Rare-earth yttrates

We now consider the rare-earth yttrates, i.e., AYO_3 where A is one of the rare-earth atoms. These are similar to the rare-earth scandates, but with yttrium on the B site instead of scandium. Only one such compound, LaYO_3 , is known to form a perovskite [136], but others might form in thin films via epitaxial stabilization. We find that LaYO_3 has a larger zz component than does LaScO_3 . In LaYO_3 the zz component of the dielectric tensor is 38.0, while in LaScO_3 it is 27.4. On the other hand, the xx and yy components are almost unchanged with respect to LaScO_3 . In LaYO_3 the xx component is 30.6 and the yy component is 25.7. We also find that heavier rare-earth atoms on the A -site tend to destabilize this $Pbnm$ structure even further. For example, we find that DyYO_3 in the $Pbnm$ structure has an unstable mode at $i70 \text{ cm}^{-1}$ that is IR-active along the z direction.

CaZrO_3 , SrZrO_3 , and SrHfO_3

According to our calculations, SrZrO_3 and SrHfO_3 show large and rather isotropic dielectric tensors. The average dielectric tensor $\bar{\epsilon}$ is 40.9 and 32.8 in SrZrO_3 and SrHfO_3 , respectively. Their x - y anisotropies have an opposite sign as compared to all of the other compounds we analyzed. Unfortunately, because of twinning (see Ref. [2]), only average dielectric constant was measured for these systems, and therefore we could not directly compare our full calculated dielectric tensors

with experiment. Still, if we make a comparison between theory and experiment for the average dielectric tensor $\bar{\epsilon}$, the agreement is reasonable.

Our calculations suggest that the CaZrO_3 compound, on the other hand, has a very high value of the z -anisotropy of 28.1. Its x - y anisotropy of 1.1, on the other hand, is quite small. The average dielectric tensor $\bar{\epsilon} = 43.6$ is the highest among the all compounds we considered, mostly because of the very large ϵ_{zz} component of the dielectric tensor. Very similar results were also obtained in other theoretical studies [156, 157].

$\text{La}_2\text{BB}'\text{O}_6$ compounds

The $\text{La}_2\text{BB}'\text{O}_6$ systems show a small, non-zero off-diagonal ϵ_{xz} component, -0.4 for $\text{La}_2\text{MgZrO}_3$ and 4 for $\text{La}_2\text{CaZrO}_3$. ϵ_{xz} is allowed because the space group is reduced from orthorhombic ($Pbnm$) to monoclinic ($P2_1/c$) for these compounds. Their isotropically-averaged dielectric tensors are larger for systems containing Ca than for those with Mg, and a bit larger for those with Zr than for those with Hf. Therefore, the dielectric response in this class of materials is largest for $\text{La}_2\text{CaZrO}_6$, with $\bar{\epsilon} = 28.5$, and smallest for $\text{La}_2\text{MgHfO}_6$, with $\bar{\epsilon} = 23.6$. All computed dielectric tensor components for these systems are given in Table 4.4.

Rare-earth rare-earth perovskites

The 11 rare-earth–rare-earth perovskites we considered show a bigger variation in the isotropically-averaged dielectric constant $\bar{\epsilon}$ than do the rare-earth scandates (LaScO_3 - DyScO_3). The largest average dielectric constant among them is 32.9 in LaHoO_3 . The largest component of a dielectric tensor is also found in LaHoO_3 , whose ϵ_{zz} is 41.7.

The measure $\Delta\epsilon_{\parallel}$ of x - y anisotropy shows little variation among the components in this series. The anisotropy is of the same sign as in the rare-earth

scandates.

Finally, the z anisotropy $\Delta\epsilon_{\perp}$ once more shows a larger variation than in the rare-earth scandates. This anisotropy is largest for LaHoO_3 and smallest for LaLuO_3 .

Table 4.3: Dielectric parameters of the perovskites with space group $Pbnm$ that we considered. Our calculated values are denoted by the letter T, experimental values with E, and other theoretical data with O. First all three non-zero dielectric constant tensor components are given. Next, the average dielectric constant tensor, the x - y anisotropy, and z anisotropy are given; see Eqs. (4.13)-(4.15). Finally, the electronic and ionic contributions are given separately.

		Dielectric tensor			Reduced variables			Electronic part			Ionic part		
		ϵ_{xx}	ϵ_{yy}	ϵ_{zz}	$\bar{\epsilon}$	$\Delta\epsilon_{\parallel}$	$\Delta\epsilon_{\perp}$	$\epsilon_{xx}^{\text{el}}$	$\epsilon_{yy}^{\text{el}}$	$\epsilon_{zz}^{\text{el}}$	$\epsilon_{xx}^{\text{ion}}$	$\epsilon_{yy}^{\text{ion}}$	$\epsilon_{zz}^{\text{ion}}$
LaScO ₃	T	30.4	26.4	27.4	28.1	4.0	-1.0	5.0	5.0	4.8	25.3	21.5	22.6
PrScO ₃	T	28.6	24.2	26.2	26.3	4.4	-0.1	5.0	5.0	4.8	23.5	19.2	21.5
	E ²	25.4	27.3	29.6	27.4	-1.9	3.3						
NdScO ₃	T	27.8	23.4	26.1	25.7	4.4	0.5	5.0	4.9	4.7	22.8	18.5	21.3
	E ²	25.5	21.5	26.9	24.6	4.0	3.4						
SmScO ₃	T	27.2	22.9	27.3	25.8	4.3	2.3	4.9	4.9	4.7	22.2	18.0	22.7
	E ²	23.1	19.9	29.0	24.0	3.2	7.5						
GdScO ₃	T	26.4	22.5	29.3	26.1	3.9	4.9	4.9	4.8	4.6	21.6	17.7	24.7
	E ²	22.8	19.2	29.5	23.8	3.6	8.5						
TbScO ₃	T	26.1	22.4	30.7	26.4	3.7	6.5	4.8	4.8	4.6	21.2	17.6	26.1
DyScO ₃	T	25.7	22.3	32.6	26.9	3.5	8.6	4.8	4.8	4.5	20.9	17.5	28.0
	O ¹	24.1	21.2	27.7	24.3	2.9	5.1	4.9	4.9	4.7	19.2	16.3	23.0
	E ²	21.9	18.9	33.8	24.9	3.0	13.4						
LaYO ₃	T	30.6	25.7	38.0	31.4	4.9	9.9	4.7	4.6	4.3	25.9	21.1	33.7
CaZrO ₃	T	34.8	33.7	62.4	43.6	1.1	28.1	4.6	4.7	4.6	30.2	29.1	57.8
	E ⁶				30								
SrZrO ₃	T	38.0	41.5	43.3	40.9	-3.4	3.5	4.6	4.6	4.6	33.4	36.9	38.7
	O ³	19.9	21.5	23.0	21.5	-1.6	2.3	5.1	4.9	4.8	14.8	16.6	18.2

Continued on next page

		Dielectric tensor			Reduced variables			Electronic part			Ionic part		
		ϵ_{xx}	ϵ_{yy}	ϵ_{zz}	$\bar{\epsilon}$	$\Delta\epsilon_{\parallel}$	$\Delta\epsilon_{\perp}$	$\epsilon_{xx}^{\text{el}}$	$\epsilon_{yy}^{\text{el}}$	$\epsilon_{zz}^{\text{el}}$	$\epsilon_{xx}^{\text{ion}}$	$\epsilon_{yy}^{\text{ion}}$	$\epsilon_{zz}^{\text{ion}}$
	E ²				32								
SrHfO ₃	T	30.0	35.1	33.2	32.8	-5.1	0.7	4.3	4.3	4.3	25.7	30.9	29.0
	O ⁴	33.1	46.8	40.8	40.2	-13.7	0.9	4.4	4.3	4.3	28.7	42.4	36.5
	E ²				26.2								
LaHoO ₃	T	31.0	26.0	41.7	32.9	5.0	13.3	4.7	4.6	4.3	26.3	21.4	37.4
LaErO ₃	T	29.9	25.0	36.4	30.4	4.9	9.0	4.7	4.6	4.3	25.2	20.4	32.1
LaTmO ₃	T	29.0	24.2	33.3	28.8	4.8	6.7	4.6	4.6	4.3	24.4	19.7	29.0
LaYbO ₃	T	28.3	23.6	30.9	27.6	4.7	5.0	4.6	4.5	4.3	23.6	19.1	26.6
LaLuO ₃ ⁵	T	27.6	23.1	29.3	26.7	4.5	3.9	4.6	4.5	4.3	23.0	18.6	25.0
CeTmO ₃	T	27.9	23.5	34.9	28.8	4.4	9.2	4.6	4.6	4.3	23.3	18.9	30.6
CeYbO ₃	T	27.2	22.8	32.1	27.4	4.4	7.0	4.6	4.6	4.3	22.6	18.3	27.8
CeLuO ₃	T	26.6	22.3	30.2	26.4	4.3	5.7	4.6	4.5	4.3	22.0	17.8	25.9
PrYbO ₃	T	26.6	22.7	34.7	28.0	4.0	10.1	4.6	4.5	4.3	22.0	18.2	30.4
PrLuO ₃	T	26.1	22.0	32.2	26.8	4.0	8.1	4.6	4.5	4.3	21.5	17.5	27.9
NdLuO ₃	T	25.3	21.7	34.5	27.2	3.6	11.0	4.6	4.5	4.3	20.8	17.2	30.2

¹Reference [129]. ²Reference [2]. ³Reference [130]. ⁴Reference [131]. ⁵Experimental data in reference [158]. ⁶Reference [155].

Table 4.4: Dielectric parameters of the $\text{La}_2\text{BB}'\text{O}_6$ perovskites we considered. First all four dielectric constant tensor components are given, followed by the average value of the dielectric constant tensor. Finally, the electronic and ionic contributions to dielectric constant tensor are given. x axis is chosen along lattice vector a , y along b , and z close to c (with the small component along a due to monoclinic cell).

	Dielectric tensor					Electronic part				Ionic part			
	ϵ_{xx}	ϵ_{yy}	ϵ_{zz}	ϵ_{xz}	$\bar{\epsilon}$	$\epsilon_{xx}^{\text{el}}$	$\epsilon_{yy}^{\text{el}}$	$\epsilon_{zz}^{\text{el}}$	$\epsilon_{xz}^{\text{el}}$	$\epsilon_{xx}^{\text{ion}}$	$\epsilon_{yy}^{\text{ion}}$	$\epsilon_{zz}^{\text{ion}}$	$\epsilon_{xz}^{\text{ion}}$
$\text{La}_2\text{MgZrO}_6$	26.5	23.6	24.4	-0.4	24.8	4.8	4.6	4.5	-0.06	21.7	18.9	19.8	-0.3
$\text{La}_2\text{MgHfO}_6$	24.9	22.9	22.9	-0.3	23.6	4.6	4.5	4.4	-0.05	20.3	18.4	18.5	-0.3
$\text{La}_2\text{CaZrO}_6$	29.9	24.9	30.6	3.8	28.5	4.7	4.6	4.4	-0.06	25.2	20.3	26.2	3.8
$\text{La}_2\text{CaHfO}_6$	27.4	23.0	26.6	2.6	25.7	4.6	4.5	4.3	-0.04	22.9	18.5	22.3	2.6

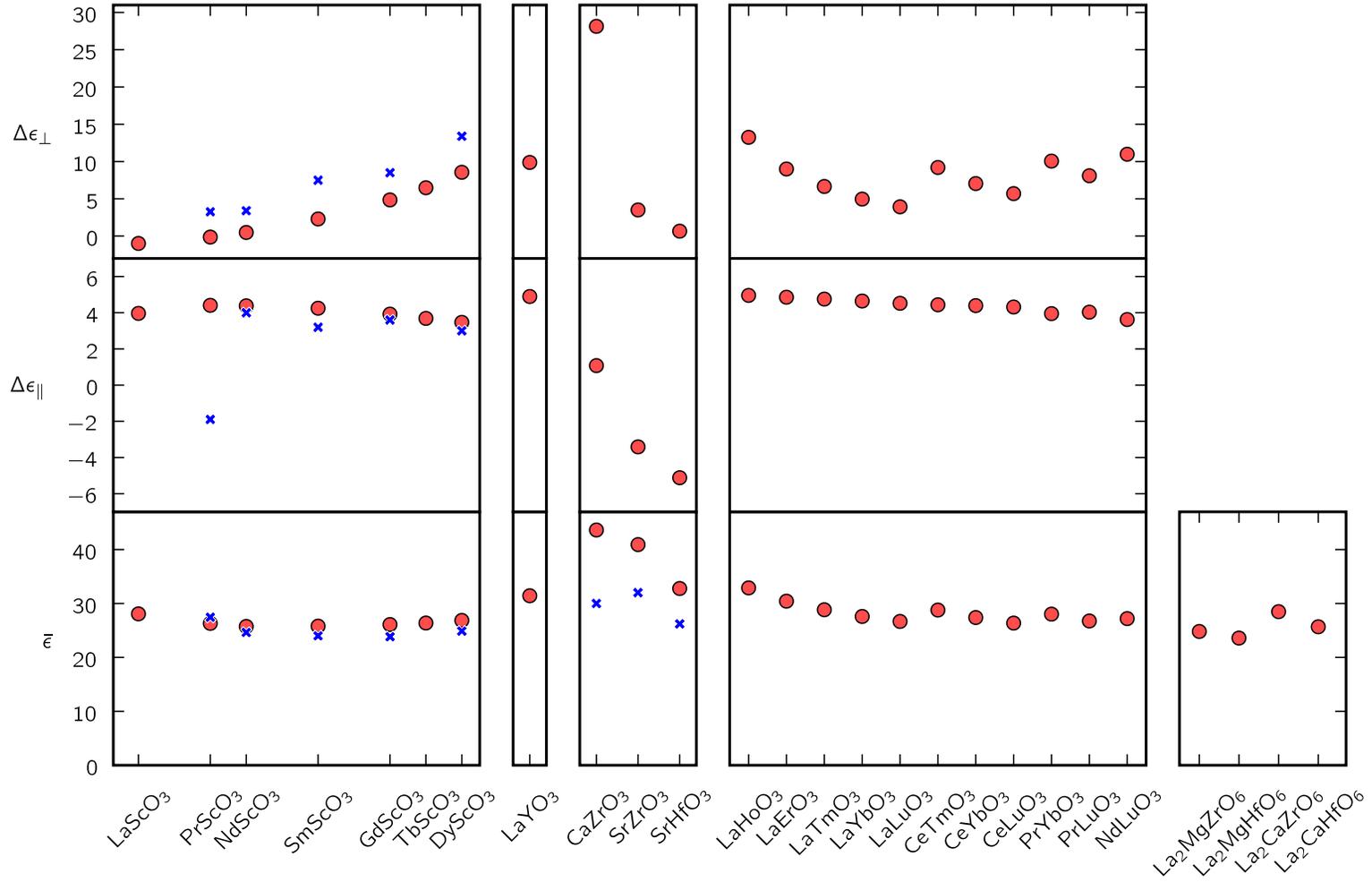


Figure 4.6: Dielectric information for all systems we considered. Bottom pane shows average dielectric tensor $\bar{\epsilon}$, mid-pane shows x - y anisotropy of the dielectric tensor, $\Delta\epsilon_{\parallel}$, and z anisotropy of the dielectric tensor $\Delta\epsilon_{\perp}$ is given in top pane. Calculated values are shown with red circles and experimental values with blue cross symbols, if available. See Table 4.3 for numerical values.

4.3.4 Decomposition of the ionic contribution to the dielectric tensor

As already mentioned, the ionic contribution to the dielectric tensor dominates in all of the systems we considered. The expression for the ionic contribution given in Eq. (4.4) provides a decomposition into contributions coming from eigenmodes of the force-constant matrix. The $Pbnm$ symmetry in perovskites, which is also approximately satisfied in $\text{La}_2\text{BB}'\text{O}_6$ compounds, allows a given eigenmode to contribute only to a single component (ϵ_{xx} , ϵ_{yy} , or ϵ_{zz}) of the dielectric tensor. This decomposition is given in Fig. 4.7 for all three components.

In rare-earth scandates, all three directions are evidently very different. The ϵ_{xx} component is dominated by a low-lying mode whose contribution is almost constant along the series (it contributes to ϵ_{xx} by 9.6 for LaScO_3 and 11.1 for DyScO_3). The ϵ_{yy} component, on the other hand, has sizable contributions coming from several modes. Finally, the ϵ_{zz} component comes mostly from a single low-lying mode. Unlike for the ϵ_{xx} component, the contribution from the mode responsible for the ϵ_{zz} component changes dramatically across the series, varying from 6.9 for LaScO_3 to 16.3 for DyScO_3 . This explains the large value of the z anisotropy in DyScO_3 as compared to LaScO_3 that is visible in Fig. 4.6.

A behavior similar to that of the rare-earth scandates is also observed in LaYO_3 and in the rare-earth perovskites. The SrZrO_3 and SrHfO_3 compounds show a quite similar behavior to each other. The ϵ_{xx} component has contributions coming from many modes, the ϵ_{yy} component is dominated by a single low-lying mode, and ϵ_{zz} is dominated by two low-lying modes. On the other hand, ϵ_{zz} component in CaZrO_3 shows very large contribution coming from a single low-lying mode. Finally, we note that the $\text{La}_2\text{BB}'\text{O}_6$ compounds containing Ca have stronger contributions to ϵ_{xx} and ϵ_{zz} from low-lying modes than do those containing Mg.

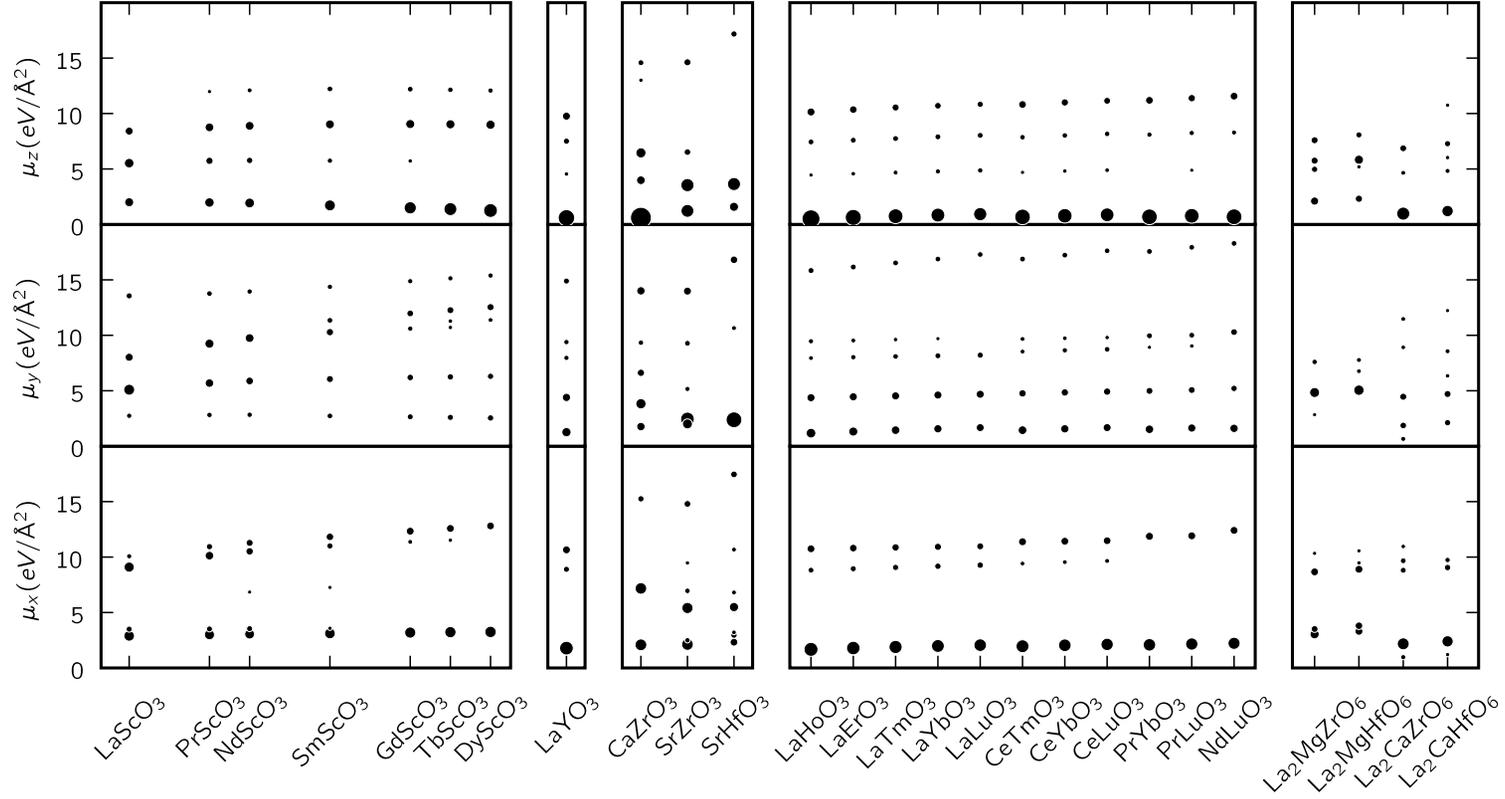


Figure 4.7: Eigenvalues of a force constant matrix $\Phi_{i\alpha,j\beta}$ of modes that contribute to $\epsilon_{xx}^{\text{ion}}$ (first horizontal pane from bottom), $\epsilon_{yy}^{\text{ion}}$ (second pane), or $\epsilon_{zz}^{\text{ion}}$ (third) by more than 1.5. For each mode the area of the circle is proportional to its contribution to $\epsilon_{ii}^{\text{ion}}$. Eigenvalues are in $\text{eV}/\text{\AA}^2$.

4.3.5 Compounds with $R\bar{3}c$ symmetry

At room temperature the ground state of BiFeO_3 is ferroelectric with polar space group $R3c$, the pattern of octahedral rotations being $(a^-a^-a^-)$ in the Glazer notation. At higher temperature, however, BiFeO_3 undergoes a phase transition in which the ferroelectricity and the $(a^-a^-a^-)$ pattern of octahedral rotations disappear simultaneously [159, 160]. This observations led us to hypothesize that rotations of octahedra around the pseudocubic $[111]$ axis, as in the $(a^-a^-a^-)$ pattern, tend to be energetically compatible with the presence of a ferroelectric distortion along the same axis. This would tend to suggest that perovskites that adopt the centrosymmetric $R\bar{3}c$ group, which also exhibits the $(a^-a^-a^-)$ pattern of oxygen octahedra, might be close to a ferroelectric instability leading to the lower-symmetry $R3c$ space group, and thus that such compounds might have an especially large component of the dielectric tensor along the pseudocubic $[111]$ axis.

To test this hypothesis, we have carried out a series of calculations on SrZrO_3 and GdScO_3 in which structural relaxation was allowed while maintaining the $R\bar{3}c$ symmetry. In both compounds we find some IR-active phonon modes that either have very low or imaginary frequency, indicating a near or actual instability. In the case of SrZrO_3 we find a mode that is active along the $[111]$ pseudocubic direction and has an extremely small frequency of only 6 cm^{-1} , while for GdScO_3 we find that the corresponding mode is unstable with an imaginary frequency of $i142 \text{ cm}^{-1}$. These calculations show that imposing the $R\bar{3}c$ structure on SrZrO_3 and GdScO_3 make them nearly or actually ferroelectric, thus confirming our hypothesis.

Incidentally, the observation that SrZrO_3 is more likely than GdScO_3 to be stabilized in the $R\bar{3}c$ structure is consistent with the fact that perovskite structures that prefer smaller rotation angles are more likely to form $R\bar{3}c$ than $Pbnm$

structures, as discussed by Woodward [141]. We find that the rotational angles for SrZrO₃ in the *Pbnm* space group are $\theta_R = 11.6^\circ$ and $\theta_M = 7.7^\circ$, while in GdScO₃ they are substantially larger, $\theta_R = 19.0^\circ$ and $\theta_M = 13.3^\circ$. More directly, we also find that the ground-state energy of SrZrO₃ having the $R\bar{3}c$ structure is only higher by 33.6 meV per formula unit than in the *Pbnm* structure. On the other hand, in GdScO₃ the $R\bar{3}c$ is higher in energy by a much larger increment of 386 meV.

Finally, we note that LaAlO₃, NdAlO₃, and BaTbO₃ may also be of interest, as these all have the $R\bar{3}c$ space-group symmetry and should also be chemically stable on silicon.

4.3.6 Correlation between structural and dielectric properties

The heuristic observation about BiFeO₃ mentioned in the previous section (Sec. 4.3.5) led us to make a more detailed analysis of the correlation between structural and dielectric anisotropies in all five groups of *Pbnm* perovskites. As can be seen from Fig. 4.3, the presence of the octahedral rotations breaks the symmetry among the three Cartesian directions in the *Pbnm* perovskites. One would therefore naively expect that the anisotropy in the dielectric tensor component should also be correlated with the size of these rotation angles, but this is not what we observe. For example, LaScO₃ and DyScO₃ both have rather substantial octahedral rotation angles (θ_R is 14.9° in LaScO₃ and 19.7° in DyScO₃), but they have very different values of the dielectric z anisotropy ($\Delta\epsilon_\perp$ is -1.0 in LaScO₃ and 8.6 in DyScO₃). An even more extreme behavior can be seen in the case of rare-earth perovskites between, e.g., LaHoO₃ and LaLuO₃.

Thus, we find no simple correlation between the dielectric tensor anisotropies and the values of the octahedral rotation angles. Instead, we find a correlation

between the dielectric tensor anisotropies and the *mismatch* of the two rotation angles θ_R and θ_M , as we explain next.

While the $Pbnm$ symmetry does not impose any relationship between the two octahedral rotation angles θ_R and θ_M , we find in practice that all the compounds we studied obey the heuristic relationship $\theta_R \simeq \sqrt{2}\theta_M$. This means that the oxygen octahedra are rotated about the three Cartesian axes by almost the same rotation angle, or equivalently, that the rotation axis is nearly $\langle 111 \rangle$. In the Glazer language, these $Pbnm$ perovskites having $(a^-a^-c^+)$ rotations can be said to be very close to an $(a^-a^-a^+)$ pattern. We can measure the mismatch between the actual $(a^-a^-a^+)$ and the hypothetical $(a^-a^-c^+)$ rotation pattern by the quantity $\theta_M - \theta_R/\sqrt{2}$, and it is this quantity that we find to be strongly correlated with the dielectric anisotropy $\Delta\epsilon_\perp$.

This is shown in Fig. 4.8, where $\Delta\epsilon_\perp$ is plotted versus $\theta_M - \theta_R/\sqrt{2}$ for all of the compounds considered in this work. It is apparent that the III-III-valent perovskites have a different behavior than the II-IV-valent ones. Nevertheless, we conclude that in both cases there is a strong correlation between the mismatch angle and the dielectric tensor anisotropy. The sign of the correlation is such that a deviation from the $(a^-a^-a^+)$ pattern having an increased rotation angle around the z axis gives a larger dielectric tensor component along the z axis, and thus a larger z anisotropy $\Delta\epsilon_\perp$.

4.3.7 Antisite substitutions

Experimentally the compositions of the perovskites that we have been describing up to now by their nominal compositions, e.g., LaLuO_3 , are in fact slightly different from the compositions of the single crystals on which the dielectric tensors were measured. This is because these crystals are grown at the congruently melting compositions, e.g., $\text{La}_{0.94}\text{Lu}_{1.06}\text{O}_3$, which differ from the nominal compositions

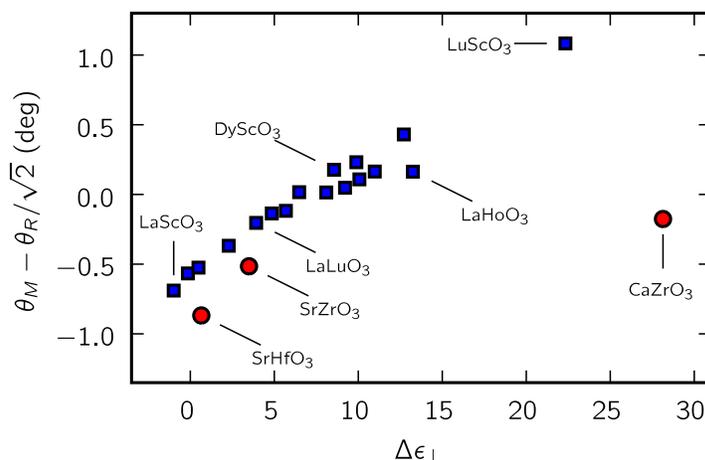


Figure 4.8: Correlation between the dielectric tensor z -anisotropy ($\Delta\epsilon_{\perp}$) and the mismatch in the oxygen octahedra rotation angle. Only some perovskites are labeled. Perovskites where both A and B ions are III-valent are indicated with blue square symbols and those where A ions are II-valent and B are IV-valent are indicated with red circles.

described up to now. The congruently melting compositions of all relevant $Pbnm$ perovskites studied have been found to be poor in the A -site cation and rich in the B -site cation composition [161, 162, 152].

For this reason, we decided to carry out a theoretical analysis of the effects of B atoms substituting at the A site on the structural and dielectric properties of the material. Detailed calculations were done only for the case of LaLuO_3 , but we expect that similar trends will be observed in the remaining rare-earth rare-earth perovskites as well as in the rare-earth scandates and yttrates. Of course, other kinds of compositional disorder might also be present, but such possibilities are not analyzed here.

Analysis of antisite defects in LaLuO_3

We studied Lu_{La} antisite defects in LaLuO_3 using a supercell approach. Specifically, in order to model a situation in which one of every 16 La atoms is substituted by Lu, which is about a 6% substitution, we constructed an 80-atom supercell

containing a single antisite defect. The supercell is enlarged with respect to the primitive 20-atom $Pbnm$ primitive cell by doubling along both the orthorhombic a and b lattice vectors. The resulting stoichiometry is



The presence of the Lu_{La} antisite in this particular 80-atom supercell reduces the crystal symmetry from orthorhombic $Pbnm$ to monoclinic Pm . After full relaxation of the crystal structure in this space group, we find that the a , b , and c lattice vectors are reduced by 0.3%, 0.2% ,and 0.1%, respectively, while the monoclinic angle between a and b lattice vectors of 90.03° deviates only very slightly from 90° .

The influence of the Lu_{La} substitution on the dielectric properties is more complex. Evaluated in the same coordinate frame as in the $Pbnm$ unit cell, the ϵ_{xx} and ϵ_{yy} dielectric tensor components remain almost unchanged, and the new ϵ_{xy} component allowed by the monoclinic symmetry is quite small, only 2.1. On the other hand, the ϵ_{zz} component is drastically altered by the presence of Lu atom on the La site. In fact, we find that the 80-atom supercell is actually just barely unstable in the Pm space group, as indicated by the presence of a phonon mode with a very small imaginary frequency of $i16 \text{ cm}^{-1}$ ². The contribution of this phonon mode to the ϵ_{zz} component (evaluated in the unstable Pm structure) is therefore negative, specifically, -33.6 . Since this phonon frequency is so close to zero, we expect that it would get renormalized to positive frequency at room temperature. For this reason, we did not follow the structural relaxation of our 80-atom supercell along the direction of the unstable mode, and a realistic estimate of

²The eigendisplacement of this unstable mode is very close to that of the S -point $(\frac{1}{2}\frac{1}{2}0)$ phonon of the 20-atom $Pbnm$ structure having a frequency of 39 cm^{-1} . However, this phonon remains inactive in the $Pbnm$ structure because it does not appear at the Γ point.

the dielectric response of the system is difficult. Nevertheless, we conclude that Lu_{La} substitutions in LaLuO_3 have the potential to increase the ϵ_{zz} dielectric tensor component substantially.

Discussion

As can be clearly seen in Fig. 4.6, our calculated z anisotropy ($\Delta\epsilon_{\perp}$) is consistently larger than the measured one for all the rare-earth scandates for which we have experimental measurements. In view of the calculations reported for LaLuO_3 above, we tentatively attribute this discrepancy to the generic tendency of B atoms to substitute on the A site in these compounds. This observation is consistent with the fact that smaller B ions that substitute for larger A ions will reside in a relatively larger cage, providing room to rattle and thereby contribute to an enhanced dielectric response.

4.4 Summary

The main focus of this work has been the application of computational methods to study the structural and dielectric properties of various $Pbnm$ perovskites that have potentially large dielectric tensor components and are chemically stable on silicon up to ~ 1000 °C [101]. Such compounds might be good candidates for future use as high- K dielectrics in microelectronics applications, e.g., as a possible replacement of hafnia-based high- K dielectrics currently used in the CMOS transistors in integrated circuits.

Of the compounds we have considered, CaZrO_3 , SrZrO_3 , LaHoO_3 , and LaYO_3 appear to be especially promising. CaZrO_3 has the largest calculated average dielectric tensor ($\bar{\epsilon} = 43.6$) among the compounds we considered, and SrZrO_3 is a close second with $\bar{\epsilon} = 40.9$. The dielectric tensor in CaZrO_3 is very anisotropic,

with its ϵ_{zz} component almost twice as large as ϵ_{xx} or ϵ_{yy} , while on the other hand SrZrO₃ has an almost isotropic dielectric tensor. Unfortunately, the full dielectric tensors of these compounds have not yet been measured due to lack of single crystals.

Of the rare-earth rare-earth *Pbnm* perovskites, only LaLuO₃ has had its dielectric tensor measured to date. Our results on this compound will be presented in detail elsewhere [158]. The theoretical calculations, however, indicate that other compounds in this series should have even larger dielectric tensor components, with LaHoO₃, having $\bar{\epsilon} = 32.9$, being the most promising among these. LaYO₃ is expected to behave very similar to LaHoO₃ since Y and Ho have almost the same ionic radii, so it may be promising as well ($\bar{\epsilon} = 31.4$). Thus, this series of compounds clearly deserves additional scrutiny.

Of course, there are good reasons for preferring amorphous over single-crystalline materials for such high-*K* applications. Certainly the ability of amorphous SiO₂ to conform to the substrate and to eliminate electrical traps played a central role in its dominance as the gate dielectric of choice for 40 years for silicon-based metal-oxide-semiconductor field-effect transistors. The present hafnia-based high-*K* dielectrics are amorphous or nanocrystalline [163]. For this reason, any eventual application of these materials for high-*K* applications would presumably require the adoption of one of two strategies. The first is the possibility of growing crystalline epitaxial oxides directly on silicon, which clearly would require a very high level of control of interface chemistry and morphology before it could become a practical solution. The second is the possibility that some of the compounds investigated here could be synthesized in amorphous or nanocrystalline form. We have not investigated these issues here, nor have we tried to calculate what (possibly very substantial) changes in the dielectric properties would occur in the amorphous counterparts, as this would take us far beyond the scope of the present

study. Nevertheless, these are important questions for future investigations.

Chapter 5

Large isosymmetric reorientation of oxygen octahedra rotation axes in epitaxially strained perovskites

Using first-principles density functional theory calculations, we discover anomalously large bi-axial strain-induced reorientation of the axis about which the oxygen octahedral framework rotates in orthorhombic perovskites with tendency towards rhombohedral symmetry. The transition between crystallographically equivalent (isosymmetric) structures with different octahedral rotation magnitudes originates from both the strong strain–octahedral rotation coupling available to perovskites and the energetic hierarchy among competing octahedral tilt patterns. By elucidating these criteria, we suggest many functional perovskites would exhibit the transition in thin film form, thus offering a new landscape in which to tailor highly anisotropic electronic responses.

5.1 Introduction

Phase transitions are a ubiquitous phenomena in nature; they describe diverse topics ranging from crystallization and growth to superconducting Cooper pair condensation. Isosymmetric phase transitions (IPT)—those which show no change

in occupied Wyckoff positions or crystallographic space group—are an intriguing class since there are relatively few examples in crystalline matter [164]: Most condensed matter systems respond to external pressures and temperatures by undergoing “conventional” symmetry-lowering displacive [165], martensitic [166] or reconstructive [58] transitions. Furthermore, the experimental characterization and identification of a suitable symmetry-preserving order parameter through such transitions is often challenging [167]. Although some *electronic* order parameters [168, 169] that include ferroelectric [170, 171, 172, 173] or orbital polarizations [174] have been proposed for IPT, which lead to subsequent changes in local cation coordinations [175, 176, 177], to the best of our knowledge, there is no case where the IPT connects two structures with essentially the same local bonding environment.

Using first-principles density functional calculations, we find an isosymmetric transition in the low energy rhombohedral phases of epitaxially strained orthorhombic perovskites and describe how to experimentally access it. We show that the transition originates from non-polar distortions that describe the geometric connectivity and relative phase of the BO_6 octahedra found in rhombohedral and orthorhombic perovskites. Although a previous IPT in a thin film perovskite that relies on strong strain–polar phonon coupling has been reported [178], we describe here a universal symmetry preserving transition that originates from the strong lattice–octahedral rotation coupling ubiquitous in nearly all perovskites, not just those with ferroelectric tendencies. For this reason, the large isosymmetric reorientation of the oxygen rotation axes should be readily observable in many rhombohedral perovskites with diverse chemistries. Since dielectric tensor studies from Chapter 4 indicate that the dielectric anisotropy in perovskites is strongly linked to the deviations in the octahedral rotation axis direction, we suggest control over the strain-induced reorientation of octahedral rotation axes

could provide for highly tunable high- K dielectric actuators and temperature-free relative permittivity resonance frequencies [179].

We choose LaGaO_3 as our model system since it has a tolerance factor of $\tau = 0.966$ indicating the perovskite structure is highly susceptible to GaO_6 octahedral rotations about the principle symmetry axes [180]: At room temperature it is orthorhombic (space group $Pbnm$) and undergoes a first-order phase transition to rhombohedral $R\bar{3}c$ around 418 K [181], with a subsequent change in the GaO_6 octahedral rotation patterns from $a^-a^-c^+$ to $a^-a^-a^-$, respectively, in Glazer notation [142]. The $+$ ($-$) superscripts indicate in- (out-of)-phase rotations of adjacent octahedra along a given Cartesian direction. The non-magnetic Ga^{3+} cations additionally allow us to eliminate possible contributions of spin and orbital degrees of freedom for driving the IPT through alternative electronic mechanisms.

5.2 Calculation details and notation

Our density functional calculations are performed within the local density approximation (LDA) as implemented in the Vienna *Ab initio* Simulation Package (VASP) [182, 183] with the projector augmented wave (PAW) method [13], a $5 \times 5 \times 5$ Monkhorst-Pack k -point mesh [80] and a 500 eV plane wave cutoff. For structural relaxations, we relax the atomic positions (forces to be less than 0.1 meV \AA^{-1}) and the out-of-plane c -axis lattice constants for the strained films ¹.

The principle difference between the ground state orthorhombic $Pbnm$ and metastable [12 meV per formula unit (f.u.) higher in energy] rhombohedral $R\bar{3}c$ phases of LaGaO_3 is that the GaO_6 octahedra rotate in-phase ($+$) along the Cartesian z -direction of the $Pbnm$ structure while they rotate out-of-phase ($-$)

¹All strain values are given relative to the hypothetical cubic equilibrium LDA lattice parameter (3.831 \AA)

about that same direction in the $R\bar{3}c$ structure. Our homoepitaxial bi-axial strain calculations simulate film growth on a (001)-terminated substrate, we therefore choose the c^+ rotations of the orthorhombic phase to be about the axis perpendicular to the epitaxial plane (Fig. 5.1); this orientation allows us to evaluate the bi-axial strain effect on the in- versus out-of-phase GaO_6 rotations present in the two phases. Note, the bi-axial constraint preserves the orthorhombic symmetry in the $a^-a^-c^+$ phase, however, we designate the *epitaxially* (e) strained phase as $e\text{-Pbnm}$ to distinguish it from the bulk structure. In contrast, the symmetry of the bulk rhombohedral phase is lowered to monoclinic (space group $C2/c$) and we therefore refer to it as such throughout ².

5.3 Strain-stabilized structures

We first compute the evolution in the total energy with bi-axial strain for the $e\text{-Pbnm}$ and $C2/c$ structures [Fig. 5.1(a)]. We find that between approximately -1 to $+3\%$ strain, the orthorhombic phase with the $a^-a^-c^+$ rotation pattern is more stable than the monoclinic $a^-a^-c^-$ structure. For now we focus on the monoclinic phases [Fig. 5.1(b)] near 0% strain and explore plausible ways later to access them. We find an abrupt discontinuity in the first derivative of the total energy with strain for the monoclinic structure between two states with the same symmetry denoted as $C2/c(1)$ and $C2/c(2)$. In contrast, we find a single continuous equation of state when we apply *uniform* hydrostatic pressure over ± 50 GPa on our bulk reference structures. The evolution in the c/a axial ratio for these structures is also qualitatively different [Fig. 5.1(c)]. The $e\text{-Pbnm}$ axial ratio continuously decreases with increasing tensile strain (consistent with

²In the monoclinic structures, we follow the convention of constraining the free inter-axial angle to be that of the fully relaxed rhombohedral structure since it is known to weakly affect the ground state properties [178, 184].

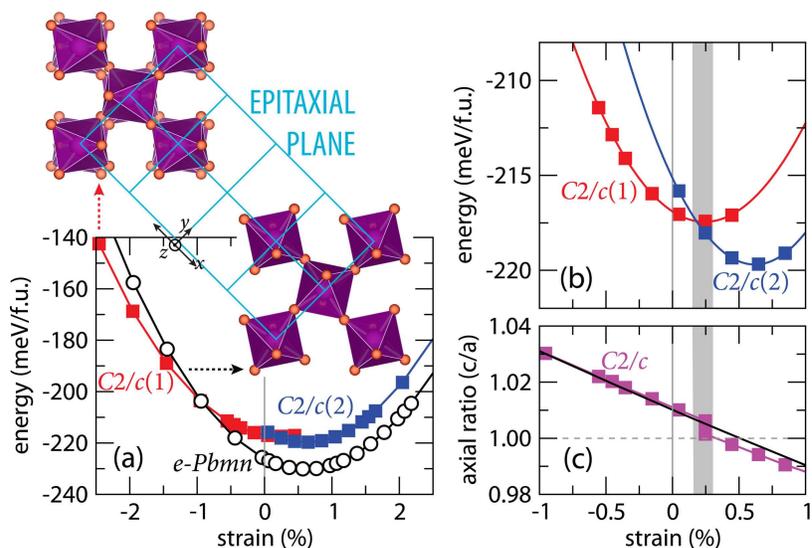


Figure 5.1: Evolution of the total energy (a) for the e - $Pbmn$ and $C2/c$ phases with in- and out-of-phase octahedral rotations (inset) along the z -direction. (b) Magnified region about the IPT (shaded) between phases $C2/c(1)$ and $C2/c(2)$. (c) The change in axial ratio with strain shows a discontinuity in the $C2/c$ phase that is absent in the e - $Pbmn$ structure.

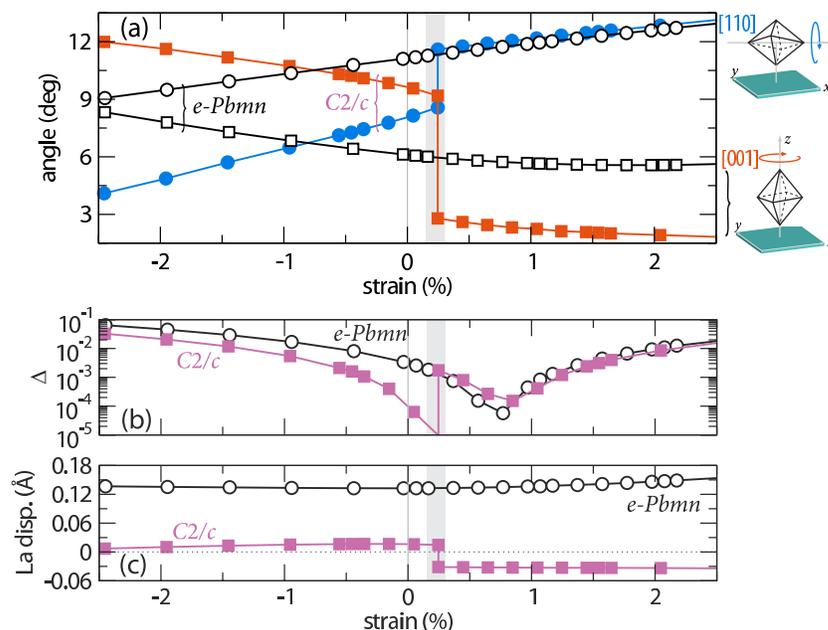


Figure 5.2: Evolution in (a) the GaO_6 rotation angles about different directions relative to the substrate, (b) the octahedral distortion parameter Δ , and (c) the La displacements about the bulk structures with epitaxial strain.

elastic theory), whereas in the $C2/c$ structures a sharp discontinuity occurs in the vicinity of $c/a \sim 1$. We find the first-order phase transition occurs at a critical strain of $\sim 0.18\%$ from intersection of the quadratics fit to the total energies for each monoclinic phase.

5.3.1 Microscopic structure evolution

To investigate if the $C2/c(1) \rightarrow C2/c(2)$ transition is indeed isosymmetric, we evaluate how the internal structural parameters – octahedral tilts and bond distortions – evolve with epitaxial strain [Fig. 5.2(a)]. We find a continuous evolution in the GaO_6 rotation angles for the $e\text{-}Pbnm$ structures (open symbols): the rotation axis changes from being along the $[001]$ -direction to mainly in-plane along $[110]$ as the strain state changes from compressive to tensile. In contrast, we find an abrupt change in the octahedral rotation angles with strain in the monoclinic phases (filled symbols). We identify that the $C2/c(1)$ and $C2/c(2)$ phases, despite possessing the same symmetry are distinguishable—each is characterized as having either mainly $[110]$ in-plane or $[001]$ out-of-plane GaO_6 octahedra rotations [Fig. 5.2(a)]. Consistent with the orthorhombic case we find that increasing tensile strain drives the the octahedral rotation axis into the $[110]$ -epitaxial plane.

The bi-axial strain is not solely accommodated by rigid octahedral rotations. It produces additional deviations in the Ga–O bond lengths and causes La cation displacements. We quantify the former effect through the octahedral distortion parameter $\Delta = \frac{1}{6} \sum_{n=1,6} [(\delta(n) - \langle \delta \rangle) / \langle \delta \rangle]^2$, where δ is a Ga–O bond length and $\langle \delta \rangle$ is the mean bond length in the GaO_6 octahedra. With increasing strain, Δ increases, indicating that bond stretching (and compression) occurs simultaneously with changes in the magnitude of the octahedral rotation angles to alleviate the substrate-induced strain [Fig. 5.2(b)]. According to our bond-valence calculations, the Ga–O bond stretching modifies the “chemical strain” imposed on the

over-bonded Ga^{3+} cations when a regular GaO_6 octahedra ($\Delta \rightarrow 0$) occurs. The IPT allows the monoclinic phase to maintain a uniform charge density distribution with the $a^-a^-c^-$ tilt pattern. Note, this chemically over-bonded structure is absent in the $e\text{-}Pbnm$ structure because the $a^-a^-c^+$ tilt pattern (D_{2h} symmetry) permits non-uniform Ga–O bonds.

We also show in Fig. 5.2(c) that the La anti-parallel displacements exhibit additional anomalous behavior and change sign in monoclinic structure about the transition region (shaded). This occurs to maintain a trigonal planar configuration, with a nearly uniform charge density, in the cavity created by the different $a^-a^-c^-$ rotation “senses.” In contrast, the $e\text{-}Pbnm$ structures exhibit a single La cation displacement direction because the octahedral rotation pattern never reverses.

5.4 Origin of the isosymmetric transition

To identify the origin of the isosymmetric transition, we first analyze the energy of the monoclinic $a^-a^-c^-$ structures under different bi-axial strain states as a function of *direction* and *magnitude* of the GaO_6 octahedron rotation axis. The direction of the GaO_6 rotation axis with the $a^-a^-c^-$ pattern is constrained to be in the $(\bar{1}10)$ -plane because the rotation pattern can be decomposed into rotations of the type $a^-a^-c^0$ and $a^0a^0c^-$ with rotation axes aligned along the $[110]$ - and $[001]$ -directions [Fig. 5.2(a)]. We show in Fig. 5.3(a-c) our first-principles results of the energy dependence on the direction (vertical axes) and magnitude (horizontal axes) of the GaO_6 rotation axis for strain values of -1.5%, 0.0% and 1.5%, respectively. For all strain states, we find a single well-defined energy minimum for each *direction* of the GaO_6 rotation axis, as shown by the dotted lines in Fig. 5.3(a-c). We therefore are able to remove the rotation angle magnitude as

a variable and to analyze the energy dependence solely in terms of the bi-axial strain and the GaO_6 rotation axis *direction* [185].

We show in Fig. 5.3(d) the calculated evolution of the extremal octahedra rotation axis directions with bi-axial strain. The local energy minima (maxima) for all strains are indicated with a heavy (broken) black line in Fig. 5.3(d). For -1.5%, 0.0% and 1.5% bi-axial strains, we additionally indicate the extrema using symbols both in Fig. 5.3(a-c) and Fig. 5.3(d). Consistent with our earlier structural analysis (Fig. 5.2), we find that the rotation axis direction smoothly approaches the [110] ([001]) direction for large tensile (compressive) strains. For the range of strains between -0.5% and 0.5% , we observe the co-existence of two energy minima separated by an energy maximum (broken line); this indicates an *inaccessible region* of rotation axis directions close to [111] for any value of strain and is consistent with a first-order transition.

Our results suggest there are two main reasons for the appearance of the isosymmetric transition in epitaxially strained rhombohedral perovskites. The first reason is that the octahedral rotations are strongly coupled to the bi-axial strain. This coupling originates from the rigidity of the GaO_6 octahedra, since the rigidity causes contraction of the crystal lattice in the direction orthogonal to the rotation axis [186, 187, 188]. The second reason is that the bulk rhombohedral $a^-a^-a^-$ structure of LaGaO_3 is *higher* in energy than the bulk orthorhombic $a^-a^-c^+$ structure.

We now show that the energy ordering of the bulk phases is responsible for the *inaccessible region* of rotation axis directions. The $a^-a^-a^-$ and $a^-a^-c^+$ structures differ only in the phase of the GaO_6 octahedra rotations about the z -axis. For this reason, each structure can be transformed into the other through a combination of rigid octahedral distortions. One distortion should deactivate the a^- rotation about the z -axis, while the other would induce the c^+ rotation about the same

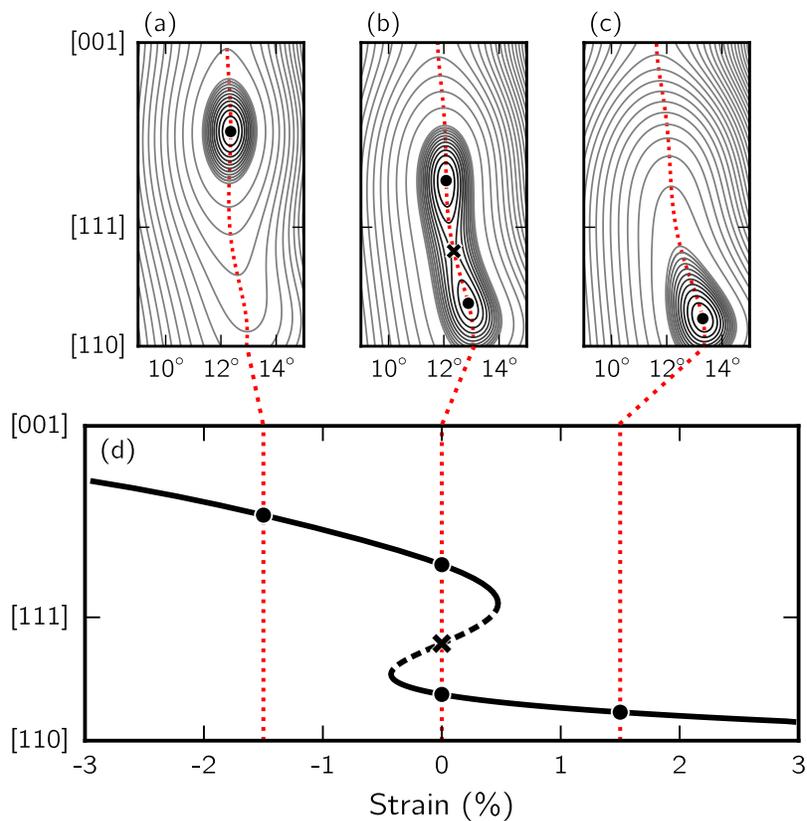


Figure 5.3: Calculated energy (a-c) of the monoclinic LaGaO_3 phases as a function of the GaO_6 rotation axis direction and angle magnitude at -1.5% , 0.0% and 1.5% strains, respectively, with energy contours at 5 meV/f.u. and additional contours close to extremal points at 0.5 meV/f.u. (d) Position of energy minimum (solid line) or maximum point (broken line) with strain; circles correspond to minima (a-c), and the cross indicates the saddle point in (b).

axis. We would expect these distortions to impose minor energetic penalties since they are nearly rigid³. In the present case, where the $a^-a^-a^-$ structure is higher in energy than $a^-a^-c^+$, we expect that introduction of either of these distortions into the *higher* energy $a^-a^-a^-$ structure will lower the total energy⁴.

Finally, smoothness of the total energy as a function of strain and rotation axis direction requires that the difference between the number of energy minima (N_{\min}) and maxima (N_{\max}), for any value of bi-axial strain, remains fixed as elaborated in Morse theory [185]. In other words, any smooth deformation which produces additional energy maximum must also produce additional energy minimum. Due to first reason for the IPT mentioned above, we anticipate that for sufficiently large compressive or tensile strains, the strain–octahedral rotation direction coupling dominates to yield a single energy minimum: $N_{\min} = 1$, $N_{\max} = 0$ and then from continuity, $N_{\min} - N_{\max} = 1$ must remain constant for all strains. From our energetic hierarchy of the bulk structures, we conclude that when strain induces structural distortions with magnitudes which nearly coincide with those of the bulk $a^-a^-a^-$ phase (near 0% strain and [111] direction), there will exist an energy maximum ($N_{\max} = 1$); from continuity, this must introduce two energy minima ($N_{\min} = 2$) at the same value of strain. These reasons together produce the energy landscape shown in Fig. 5.3(d) and require an IPT in the LaGaO₃ system. For comparison, the orthorhombic phase of LaGaO₃ does not show an IPT as one varies bi-axial strain, since the second condition for the transition described above does not apply, ie. the orthorhombic structure is the global ground state. Therefore, N_{\min} is fixed to 1 ($N_{\max} = 0$) for all strain values.

³If a change in the $a^-a^-a^0$ rotation also occurs, so as to keep the total octahedron rotation angle magnitude nearly constant, the energy penalty is even smaller and fully consistent with the energy landscape in Fig. 5.3(a-c).

⁴Indeed, we calculate an unstable phonon at Γ with E_g symmetry ($\omega = 25i$ cm⁻¹) and an unstable phonon at the F -point ($\omega = 43i$ cm⁻¹) in the rhombohedral structure corresponding to these kind of distortions.

5.5 Accessing and applications of the isosymmetric transition

We first note that we obtain a $C2/c \rightarrow e\text{-}Pbmn$ transition near -1% compressive strain with respect to our hypothetical cubic LaGaO_3 phase. In the vicinity of the IPT at 0 K, however, the $e\text{-}Pbmn$ phase is the global ground state. Although our minimal model for the IPT relies on this energetic ordering of the competing rotational phases ($a^-a^-a^-$ versus $a^-a^-c^+$), we anticipate three experimental routes by which to access the essential signature of the isosymmetric transition—large strain-induced reorientation of the octahedral rotation axis direction. First, the monoclinic phases could be stabilized in thin films through the substrate coherency effect [189, 190], where the film’s tilt pattern adopts that of the substrate: Perovskite substrates with the $a^-a^-a^-$ (LaAlO_3) or the $a^0a^0c^-$ (tetragonal- SrTiO_3) tilt pattern are promising candidates. Second, additional electronic degrees of freedom (first- and second-order Jahn-Teller effects), introduced through cation substitution, could be exploited to stabilize the IPT because they often energetically compete with the octahedra rotations [178, 191]. The next route to access the monoclinic phases for all strains would be to perform experiments above the bulk LaGaO_3 structural transition temperature ($\sim 100^\circ\text{C}$). In this case, the IPT would exhibit a weak-first order transition while still providing strong strain–octahedral rotation axis direction coupling. At sufficiently high temperatures, the IPT could be suppressed, making the monoclinic phases indistinguishable and terminating the IPT boundary at a critical point [192].

5.6 Conclusion

We have shown that strain–octahedral rotation axis directions are strongly coupled in epitaxial perovskite thin films. For this reason, we suggest similar large

reorientations of coordinating polyhedra frameworks could be achieved in alternative structural families: thin films with the garnet, apatite or spinel structures are particularly promising. However, the functional materials design challenge remains: how does one couple the rotation axis direction to additional *electronic* degrees of freedom? For this reason, we advocate for detailed epitaxial film studies on perovskites close to the $R\bar{3}c \leftrightarrow Pnma$ phase transition ($0.96 < \tau < 1.01$). Controlling the IPT in LaCrO_3 , LaNiO_3 and LaCuO_3 perovskites could yield unknown, and potentially functional, orbitally-, spin- and charged-ordered phases.

Part II

Geometrical aspects of electronic structure theory

Chapter 6

Electric polarization in a Chern insulator

We extend the Berry-phase concept of polarization to insulators having a non-zero value of the Chern invariant. The generalization to such Chern insulators requires special care because of the partial occupation of chiral edge states. We show how the integrated bulk current arising from an adiabatic evolution can be related to a difference of bulk polarizations. We also show how the surface charge can be related to the bulk polarization, but only with a knowledge of the wavevector at which the occupancy of the edge state is discontinuous. Furthermore we present numerical calculations on a model Hamiltonian to provide additional support for our analytic arguments.

6.1 Introduction and motivation

In 1988 Haldane pointed out that an insulating crystal with broken time-reversal symmetry may exhibit a quantized Hall conductance even in the absence of a macroscopic magnetic field [32]. We shall refer to such a material as a “Chern insulator” (CI) because it necessarily would have a non-zero Chern invariant associated with its manifold of occupied Bloch states [193, 194]. While no CI has yet been discovered experimentally, there appears to be no reason why one could not exist, and theoretical models that behave as CIs are not difficult to construct.

It seems plausible that the current blossoming of interest in exotic non-collinear magnets and multiferroics could yield an experimental example before long.

CI's occupy a middle ground between metals and ordinary insulators. Like metals, their conductivity tensor $\sigma_{\alpha\beta}$ is non-zero, their surfaces are metallic (as a result of topological edge states crossing the Fermi energy), and it is impossible to construct exponentially localized Wannier functions (WFs) for them [195]. On the other hand, only the off-diagonal (dissipationless) elements of $\sigma_{\alpha\beta}$ can be non-zero, the chiral edge states decay exponentially into the bulk, the one-particle density matrix decays exponentially in the interior [196], and the localization measure Ω_I [197, 30] is finite [196] as in other insulators. Overall it appears natural to regard a CI as an unusual species of insulator, but many aspects of its behavior remain open to investigation.

As is well known, the electric polarization \mathbf{P} is not well-defined in a metal. For an ordinary insulator, its definition alternatively in terms of Berry phases or WFs is by now well established [24, 198, 199]. For a CI, the absence of a Wannier representation removes the possibility of using it to define the polarization, and we shall show below that there is a fundamental difficulty with the Berry-phase definition as well. In view of the presence of dissipationless currents and metallic edge states, one might be tempted to conclude that \mathbf{P} is not well-defined at all in a CI. On the other hand, Ω_I is related to the fluctuations of \mathbf{P} [31], and the finiteness of this quantity [196] suggests that the polarization might be well-defined after all.

The purpose of this chapter is to discuss whether, and in what sense, a definition of electric polarization is possible in a CI. We demonstrate that the usual Berry-phase definition does remain viable if it is interpreted with care when connecting it to observables such as the internal current that flows in response to an adiabatic change of the crystal Hamiltonian, or to the surface charge at the edge

of a bounded sample.

6.2 Results

For the remainder of this chapter we restrict ourselves to the case of a 2D crystalline insulator having a single isolated occupied band. The generalization to the case of a 3D multiband insulator is not difficult, but would complicate the presentation. We also restrict ourselves to a single-particle Hamiltonian, noting that the principal difficulties in understanding CIs occur already at the one-particle level. The lattice vectors \mathbf{a}_1 and \mathbf{a}_2 are related to the reciprocal lattice vectors \mathbf{b}_1 and \mathbf{b}_2 in the usual way ($\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$) and the cell area is $S = |\mathbf{a}_1 \times \mathbf{a}_2|$.

The Berry-phase expression for the electric polarization can be written as

$$\mathbf{P}_{[\mathbf{k}_0]} = \frac{e}{(2\pi)^2} \text{Im} \int_{[\mathbf{k}_0]} d\mathbf{k} \langle u_{\mathbf{k}} | \nabla_{\mathbf{k}} | u_{\mathbf{k}} \rangle \quad (6.1)$$

where e is the charge quantum ($e > 0$), $|u_{\mathbf{k}}\rangle$ are the cell-periodic Bloch functions, and $[\mathbf{k}_0]$ indicates the parallelogram reciprocal-space unit cell with origin at \mathbf{k}_0 (that is, with vertices \mathbf{k}_0 , $\mathbf{k}_0 + \mathbf{b}_1$, $\mathbf{k}_0 + \mathbf{b}_1 + \mathbf{b}_2$, and $\mathbf{k}_0 + \mathbf{b}_2$). In an ordinary insulator one insists on a smooth and periodic choice of gauge (relative phases of the $|u_{\mathbf{k}}\rangle$) in Eq. (6.1), and \mathbf{P} is well defined (modulo $e\mathbf{R}/S$, where \mathbf{R} is a lattice vector [24]) independent of \mathbf{k}_0 . However, in a CI such a gauge choice is no longer possible. To see this, we decompose $\mathbf{P}_{[\mathbf{k}_0]} = P_1\mathbf{a}_1 + P_2\mathbf{a}_2$, $\mathbf{k} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2$, and $\mathbf{k}_0 = \kappa_1\mathbf{b}_1 + \kappa_2\mathbf{b}_2$, and rewrite Eq. (6.1) as

$$P_1^{[\kappa_2]} = \frac{-e}{S} \int_{\kappa_2}^{\kappa_2+1} dk_2 \frac{\theta_1(k_2)}{2\pi}, \quad (6.2)$$

$$\theta_1(k_2) = -\text{Im} \int_{\kappa_1}^{\kappa_1+1} dk_1 \langle u_{k_1, k_2} | \partial_{k_1} | u_{k_1, k_2} \rangle. \quad (6.3)$$

Eq. (6.3) is a Berry phase and is gauge independent modulo 2π (independent of κ_1). This allows us to make an arbitrary choice of branch for $\theta_1(k_2 = \kappa_2)$ and to insist, as part of the definition of $P_1^{[\kappa_2]}$, that $\theta_1(k_2)$ should remain continuous as k_2 is increased from κ_2 to $\kappa_2 + 1$. Since states at (k_1, κ_2) and $(k_1, \kappa_2 + 1)$ are equivalent, it follows that

$$\theta_1 \Big|_{\kappa_2}^{\kappa_2+1} = -2\pi C \quad (6.4)$$

where C is an integer. In fact C just defines the Chern number, and the insulator is a CI if $C \neq 0$. For simplicity we focus henceforth on a CI having $C = \pm 1$.

Using Eqs. (6.2)-(6.3) and similar equations for P_2 , we have arrived at a definition $\mathbf{P}_{[\mathbf{k}_0]}$ that is well-defined, modulo $e\mathbf{R}/S$ as usual, even for a CI. However, as illustrated in Fig. 6.1(a),

$$\mathbf{P}_{[\mathbf{k}_0+\Delta\mathbf{k}]} = \mathbf{P}_{[\mathbf{k}_0]} - \frac{eC}{2\pi} \hat{\mathbf{z}} \times \Delta\mathbf{k} \quad (6.5)$$

where $\hat{\mathbf{z}}$ is the unit vector along $\mathbf{a}_1 \times \mathbf{a}_2$. This dependence on \mathbf{k}_0 clearly presents a problem for the interpretation of Eq. (6.2) as a “physical” polarization in the case of a CI.

However, let us recall how the concept of polarization is *used*. For a normal insulator at least [24], the change of polarization during an adiabatic change of some internal parameter of the system from time t_i to t_f is given by

$$\int_{t_i}^{t_f} dt \mathbf{J}(t) = \mathbf{P}_{[\mathbf{k}_0]}^{(f)} - \mathbf{P}_{[\mathbf{k}_0]}^{(i)} \quad (\text{modulo } e\mathbf{R}/S), \quad (6.6)$$

where $\mathbf{J}(t)$ is the cell-averaged adiabatic current flowing in the bulk. A related statement, connected with the requirement that the charge pumped to the surface must be consistent with Eq. (6.6), is that the charge on an insulating surface

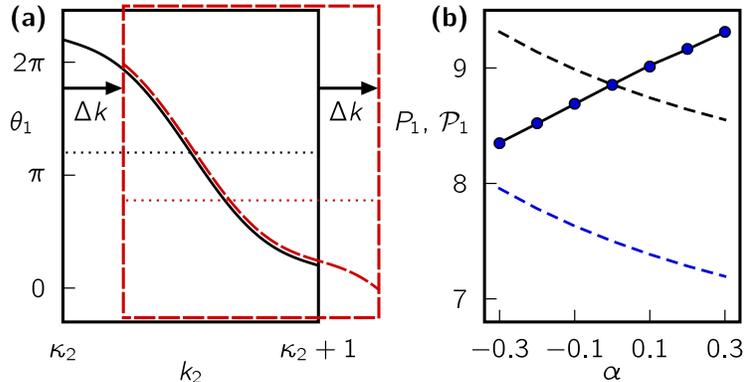


Figure 6.1: (a) Sketch of $\theta_1(k_2)$ in a Chern insulator ($C = +1$). Solid black and dashed red frames indicate reciprocal-cell origin chosen at κ_2 and $\kappa_2 + \Delta k$ respectively. Dotted lines indicate corresponding averages, proportional to P_1 . (b) Computed $P_1(\alpha)$ and $\mathcal{P}_1(\alpha)$ for the modified Haldane model, in units of $-0.01e/S$, for adiabatic (dashed lines) and thermal (solid line and symbols) filling. See text.

normal to reciprocal vector \mathbf{b}_1 is [198]

$$\sigma = \mathbf{P} \cdot \hat{\mathbf{b}}_1 \quad (\text{modulo } e/a_2). \quad (6.7)$$

Eqs. (6.6) and (6.7) embody the attributes of a useful definition of \mathbf{P} . In the remainder of this chapter, we demonstrate that a generalized definition of \mathbf{P} , having similar attributes, can be given in a CI. We first show that Eq. (6.6) remains correct, provided that the *same* \mathbf{k}_0 (i.e., the same reciprocal-space cell) is used for $\mathbf{P}^{(i)}$ and $\mathbf{P}^{(f)}$ in Eq. (6.6). We also show that Eq. (6.7) must be modified and explain how. We provide numerical tests as well as analytic arguments for both claims.

We begin by giving two arguments for the correctness of Eq. (6.6) in the CI case. First, it is straightforward to see that the contribution to $J_1(t)$ can be computed independently for each k_2 [198], with the problem in (k_1, t) space effectively corresponding to that of an ordinary 1D crystal. Thus, the derivation of Eq. (6.6) given in Ref. [24] goes through unchanged for the CI case. Second, we note that the expected result is obtained for the special case that the parameter of

interest is a spatially uniform but time-dependent vector potential $\mathbf{A}(t)$. Since a slow turning on of $\mathbf{A}(t)$ causes state $u_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}$ to evolve into $u_{\mathbf{k}+(e/hc)\mathbf{A}}e^{i\mathbf{k}\cdot\mathbf{r}}$, it follows that

$$\mathbf{P}_{[\mathbf{k}_0]}^{[\mathbf{A}]} = \mathbf{P}_{[\mathbf{k}_0]}^{[\mathbf{A}=0]} - \frac{e^2 C}{hc} \hat{\mathbf{z}} \times \mathbf{A}. \quad (6.8)$$

But a time varying vector potential generates an electric field $\mathbf{E} = (-1/c)d\mathbf{A}/dt$, so that $\mathbf{J} = (Ce^2/h)\hat{\mathbf{z}} \times \mathbf{E}$. The transverse conductivity σ_{xy} is thus quantized in units of e^2/h , expressing the fact that a CI is a realization of the integer quantum Hall effect [32].

We further confirm the validity of Eq. (6.6) by numerically testing our prediction on the Haldane model [32], a tight-binding model for spinless electrons on a honeycomb lattice at half filling with staggered site energies and complex second-neighbor hoppings chosen so that $C=1$. Using the notation of Ref. [32], we adopt parameters $t_1 = 1$, $t_2 = 1/3$, $\phi = \pi/4$, $\Delta = 2/3$ and the lattice vectors $\mathbf{a}_1 = a_0(\sqrt{3}\hat{x} + \hat{y})/2$ and $\mathbf{a}_2 = a_0\hat{y}$ (so that $a_1 = a_2 = a_0$). Furthermore, we modify the first-neighbor hopping $t_1 \rightarrow t_1(1 + \alpha)$ on the bonds parallel to $\mathbf{a}_1 + \mathbf{a}_2$ so as to break the threefold rotational symmetry and allow an adiabatic current to flow as α is varied. The compensating ionic charge is assumed to sit on the site with lower site energy.

We consider an infinite strip of the Haldane model N_1 cells wide and extending to $\pm\infty$ along y , as sketched in the inset of Fig. 6.2. States $\psi_{nk_2}(\mathbf{r})$ are labeled by k_2 , which remains a good quantum number, and an additional index $n = 1, \dots, 2N_1$. The dipole moment across the strip, per unit length, is

$$\mathcal{P}_1 = \frac{-e}{N_1 S} \int_0^1 dk_2 \sum_{n \in \mathcal{N}(k_2)} \langle \psi_{nk_2} | r_1 | \psi_{nk_2} \rangle, \quad (6.9)$$

where position vector \mathbf{r} is decomposed as $\mathbf{r} = r_1\mathbf{a}_1 + r_2\mathbf{a}_2$ and $\mathcal{N}(k_2)$ is the set of occupied states to be discussed shortly. In the limit of large N_1 , we associate

the integrated current that flows along \hat{x} in the interior of the strip during an adiabatic evolution from $\alpha = \alpha_i$ to $\alpha = \alpha_f$ with the corresponding change in \mathcal{P}_1 , since by continuity the charge must arrive at the surface. We then compare this with the change of P_1 evaluated using a single bulk unit cell via Eqs. (6.2)-(6.3) to validate the theory.

There is a subtlety, however. Neutrality implies that $\mathcal{N}(k_2)$ contains N_1 states, but which ones? The problem arises because a CI is topologically required to have chiral metallic edge states. Our ribbon of CI therefore has one band of edge states along its left (L) edge and one along its right (R) edge (see inset of Fig. 6.2). For any given α , let $k_2^\times(\alpha)$ be the value of k_2 at which L-edge and R-edge bands cross. A *thermalized* filling of the edge states would correspond to the thick black curve for case α_i in Fig. 6.2, where the N_1 lowest-energy states are occupied at each k_2 and $\epsilon_F = \epsilon(k_2^\times)$. Defining k_2^* to be the point at which the occupation switches between L and R edge states, we have $k_2^* = k_2^\times$ for the thermalized case.

In general $k_2^\times(\alpha)$ varies with α . However, k_2^* cannot change during an *adiabatic* evolution. Because we want to “measure” the polarization by the charge that accumulates at the surface, we specify that the adiabatic evolution is fast compared to the tunneling time between edge states but slow compared to all other processes, so that electrons cannot scatter between edges. Thus if we thermalize the system at α_i and then adiabatically carry the system from α_i to α_f , we arrive at the *adiabatic* filling illustrated by the thick red curve for case α_f in Fig. 6.2.

We thus expect that the change in polarization calculated from the right-hand side of Eq. (6.6) from the bulk bandstructure using Eqs. (6.1)-(6.3) should match that given by the change of Eq. (6.9) *only* if the adiabatic filling is maintained. We have confirmed this numerically for our modified Haldane model. The polarization as a function of α calculated using Eq. (6.9) and using the right-hand side of

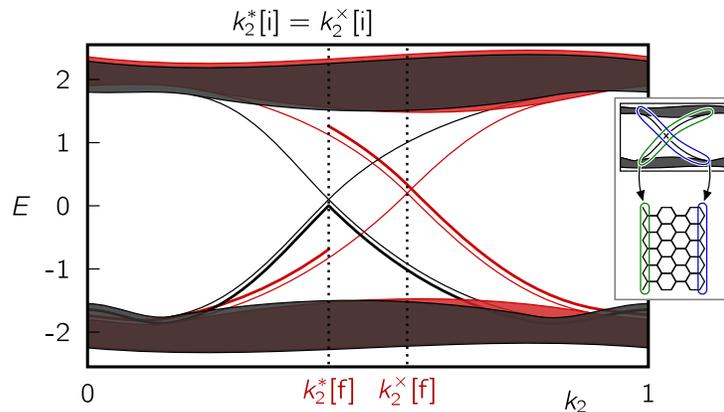


Figure 6.2: Sketch of a band structure of a finite ribbon of a Chern insulator. Solid regions indicate projected bulk bands; thin solid lines are edge states. Black and red correspond to $\alpha = \alpha_i$ and $\alpha = \alpha_f$ respectively; corresponding values of k_2^x are indicated. Thick lines indicate filling of edge states as dictated by k_2^* , chosen to illustrate system thermalized at α_i and then carried adiabatically to α_f . Inset: edge states associated with left (green) and right (blue) surfaces.

Eq. (6.6) is indicated in Fig. 6.1(b) with black and blue dashed lines respectively¹. Eqs. (6.2)-(6.3) were evaluated on a 300×300 k-point mesh. Eq. (6.9) was calculated using five values of $N_1 \in [25, 70]$ and then extrapolating to infinity, while the k_2 integral was discretized with 5000 k-points. While there is a vertical offset between these curves that depends on the choice of \mathbf{k}_0 in Eq. (6.6), the differences ΔP_1 between different α are correct at the level of 10^{-5} . On the other hand, the results obtained with the thermalized filling in Eq. (6.9), shown by the solid line in Fig. 6.1(b), are drastically different. These results confirm that the appropriate comparison is with the adiabatic filling, and provide numerical confirmation that Eq. (6.6) is indeed satisfied even in a CI.

We now turn to Eq. (6.7). A naive generalization to the CI case might be that $\sigma = \mathbf{P}_{[\mathbf{k}_0]} \cdot \hat{\mathbf{b}}_1$ (modulo e/a_2), but this cannot be correct. First, the left-hand side should be independent of \mathbf{k}_0 , but the right-hand side is not. Second, the usual proof for ordinary insulators of the connection between surface charge and bulk

¹Note that $P_1 \neq 0$ even for $\alpha = 0$, since $\Delta \neq 0$ allows for an asymmetric population of the edge states.

polarization assumes that the surface is insulating, with the Fermi level lying in a gap common to both the bulk and surface [198]. When chiral edge states are present, the surfaces cannot be insulating, so the usual conditions are violated.

To show how Eq. (6.7) can be corrected for the case of a CI, let us again consider our Haldane-model ribbon at some fixed α . Its surface charge σ can be calculated from $\sigma = \mathcal{P} \cdot \hat{\mathbf{b}}_1 = (S/a_2)\mathcal{P}_1$ with \mathcal{P}_1 evaluated using Eq. (6.9), but its value will depend on the choice of the k_2^* at which the occupation of the edge state has its discontinuity, so that

$$\sigma^{[k_2^*]} = \frac{-e}{N_1 a_2} \int_0^1 dk_2 \sum_{n \in \mathcal{N}} \langle \psi_{nk_2} | r_1 | \psi_{nk_2} \rangle, \quad (6.10)$$

where \mathcal{N} is the set of N_1 occupied states at k_2 given the specified k_2^* (i.e., the choice whether the L or R edge state is included in \mathcal{N} flips as k_2 passes through k_2^*).

Since the surface charge theorem of Eq. (6.7) for ordinary insulators was demonstrated via the Wannier representation [198], we take the same approach here. However, well-localized bulk WFs do not exist in a CI [195], so we focus instead on “hybrid Wannier functions” (HWFs) [200] in which the Fourier transform from Bloch functions is carried out in the r_1 direction only. Thus k_2 remains a good quantum number and the HWF

$$W_{k_2}(r_1, r_2) = \sqrt{N_1} \int_0^1 dk_1 \Psi_{k_1 k_2}(r_1, r_2) \quad (6.11)$$

is well localized only in the \mathbf{a}_1 direction. Using these we can represent the polarization

$$P_1^{[\kappa_2]} = \frac{-e}{S} \int_{\kappa_2}^{\kappa_2+1} dk_2 \rho_{k_2}^{[\kappa_2]} \quad (6.12)$$

in terms of the HWF center $\rho_{k_2}^{[\kappa_2]} = \langle W_{k_2} | r_1 | W_{k_2} \rangle$. We require ρ to be a continuous

function of $k_2 \in [\kappa_2, \kappa_2 + 1]$ so as to guarantee a result that is equivalent to Eqs. (6.1)-(6.3).

To make the connection between Eqs. (6.10) and (6.12), we recast the former by constructing Wannier-like functions along the \mathbf{a}_1 direction for the finite-width strip, starting from the $N_1 \times N_1$ matrix $\mathcal{R}_{mn,k_2}^{[k_2^*]} = \langle \psi_{mk_2} | r_1 | \psi_{nk_2} \rangle$, where $m, n \in \mathcal{N}$ as specified by k_2^* . The N_1 eigenvectors of $\mathcal{R}_{k_2}^{[k_2^*]}$ correspond to states that are Bloch-like along r_2 but localized along r_1 , which we refer to as ribbon HWFs, and the eigenvalues $\varrho_{jk_2}^{[k_2^*]}$ locate their centers of charge. Using the basis-independence of the trace, Eq. (6.10) can now be rewritten as

$$\sigma^{[k_2^*]} = \frac{-e}{N_1 a_2} \int_0^1 dk_2 \sum_j \varrho_{jk_2}^{[k_2^*]}. \quad (6.13)$$

The similarity between Eqs. (6.12) and (6.13) suggests that these can be connected. Since k_2 is a good quantum number, each k_2 can be treated independently. For each k_2 we can compare the infinite (bulk) 1D system described by Eq. (6.12) with the finite (ribbon) 1D system described by Eq. (6.13). The essential observation is that, in the limit of large N_1 , the HWF centers ϱ_{jk_2} deep inside the ribbon converge to the bulk ρ_{k_2} , modulo an integer [198]. This is illustrated in Fig. (6.3), where both sets of HWF centers are plotted as a function of k_2 for a ribbon of width $N_1 = 6$. Furthermore, the fact that the occupation of edge states switches between L and R edge at k_2^* is reflected in the discontinuity of ribbon HWF centers ϱ_{jk_2} at k_2^* . On the other hand, the bulk HWF centers ρ_{k_2} are chosen to be continuous across k_2^* . We can account for this discrepancy either by including a correction term proportional to $(k_2^* - \kappa_2)$,

$$\sigma^{[k_2^*]} = \frac{1}{a_2} \left[SP_1^{[\kappa_2]} + eC(k_2^* - \kappa_2) \right] \pmod{e/a_2}, \quad (6.14)$$

or by realizing that by the virtue of Eq. (6.5) this is equivalent to shifting the

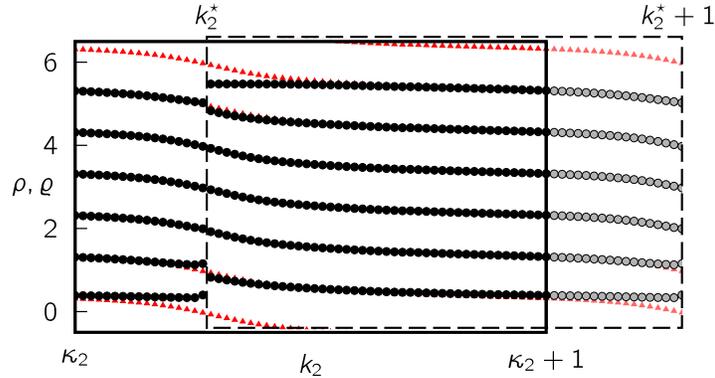


Figure 6.3: Black dots show ribbon HWF centers $\varrho_{jk_2}^{[k_2^*]}$ and red triangles bulk HWF centers $\rho_{k_2}^{[k_2^*]}$ and its periodic images as a function of k_2 . Dashed frame corresponds to choice of origin at discontinuity in $\varrho_{jk_2}^{[k_2^*]}$, k_2^* .

reciprocal space origin to k_2^* ,

$$\sigma^{[k_2^*]} = \frac{S}{a_2} P_1^{[k_2^*]} \pmod{e/a_2}, \quad (6.15)$$

as can be seen from the dashed frame in Fig. 6.3. Eq. (6.14) or (6.15) is the appropriate generalization of the surface charge theorem, Eq. (6.7), to the case of a CI, and should be correct in large N_1 limit for *both* thermalized and adiabatic fillings as long as the appropriate k_2^* is used.

We have also tested the correctness of this formula using our numerical calculations on the modified Haldane model. Recall that the solid curve in Fig. 6.1(b) represents the surface charge as computed from Eq. (6.9) for the thermalized case. For each α , we first locate k_2^\times using 1000 k-points on a ribbon of width $N_1 = 70$ and evaluate Eq. (6.15) with $k_2^* = k_2^\times$ using Eqs. (6.2)-(6.3) on a 250×250 k-point mesh. The resulting values are plotted as blue dots in Fig. 6.1(b). The agreement is excellent.

6.3 Conclusion

In summary, we have generalized the Berry-phase concept of polarization to the case of a Chern insulator. The integrated current flow during adiabatic evolution is given by Eq. (6.6), where the reciprocal-space cell must be the same in both terms on the right-hand side. The surface charge at an edge of a bounded sample is given by Eq. (6.15), where k_2^* specifies the wavevector at which the occupation discontinuity occurs in the chiral edge state. These results may be of use in understanding the physical properties of these topological insulators, and perhaps in searching for experimental realizations.

Chapter 7

Chern-Simons orbital magnetoelectric coupling in generic insulators

In this chapter we present a Wannier-based method to calculate the Chern-Simons orbital magnetoelectric coupling in the framework of first-principles density-functional theory. In view of recent developments in connection with strong \mathbb{Z}_2 topological insulators, we anticipate that the Chern-Simons contribution to the magnetoelectric coupling could, in special cases, be as large or larger than the total magnetoelectric coupling in known magnetoelectrics like Cr_2O_3 . The results of our calculations for the ordinary magnetoelectrics Cr_2O_3 , BiFeO_3 and GdAlO_3 confirm that the Chern-Simons contribution is quite small in these cases. On the other hand, we show that if the spatial inversion and time-reversal symmetries of the \mathbb{Z}_2 topological insulator Bi_2Se_3 are broken by hand, large induced changes appear in the Chern-Simons magnetoelectric coupling.

7.1 Introduction

In recent years there has been a significant revival of interest in magnetoelectric effects in solids, as surveyed in several reviews [201, 202, 203, 204]. Potential applications of these materials have long been discussed [205, 206] in areas ranging

from the optical manipulation and frequency conversion to magnetoelectric memories. Of the various quantities that can be discussed, the linear magnetoelectric coupling tensor α_{ij} is clearly of primary interest, as it quantifies the leading-order term in the coupling at small fields. We define it as

$$\alpha_{ij} = \left(\frac{\partial \mathcal{P}_i}{\partial B_j} \right)_{\mathcal{E}} = \left(\frac{\partial M_j}{\partial \mathcal{E}_i} \right)_B, \quad (7.1)$$

where \mathcal{P}_i is the electric polarization induced by the magnetic field B_j , or equivalently, M_j is the magnetization induced by the electric field \mathcal{E}_i . We use SI units (see Sec. 7.2.1) and the derivatives are to be evaluated at zero electric and magnetic field. In the special case that the induced response (\mathcal{P} or \mathbf{M}) remains parallel to the applied field (\mathbf{B} or \mathcal{E}), the tensor α is purely diagonal with equal diagonal elements, and its strength can be measured by a dimensionless scalar parameter θ defined via

$$\alpha_{ij}^{\text{iso}} = \frac{\theta e^2}{2\pi\hbar} \delta_{ij}. \quad (7.2)$$

More generally, depending on the magnetic point group of the crystal, α_{ij} can have distinct diagonal components as well as non-zero off-diagonal ones.

The linear magnetoelectric response α_{ij} can be decomposed into two contributions coming from purely electronic and from ionic responses respectively. The former is defined as the magnetoelectric response that occurs when atoms are not allowed to displace in response to the applied field, while the latter is defined as the remaining lattice-mediated response. One generally expects ionic effects to dominate over electronic responses, as for example was shown recently in Ref. [207, 208] for the case of Cr_2O_3 . Moreover, each of these components can be decomposed further into spin and orbital parts, since the magnetization induced by the electric field can be decomposed in that way. Here one would naively expect that the spin contribution will dominate with respect to the orbital one,

since orbital moments are usually strongly quenched by crystal fields. Mostly for this reason, realistic theoretical calculations of magnetoelectric coupling have been developed [207, 209, 208] only for the spin component.

As shown in Refs. [33] and [34] using two complementary approaches, the orbital magnetoelectric polarizability (OMP), defined as the contribution of orbital currents to the magnetoelectric coupling α_{ij} , can be written as the sum of three gauge-invariant contributions. One of these, first discussed by Qi *et al.* [28] and Essin *et al.* [29], is the Chern-Simons term (CSOMP). Since this contribution is purely isotropic it contributes only to θ , as in Eq. (7.2). In this chapter we will focus mostly on the CSOMP component of α_{ij} . From an implementation viewpoint, the CSOMP component is quite different from the other two components of the OMP: it can be calculated from a knowledge of the ground-state electron wavefunctions alone, but only after careful attention is given to the need to choose a smooth gauge in discretized k -space.

One of the motivations for the current work is the possibility of finding a material whose CSOMP component of the linear magnetoelectric tensor will be large compared to the total coupling in known magnetoelectric materials. As elaborated in more detail in Sec. 7.2, the basis for this possibility arises from the before-mentioned theoretical developments [210] and the experimental verification of the existence of \mathbb{Z}_2 topological insulators such as $\text{Bi}_{1-x}\text{Sb}_x$, Bi_2Se_3 , Bi_2Te_3 and Sb_2Te_3 [45, 46, 47]. Roughly speaking, we seek a material that is similar to a \mathbb{Z}_2 topological insulator, but having broken inversion and time-reversal symmetries. In order to take the first steps toward searching for such materials, we have set out to calculate the CSOMP component of the magnetoelectric tensor in several compounds of interest using density-functional theory.

This chapter is organized as follows. In Sec. 7.2 we provide theoretical background by reviewing the previously-derived [33, 34] expression for the $\boldsymbol{\alpha}$ tensor,

and by discussing the connection between bulk and surface properties in a way that is analogous to the theory of surface charge and bulk electric polarization. We also review the connection to \mathbb{Z}_2 topological insulators and make some general comments about symmetry. In Sec. 7.3 we discuss the gauge-fixing issues that arise when discretizing the CSOMP expression on a k -point mesh, and show how these can be resolved using Wannier-based methods. By this route, we arrive at an explicit expression for the CSOMP in terms of position matrix elements between Wannier functions. We evaluate this expression in the density-functional context for several materials of interest in Sec. 7.4. Finally, we summarize and give an outlook in Sec. 7.5.

7.2 Background and motivation

In this section we briefly summarize previous work from Refs. [33] and [34] on the orbital magnetoelectric coupling (OMP), describe relationships between bulk and surface properties, discuss motivations for this work based on the discovery of strong \mathbb{Z}_2 topological insulators, and present a brief symmetry analysis.

7.2.1 Units and conventions

In this chapter we use SI units and define α according to Eq. (7.1) using independent field variables \mathcal{E} and B . It follows that α has the same units as the vacuum admittance $1/c\mu_0$ [211]. While this is convenient from the point of view of first-principles theory, where B is fixed to zero in practice, the more conventional definition in the literature is in terms of fixed \mathcal{E} and H fields, in which case one has

$$\alpha_{ij}^{\text{EH}} = \left(\frac{\partial \mathcal{P}_i}{\partial H_j} \right)_{\mathcal{E}} = \mu_0 \left(\frac{\partial M_j}{\partial \mathcal{E}_i} \right)_H \quad (7.3)$$

and α^{EH} has units of inverse velocity [212]. In the typical case that the magnetic susceptibility of the material is negligible, these are related by $\alpha^{\text{EH}} = \alpha\mu_0$, and one can define a reduced (dimensionless) quantity $\alpha_r = c\mu_0\alpha = c\alpha^{\text{EH}}$ [211]. Defined in this way, α_r is numerically equal to the value of the magnetoelectric coupling in Gaussian units using the conventions of Rivera [212], which in turn corresponds to the notation “g.u.” (“Gaussian units”) in some recent papers [207, 209]. Furthermore, using the notation of Eq. (7.2) for the isotropic magnetoelectric coupling, it follows that the diagonal component of α_r is just θ/π times the fine structure constant (which is $e^2c\mu_0/2h$ in SI units).

7.2.2 Theory of orbital magnetoelectric coupling

The purely electronic orbital magnetoelectric coupling α_{ij} can be written in terms of three gauge-invariant contributions

$$\alpha_{ij} = \alpha_{ij}^{\text{CS}} + \tilde{\alpha}_{ij}^{\text{LC}} + \tilde{\alpha}_{ij}^{\text{IC}}, \quad (7.4)$$

where $\alpha_{ij}^{\text{CS}} = \delta_{ij}\alpha^{\text{CS}}$ is the above-mentioned (isotropic) CSOMP, while $\tilde{\alpha}_{ij}^{\text{LC}}$ and $\tilde{\alpha}_{ij}^{\text{IC}}$ are two additional contributions. The isotropic part of the OMP tensor has contributions from the two $\tilde{\alpha}$ terms as well as from the CSOMP term. The three contributions to the OMP can compactly be expressed as

$$\alpha^{\text{CS}} = \eta \frac{e}{2} \int d^3k \epsilon_{ijk} \text{tr} \left[\mathcal{A}_i \partial_j \mathcal{A}_k - \frac{2i}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k \right], \quad (7.5)$$

$$\tilde{\alpha}_{ij}^{\text{LC}} = \eta \epsilon_{jkl} \text{Im} \int d^3k \langle \tilde{\partial}_k u_{n\mathbf{k}} | (\partial_l H_{\mathbf{k}}) | \tilde{D}_i u_{n\mathbf{k}} \rangle, \quad (7.6)$$

$$\tilde{\alpha}_{ij}^{\text{IC}} = \eta \epsilon_{jkl} \text{Im} \int d^3k \langle \tilde{\partial}_k u_{n\mathbf{k}} | \tilde{D}_i u_{m\mathbf{k}} \rangle \langle u_{m\mathbf{k}} | (\partial_l H_{\mathbf{k}}) | u_{n\mathbf{k}} \rangle, \quad (7.7)$$

where the notations are defined as follows. An implied sum notation applies to repeated Cartesian ($ijkl$) and band (mn) indices, corresponding to a trace over

occupied bands in the latter case (written explicitly as ‘tr’). A common prefactor $\eta = -e/\hbar(2\pi)^3$ appears in each equation, with $e > 0$ being the magnitude of the electron charge. The Berry connection

$$\mathcal{A}_{mn\mathbf{k}j} = \langle u_{n\mathbf{k}} | i\partial_j | u_{m\mathbf{k}} \rangle \quad (7.8)$$

is defined in terms of the cell-periodic Bloch functions

$$|u_{n\mathbf{k}}\rangle = e^{-i\mathbf{k}\cdot\mathbf{r}} |\psi_{n\mathbf{k}}\rangle, \quad (7.9)$$

which are the eigenvectors of $H_{\mathbf{k}} = e^{-i\mathbf{k}\cdot\mathbf{r}} \mathcal{H} e^{i\mathbf{k}\cdot\mathbf{r}}$, where \mathcal{H} is the bulk periodic Hamiltonian of the crystal at zero electric and magnetic field. ∂_j and D_j are the partial derivatives with respect to the j -th component of the wavevector \mathbf{k} and the electric field $\boldsymbol{\mathcal{E}}$ respectively. Finally, the tilde indicates a covariant derivative, $\tilde{\partial}_j = Q_{\mathbf{k}} \partial_j$ and $\tilde{D}_j = Q_{\mathbf{k}} D_j$, where $Q_{\mathbf{k}} = 1 - |u_{n\mathbf{k}}\rangle \langle u_{n\mathbf{k}}|$ (sum implied over n). Additional screening contributions to $\tilde{\alpha}_{ij}^{\text{LC}}$ and $\tilde{\alpha}_{ij}^{\text{IC}}$ that occur in the context of self-consistent field calculations, not given here, can be found in Ref. [34].

As in the case of electronic polarization, one needs to be careful about relating the above bulk expressions to experimentally measurable physical quantities, since arbitrary surface modifications can contribute to the effective measurable OMP. The relationship between the OMP and experimentally measurable responses are explained in more detail in the next section.

7.2.3 Relation between bulk and surface properties

In order to discuss the relationship between bulk and surface quantities in connection with the OMP, it is instructive first to review the corresponding connections in the theory of electric polarization.

Electric polarization and surface charge

We first review the relationship between the bulk electric polarization, as obtained from the crystal bandstructure according to the Berry-phase theory [24, 198], and a measurable quantity which is the macroscopic dipole moment of a finite sample cut from this crystal. Given the set of valence Bloch wavefunctions $|\psi_{n\mathbf{k}}\rangle$ of an insulating crystal, one can readily calculate the electronic contribution to the polarization as the integral

$$\mathcal{P}_i = -\frac{e}{(2\pi)^3} \sum_n \int d^3k \langle u_{n\mathbf{k}} | i\partial_{k_i} | u_{n\mathbf{k}} \rangle \quad (7.10)$$

over the Brillouin zone (BZ). Gauge changes ($|u_{n\mathbf{k}}\rangle \rightarrow e^{-i\beta(\mathbf{k})}|u_{n\mathbf{k}}\rangle$) can change the value of this integral only by $\mathbf{R}e/\Omega$, where \mathbf{R} is a lattice vector and Ω is the unit cell volume. The value of this integral is therefore only well-defined modulo $\mathbf{R}e/\Omega$. In what follows we assume that a definite choice of gauge has been made so that a definite value of \mathcal{P} has been established. We now analyze how, and under what circumstances, one can relate this \mathcal{P} to the (experimentally measurable) dipole moment \mathbf{d} of an arbitrarily faceted finite sample of this crystal.

At each local region on the surface of this finite sample, assuming a perfect surface preparation (defect-free with ideal periodicity), we can relate \mathcal{P} to the surface charge density σ at that same point via [198]

$$\sigma = \left(\mathcal{P} + \frac{e}{\Omega} \mathbf{R} \right) \cdot \hat{n} + \Delta. \quad (7.11)$$

Here \hat{n} is the surface normal unit vector, \mathbf{R} is a lattice vector, and Δ is an additional contribution present only for metallic surfaces. The term involving \mathbf{R} , which corresponds to an integer number of electrons per surface unit cell, is required because, for a given surface \hat{n} , it may be possible to prepare the surface

in different ways (e.g., by adding or subtracting a layer of ions, or by filling or emptying a surface band) such that the surface charge per cell changes by a quantum. Thus, \mathbf{R} is in general a surface-dependent quantity in Eq. (7.11). If the surface patch under consideration is not insulating, then Δ is a term which measures the contribution of the partially occupied surface bands to the surface charge, and is proportional to the area fraction of occupied band in k space. (In the case of an insulator with non-zero first Chern number, this fraction has to be calculated with special care [4], but we shall not consider this case in what follows.)

Now, let us consider the special case that all surfaces are insulating ($\Delta = 0$) and that the surface charges of *all* surface patches are consistent with a *single* vector value of \mathbf{R} (“global consistency”). Under these circumstances, the macroscopic dipole moment \mathbf{d} of the crystallite is given by

$$\mathbf{d} = \mathcal{V} \left(\mathcal{P} + \frac{e}{\Omega} \mathbf{R} \right), \quad (7.12)$$

which can be obtained trivially by integrating Eq. (7.11). Here \mathcal{V} is the volume of entire finite sample. As could be anticipated, \mathbf{d}/\mathcal{V} has a component depending only on the bulk wavefunctions and our gauge choice, and an additional component $e\mathbf{R}/\Omega$ reflecting the preparation of the surfaces.

OMP and surface anomalous Hall conductivity

We now discuss a corresponding set of relationships between the bulk-calculated OMP and the surface anomalous Hall conductivity.

Using Eqs. (7.5), (7.6), and (7.7) one can calculate the tensor $\boldsymbol{\alpha}$ from the knowledge of bulk Hamiltonian of an insulating crystal. Analogously as in the

case of polarization, one can again show that a gauge change¹ must either leave $\boldsymbol{\alpha}$ invariant or change it by a quantum $m(e^2/h)\mathbf{I}$, where m is an integer and \mathbf{I} is the unit matrix. More precisely, this gauge transformation will only affect the CSOMP component $\boldsymbol{\alpha}^{\text{CS}}$ of the OMP, since the other two contributions $\tilde{\boldsymbol{\alpha}}^{\text{LC}}$ and $\tilde{\boldsymbol{\alpha}}^{\text{IC}}$ are fully gauge-invariant (see Ref. [34] for details).

We now imagine cutting a finite crystallite from this infinite crystal, and we wish to relate $\boldsymbol{\alpha}$ to its physically observable linear magnetoelectric coupling $\boldsymbol{\beta}$, defined for a finite sample by

$$\beta_{ij} = \frac{\partial d_i}{\partial B_j} = \frac{\partial \mu_j}{\partial \mathcal{E}_i}, \quad (7.13)$$

where d_i is the dipole moment of the finite sample and μ_j is its magnetic dipole moment. We want to discuss this relationship in a way that is analogous to that between the bulk \mathcal{P} and sample dipole moment \mathbf{d} in Sec. 7.2.3.

As follows from Eq. (7.1), the application of an electric field \mathcal{E}_j to the insulating crystal induces the magnetization

$$M_k = \alpha_{jk}\mathcal{E}_j, \quad (7.14)$$

where $\boldsymbol{\alpha}$ is given by Eq. (7.4) and is only determined modulo the quantum $m(e^2/h)\mathbf{I}$. Having a homogeneous M_k inside the sample and $M_k = 0$ outside is equivalent to having a surface current K_i equal to

$$K_i = \epsilon_{ikl}M_k n_l, \quad (7.15)$$

where n_l is the surface unit normal. By eliminating M_k from these equations,

¹Here we refer to a multiband gauge transformation having the form of Eq. (7.18), since $\boldsymbol{\alpha}$ can be shown to be fully invariant under a single-band phase twist.

we see that having a magnetoelectric tensor $\boldsymbol{\alpha}$ is equivalent to having a surface anomalous Hall conductivity $\sigma_{ij}^{\text{AH}} = \epsilon_{ikl}\alpha_{jk}n_l$. If the surface patch in question is insulating, then its anomalous Hall conductivity should just be given, modulo $m(e^2/h)\mathbf{I}$, by this equation. If instead the surface patch is metallic, then an additional surface contribution Δ_{ij} should be present, leading to the relation

$$\sigma_{ij}^{\text{AH}} = \epsilon_{ikl} \left(\alpha_{jk} + m \frac{e^2}{h} \delta_{jk} \right) n_l + \Delta_{ij}. \quad (7.16)$$

This equation is in precise analogy to Eq. (7.11) relating the polarization to the surface charge. Here Δ_{ij} may in general contain dissipative contributions, but in the dirty limit it will be dominated by the intrinsic surface contribution that can be calculated as a 2D BZ integral of the Berry curvature of the occupied surface states [26]. The integer quantum m appearing in Eq. (7.16) corresponds to the theoretical possibility that the surface preparation can be changed in such a way that a surface band having a nonzero Chern number may become occupied. For example, this could be done in principle by constructing a 2D quantum anomalous Hall layer (as described, e.g., by the Haldane model [32]), straining it to be commensurate with the surface, and adiabatically turning on hopping matrix elements to “stitch it” onto the surface.

In the special case that all surface patches are insulating ($\Delta_{ij} = 0$), and all surface patches have an anomalous Hall conductivity given by Eq. (7.16) with the *same* value of m (“global consistency”), we can relate the experimentally measurable magnetoelectric response $\boldsymbol{\beta}$ of the finite crystallite to the bulk-calculated $\boldsymbol{\alpha}$ via

$$\boldsymbol{\beta} = \mathcal{V} \left(\boldsymbol{\alpha} + m \frac{e^2}{h} \mathbf{I} \right), \quad (7.17)$$

which follows by integrating Eq. (7.16) over all surfaces. This equation is in close analogy to Eq. (7.12) for the case of electric polarization. In particular, we see

that β/\mathcal{V} has a component α depending only on the bulk wavefunctions and our gauge choice, and an additional component that is an integer multiple of $(e^2/h)\mathbf{I}$, reflecting the preparation of the surfaces.

As will be discussed in the next section, time-reversal symmetry imposes additional constraints on α , and some care is needed in the interpretation of Eq. (7.17) for the case of \mathbb{Z}_2 topological insulators.

7.2.4 Motivation and relationship to strong \mathbb{Z}_2 topological insulators

In this Section, we give arguments to motivate our hope that in certain materials the CSOMP might be on the order of, or even much larger than, the total magnetoelectric coupling in typical known magnetoelectric materials. For simplicity, we focus henceforth only on the CSOMP part of the total OMP response, even though there are additional contributions coming from $\tilde{\alpha}^{\text{LC}}$ and $\tilde{\alpha}^{\text{IC}}$. Thus, from now on, the quantity θ measures the strength of the CSOMP through the relation $\alpha^{\text{CS}} = \theta e^2/2\pi h$.

Time-reversal symmetry constraints on θ

Let us analyze the allowed values of θ for an infinite bulk insulating system that respects time-reversal (T) symmetry. Since T flips the sign of the magnetic field, it will also reverse the sign of θ . As mentioned earlier in Sec. 7.2.3, however, the value of θ can be changed by 2π under a gauge transformation. Therefore one concludes [28, 29] that the allowed values of θ consistent with T symmetry are 0 and π (each modulo 2π), and that these two cases provide a topological classification of all T -invariant insulators. Indeed, this classification has been shown [28, 29] to be identical to the one based on the \mathbb{Z}_2 index, with \mathbb{Z}_2 -odd or “strong topological” insulators having $\theta = \pi$, while \mathbb{Z}_2 -even or “normal” insulators

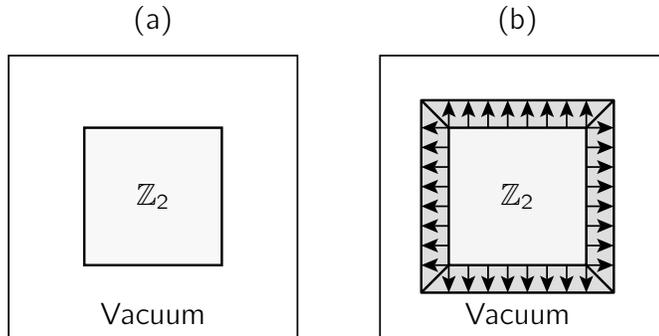


Figure 7.1: Identical samples cut from a strong \mathbb{Z}_2 topological insulator, but with two different surface preparations. (a) Time-reversal symmetry is preserved at vacuum-terminated surfaces; the net magnetoelectric coupling of this sample is zero. (b) Time-reversal symmetry is broken at the surface as a result of exchange coupling to an insulating ferromagnetic adlayer; if this opens a gap in the surface-state spectrum, the entire sample will behave as if it has a magnetoelectric coupling of exactly $\theta = \pi$.

have $\theta = 0$, even though the \mathbb{Z}_2 index is most often introduced in a different context [36]. (Incidentally, $\tilde{\alpha}^{\text{LC}} = \tilde{\alpha}^{\text{IC}} = 0$ in both cases since these terms are fully gauge-independent, unlike the CSOMP term which can be changed by 2π .)

Consider now a finite sample of a normal (\mathbb{Z}_2 -even) T -symmetric insulator ($\theta = 0$ in the bulk) with insulating surfaces ($\Delta_{ij} = 0$) prepared in a way that the integer m is nonzero, and the same on every surface. From Eq. (7.17) we conclude that this sample will have a non-zero magnetoelectric response, β , proportional to m . Obviously a sample that has T symmetry both in the bulk and on the surface must have $\beta = 0$, and therefore we conclude that this system needs to have broken T reversal symmetry at the surface. As mentioned earlier, one could, at least formally, prepare such a surface by starting from the one that has $m = 0$ and then absorbing to each surface a layer of anomalous Hall insulator [32] with Chern index m . Such a procedure will keep the surfaces insulating but it will necessarily break the T -reversal symmetry.

Next we analyze the case of a strong \mathbb{Z}_2 topological insulator having $\theta = \pi$, or equivalently, $\alpha = \alpha^{\text{CS}} = (e^2/2h)\mathbf{I}$. We first consider a sample of such a system

that has T symmetry conserved at its surfaces, as in Fig. 7.1(a). Again, since the entire sample is T -symmetric, its experimentally measurable magnetoelectric coupling tensor $\boldsymbol{\beta}$ clearly has to vanish. Using Eq. (7.16) and the fact that m can take on only integer, and not half-integer, values, we conclude that the only way to make the response of the entire sample vanish is to have Δ_{ij} be non-zero. This requires that the surfaces of such a system must be metallic. Moreover, since the contribution Δ_{ij} of the metallic surface band to the surface anomalous Hall conductivity is just given by the Berry phase around the Fermi loop [26], the needed cancellation requires this Berry phase to be exactly $\pm\pi$. All this is in precise accord with the known properties of \mathbb{Z}_2 -odd insulators and their topologically protected surface states [36].

The Kramers degeneracy at the Dirac cone in the surface bandstructure can be removed by the application of a T -breaking perturbation to the surface. In principle, this could be accomplished, for example, by applying a local magnetic field to the surface or by interfacing the surface to an insulating magnetic overlayer. In the latter case, the interatomic exchange couplings provide a kind of effective magnetic field acting on the surface layer of the topological insulator. If the local Fermi level resides in the gap opened by field, then the surface becomes insulating. If the field can be consistently oriented (see Ref. [28]) on each patch of the surface, either along or opposite the direction of surface normal vector \mathbf{n} (as shown in Fig. 7.1(b)), then the entire surface becomes insulating. It is important that the field is applied consistently in the same direction with respect to \mathbf{n} , since conducting channels will otherwise appear at domain boundaries [36].

If all of these requirements are met, the surface contribution Δ_{ij} to $\boldsymbol{\beta}$ vanishes, so that $\boldsymbol{\beta} = \mathcal{V}\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}$ given only by bulk value of $\theta = \pi$ (assuming $m = 0$ for simplicity). Therefore such a sample of a strong \mathbb{Z}_2 topological insulator would behave as if the entire sample has exactly half a quantum of magnetoelectric

coupling ($\theta = \pi$), even though its bulk is time-reversal symmetric!

Prospects for large- θ materials

Recently surface-sensitive ARPES measurements have experimentally confirmed that several compounds [45, 46, 47], including $\text{Bi}_{1-x}\text{Sb}_x$, Bi_2Se_3 , Bi_2Te_3 and Sb_2Te_3 , do indeed behave as strong \mathbb{Z}_2 topological insulators. Therefore their bulk wavefunctions must be characterized by $\theta = \pi$. Up to now, the corresponding magnetoelectric response has not been measured experimentally, in part because of the difficulties in obtaining truly insulating behavior in the bulk, as well as the need to gap the surfaces by putting them in contact with magnetic overlayers as described earlier.

We believe that a more promising approach to observing a large CSOMP (i.e., θ comparable to π) is to consider an insulator that has neither T nor spatial inversion symmetry. In this case the \mathbb{Z}_2 classification does not apply, and the surface can be gapped without any need to apply a T -breaking perturbation. (A more precise statement of the symmetry considerations will be given in Sec. 7.2.5.) The sample can then display a bulk magnetoelectric coupling of the simple form $\boldsymbol{\beta} = \mathcal{V}\boldsymbol{\alpha}$. We note that an orbital magnetoelectric coupling of $\theta \simeq \pi$ (i.e., $\alpha_r \simeq 1/137$) would correspond to $\alpha^{\text{EH}} \simeq 24.3$ ps/m, a value that is significantly larger than the observed coupling in Cr_2O_3 , one of the best-studied magnetoelectric materials. For comparison, the reported experimental values for $\alpha_{\perp}^{\text{EH}}$ in Cr_2O_3 , which are presumably dominated by spin-lattice coupling, range between 0.7 and 1.6 ps/m at 4.2 K [213, 214].

Of course, in order to have a good chance of finding a material with a large θ , it may be advisable to look for materials with some of the same characteristics as the known \mathbb{Z}_2 -odd insulators, of which the most important is probably the presence of heavy atoms with strong spin-orbit coupling. We see no strong reason

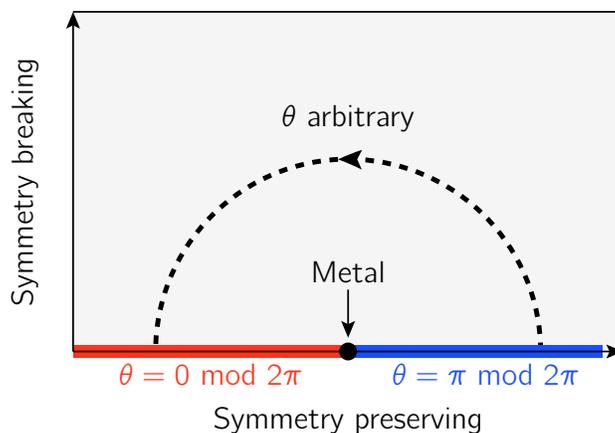


Figure 7.2: Schematic view of the allowable values of θ in different parts of the two-parameter space of some unspecified model Hamiltonian. Horizontal axis corresponds to the perturbation that preserves at least one of the symmetries that render θ to be 0 or π (see Sec. 7.2.5). Vertical axis parameterizes a perturbation that breaks those symmetries and allows θ to be arbitrary. See text for the details.

why such a search might not reveal a material having a large OMP in the above sense.

To illustrate the kind of a search we have in mind, consider some model Hamiltonian that depends on two parameters, one that preserves either the T or spatial inversion symmetry (or both), and another that breaks symmetry such that θ takes a generic value. The possible behavior of such a model is sketched in Fig. 7.2, where these two parameters are plotted along the horizontal and vertical axes respectively. The figure also indicates the generic value of θ in each region of parameter space. Along the horizontal axis, where the extra symmetry is present, three regions are indicated. The black dot indicates a point of gap closure forming the boundary between a normal T -symmetric insulator regime on the left ($\theta = 0$) and a strong \mathbb{Z}_2 topological insulator regime on the right ($\theta = \pi$). If the system is carried along the horizontal axis, θ must be either 0 or π except at the critical point, and it must therefore jump discontinuously when passing through this point of metallic behavior. On the other hand, if we now imagine passing from the \mathbb{Z}_2 -odd to the \mathbb{Z}_2 -even phase along the dashed curve

in Fig. 7.2, θ can vary smoothly and continuously from π to 0 without any gap closure anywhere along the path. If we can identify a material lying near, but not at, the right end of this dashed path, it could be the kind of large- θ material we seek.

Thus, our ultimate goal is to use first-principles calculations to search for a large θ , not in a topological insulator, but in an “ordinary” (but presumably strongly spin-orbit coupled) insulating magnetic material. While our work has yet to result in the identification of a large- θ material of this kind, it represents a first step in the desired direction.

7.2.5 General symmetry considerations

Recall that θ is a pseudoscalar that changes sign under time-reversal and spatial-inversion symmetries (since \mathbf{B} changes sign under T while $\boldsymbol{\mathcal{E}}$ changes sign under inversion). On the other hand, θ is invariant under any translation or proper rotation of a crystal. Therefore if the magnetic point group of a crystal contains an element that involves T , possibly combined with a proper rotation, the value of θ is constrained to be 0 or π (modulo 2π) as discussed earlier. The same happens if the magnetic point group contains inversion symmetry or any other improper rotation.

All 32 of the 122 magnetic point groups that do not contain such symmetry elements, and which therefore allow for an arbitrary value of θ , are listed in Table 7.1. (The bold entries in the table are those magnetic groups for which the tensor $\boldsymbol{\alpha}$ must be isotropic, i.e., a constant times the identity matrix; the same magnetic groups were also analyzed in Ref. [211]). Clearly we can constrain our search for interesting materials to the cases listed in the Table.

Table 7.1: Magnetic point groups for which a generic non-zero CSOMP is allowed by symmetry. Notation follows Ref. [215]. Point groups in bold allow only for a purely isotropic magnetoelectric tensor.

1	$\bar{1}'$	2	m'	$2/m'$	222	$m'm'2$
$m'm'm'$	4	$\bar{4}'$	$4/m'$	3	$\bar{3}'$	6
$\bar{6}'$	$6/m'$	422	$4m'm'$	$\bar{4}'2m'$	$4/m'm'm'$	32
$3m'$	$\bar{3}'m'$	622	$6m'm'$	$\bar{6}'m'2$	$6/m'm'm'$	
23	$m'3$	432	$\bar{4}'3m'$	$m'3m'$		

7.3 Methods

In this section we present our methods for calculating the CSOMP in the framework of density-functional theory, and analyze in more detail its mathematical properties and the formal similarities to the formulas used to calculate electric polarization and anomalous Hall conductivity.

7.3.1 Review of Berry formalism

Assume we are given the Bloch wavefunctions $|\psi_{n\mathbf{k}}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{n\mathbf{k}}\rangle$ as a function of wavevector \mathbf{k} in the d -dimensional BZ ($d = 1, 2$, or 3) for an insulator having valence bands indexed by $n \in \{1, \dots, N\}$. We work with the cell-periodic Bloch functions $u_{n\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{n\mathbf{k}}(\mathbf{r})$ and allow them to be mixed at each k point by an arbitrary k -dependent unitary matrix

$$|u_{n\mathbf{k}}\rangle \rightarrow |u_{m\mathbf{k}}\rangle U_{mn\mathbf{k}} \quad (7.18)$$

(sum on m implied). After this gauge transformation the wavefunctions are no longer eigenfunctions of the Hamiltonian, but they span the same N -dimensional subset of the Hilbert space as the true eigenfunctions. For any given choice of

gauge, we define the Berry connection

$$\mathcal{A}_{mn\mathbf{k}j} = \langle u_{m\mathbf{k}} | i \frac{\partial}{\partial k_j} | u_{n\mathbf{k}} \rangle, \quad (7.19)$$

which is a k -dependent $N \times N \times d$ matrix that measures, at each k point, the infinitesimal phase difference between the m -th and n -th wavefunctions associated with neighboring points along Cartesian direction j in k space. This object was already briefly introduced in Eq. (7.8).

In the context of electronic structure calculations, we can now list three material properties that can be evaluated knowing only the Berry connection: the electric polarization, the intrinsic anomalous Hall conductivity, and the CSOMP.

The electric polarization \mathcal{P} already appears in dimension $d = 1$ and it can be evaluated as an integral of the Berry connection over the one-dimensional BZ as [24]

$$\mathcal{P} = -\frac{e}{2\pi} \int_{\text{BZ}} dk \text{tr} \mathcal{A}_k, \quad (7.20)$$

where the trace is performed over the band indices of the Berry connection, as in Eq. (7.10). The integrand is also referred to as the Chern-Simons 1-form, and its integral over the BZ is well known to be defined only modulo 2π . Any periodic adiabatic evolution of the Hamiltonian $\mathcal{H}(\lambda)$ whose first Chern number in (k, λ) space is non-zero will change the integral above by a multiple of 2π [24].

Unlike one-dimensional systems, crystals in $d = 2$ can have an anomalous Hall conductivity. For a metal, the intrinsic contribution from a band crossing the Fermi level can be evaluated as a line integral [26, 27]

$$\sigma^{\text{AH}} = \frac{e^2}{h} \frac{1}{2\pi} \oint_{\text{FL}} d\mathbf{k} \cdot \mathcal{A}_{\mathbf{k}} \quad (7.21)$$

(a)



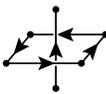
$P \sim \int \mathcal{A}$

(b)



$\sigma^{\text{AH}} \sim \oint \mathcal{A}$

(c)



$\theta \sim \int \mathcal{A}_i \partial_j \mathcal{A}_k$

Figure 7.3: Graphical interpretation of Eqs. (7.20) (a), (7.21) (b) and (7.22) (c) in the case of one occupied electron band and for cubic crystal symmetry, for simplicity. See text for more detail.

over the Fermi loop. Fully-filled deeper bands can also make a quantized contribution given by a similar integral, but around the entire BZ; this is the only contribution in the case of a quantum anomalous Hall insulator [32]. (In both cases, the gauge choice on the boundary of the region should be consistent with a continuous, but not necessarily k -periodic, gauge in its interior; alternatively, each expression can be converted to an area integral of a Berry curvature to resolve any uncertainty about branch choice. See Ref. [216] for more details.)

Finally, unlike one- or two-dimensional systems, three-dimensional systems can have an isotropic magnetoelectric coupling. The CSOMP can be evaluated in $d = 3$ as a BZ integration of a quantity involving the Berry connection:

$$\theta = -\frac{1}{4\pi} \int_{\text{BZ}} d^3k \epsilon_{ijk} \text{tr} \left[\mathcal{A}_i \partial_j \mathcal{A}_k - \frac{2i}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k \right]. \quad (7.22)$$

The integrand in this expression is known as the Chern-Simons 3-form, and its integral over the entire BZ is again ill-defined modulo 2π , since any periodic adiabatic evolution of the Hamiltonian $\mathcal{H}(\lambda)$ whose second Chern number in (\mathbf{k}, λ) space is non-zero will change θ by an integer multiple of 2π [28, 29].

The sketches in Fig. 7.3 compare the geometrical characters of the operations needed to evaluate Eqs. (7.20)-(7.22) in practice. We consider the case of one occupied electron band for simplicity. The polarization of Eq. (7.20) is calculated by a line integral; on a discrete k -mesh, the integral of the Berry connection \mathcal{A}

over each of line segment, as in Fig. 7.3(a), is converted to a discretized form (see Eq. (7.23)). Similarly, in two dimensions the anomalous Hall conductivity of Eq. (7.21) can be calculated as suggested in Fig. 7.3(b) by dividing the occupied part of the BZ into small square segments and then integrating \mathcal{A} around each square. (Equivalently, one can integrate \mathcal{A} along the Fermi loop [216].) In three dimensions, Fig. 7.3(c), Eq. (7.22) can be evaluated by dividing the BZ into small cubes. In each, one needs to multiply the integral of \mathcal{A} along one of the Cartesian directions (as in Eq. (7.20)) with the integral of Berry connection in the square orthogonal to that direction (as in Eq. (7.21)), followed by a symmetrization over the three Cartesian directions.

7.3.2 Numerical evaluation of θ

In electronic-structure calculations, the cell-periodic wavefunctions $|u_{n\mathbf{k}}\rangle$ are typically calculated on a uniform k -space grid with no special gauge choice; in general, one should assume that the phases have been randomly assigned. Nevertheless, it is straightforward to construct a gauge-invariant polarization formula that is immune to this kind of scrambling of the gauge [30]. In one dimension with k_j for $j \in \{1, \dots, M\}$ (where k_M is the periodic image of point k_1), the electronic polarization is calculated as

$$P = \frac{e}{2\pi} \text{Im} \ln \det [M_{k_1 k_2} M_{k_2 k_3} \dots M_{k_{M-1} k_M}] \quad (7.23)$$

where the overlap matrix $M_{kk'}$ is defined as

$$[M_{kk'}]_{mn} = \langle u_{mk} | u_{nk'} \rangle. \quad (7.24)$$

The reason for using Eq. (7.23) is that the determinant of the matrix

$$M_{k_1 k_2} M_{k_2 k_3} \dots M_{k_{M-1} k_M}$$

is gauge-invariant under any transformation in the form of Eq. (7.18). Additionally, the implementation of Eq. (7.23) is numerically stable even when there are band crossings. A similar gauge-invariant discretization can also be used to calculate the anomalous Hall conductivity σ^{AH} [216].

Unfortunately, except in the single-band (“Abelian”) case, we are unaware of any corresponding gauge-invariant discretized formula for the integral of the Chern-Simons 3-form. As a result, we have no prescription for computing the CSOMP that is exactly gauge-invariant for a given choice of k mesh. This is a serious problem. Unlike the calculation of the polarization, which is straightforward even if the gauge is randomly scrambled at each mesh point, the calculation of the CSOMP requires that we first identify a reasonably smooth gauge on the discrete mesh.

The problem of finding a smooth gauge in \mathbf{k} is essentially the same as that of finding well-localized Wannier functions. For this reason, we have adopted here the approach of first constructing a Wannier representation for the valence bands, and then using it to compute the CSOMP. In fact, starting from Eq. (7.22), we derive an expression that allows us to compute θ directly in the Wannier representation. Once we have well-localized Wannier functions, this guarantees smoothness of the gauge and avoids problems with band crossings. Admittedly, such a formula still depends on the gauge choice, meaning that different choices of Wannier functions will lead to slightly different results. However, this difference will vanish as one increases the density of the k -point mesh, since in the continuum limit the k -space expression for θ is gauge-invariant (modulo 2π). More precisely,

we expect the calculation of θ to converge once the inverse of the k -point mesh spacing becomes much larger than the spread of the Wannier functions.

Therefore, we adopt the strategy of calculating θ on k meshes of different density, and extrapolating θ to the limit of an infinitely dense mesh. Furthermore, we construct maximally-localized Wannier functions (MLWF) following Ref. [30], expecting this to give relatively rapid convergence as a function of the k mesh density.

Recall that the Wannier function associated with (generalized) band index n in unit cell \mathbf{R} is defined in terms of the rotated Bloch states (7.18) as

$$|\mathbf{R}n\rangle = \frac{\Omega}{(2\pi)^3} \int d^3k e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} |u_{m\mathbf{k}}\rangle U_{m\mathbf{k}}. \quad (7.25)$$

In the case of MLWFs, the $U_{m\mathbf{k}}$ are chosen in such a way that the total quadratic spread of the Wannier function is minimized [30]. (In practice the BZ integral is replaced by a summation over a uniform grid of k points.)

Using Eq. (7.25), one can relate the Berry-connection matrix $\mathcal{A}_{m\mathbf{n}k_j}$ in the smooth gauge to the Wannier matrix elements of the position operator through [30].

$$A_{m\mathbf{n}k_j} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \mathbf{0}m | r_j | \mathbf{R}n \rangle. \quad (7.26)$$

Replacing each occurrence of A_j in Eq. (7.22) with the above gives, after some algebra,

$$\begin{aligned} \theta = & \frac{1}{4\pi} \frac{(2\pi)^3}{\Omega} \epsilon_{ijk} \text{Im} \left[\frac{1}{3} \sum_{\mathbf{R}} \langle \mathbf{0}m | r_i | \mathbf{R}n \rangle \langle \mathbf{R}n | r_j | \mathbf{0}m \rangle R_k \right. \\ & \left. - \frac{2}{3} \sum_{\mathbf{R}\mathbf{P}} \langle \mathbf{0}l | r_i | \mathbf{R}m \rangle \langle \mathbf{R}m | r_j | \mathbf{P}n \rangle \langle \mathbf{P}n | r_k | \mathbf{0}l \rangle \right], \end{aligned} \quad (7.27)$$

where the sum is implied over band (lmn) and Cartesian (ijk) indices. (Even

though Eq. (7.22) and Eq. (7.27) are equivalent as a whole, they do not match term by term.)

To obtain a more symmetric form, we introduce a modified position-operator matrix element between WFs defined as

$$\langle \mathbf{R}m | \tilde{r}_i | \mathbf{P}n \rangle = \langle \mathbf{R}m | r_i | \mathbf{P}n \rangle (1 - \delta_{mn} \delta_{\mathbf{R}\mathbf{P}}) \quad (7.28)$$

and a notation for the Wannier center

$$\tau_{ni} = \langle \mathbf{0}n | r_i | \mathbf{0}n \rangle. \quad (7.29)$$

Then Eq. (7.27) becomes

$$\begin{aligned} \theta = & \frac{1}{4\pi} \frac{(2\pi)^3}{\Omega} \epsilon_{ijk} \times \quad (7.30) \\ & \text{Im} \left[\sum_{\mathbf{R}} \langle \mathbf{0}m | \tilde{r}_i | \mathbf{R}n \rangle \langle \mathbf{R}n | \tilde{r}_j | \mathbf{0}m \rangle (R_k + \tau_{nk} - \tau_{mk}) \right. \\ & \left. - \sum_{\mathbf{R}\mathbf{P}} \frac{2}{3} \langle \mathbf{0}l | \tilde{r}_i | \mathbf{R}m \rangle \langle \mathbf{R}m | \tilde{r}_j | \mathbf{P}n \rangle \langle \mathbf{P}n | \tilde{r}_k | \mathbf{0}l \rangle \right]. \quad (7.31) \end{aligned}$$

We find this form more convenient because it separates the contributions of diagonal and off-diagonal elements of position operators. (It is also manifestly invariant to the reassignment of a Wannier function to a neighboring cell. Furthermore, note that while Eqs. (7.22), (7.27), and (7.31) are all equivalent as a whole, the division of contributions between the first and second term is different in each case.) The validity of Eqs. (7.27) and (7.31) has been tested numerically by comparing with the evaluation of Eq. (7.22) for the case of a tight-binding model introduced in Ref. [34]. The evaluated expressions agreed to numerical accuracy after extrapolation to the infinitely dense mesh. These expressions can also be shown to be gauge-invariant by working directly within the Wannier representation.

7.3.3 Computational details

Calculations of the electronic ground state and of structural relaxations were performed using the QUANTUM-ESPRESSO package [143], and the WANNIER90 code [217] was used for constructing maximally localized Wannier functions. We used radial-grid discretized HGH [218] norm-conserving pseudopotentials. Calculations were performed in the noncollinear spin framework. Quantum-ESPRESSO incorporates the spin-orbit interaction at the level of the pseudopotentials, which is a good approximation since the relativistic effects arise predominantly from the core region. The pseudopotentials used for Cr, Fe and Gd contain semi-core states in the valence, while the ones for Al, Bi, Se and O do not. In all calculations we used the Perdew-Wang [219] LDA energy functional.

The self-consistent calculations on Cr_2O_3 were performed on a $4 \times 4 \times 4$ Monkhorst-Pack [80] grid in k space. Non-self-consistent calculations for the Wannier-function construction were performed on k -space grids containing the origin and ranging in size from $6 \times 6 \times 6$ to $12 \times 12 \times 12$. The plane-wave energy cutoff was chosen to be 150 Ry.

In the case of Bi_2Se_3 , the self-consistent calculations were performed on a $6 \times 6 \times 6$ grid with energy cutoff of 60 Ry, while the non-selfconsistent calculation was done on grids between $6 \times 6 \times 6$ and $11 \times 11 \times 11$.

The position-operator matrix elements $\langle \mathbf{0}m | r_j | \mathbf{R}n \rangle$ needed to evaluate Eq. (7.31) were calculated in k space by inverting the Fourier sum in Eq. (7.26) over the non-self-consistent k -point mesh, and then approximating the k derivative in Eq. (7.19) by finite differences on that mesh, as detailed in Ref. [27].

7.4 Results and discussion

7.4.1 Conventional magnetoelectrics

In this section we present the results of our first-principles electronic-structure calculations of θ . We begin with conventional magnetoelectrics, i.e., materials that are already experimentally known to have a non-zero magnetoelectric tensor. Some of these materials do not allow all diagonal components of the magnetoelectric tensor to be non-zero. We omit those materials from our analysis here, since we are interested in calculating the CSOMP part of the magnetoelectric coupling, which would vanish in such cases. We first present our results on Cr_2O_3 in some detail, and then briefly discuss our results for BiFeO_3 and GdAlO_3 .

Calculation of θ in Cr_2O_3

We first fully relax the structure in the $R\bar{3}c$ space group and obtain the Wyckoff position to be $x = 0.1575$ for Cr atoms (4c orbit) and $x = -0.0690$ for O (6e orbit). The length of the rhombohedral lattice vector is $a = 5.3221 \text{ \AA}$ while the rhombohedral angle is 53.01° . The Cr atoms have magnetic moments pointing along the rhombohedral axis as illustrated in Fig. 7.4(a) in an antiferromagnetic arrangement. The value of the magnetic moment is $2.0 \mu_B$ per Cr atom and the electronic gap is 1.3 eV, which agrees well with previous LDA+U calculations [220, 221] in the limit where the on-site Coulomb parameter U is set to zero.

Neglecting for a moment the magnetic spins on the Cr sites, the space-group generators are a three-fold rotation, a two-fold rotation, and an inversion symmetry as indicated in Fig. 7.4(a). Its point group is therefore $\bar{3}m$. If we now include the spins on the Cr atoms in the analysis, we find that the three-fold and two-fold rotations remain, while the inversion becomes a symmetry only when combined

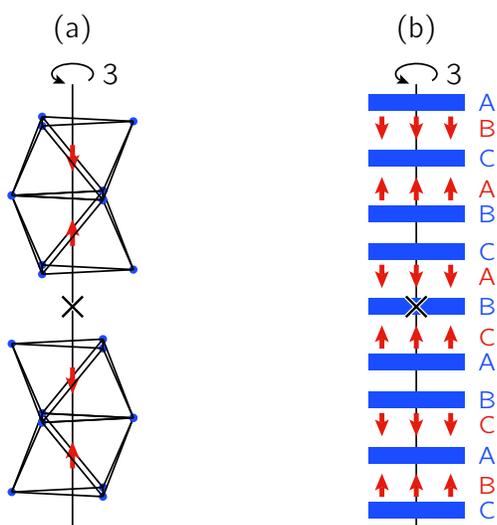


Figure 7.4: (a) Rhombohedral unit cell of Cr_2O_3 . Magnetic moments on Cr atoms are indicated by red arrows and oxygen octahedra are drawn around each Cr atom. (b) Schematic of hexagonal unit cell of Bi_2Se_3 with imposed local Zeeman field on Bi atoms. Induced magnetic moments are shown by red arrows. Thick blue lines indicate Se layers; letters (ABC) indicate stacking sequence of the hexagonal layers. In both panels, the vertical line indicates the 3-fold rhombohedral axis, and the cross designates a 2-fold rotation axis orthogonal to the plane of the figure (also a center of inversion coupled with time reversal).

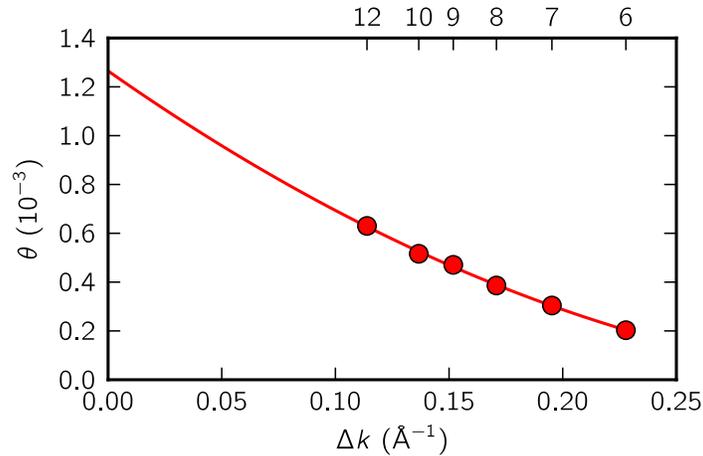


Figure 7.5: Calculated value of θ in Cr_2O_3 for varying densities of k -space grids, where Δk is the nearest-neighbor distance on the grid. Top axis specifies the size of the corresponding uniform Monkhorst-Pack grid. Line indicates a quadratic extrapolation of θ to the infinitely dense k mesh.

with time-reversal. Therefore the magnetic point group of Cr_2O_3 is $\bar{3}'m'$.² This magnetic point group allows θ to be different from 0 or π , as discussed in Sec. 7.2.5.

Figure 7.5 shows the calculated values of θ using Eq. (7.31) for Cr_2O_3 with k -space meshes of various densities. The line indicates the second-order polynomial extrapolation to an infinitely dense mesh. The extrapolated value of θ is 1.3×10^{-3} , which is a small fraction of the quantum of OMP $\theta = 2\pi$ and corresponds to $\alpha_{xx}^{\text{EH}} = \alpha_{yy}^{\text{EH}} = \alpha_{zz}^{\text{EH}} = 0.01$ ps/m. The positive sign of θ pertains to the pattern of Cr magnetic moments shown in Fig. 7.4(a); reversal of all magnetic moments would flip the sign of θ .

In order to compare this purely isotropic component of the magnetoelectric coupling with experimental values and other theoretical calculations of the full magnetoelectric response, which is not entirely isotropic, we somewhat arbitrarily define

$$\alpha^{\text{eff}} = \frac{|\alpha_{xx}| + |\alpha_{yy}| + |\alpha_{zz}|}{3}. \quad (7.32)$$

²Throughout the chapter, the notation for magnetic point groups follows Ref. [215].

The value of α^{eff} obtained from the results of Ref. [208] is 0.23 ps/m for the purely electronic part of the spin-mediated component. Therefore, our calculated CSOMP contribution in Cr_2O_3 amounts to only 4% of this electronic spin component. The ionic component of the spin response calculated by the same authors results in $\alpha^{\text{eff}} = 0.74$ ps/m, while the one calculated in Ref. [207] is about 2.6 times smaller, 0.29 ps/m. (In both of these calculations, α_{zz} is zero.) Finally, experimental measurements of the magnetoelectric tensor in Cr_2O_3 at 4.2 K vary between $\alpha^{\text{eff}} = 0.55$ ps/m and 1.17 ps/m (see Refs. [213] and [214] respectively).

Clearly, our computed CSOMP contribution for Cr_2O_3 is negligible, being two orders of magnitude smaller than the dominant lattice-mediated spin contribution. This is probably not surprising, since the spin-orbit coupling is relatively weak in this material. Given that it is weak, we can guess that that magnitude of the CSOMP should be linear in the strength of the spin-orbit interaction in Cr_2O_3 . Our calculations allow us to check this by varying the spin-orbit interaction strength λ_{SO} between 0 (no spin orbit) and 1 (full spin-orbit interaction). As shown in Fig. 7.6, if we calculate θ for various intermediate values of λ_{SO} , we see that the CSOMP does indeed depend roughly linearly on λ_{SO} .

Other conventional magnetoelectrics

We have also carried out calculations of θ in BiFeO_3 and GdAlO_3 , but with a smaller number of k -point grids than in the case of Cr_2O_3 . Therefore, our results are less accurate, but should still give a correct order-of-magnitude estimate of θ .

For BiFeO_3 we perform the calculation in the 10-atom antiferromagnetic unit cell (the long-wavelength spin spiral was suppressed). We obtain an electronic band gap of 0.95 eV with magnetic moments of $3.5 \mu_{\text{B}}$ on each Fe atom, and with a net magnetization of $0.1 \mu_{\text{B}}$ per 10-atom primitive unit cell due to the canting of the Fe magnetic moments. Extrapolating θ to an infinitely dense mesh using just

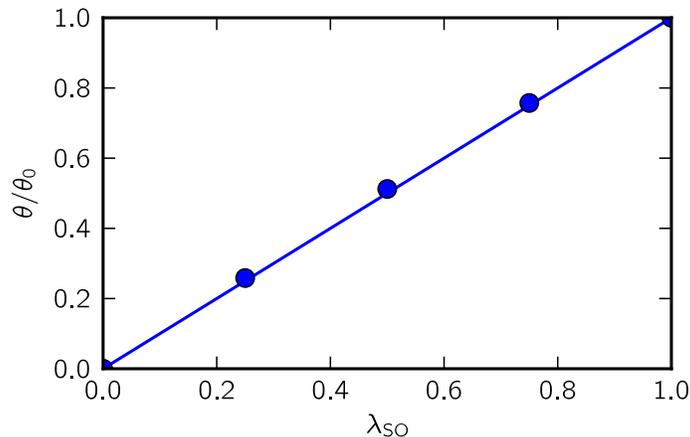


Figure 7.6: Calculated θ in Cr_2O_3 as a function of spin-orbit coupling strength, scaled such that $\lambda_{\text{SO}} = 1$ corresponds to the full spin-orbit coupling strength and $\theta_0 = \theta(\lambda_{\text{SO}} = 1)$.

$6 \times 6 \times 6$ and $8 \times 8 \times 8$ k -point meshes, we obtain $\theta = 0.9 \times 10^{-4}$. In the case of GdAlO_3 we calculate the electronic band gap to be 5.0 eV and the Gd magnetic moment to be $6.7 \mu_{\text{B}}$. We obtain a value of $\theta = 1.1 \times 10^{-4}$ after extrapolating calculations using $4 \times 4 \times 4$ and $6 \times 6 \times 6$ k -space meshes. Thus, it is clear that the CSOMP is very small in both materials.

7.4.2 Strong \mathbb{Z}_2 topological insulators

We now investigate the CSOMP in the case of Bi_2Se_3 , which is known experimentally [47] and theoretically [222] to belong to the class of strong \mathbb{Z}_2 topological insulators. In the absence of broken T symmetry, such a material should have a θ of exactly π (modulo 2π). We first confirm this numerically. Then, in Sec. (7.4.3), we also study what happens when T is broken artificially by inducing antiferromagnetic order on the Bi atoms and tracking the resulting variation of θ .

Bi_2Se_3 is known to belong to space group $\text{R}\bar{3}\text{m}$, with Bi at a 2c site and Se at the high-symmetry 1a site as well as at a 2c site. In our calculations we find that the Wyckoff parameters for Bi and Se are $x = 0.4013$ and 0.2085 respectively. We

also find the length of the rhombohedral lattice vector to be $a = 9.5677 \text{ \AA}$ and the rhombohedral angle to be only 24.77° . The electronic gap is calculated to be 0.4 eV .

The generators of the $R\bar{3}m$ space group are again three-fold and two-fold rotations and inversion (point group $\bar{3}m$). Since the system is nonmagnetic, the magnetic space group also contains the T symmetry operator, and its magnetic point group is $\bar{3}m1'$. According to the analysis given in Sec. 7.2.5, it is clear that θ must therefore be zero or π (modulo 2π).

Since we know that Bi_2Se_3 is a strong \mathbb{Z}_2 topological insulator, we expect that θ should be equal to π (modulo 2π). However, special care needs to be taken in order to evaluate θ in such a case, because the choice of a smooth gauge becomes problematic. Specifically, it is known that the \mathbb{Z}_2 topology presents an obstruction to the construction of a Wannier representation (or equivalently, a smooth gauge in k space) that respects T symmetry [223, 224]. Therefore, during the maximal localization procedure, one needs to choose trial Wannier functions that do *not* take the form of Kramers pairs, thereby explicitly breaking the T symmetry [225]. (It is important to note that this choice of Wannier functions does not bias our calculation towards having $\theta = \pi$, since the same starting choice of T -symmetry-broken Wannier functions for a normal T -symmetric insulator would result in $\theta = 0$ up to the numerical accuracy of the calculation.)

Our results for θ in Bi_2Se_3 are given in Fig. 7.7 for various densities of k meshes, ranging from $6 \times 6 \times 6$ to $11 \times 11 \times 11$. A quadratic polynomial extrapolation to the infinitely dense mesh limit gives $\theta = 1.07\pi$. This is in reasonable agreement with the expected value of $\theta = \pi$, given the uncertainties in the extrapolation. (Of course, if we make a time-reversed choice of starting Wannier functions, we obtain $\theta = -1.07\pi$, which is consistent, within the errors, with $\theta = -\pi$ and modulo 2π to $\theta = \pi$.) Clearly the convergence with respect to mesh density is

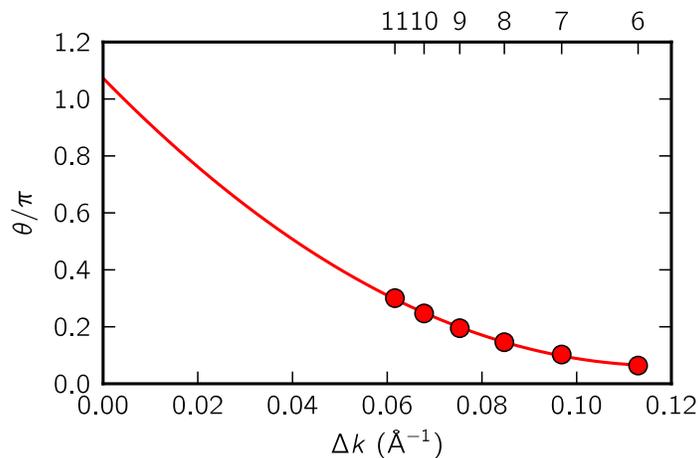


Figure 7.7: Calculated value of θ in Bi_2Se_3 for varying densities of k -space grids, where Δk is the nearest-neighbor distance on the grid. Top axis specifies the size of the corresponding uniform Monkhorst-Pack grid. Line indicates a quadratic extrapolation of θ to the infinitely dense k mesh.

somewhat slow, making a precise extrapolation difficult. The reasons for this, and some possible paths to improvement, will be discussed in Sec. 7.5.

7.4.3 Z_2 -derived nontopological insulators with broken symmetries

Even though $\theta = \pi$ in Bi_2Se_3 , a finite sample with T symmetry preserved everywhere, including at the surfaces, will not exhibit any magnetoelectric coupling. From the point of view of the discussion in Sec. 7.2.4, this happens because of an exact cancellation between $\theta = \pm\pi$ contributions coming from the bulk (α) and metallic surface (Δ) terms in Eq. (7.16). However, if one breaks the T symmetry in the bulk (and possibly some other bulk symmetries, as detailed in Sec. 7.2.5), the CSOMP term can become allowed.

The magnetic space group of Bi_2Se_3 contains both T and spatial inversion symmetries. The presence of either by itself is enough to insure that $\theta = 0$ or π (modulo 2π). Now let us consider turning on, “by hand,” a local Zeeman field on each Bi atom in the staggered arrangement shown in Fig. 7.4(b), i.e., with fields

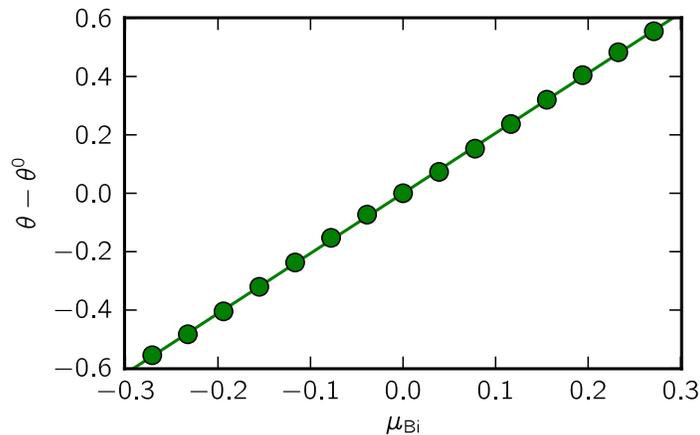


Figure 7.8: Calculated value of θ (vertical axis) and induced magnetic moment on the Bi atom (horizontal axis) for Bi_2Se_3 with artificially applied staggered Zeeman field on Bi atoms, as described in the text. θ^0 is the value of CSOMP when magnetic field is not present.

oriented parallel to the rhombohedral axis and alternating in sign. The induced magnetic moments along the three-fold axis preserve both three-fold and two-fold rotation symmetries; both inversion and T symmetries are broken, but T taken together with inversion is still a symmetry. The resulting magnetic point group of the system is again $\bar{3}'m'$, as it was for Cr_2O_3 , and it does allow for a CSOMP (the same magnetic arrangement has also been discussed in Ref. [226] in a different context).

In the density functional calculation one can easily apply a local Zeeman field to individual atoms in an arbitrary direction.³ Using this method, we have calculated the CSOMP in Bi_2Se_3 with the pattern of local fields described previously and illustrated in Fig. 7.4(b). Fig. 7.8 presents the calculated values of θ as a function of induced magnetic moment on Bi, where a positive μ_{Bi} corresponds to the pattern of magnetic moments indicated in Fig. 7.4(b). (Actually this was done by

³This is done by adding to the Kohn-Sham energy functional an energy penalty term of the form $\lambda \sum_i (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^2$, where $\boldsymbol{\mu}_i$ is the actual value of the magnetic moment of the i -th atom in the unit cell while $\bar{\boldsymbol{\mu}}_i$ and λ are adjustable parameters. The moments $\boldsymbol{\mu}_i$ are calculated by integrating the spin density within atom-centered spheres.

applying the full extrapolation procedure of Fig. 7.7 for one case, $\mu_{\text{Bi}} = 0.16 \mu_{\text{B}}$, and using this to scale the results calculated on the $10 \times 10 \times 10$ grid at other μ_{Bi} .) The dependence of the change in CSOMP on the magnetic moment is linear over a wide range. One can see that for a relatively moderate magnetic moment of $\pm 0.27 \mu_{\text{B}}$, the value of θ is changed from π to $\pi \pm 0.55$. (For much higher local magnetic fields, Bi_2Se_3 becomes metallic and the CSOMP becomes ill-defined.)

These results indicate that it is possible, at least in principle, for a magnetic material to have a large but unquantized value of θ , thereby providing an incentive for future searches for materials in which such a state arises spontaneously, without the need to apply perturbations by hand as done here.

7.5 Summary and outlook

In this manuscript, we have presented a first-principles method for calculating the Chern-Simons orbital magnetoelectric coupling in the framework of density-functional theory. We have also carried out calculations of this coupling for a few well-known magnetoelectric materials, namely Cr_2O_3 , BiFeO_3 and GdAlO_3 . Unfortunately, in these materials the CSOMP contribution to the total magnetoelectric coupling is quite small. This is not surprising, since in most magnetoelectric materials the coupling is expected to be dominated by the lattice-mediated response, whereas the CSOMP is a purely electronic (frozen-ion) contribution. Moreover, the CSOMP is part of the orbital frozen-ion response, which is again expected to be smaller than the spin response, except perhaps in systems with very strong spin-orbit coupling, as discussed in Sec. 7.1. For example, in Cr_2O_3 the CSOMP is about 4% of the frozen-ion spin contribution to the magnetoelectric coupling.

On the other hand, we have reasons to believe that in special cases the CSOMP

contribution to the magnetoelectric coupling could be large compared to the total magnetoelectric coupling in known magnetoelectrics such as Cr_2O_3 . After all, as already pointed out in Sec. 7.2.4, \mathbb{Z}_2 topological insulators are predicted to display a large magnetoelectric effect of purely orbital origin when their surfaces are gapped in an appropriate way. If this is so, why shouldn't a similar effect occur in certain T -broken systems?

As a proof of concept for the existence of those special cases, we have considered Bi_2Se_3 with inversion and time-reversal symmetries explicitly broken “by hand.” Here we find that with a relatively modest induced magnetic moment on the Bi atoms, one can still achieve quite a large change in the CSOMP.

On the computational side, there still remain several challenges. For example, the convergence of our calculations of the CSOMP with respect to the k -point mesh density is disappointingly slow. A direct calculation of θ in Bi_2Se_3 using a very dense mesh of $11 \times 11 \times 11$ k points only manages to recover about 30% of the converged value of $\theta = \pi$, and an extrapolation procedure is needed to bring us within 10% of that value. This clearly points to the need for methodological improvements, and we now comment briefly on some possible paths for future work.

The slow convergence that we observe is related in part to the way in which we evaluate the position-operator matrix elements $\langle \mathbf{0}m | r_j | \mathbf{R}n \rangle$. As discussed in Ref. [27], the k -space procedure we adopted (see Sec. 7.3.3) entails an error of $\mathcal{O}(\Delta k^2)$. Preliminary tests on a tight-binding model suggest that an exponentially fast convergence of θ can be achieved by an alternative procedure, in which the WFs are first constructed on a real-space grid over a supercell (whose size scales with the k -mesh density), and the position matrix elements are then evaluated directly on that grid, as in Ref. [227]. It may also be possible to improve the k -space calculation by using higher-order finite-difference formulas that have a

more rapid convergence with respect to mesh density.

An alternative approach would be to develop a formula for the CSOMP that is exactly gauge invariant in the case of a discretized k -space grid. Such an expression already exists for the case of electronic polarization, Eq. (7.23), but we are aware of no counterpart for the CSOMP. Even though such an approach would not necessarily provide much faster convergence with respect to the k -space sampling, it would still be a significant improvement. For example, one would not need to construct a smooth gauge in k space, which is a particular problem in the case of \mathbb{Z}_2 insulators (or for a symmetry-broken insulator in the vicinity of a \mathbb{Z}_2 phase). Another use of such a formula would be to calculate with relative ease the \mathbb{Z}_2 index of any insulator, even in the cases when other methods [42, 43, 44] cannot be applied (for example, when inversion symmetry is not present).

Furthermore, a full calculation of the electronic contribution to the orbital magnetoelectric response should also include the remaining two contributions given in Eqs. (7.6) and (7.7). This calculation would also require a knowledge of the first derivatives of the electronic wavefunctions with respect to electric field. While these derivatives are available as part of the linear-response capabilities of the QUANTUM-ESPRESSO package [143], some care is needed to arrive at a robust implementation of Eqs. (7.6) and (7.7), as will be reported in a future communication.

Finally, recall that our calculations have all been carried out in the context of ordinary density-functional theory. In cases where orbital currents play a role, it is possible that current-density functionals [228, 229] could give an improved description. However, such functionals are still in an early stage of development and testing, and we prefer to focus first on exploring the extent to which conventional density functionals can reproduce experimental properties of systems in which orbital currents are present.

Overall, significant progress has been made in the ability to calculate the magnetoelectric coupling of real materials in the context of density-functional theory. The methods described in Ref. [207] and [208] allow for the calculation of both the electronic and lattice components of the spin (i.e., Zeeman) contribution to the magnetoelectric coupling. In principle at least, the lattice component of the orbital contribution could be computed using the methods of Ref. [230], while the remaining orbital electronic contributions can be computed from the formulas derived in Refs. [33] and [34] following the developments discussed here. While we have not focused here on the contributions of Eqs. (7.5)-(7.6), we plan to present calculations of these terms in a forthcoming publication. We thus expect that the computation of all of the various contributions to the magnetoelectric coupling will soon be accessible to modern density-functional methods.

Chapter 8

Conclusion

In this thesis we have discussed several different applications of electronic structure methods based on the theory of pseudopotentials and the density-functional theory. Nevertheless, there are many common themes that unite this thesis.

First of all, a common theme of this work is the idea of using electronic structure methods to make specific predictions of materials exhibiting some new or improved property. More specifically, in Chapter 4 we have proposed new high- K dielectric materials, and in Chapter 5 we discuss a new mechanism to control distorted perovskite structures using epitaxial strain. Finally, in Chapter 7 we speculate about the existence of “large- θ materials.” We hope that eventually some or all of these predictions will be experimentally verified.

A second underlying theme of this thesis is geometry and topology. These concepts were explored either arising from the geometry of the crystal structure or more subtly from the geometry and topology of the electron wavefunctions. For example, in Chapter 3, geometrical constraints arising from the relative rigidity of the Si-O tetrahedra result in the intriguing energy landscape for this compound. Similarly, in Chapter 5, the interplay between the epitaxial strain and geometrical constraints of the rigid octahedral network results in the anomalous behaviour of the strained perovskite thin film with strain. Additionally, in Chapters 6

and 7 we analyze the geometry of electron wavefunctions in periodic crystals. More specifically, in Chapter 6 we extend the Berry phase notion of electrical polarization to the case of a Chern insulator. Closely related to the concept of the Berry phase is a Chern-Simons component of the orbital magnetoelectric coupling, which we discuss in Chapter 7.

There still remain some unresolved questions in the context of the geometrical aspects of electronic structure theory. For example, it is natural to ask what other kinds of geometrical or topological terms can be defined for an electron wavefunction in a periodic potential, and to which physically measurable quantity do they correspond. Some such quantities have been discussed at length in this thesis and others have been mentioned in Sec. 2.2. Nevertheless, the complete list of such quantities and corresponding measurable properties has not yet been established. Furthermore, it would be insightful to have a unified theoretical framework through which one could derive and understand all of these quantities on the same footing. Moreover, in order to allow for efficient calculation of these quantities in the framework of modern computation electronic structure techniques, it is important to arrive at properly discretized version of these equations. Discretization of the polarization formula has been known for a while [24], and it is given also in Eq. 7.23, but we are still not aware of a similar discretization of a Chern-Simons orbital magnetoelectric coupling. We leave these topics for a future work.

References

- [1] Sinisa Coh and David Vanderbilt. Structural stability and lattice dynamics of *sio2* cristobalite. *Phys. Rev. B*, 78(5):054117, Aug 2008.
- [2] Sinisa Coh, Tassilo Heeg, J. H. Haeni, M. D. Biegalski, J. Lettieri, L. F. Edge, K. E. O'Brien, M. Bernhagen, P. Reiche, R. Uecker, S. Trolier-McKinstry, Darrell G. Schlom, and David Vanderbilt. Si-compatible candidates for high- κ dielectrics with the *Pbnm* perovskite structure. *Phys. Rev. B*, 82(6):064101, 2010.
- [3] James M. Rondinelli and Sinisa Coh. Large isosymmetric reorientation of oxygen octahedra rotation axes in epitaxially strained perovskites. *Phys. Rev. Lett.*, 106(23):235502, Jun 2011.
- [4] Sinisa Coh and David Vanderbilt. Electric polarization in a chern insulator. *Phys. Rev. Lett.*, 102(10):107603, Mar 2009.
- [5] Sinisa Coh, David Vanderbilt, Andrei Malashevich, and Ivo Souza. Chern-simons orbital magnetoelectric coupling in generic insulators. *Phys. Rev. B*, 83(8):085108, Feb 2011.
- [6] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964.
- [7] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965.
- [8] Mel Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Nat. Acad. Sci.*, 76(12):6062–6065, 1979.
- [9] Richard M. Martin. *Electronic structure*. Cambridge University Press, Cambridge, 2004.
- [10] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45(7):566–569, Aug 1980.
- [11] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43(20):1494–1497, Nov 1979.

- [12] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41(11):7892–7895, Apr 1990.
- [13] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953–17979, Dec 1994.
- [14] P. A. M. Dirac. On the theory of quantum mechanics. *Proceedings of the Royal Society of London. Series A*, 112(762):661–677, 1926.
- [15] P. A. M. Dirac. The quantum theory of the electron. part ii. *Proceedings of the Royal Society of London. Series A*, 118(779):351–361, 1928.
- [16] A. K. Rajagopal and J. Callaway. Inhomogeneous electron gas. *Phys. Rev. B*, 7(5):1912–1919, Mar 1973.
- [17] James D. Bjorken and Sidney D. Drell. *Relativistic Quantum Mechanics*. McGraw-Hill Book Company, New York, 1964.
- [18] Leonard Kleinman. Relativistic norm-conserving pseudopotential. *Phys. Rev. B*, 21(6):2630–2631, Mar 1980.
- [19] Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73(2):515–562, Jul 2001.
- [20] N. E. Zein. *Sov. Phys. Solid State*, 26:1825, 1984.
- [21] Stefano Baroni, Paolo Giannozzi, and Andrea Testa. Green’s-function approach to linear response in solids. *Phys. Rev. Lett.*, 58(18):1861–1864, May 1987.
- [22] Xavier Gonze. Adiabatic density-functional perturbation theory. *Phys. Rev. A*, 52(2):1096–1114, Aug 1995.
- [23] R. Resta. The insulating state of matter: a geometrical theory. *Eur. Phys. J. B*, 79(2):121–137, 2011.
- [24] R. D. King-Smith and David Vanderbilt. Theory of polarization of crystalline solids. *Phys. Rev. B*, 47:1651–1654, 1993.
- [25] Ganesh Sundaram and Qian Niu. Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and berry-phase effects. *Phys. Rev. B*, 59(23):14915–14925, Jun 1999.
- [26] F. D. M. Haldane. Berry curvature on the fermi surface: Anomalous hall effect as a topological fermi-liquid property. *Phys. Rev. Lett.*, 93(20):206602, Nov 2004.

- [27] Xinjie Wang, Jonathan R. Yates, Ivo Souza, and David Vanderbilt. Ab initio calculation of the anomalous hall conductivity by wannier interpolation. *Phys. Rev. B*, 74(19):195118, Nov 2006.
- [28] X. L. Qi, T. L. Hughes, and S.-C. Zhang. *Phys. Rev. B*, 78:195424, 2008.
- [29] Andrew M. Essin, Joel E. Moore, and David Vanderbilt. *Phys. Rev. Lett.*, 102:146805, 2009.
- [30] N. Marzari and D. Vanderbilt. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12865, 1997.
- [31] Ivo Souza, Tim Wilkens, and Richard M. Martin. Polarization and localization in insulators: Generating function approach. *Phys. Rev. B*, 62(3):1666–1683, Jul 2000.
- [32] F. D. M. Haldane. Model for a quantum hall effect without landau levels: Condensed-matter realization of the "parity anomaly". *Phys. Rev. Lett.*, 61(18):2015–2018, Oct 1988.
- [33] Andrew M. Essin, Ari M. Turner, Joel E. Moore, and David Vanderbilt. Orbital magnetoelectric coupling in band insulators. *Phys. Rev. B*, 81(20):205104, May 2010.
- [34] Andrei Malashevich, Ivo Souza, Sinisa Coh, and David Vanderbilt. Theory of orbital magnetoelectric response. *New J. Phys.*, 12(5):053032, 2010.
- [35] Shou-Cheng Zhang and Jiangping Hu. A Four-Dimensional Generalization of the Quantum Hall Effect. *Science*, 294(5543):823–828, 2001.
- [36] M. Z. Hasan and C. L. Kane. Colloquium: Topological insulators. *Rev. Mod. Phys.*, 82(4):3045–3067, Nov 2010.
- [37] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig. Classification of topological insulators and superconductors in three spatial dimensions. *Phys. Rev. B*, 78(19):195125, Nov 2008.
- [38] Alexei Kitaev. Periodic table for topological insulators and superconductors. *AIP Conference Proceedings*, 1134(1):22–30, 2009.
- [39] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig. Classification of topological insulators and superconductors. *AIP Conference Proceedings*, 1134(1):10–21, 2009.
- [40] Shinsei Ryu, Andreas P Schnyder, Akira Furusaki, and Andreas W W Ludwig. Topological insulators and superconductors: tenfold way and dimensional hierarchy. *New Journal of Physics*, 12(6):065010, 2010.
- [41] Taylor L. Hughes, Emil Prodan, and B. Andrei Bernevig. Inversion-symmetric topological insulators. *Phys. Rev. B*, 83(24):245132, Jun 2011.

- [42] Liang Fu, C. L. Kane, and E. J. Mele. Topological insulators in three dimensions. *Phys. Rev. Lett.*, 98(10):106803, Mar 2007.
- [43] J. E. Moore and L. Balents. Topological invariants of time-reversal-invariant band structures. *Phys. Rev. B*, 75(12):121306, Mar 2007.
- [44] Rahul Roy. Topological phases and the quantum spin hall effect in three dimensions. *Phys. Rev. B*, 79(19):195322, May 2009.
- [45] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan. A topological dirac insulator in a quantum spin hall phase. *Nature*, 452(7190):970–974, 2008.
- [46] D. Hsieh, Y. Xia, D. Qian, L. Wray, F. Meier, J. H. Dil, J. Osterwalder, L. Patthey, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan. Observation of time-reversal-protected single-dirac-cone topological-insulator states in bi_2te_3 and sb_2te_3 . *Phys. Rev. Lett.*, 103(14):146401, Sep 2009.
- [47] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan. Observation of a large-gap topological-insulator class with a single dirac cone on the surface. *Nat. Phys.*, 5:398–402, 2008.
- [48] R. W. G. Wyckoff. *Crystal Structures*. Interscience, New York, 2nd edition, 1965.
- [49] D.R. Peacor. *Z. Kristallogr.*, 138:274, 1973.
- [50] A. F. Wright and A. J. Leadbetter. *Philos. Mag.*, 31:1391–1401, 1975.
- [51] M. O’Keeffe and B. G. Hyde. Cristobalites and topologically related structures. *Acta Cryst. B*, 32(11):2923–2936, 1976.
- [52] F. Liu, S. H. Garofalini, R. D. King-Smith, and D. Vanderbilt. First-principles studies on structural properties of β -cristobalite. *Phys. Rev. Lett.*, 70:2750–2753, 1993.
- [53] M. Zhang and J. F. Scott. *J. Phys.: Condens. Matter*, 19(27):275201, 2007.
- [54] I. P. Swainson and M. T. Dove. Low-frequency floppy modes in β -cristobalite. *Phys. Rev. Lett.*, 71(1):193–196, Jul 1993.
- [55] I. P. Swainson and M. T. Dove. Comment on first-principles studies on structural properties of beta-cristobalite. *Phys. Rev. Lett.*, 71:3610, 1993.
- [56] F. Liu, S. H. Garofalini, R. D. King-Smith, and D. Vanderbilt. Liu et al. reply. *Phys. Rev. Lett.*, 71:3611, 1993.

- [57] J. J. Pluth, J. V. Smith, and Jr. J. Faber. *J. Appl. Phys.*, 57(4):1045–1049, 1985.
- [58] P. Tolédano and V. Dimitrev. *Reconstructive Phase Transitions in Crystals and Quasicrystals*. World Scientific Publishing Co., 1996.
- [59] M.T. Dove, private communication.
- [60] I. P. Swainson, M. T. Dove, and D. C. Palmer. Infrared and raman spectroscopy studies of the $\alpha - \beta$ phase transition in cristobalite. *Phys. Chem. Minerals*, 30:353–365, 2003.
- [61] J. H. Nicola, J. F. Scott, and H. N. Ng. Raman study of the $\alpha - \beta$ cristobalite phase transition in aipo4. *Phys. Rev. B*, 18(4):1972–1976, Aug 1978.
- [62] W. Dultz, M. Quilichini, J. F. Scott, and G. Lehmann. Phonon spectra of quartz isomorphs. *Phys. Rev. B*, 11(4):1648–1653, Feb 1975.
- [63] MT Dove, MS Craig, DA Keen, WG Marshall, SAT Redfern, KO Trachenko, and MG Tucker. Crystal structure of the high-pressure monoclinic phase-II of cristobalite, SiO₂. *Mineralogical Magazine*, 64:569–576, 2000.
- [64] D. C. Allan and M. P. Teter. *Phys. Rev. Lett.*, 59:1136–1139, 1987.
- [65] D. C. Allan and M. P. Teter. *J. Am. Ceram. Soc.*, 73:3247–3250, 1990.
- [66] F. Liu, S. H. Garofalini, D. King-Smith, and D. Vanderbilt. 1st-principles study of crystalline silica. *Phys. Rev. B*, 49:12528–12534, 1994.
- [67] J. S. Tse, D. D. Klug, and D. C. Allan. Structure and stability of several high-pressure crystalline polymorphs of silica. *Phys. Rev. B*, 51:16392–16395, 1995.
- [68] A. A. Demkov, J. Ortega, O. F. Sankey, and M. P. Grumbach. Electronic-structure approach for complex silicas. *Phys. Rev. B*, 52:1618–1630, 1995.
- [69] D. R. Hamann. Generalized gradient theory for silica phase transitions. *Phys. Rev. Lett.*, 76:660–663, 1996.
- [70] Th. Demuth, Y. Jeanvoine, J. Hafner, and J. G. Ángyán. Polymorphism in silica studied in the local density and generalized-gradient approximations. *J. Phys.: Condens. Matter*, 11(19):3833–3874, 1999.
- [71] M. Catti, B. Civalleri, and P. Ugliengo. Structure and energetics of SiO₂ polymorphs by quantum-mechanical and semiclassical approaches. *J. Phys. Chem. B*, 104:7259–7265, 2000.
- [72] B. Civalleri and N. M. Harrison. New ultrasoft pseudopotentials for the study of silicates. *Mol. Simulation*, 28:213–237, 2002.

- [73] D. Donadio, M. Bernasconi, and F. Tassone. Photoelasticity of crystalline and amorphous silica from first principles. *Phys. Rev. B*, 68:134202, 2003.
- [74] E. Chagarov, A. A. Demkov, and J. B. Adams. Ab initio calculations of surface phase diagrams of silica polymorphs. *Phys. Rev. B*, 71:075417, 2005.
- [75] N. Flocke, W. M. Zhu, and S. B. Trickey. Density functional energetics of alpha-quartz for calibration of SiO₂ interatomic potentials. *J. Phys. Chem. B*, 109:4168–4171, 2005.
- [76] R. Martonak, D. Donadio, A. R. Oganov, and M. Parrinello. Crystal structure transformations in SiO₂ from classical and ab initio metadynamics. *Nature Materials*, 5:623–626, 2006.
- [77] J. G. Yu, S. R. Phillpot, and S. B. Sinnott. Interatomic potential for the structure and energetics of tetrahedrally coordinated silica polymorphs. *Phys. Rev. B*, 75:233203, 2007.
- [78] X. Gonze. First-principles computation of material properties: The ABINIT software project. *Comp. Mat. Sci.*, 25:478–492, 2002. See also <http://www.abinit.org>.
- [79] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, Oct 1996.
- [80] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, Jun 1976.
- [81] Xavier Gonze and Changyol Lee. Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B*, 55(16):10355–10368, Apr 1997.
- [82] N. Troullier and J. L. Martins. A straightforward method for generating soft transferable pseudopotentials. *Solid State Commun.*, 74:613, 1990.
- [83] M. I. Aroyo, J. M. Perez-Mato, C. Capillas, E. Kroumova, S. Ivantchev, G. Madariaga, A. Kirov, and H. Wondratschek. *Zeit. fuer Krist.*, 221(1):15–27, 2006.
- [84] M. I. Aroyo, A. Kirov, C. Capillas, J. M. Perez-Mato, and H. Wondratschek. *Acta Cryst.*, A62(2):115–128, 2006.
- [85] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, May 1981.

- [86] V. N. Sigaev, E. N. Smelyanskaya, V. G. Plotnichenko, V. V. Koltashev, A. A. Volkov, and P. Pernice. Low-frequency band at 50 cm⁻¹ in the Raman spectrum of cristobalite: identification of similar structural motifs in glasses and crystals of similar composition. *J. Non-crystalline Solids*, 248:141–146, 1999.
- [87] J. B. Bates. Raman spectra of α and β cristobalite. *J. Chem. Phys.*, 57:4042–4047, 1972.
- [88] K. S. Finnie, J. G. Thomson, and R. L. Withers. Phase transitions in cristobalite and related structures studied by variable temperature infrared emission spectroscopy. *J. Phys. Chem. Solids*, 55:23–29, 1994.
- [89] A. A. Volkov, G. V. Kozlov, S. P. Lebedev, J. Petzelt, V. N. Sigaev, and E. N. Smelyanskaya. *Fizika i Khimiya Stekla*, 16:587–592, 1990.
- [90] A. J. Leadbetter. Inelastic cold neutron scattering from different forms of silica. *The Journal of Chemical Physics*, 51(2):779–786, 1969.
- [91] Martin T. Dove, Alexandra K. A. Pryde, Volker Heine, and Kenton D. Hammonds. *J. Phys.: Condens. Matter*, 19:275209, 2007.
- [92] Kenton D. Hammonds, Martin T. Dove, Andrew P. Giddy, Volker Heine, and Bjoern Winkler. Rigid-unit phonon modes and structural phase transitions in framework silicates. *American Mineralogist*, 81(9-10):1057–1079, 1996.
- [93] See EPAPS Document No. E-PRBMDO-78-097829 for a video animation showing the evolution of the structure along the minimum-energy path of Fig. 3.7. For more information on EPAPS, see <http://www.aip.org/pubserve/epaps.html>. Animation also available here <http://physics.rutgers.edu/~sinisa/sio2/trans.mpg>.
- [94] Kenton D. Hammonds, Martin T. Dove, Andrew P. Giddy, and Volker Heine. *American Mineralogist*, 79(11-12):1207–1209, 1994.
- [95] W. Nieuwenkamp. *Z. Kristallogr.*, 96:454, 1937.
- [96] W.W. Schmahl, I. P. Swainson, M. T. Dove, and A. Graeme-Barber. *Z. Kristallogr.*, 201:125–145, 1992.
- [97] M. Gambhir, M. T. Dove, and V. Heine. *Phys. Chem. Miner.*, 26:484–495, 1999.
- [98] D.M. Hatch and S. Ghose. *Phys. Chem. Miner.*, 17:544, 1991.
- [99] J.F. Scott, private communication.

- [100] G.E. Moore. No exponential is forever: but "forever" can be delayed! [semiconductor industry]. In *2003 IEEE International Solid-State Circuits Conference.*, volume 1, pages 20–23, Piscataway, NJ, USA, 2003.
- [101] D. G. Schlom, C. A. Billman, J. H. Haeni, J. Lettieri, P. H. Tan, R. R. M. Held, S. Völkl, and K. J. Hubbard. High-k candidates for use as the gate dielectric in silicon mosfets. In S. B. Ogale, editor, *Thin Films and Heterostructures for Oxide Electronics*, pages 31–78. Springer, 2005.
- [102] D. G. Schlom, S. Guha, and S. Datta. Gate oxides beyond SiO_2 . *Mater. Res. Bull.*, 33:1017–1025, Nov 2008.
- [103] John Robertson. Maximizing performance for higher k gate dielectrics. *J. Appl. Phys.*, 104(12):124111, 2008.
- [104] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neiryneck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki. A 45nm logic technology with high-k+metal gate transistors, strained silicon, 9 cu interconnect layers, 193nm dry patterning, and 100% pb-free packaging. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 247–250, Dec. 2007.
- [105] J. Hicks, D. Bergstrom, M. Hattendorf, J. Jopling, J. Maiz, S. Pae, C. Prasad, and J. Wiedemer. 45nm transistor reliability. *Intel Technol. J.*, 12:131–144, June 2008.
- [106] D. Scansen. Under the hood: 45 nm: What intel didn't tell you, 21 Jan. 2008. www.techonline.com/product/underthehood/205918004.
- [107] Supratik Guha and Vijay Narayanan. Oxygen vacancies in high dielectric constant oxide-semiconductor films. *Phys. Rev. Lett.*, 98(19):196101, May 2007.
- [108] H. M. Christen, Jr. G. E. Jellison, I. Ohkubo, S. Huang, M. E. Reeves, E. Cicerrella, J. L. Freeouf, Y. Jia, and D. G. Schlom. Dielectric and optical properties of epitaxial rare-earth scandate films and their crystallization behavior. *Appl. Phys. Lett.*, 88(26):262906, 2006.
- [109] V. V. Afanas'ev, A. Stesmans, C. Zhao, M. Caymax, T. Heeg, J. Schubert, Y. Jia, D. G. Schlom, and G. Lucovsky. Band alignment between (100)si and complex rare earth/transition metal oxides. *Appl. Phys. Lett.*, 85(24):5917–5919, 2004.

- [110] C. Zhao, T. Witters, B. Brijs, H. Bender, O. Richard, M. Caymax, T. Heeg, J. Schubert, V. V. Afanas'ev, A. Stesmans, and D. G. Schlom. Ternary rare-earth metal oxide high-k layers on silicon oxide. *Appl. Phys. Lett.*, 86(13):132903, 2005.
- [111] T. Heeg, J. Schubert, C. Buchal, E. Cicerrella, J.L. Freeouf, W. Tian, Y. Jia, and D.G. Schlom. Growth and properties of epitaxial rare-earth scandate thin films. *Appl. Phys. A*, 83:103–106, 2006.
- [112] L.F. Edge, D.G. Schlom, S. Stemmer, G. Lucovsky, and J. Luning. Detection of nanocrystallinity by x-ray absorption spectroscopy in thin film transition metal/rare-earth atom, elemental and complex oxides. *Radiat. Phys. Chem.*, 75(11):1608 – 1612, 2006.
- [113] E. Cicerrella, J. L. Freeouf, L. F. Edge, D. G. Schlom, T. Heeg, J. Schubert, and S. A. Chambers. Optical properties of la-based high-k dielectric films. *J. Vac. Sci. Technol., A*, 23(6):1676–1680, 2005.
- [114] V. V. Afanas'ev, A. Stesmans, L. F. Edge, D. G. Schlom, T. Heeg, and J. Schubert. Band alignment between (100) si and amorphous laalo₃, lasco₃, and sc₂o₃: Atomically abrupt versus interlayer-containing interfaces. *Appl. Phys. Lett.*, 88(3):032104, 2006.
- [115] L. F. Edge, D. G. Schlom, S. Rivillon, Y. J. Chabal, M. P. Agustin, S. Stemmer, T. Lee, M. J. Kim, H. S. Craft, J.-P. Maria, M. E. Hawley, B. Holländer, J. Schubert, and K. Eisenbeiser. Thermal stability of amorphous lasco₃ films on silicon. *Appl. Phys. Lett.*, 89(6):062902, 2006.
- [116] P. Sivasubramani, T. H. Lee, M. J. Kim, J. Kim, B. E. Gnade, R. M. Wallace, L. F. Edge, D. G. Schlom, F. A. Stevie, R. Garcia, Z. Zhu, and D. P. Griffis. Thermal stability of lanthanum scandate dielectrics on si(100). *Appl. Phys. Lett.*, 89(24):242907, 2006.
- [117] J.M.J. Lopes, M. Roeckerath, T. Heeg, U. Littmark, J. Schubert, S. Mantl, Y. Jia, and D.G. Schlom. La-based ternary rare-earth oxides as alternative high- κ dielectrics. *Microelectron. Eng.*, 84(9-10):1890 – 1893, 2007.
- [118] M. Wang, W. He, T. P. Ma, L. F. Edge, and D. G. Schlom. Electron tunneling spectroscopy study of amorphous films of the gate dielectric candidates laalo₃ and lasco₃. *Appl. Phys. Lett.*, 90(5):053502, 2007.
- [119] L.F. Edge, W. Tian, V. Vaithyanathan, T. Heeg, D.G. Schlom, D.O. Klenov, S. Stemmer, J.G. Wang, and M.J. Kim. Growth and characterization of alternative gate dielectrics by molecular-beam epitaxy. In *ECS Trans.*, volume 16, pages 213 – 227, Honolulu, HI, United states, 2008.

- [120] E. Durgun Özben, J. M. J. Lopes, M. Roeckerath, St. Lenk, B. Holländer, Y. Jia, D. G. Schlom, J. Schubert, and S. Mantl. Smsco[sub 3] thin films as an alternative gate dielectric. *Appl. Phys. Lett.*, 93(5):052902, 2008.
- [121] M. Roeckerath, J. Lopes, E. Durgun Ozben, C. Sandow, S. Lenk, T. Heeg, J. Schubert, and S. Mantl. Gadolinium scandate as an alternative gate dielectric in field effect transistors on conventional and strained silicon. *Appl. Phys. A*, 94(3):521 – 4, 2009.
- [122] M. Wagner, T. Heeg, J. Schubert, C. Zhao, O. Richard, M. Caymax, V.V. Afanas'ev, and S. Mantl. Preparation and characterization of rare earth scandates as alternative gate oxide materials. *Solid-State Electron.*, 50(1):58 – 62, 2006.
- [123] P. Myllymaki, M. Roeckerath, M. Putkonen, S. Lenk, J. Schubert, L. Niinisto, and S. Mantl. Characterization and electrical properties of high-k gdsc₃ thin films grown by atomic layer deposition. *Appl. Phys. A*, A88(4):633 – 7, Sept. 2007.
- [124] Kyoung H. Kim, Damon B. Farmer, Jean-Sebastien M. Lehn, P. Venkateswara Rao, and Roy G. Gordon. Atomic layer deposition of gadolinium scandate films with high dielectric constant and low leakage current. *Appl. Phys. Lett.*, 89(13):133512, 2006.
- [125] R. Thomas, P. Ehrhart, M. Roeckerath, S. van Elshocht, E. Rije, M. Luysberg, M. Boese, J. Schubert, M. Caymax, and R. Waser. Liquid injection mcvd of dysprosium scandate films. *J. Electrochem. Soc.*, 154(7):147 – 54, 2007.
- [126] T. Heeg, M. Roeckerath, J. Schubert, W. Zander, Ch. Buchal, H. Y. Chen, C. L. Jia, Y. Jia, C. Adamo, and D. G. Schlom. Epitaxially stabilized growth of orthorhombic lusco[sub 3] thin films. *Appl. Phys. Lett.*, 90(19):192901, 2007.
- [127] Wolfram Wersing. Microwave ceramics for resonators and filters. *Curr. Opin. Solid State Mater. Sci.*, 1(5):715 – 731, 1996.
- [128] Terrell A. Vanderah. MATERIALS SCIENCE: Talking Ceramics. *Science*, 298(5596):1182–1184, 2002.
- [129] Pietro Delugas, Vincenzo Fiorentini, Alessio Filippetti, and Geoffrey Pourtois. Cation charge anomalies and high-kappa dielectric behavior in dysco[sub 3]: Ab initio density-functional and self-interaction-corrected calculations. *Phys. Rev. B*, 75(11):115126, 2007.
- [130] R. Vali. Band structure and dielectric properties of orthorhombic srzro₃. *Solid State Commun.*, 145(9-10):497 – 501, 2008.

- [131] R. Vali. Lattice dynamics of room temperature orthorhombic srhfo3. *Solid State Commun.*, 149(13-14):519 – 522, 2009.
- [132] J.M. Badie. *Rev. Int. Hautes Temp. Réfact., Fr*, 15:183, 1978.
- [133] J.M. Badie and M. Foex. *J. Solid State Chem.*, 26:311, 1978.
- [134] J.-P. Coutures, J.M. Badie, R. Berjoan, J. Coutures, R. Flamand, and A. Rouanet. *High Temp. Sci.*, 13:331, 1980.
- [135] J. Schubert (private communication).
- [136] U. Berndt, D. Maier, and C. Keller. *J. Solid State Chem.*, 13:131, 1975.
- [137] A Rabenau. Perowskit und fluoritphasen in den systemen zro₂-lao_{1.5}-mgo und zro₂-lao_{1.5}-cao. *Z. Anorg. Allg. Chem.*, 288:221–234, 1956.
- [138] K. Ito, K. Tezuka, and Y. Hinatsu. *J. Solid State Chem.*, 157:173, 2001.
- [139] M. Bharathy, A.H. Fox, S.J. Mugavero, and H.-C. zur Loye. *Solid State Sci.*, 11:651, 2009.
- [140] V.M. Goldschmidt. Die gesetze der krystallochemie. *Naturwissenschaften*, 21:477, 1926.
- [141] P. M. Woodward. Octahedral Tilting in Perovskites. I. Geometrical Considerations. *Acta Crystallogr. Sect. B*, 53(1):32–43, Feb 1997.
- [142] A. M. Glazer. The classification of tilted octahedra in perovskites. *Acta Crystallographica Section B*, 28(11):3384–3392, 1972.
- [143] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauszero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502 (19pp), 2009.
- [144] Viktor N. Staroverov, Gustavo E. Scuseria, Jianmin Tao, and John P. Perdew. Tests of a ladder of density functionals for bulk solids and surfaces. *Phys. Rev. B*, 69(7):075102, Feb 2004.
- [145] W. Pies and A. Weiss. *Landolt-Börnstein - Group III Condensed Matter*, volume 7c1, chapter VI.1.5.1, pages 14–34. Springer-Verlag, 2006.

- [146] Eric Cockayne and Benjamin P. Burton. Phonons and static dielectric constant in catio_3 from first principles. *Phys. Rev. B*, 62(6):3735–3743, Aug 2000.
- [147] Mark T. Anderson, Kevin B. Greenwood, Gregg A. Taylor, and Kenneth R. Poeppelmeier. B-cation arrangements in double perovskites. *Prog. Sol. St. Chem.*, 22(3):197 – 233, 1993.
- [148] Ruslan P. Liferovich and Roger H. Mitchell. A structural study of ternary lanthanide orthoscamdate perovskites. *J. Phys. Chem. Solids*, 177(6):2188 – 2197, 2004.
- [149] B Velickov, V Kahlenberg, R Bertram, and M Bernhagen. Crystal chemistry of gdsc_3 , dysco_3 , smsco_3 and ndsc_3 . *Z. Kristallogr.*, 9:466–473, 2007.
- [150] Brendan J. Kennedy, Christopher J. Howard, and Bryan C. Chakoumakos. High-temperature phase transitions in srzro_3 . *Phys. Rev. B*, 59(6):4023–4027, Feb 1999.
- [151] Brendan J. Kennedy, Christopher J. Howard, and Bryan C. Chakoumakos. High-temperature phase transitions in srhfo_3 . *Phys. Rev. B*, 60(5):2972–2975, Aug 1999.
- [152] Thorsten M. Gesing, Reinhard Uecker, and J.-Christian Buhl. *Z. Kristallogr. NCS*, 224:365–366, 2009.
- [153] B. Velickov, V. Kahlenberg, R. Bertram, and R. Uecker. *Acta Crystallogr. Sect. E*, 64:i79, 2008.
- [154] E. Ruiz-Trejo, M. S. Islam, and J. A. Kilner. *Solid State Ionics*, 123(1-4):121 – 129, 1999.
- [155] Igor Levin, Tammy G. Amos, Steven M. Bell, Leon Farber, Terrell A. Vanderah, Robert S. Roth, and Brian H. Toby. Phase equilibria, crystal structures, and dielectric anomaly in the bazro_3 - cazro_3 system. *J. Solid State Chem.*, 175(2):170 – 181, 2003.
- [156] J. W. Bennett, private communication.
- [157] J. W. Bennett, I. Grinberg, and A. M. Rappe. *Chem. Mater*, 20(16):5134–5138, 2008.
- [158] T. Heeg, K. Wiedenmann, M. Roeckerath, S. Coh, D. Vanderbilt, J. Schubert and D. G. Schlom, unpublished.
- [159] Donna C. Arnold, Kevin S. Knight, Finlay D. Morrison, and Philip Lightfoot. Ferroelectric-paraelectric transition in bifeo_3 : Crystal structure of the orthorhombic β phase. *Phys. Rev. Lett.*, 102(2):027602, Jan 2009.

- [160] R. Haumont, Igor A. Kornev, S. Lisenkov, L. Bellaiche, J. Kreisel, and B. Dkhil. Phase stability and structural temperature dependence in powdered multiferroic *bifeo3*. *Phys. Rev. B*, 78(13):134108, Oct 2008.
- [161] G. W. Berkstresser, A. J. Valentino, and C. D. Brandle. *J. Cryst. Growth*, 128:684, 1993.
- [162] K. L. Ovanesyan, A.G. Petrosyan, G. O. Shirinyan, C. Pedrini, and L. Zhang. *J. Cryst. Growth*, 198:497, 1999.
- [163] D. A. Muller, private communication.
- [164] R. A. Cowley. Acoustic phonon instabilities and structural phase transitions. *Phys. Rev. B*, 13(11):4877–4885, 1976.
- [165] W. Cochran. Crystal stability and the theory of ferroelectricity. *Phys. Rev. Lett.*, 3(9):412–414, 1959.
- [166] A.G. Khachatryan. *Theory of Structural Transformations in Solids*. Wiley, New York, NY, 1983.
- [167] A. G. Christy. Isosymmetric structural phase transitions: phenomenology and examples. *Acta Crystallographica Section B*, 51(5):753–757, 1995.
- [168] A. W. Lawson and Ting-Yuan Tang. Concerning the high pressure allotropic modification of cerium. *Phys. Rev.*, 76(2):301–302, 1949.
- [169] Razvan Caracas and Xavier Gonze. Structural, electronic, and dynamical properties of calaverite AuTe_2 under pressure. *Phys. Rev. B*, 69(14):144114, 2004.
- [170] J. F. Scott. Iso-Structural Phase Transitions in BiFeO_3 . *Adv. Mater.*, 22:2106–2107, 2010.
- [171] R. J. Zeches, M. D. Rossell, J. X. Zhang, A. J. Hatt, Q. He, C.-H. Yang, A. Kumar, C. H. Wang, A. Melville, C. Adamo, G. Sheng, Y.-H. Chu, J. F. Ihlefeld, R. Erni, C. Ederer, V. Gopalan, L. Q. Chen, D. G. Schlom, N. A. Spaldin, L. W. Martin, and R. Ramesh. A Strain-Driven Morphotropic Phase Boundary in BiFeO_3 . *Science*, 326(5955):977–980, 2009.
- [172] Silvia Tinte, Karin M. Rabe, and David Vanderbilt. Anomalous enhancement of tetragonality in PbTiO_3 induced by negative pressure. *Phys. Rev. B*, 68(14):144105, 2003.
- [173] J. S. Tse and D. D. Klug. Anomalous Isostructural Transformation in Ice VIII. *Phys. Rev. Lett.*, 81(12):2466–2469, 1998.
- [174] Karen Friese, Yasushi Kanke, Andy N. Fitch, and Andrzej Grzechnik. Isosymmetrical phase transition and charge ordering in the mixed valence vanadate -ybv4o8 . *Chemistry of Materials*, 19(20):4882–4889, 2007.

- [175] I. P. Swainson, R. P. Hammond, J. K. Cockcroft, and R. D. Weir. Apparently continuous isosymmetric transition in ammonium hexafluorophosphate NH_4PF_6 . *Phys. Rev. B*, 66(17):174109, 2002.
- [176] J. Haines, J. M. Léger, and O. Schulte. High-pressure isosymmetric phase transition in orthorhombic lead fluoride. *Phys. Rev. B*, 57(13):7551–7555, 1998.
- [177] Stefan Carlson, Yiqiu Xu, Ulf Halenius, and Rolf Norrestam. A reversible, isosymmetric, high-pressure phase transition in Na_3MnF_6 . *Inorganic Chemistry*, 37(7):1486–1492, 1998.
- [178] Alison J. Hatt, Nicola A. Spaldin, and Claude Ederer. Strain-induced isosymmetric phase transition in BiFeO_3 . *Phys. Rev. B*, 81(5):054109, 2010.
- [179] E. L. Colla, I. M. Reaney, and N. Setter. Effect of structural changes in complex perovskites on the temperature coefficient of the relative permittivity. *J. Appl. Phys.*, 74(5):3414–3425, 1993.
- [180] P. M. Woodward. Octahedral Tilting in Perovskites. II. Structure Stabilizing Forces. *Acta Crystallographica Section B*, 53(1):44–66, 1997.
- [181] Christopher J Howard and Brendan J Kennedy. The orthorhombic and rhombohedral phases of LaGaO_3 —a neutron powder diffraction study. *Journal of Physics: Condensed Matter*, 11(16):3229, 1999.
- [182] G. Kresse and J. Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169–11186, 1996.
- [183] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59(3):1758–1775, 1999.
- [184] C. J. M. Daumont, S. Farokhipoor, A. Ferri, J. C. Wojdeł, Jorge Íñiguez, B. J. Kooi, and B. Noheda. Tuning the atomic and domain structure of epitaxial films of multiferroic BiFeO_3 . *Phys. Rev. B*, 81(14):144115, 2010.
- [185] David Vanderbilt and Morrel H. Cohen. Monoclinic and triclinic phases in higher-order devonshire theory. *Phys. Rev. B*, 63(9):094108, 2001.
- [186] S. J. May, J.-W. Kim, J. M. Rondinelli, E. Karapetrova, N. A. Spaldin, A. Bhattacharya, and P. J. Ryan. Quantifying octahedral rotations in strained perovskite oxide films. *Phys. Rev. B*, 82(1):014110, 2010.
- [187] A. J. Hatt and N. A. Spaldin. Structural phases of strained LaAlO_3 driven by octahedral tilt instabilities. *Phys. Rev. B*, 82:195402, 2010.
- [188] A. T. Zayak, X. Huang, J. B. Neaton, and Karin M. Rabe. Structural, electronic, and magnetic properties of SrRuO_3 under epitaxial strain. *Phys. Rev. B*, 74(9):094104, 2006.

- [189] James M. Rondinelli and Nicola A. Spaldin. Substrate coherency driven octahedral rotations in perovskite oxide films. *Phys. Rev. B*, 82(11):113402, 2010.
- [190] Jun He, Albina Borisevich, Sergei V. Kalinin, Stephen J. Pennycook, and Sokrates T. Pantelides. Control of octahedral tilts and magnetic properties of perovskite oxide heterostructures by substrate symmetry. *Phys. Rev. Lett.*, 105(22):227203, 2010.
- [191] T. Mizokawa, D. I. Khomskii, and G. A. Sawatzky. Interplay between orbital ordering and lattice distortions in LaMnO_3 , YVO_3 , and YTiO_3 . *Phys. Rev. B*, 60(10):7309–7313, 1999.
- [192] Yoshihiro Ishibashi and Yoshiki Hidaka. On an isomorphous transition. *J. Phys. Soc. Japan*, 60(5):1634–1637, 1991.
- [193] David J. Thouless. *Topological Quantum Numbers in Nonrelativistic Physics*. World Scientific, Singapore, 1998.
- [194] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. *Phys. Rev. Lett.*, 49:405, 1982.
- [195] Christian Brouder, Gianluca Panati, Matteo Calandra, Christophe Mourougane, and Nicola Marzari. Exponential localization of wannier functions in insulators. *Phys. Rev. Lett.*, 98(4):046402, 2007.
- [196] T. Thonhauser and D. Vanderbilt. Insulator/chern-insulator transition in the Haldane model. *Phys. Rev. B*, 74:235111, 2006.
- [197] Raffaele Resta and Sandro Sorella. Electron localization in the insulating state. *Phys. Rev. Lett.*, 82(2):370–373, Jan 1999.
- [198] David Vanderbilt and R. D. King-Smith. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B*, 48:4442–4455, 1993.
- [199] Raffaele Resta and David Vanderbilt. Theory of polarization: A modern approach. In Charles Ahn and K.M. Rabe, editors, *Modern Ferroelectrics.*, pages 31–68. Springer-Verlag, Berlin, 2007.
- [200] Claudia Sgierovello, Maria Peressi, and Raffaele Resta. Electron localization in the insulating state: Application to crystalline semiconductors. *Phys. Rev. B*, 64(11):115202, Aug 2001.
- [201] Manfred Fiebig. Revival of the magnetoelectric effect. *J. Phys. D: Appl. Phys.*, 38(8):R123, 2005.
- [202] W. Eerenstein, N. D. Mathur, and J. F. Scott. Multiferroic and magnetoelectric materials. *Nature*, 442:759–765, 2006.

- [203] M. Fiebig and N. A. Spaldin. Current trends of the magnetoelectric effect. *Eur. Phys. J. B*, 71:293–297, 2009.
- [204] J.-P. Rivera. A short review of the magnetoelectric effect and related experimental techniques on single phase (multi-) ferroics. *Eur. Phys. J. B*, 71:299–313, 2009.
- [205] V. E. Wood and A. E. Austin. In A. J. Freeman and H. Schmid, editors, *Magnetoelectric Interaction Phenomena in Crystals*, pages 181–194. Gordon and Breach, London, 1975.
- [206] G. A. Smolenski and I. E. Chupis. Ferroelectromagnets. *Sov. Phys. Usp.*, 25(7):475, 1982.
- [207] J. Íñiguez. *Phys. Rev. Lett.*, 101:117201, 2008.
- [208] Kris T. Delaney, Eric Bousquet, and Nicola A. Spaldin. *arXiv:0912.1335*, 2009.
- [209] J. C. Wojdel and J. Íñiguez. *Phys. Rev. Lett.*, 103:267205, 2009.
- [210] Liang Fu and C. L. Kane. Topological insulators with inversion symmetry. *Phys. Rev. B*, 76(4):045302, Jul 2007.
- [211] Hehl, F. W., Obukhov, Y. N., Rivera, J.-P., and Schmid, H. Magnetoelectric effect and relativity theory. *Eur. Phys. J. B*, 71(3):321–329, 2009.
- [212] J.-P. Rivera. *Ferroelectrics*, 161:165–180, 1994.
- [213] H. Wiegmann, A. G. M. Jansen, P. Wyder, J. P. Rivera, and H. Schmid. Magnetoelectric effect of Cr_2O_3 in strong static magnetic fields. *Ferroelectrics*, 162:141–146, 1994.
- [214] E. Kita, K. Siratori, and A. Tasaki. Experimental determination of the mechanism of me effect of Cr_2O_3 from me susceptibility and electric shift in the antiferromagnetic resonance. *J. Appl. Phys.*, 50(B11):7748–7750, 1979.
- [215] A. P. Cracknell. *Magnetism in crystalline materials*. Pergamon press, Oxford, 1975.
- [216] Xinjie Wang, David Vanderbilt, Jonathan R. Yates, and Ivo Souza. Fermi-surface calculation of the anomalous hall conductivity. *Phys. Rev. B*, 76(19):195109, Nov 2007.
- [217] Arash A. Mostofi, Jonathan R. Yates, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. wannier90: A tool for obtaining maximally-localised wannier functions. *Comput. Phys. Commun.*, 178(9):685 – 699, 2008.

- [218] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space gaussian pseudopotentials from h to rn. *Phys. Rev. B*, 58(7):3641–3662, Aug 1998.
- [219] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244–13249, Jun 1992.
- [220] Nicholas J. Mosey and Emily A. Carter. Ab initio evaluation of coulomb and exchange parameters for *dft + u* calculations. *Phys. Rev. B*, 76(15):155123, Oct 2007.
- [221] Siqi Shi, A. L. Wysocki, and K. D. Belashchenko. Magnetism of chromia from first-principles calculations. *Phys. Rev. B*, 79(10):104404, Mar 2009.
- [222] Haijun Zhang, Chao-Xing Liu, Xiao-Liang Qi, Xi Dai, Zhong Fang, and Shou-Cheng Zhang. Topological insulators in bi_2se_3 , bi_2te_3 and sb_2te_3 with a single dirac cone on the surface. *Nat. Phys.*, 5:438–442, 2008.
- [223] L. Fu and C. L. Kane. Time reversal polarization and a Z_2 adiabatic spin pump. *Phys. Rev. B*, 74:195312, 2006.
- [224] Rahul Roy. Z_2 classification of quantum spin hall systems: An approach using time-reversal invariance. *Phys. Rev. B*, 79(19):195321, 2009.
- [225] Alexey A. Soluyanov and David Vanderbilt. *arXiv:1009.141*, 2010.
- [226] Rundong Li, Jing Wang, Xiao-Liang Qi, and Shou-Cheng Zhang. Dynamical axion field in topological magnetic insulators. *Nat. Phys.*, 6:284–288, 2010.
- [227] Massimiliano Stengel and Nicola A. Spaldin. Accurate polarization within a unified wannier function formalism. *Phys. Rev. B*, 73(7):075121, Feb 2006.
- [228] G. Vignale and Mark Rasolt. Current- and spin-density-functional theory for inhomogeneous electronic systems in strong magnetic fields. *Phys. Rev. B*, 37(18):10685–10696, Jun 1988.
- [229] Giovanni Vignale. Mapping from current densities to vector potentials in time-dependent current density functional theory. *Phys. Rev. B*, 70(20):201102, Nov 2004.
- [230] Davide Ceresoli, T. Thonhauser, David Vanderbilt, and R. Resta. Orbital magnetization in crystalline solids: Multi-band insulators, chern insulators, and metals. *Phys. Rev. B*, 74(2):024408, Jul 2006.