A STUDY OF III-V/HFO$_2$ INTERFACES AND SILICON STRUCTURE OPTIMIZATION

Antti Lahti

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. AI OSA - TOM. 683
ASTRONOMICA - CHEMICA - PHYSICA - MATHEMATICA | TURKU 2023

University of Turku

Faculty of Science
Department of Physics and Astronomy
Materials Physics
Doctoral Programme in Physical and Chemical Sciences

Supervised by

Professor, Kalevi Kokko
University of Turku

Docent, Marko Punkkinen
University of Turku

Docent, Pekka Laukkanen
University of Turku

Professor, Ralf Östermark
Åbo Akademi

Reviewed by

Professor, Hannu-Pekka Komsa
University of Oulu

Professor, Yoshitada Morikawa
Osaka University

Opponent

Professor, Adam Foster
Aalto University

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-9140-2 (PRINT)
ISBN 978-951-29-9141-9 (PDF)
ISSN 0082-7002 (PRINT)
ISSN 2343-3175 (ONLINE)
Painosalama Oy, Turku, Finland, 2023

UNIVERSITY OF TURKU
Faculty of Science
Department of Physics and Astronomy
Materials Physics
ANTTI, LAHTI: A study of III-V/HfO_2 interfaces and silicon structure optimization
Doctoral dissertation, 99 pp.
Doctoral Programme in Physical and Chemical Sciences
January 2023

ABSTRACT

Computational materials physics is a field where we try to understand the phenomena related to materials all around us. However we can only model a small sections portion of these materials in our computations, quantum mechanical or classical. In quantum mechanical calculations we are often limited to systems that are sub thousand atoms and classical simulations can reach millions atoms, which is still no where near the ballpark of 10^{20} atoms that will be present in a gram of any material.

The III-V semiconductors are interesting topic due to their potential in device applications [1; 2]. In this research the main issue we have studied is the interface of III-Vs and HfO_2. We have been interested in the structure, electrical properties and characterization of this system. The sub issue of this is a more technical issue that comes with studying these systems: finding the energy minimum structure of any atomic system. We have approached the main issue through quantum mechanical inspection of the system done through the VASP software on different constructed models. The structure optimization was worked on classically with an algorithm built on Genetic Hybrid Algorithm(GHA) and LAMMPS.

For the first issue, we started with the few existing models and studied them. We then created our models with several different crystal forms of HfO_2 and created multiple interface models. Then we studied the effects of different configurations and defects in these models. We found out that the dangling bonds were one of causes for the unwanted defect states in many of the models. We also explored the relative stability of the different models we presented by comparing to each others. In the second paper we continue working on these models and creating new models with native oxides involved. However this time the focus is on providing reference values for interpreting the X-ray photoelectron spectroscopy data from the experiments.

The optimization branch is a more exploratory research line about our approach on optimizing the structure with an algorithm built on the GHA-platform. The first silicon bulk research especially is more of a test case for the used GHA platform and its compatibility with the task. With the silicon dioxide research we dwell more into the different ways we can try to aid the algorithm and what are the pitfalls during the optimization.
KEYWORDS: materials physics, semiconductors, computational research, VASP, LAMMPS

TURUN YLIOPISTO
Matemaattis-luonnontieteellinen tiedekunta
Fysiikan ja tähtitieteen laitos
Materiaalifysiikka
ANTTI, LAHTI: A study of III-V/HfO_2 interfaces and silicon structure optimization
Väitöskirja, 99 s.
Fysikaalisten ja kemiallisten tieteiden tohtoriohjelma
Tammikuu 2023

TIIVISTELMÄ

Laskennallinen materiaalifysiikka on fysiikan haara, jossa pyritään ymmärtämään ympärillämme oleviin materiaaleihin liittyviä ilmiöitä. Emme kuitenkaan voi mallintaa oikeita materiaaleja kokonaisuudessaan laskennallisesti, sillä näiden kokoluokka on aivan liian suuri. Kvanttimekaanisesti olemme rajoittuneet systeemehin, jotka ovat korkeintaan tuhansia atomeja. Klassisessa mallinnuksessa voidaan päästä miljooniinkin atomeihin, mutta kaikkia ilmiöitä ei voida selittää klassisissa malleissa. Miljoona atomia ei kuitenkaan ole lähellä edes yhdessä grammassa materiaa olevaa atomimäärää, joka on kokoluokkaa 10^{20}.

Tässä väitöskirjassa esitetyn tutkimuksen tarkoitus on ollut kartoittaa yhdistepuolijohteisiin liittyviä ilmiöitä: erityisesti III-V:sten ja HfO_2:n muodostamaan systeemiin, jossa rajapinta näiden kahden materiaalin välissä on tärkeä. Sovellutuksien kannalta tämän systeemin karakterisointi, rakenteen ymmärtäminen ja elektronisten ominaisuuksien selvittäminen on tärkeää. Luodessamma malleja tällä systeemille törmä-simme toiseen ongelmaan tässä prosessissa: matala-energisten rakenteiden luontiin. Toisessa tutkimushaarassa keskitymmekin atomirakenteiden optimoimiseen. Tavoitteena oli kehittää algoritmi, joka voisi auttaa meitä rajapintarakenteiden etsimisessä. III-V:sten rajapintatutkimusta lähestyttiin kvanttimekaanisesti VASP-ohjelmiston kautta luomalla erilaisia malleja. Rakenteiden optimoiminen suoritettiin klassisella fysiikalla käyttäen GHA-alustalle luomaamme algoritmia.

Aloitimme rajapintatutkimuksen tutustumalla olemassa oleviin malleihin. Tämän jälkeen loimme useita omia malleja käyttäen HfO_2:sen eri kiderakenteita. Vertailimme malleja keskenään ja keskityimme erityisesti erilaisten rakennevirheiden vaikutuksiin. Erityisesti vapaiden sidosten, dimeerien ja erilaisten atomiympäristöiden vaikutukseen. Vertailimme myös eri mallien stabiiliutta keskenään. Toisessa julkaisussa keskityimme röntgensäteillä tehdyn fotoelektronispektroskopian kautta saatujen tulosten tulkintaan. Laskennallisella puolella otimme aikaisemmat mallit ja muuntelimme niitä, sekä otimme tarkasteluun uusia malleja, joissa oli mukana myös natiivioksiideja. Laskimme näistä malleista kuoritasosiirtymiä ryhmien III ja V atomien elektroneille, mikä auttoi kokeellisen datan tulkinnassa.

Tutkimuksemme optimointihaara oli vapaampaa menetelmän kehittämistä GHA-alustan päälle. Erityisesti ensimmäisessä tapauksessa, piin rakenteen optimoinnissa, keskityimme lähinnä GHA-alustan soveltuvuuteen atomirakenteiden optimoimisessa. Seuraavassa piioksidin rakenteen optimoinnissa tutkimus kääntyi itse algoritmin parantamisen suuntaan. Tutkimme erityisesti eri keinoja optimoinnin avustamiseksi ja

nopeuttamiseksi, sekä tuomme esille suurimmat ongelmat tällaisen rakenteen optimoinnissa.

ASIASANAT: materiaalifysiikka, puolijohteet, laskennallinen tutkimus, VASP, LAMMPS

Acknowledgements

The research presented here was carried out at the Materials research laboratory at Turku University during the years 2015-2023. To begin with I would like to thank Prof. Kalevi Kokko for all the guidance and mentoring along the years. He would always be open for talking about anything. Same goes for my supervisor Marko Punkkinen and Pekka Laukkanen, who were always ready to discuss and help with anything I had problems with. I especially thank Marko for all the helps with the numerous difficulties with the computational aspects of the research, and Pekka for the experimental viewpoint on many of the issues. I am also thankful to Ralf Östermark, who was a tremendous help in the optimization part of my research. His knowledge and support on our work with his GHA platform made it possible for the structure optimization research come true.

I would also like to thank all the people, Jaakko Mäkelä, Muhammad Yasir, Matti Ropo, Masoud Ebrahimzadeh and Mikael Santonen, I have shared my working space with for providing a comfortable and relaxed enviroment, where I always felt welcome. I also thank my mother and friends, both old and new ones made during the years at the university, for providing the support and a chance to relax amidst the work.

The services and resources of the IT Center for Science (CSC) are also acknowledged; most of the computation was done on their computers.

11.1.2023
Antti Lahti

Table of Contents

Acknowledgements . . . . **vi**

Table of Contents . . . . **vii**

List of Original Publications . . . . **ix**

1 Introduction **1**
1.1 What is DFT 3
1.1.1 The Schrödinger equation 3
1.1.2 Kohn-Sham scheme 4
1.1.3 The exchange-correlation 5
1.2 Used implementation: VASP 6
1.2.1 Pseudopotentials 7
1.2.2 The density of states (DOS) and charge density . . 8
1.2.3 Core level shifts 9
1.3 Representing the system with supercells 10
1.3.1 Bulk crystals 10
1.3.2 Slab system 11
1.3.3 Slab and double interface system 11
1.4 The classical semi-empirical model 12

2 Studied materials **15**
2.1 What is a semiconductor 16
2.2 Doping semiconductors to alter their properties 18
2.3 The material interface with semiconductors 19
2.4 Interface and surface stability and energy 21

3 The basics of optimization **22**
3.1 Local optimization 22
3.2 Global optimization 24
3.3 Optimization in materials physics 25

4 GHA optimization **27**
4.1 The initial phases with the GHA project 27

4.2 The algorithm 28
4.3 How do we find the global minimum 29

5 Publications **32**
5.1 The first interface studies with III-Vs 32
5.2 The continuation and collaboration with experiments 32
5.3 First steps to optimization 33
5.4 The difficult task of optimizing SiO_2 34

6 Conclusions **36**

List of References **37**

Original Publications **43**

List of Original Publications

This theses includes the following publications:

I Lahti A., Levämäki H., Mäkelä J., Tuominen M., Yasir M., Dahl J., Kuzmin M., Laukkanen P., Kokko K., Punkkinen M. ”Electronic structure and relative stability of the coherent and semi-coherent HfO2/III-V interfaces”, Applied Surface Science, 427 (2018)

II Mäkelä J., Lahti A., Tuominen M., Yasir M., Kuzmin M., Laukkanen P., Kokko K., Punkkinen M., Dong H., Brennan B., Wallace R.M ”Unusual oxidation-induced core-level shifts at the HfO2/InP interface” Scientific Reports, 9 (2019)

III Lahti A., Östermark R., Kokko K. ”Optimizing atomic structures through geno-mathematical programming”, Communications in Computational Physics, 25, 3(2019)

IV Lahti A., Östermark R., Kokko K. ”Optimization of SiO2 with GHA and basin hopping” Computational Materials Science, 0927-0256: (2021).

The original publications have been reproduced with the permission of the copyright holders.

Publications not included in the thesis:

I Punkkinen M., Lahti A., Laukkanen P., Kuzmin M., Tuominen M. , Yasir M., Dahl J., Mäkelä J., Zhang H.L., Vitos L., Kokko K. ”Thermodynamics of the pseudobinary $GaAs_{1-x}Bi_x$ (0< x <1) alloys studied by different exchange-correlation functionals, special quasi-random structures and Monte Carlo simulations”, Computational Condensed Matter, 5 (2015)

II R.-R. Uusitalo, A. Lahti, H. Levämäki, M. Punkkinen, I. Vilja, K. Kokko, L. Vitos ”Order–disorder transition of $Pd_{0.5}Ag_{0.5}$ alloys”, Philosophical Magazine, 96 (2016)

III M. Kuzmin, J. Mäkelä, J.-P. Lehtiö, M. Yasir, M. Tuominen, Z.S. Jahanshah Rad, A. Lahti, M.P.J. Punkkinen, P. Laukkanen, K. Kokko ”Imaging empty states on the Ge(100) surface at 12 K”, Physical Review B, 98 (2018)

IV P. Laukkanen, M.P.J. Punkkinen, A. Lahti, J. Puustinen, M. Tuominen, J. Hilska, J. Mäkelä, J. Dahl, M. Yasir, M. Kuzmin, J.R. Osiecki, K. Schulte, M. Guina, K. Kokko ”Local variation in Bi crystal sites of epitaxial GaAsBi studied by photoelectron spectroscopy and first-principles calculations”, Applied Surface Science, 396 (2017)

V Kuzmin M., Lehtiö J.-P., Mäkelä J., Yasir M., Rad Z.J., Vuorinen E., Lahti A., Punkkinen M., Laukkanen P., Kokko K., Hedman H.-P., Punkkinen R., Lastusaari M., Repo P., Savin H. ”Observation of Crystalline Oxidized Silicon Phase”, Advanced Materials Interfaces, 6 (2019)

VI M. Kuzmin, J. Mäkelä, J.-P. Lehtiö, M. Yasir, M. Tuominen, Z.S. Jahanshah Rad, A. Lahti, M.P.J. Punkkinen, P. Laukkanen, K. Kokko ”Dimer-vacancy defects on Si(1 0 0): The role of nickel impurity”, Applied Surface Science, 506 (2020)

VII M.P.J. Punkkinen, A. Lahti, J. Huhtala, J.-P.Lehtiö, Z.J. Rad, M. Kuzmin, P. Laukkanen, K. Kokko ”Stabilization of unstable and metastable InP native oxide thin films by interface effects”, Applied Surface Science, 567 (2021)

1 Introduction

The effects of materials research are evident all around us, even if it does not appear so on the first glance. The way everything around us conducts heat, electricity and interacts with light and matter for example can be all explained through materials physics. Many of the materials used around us have been developed through material research at some point. The search for even better materials is everlasting, and there are still applications that require a functioning material. Functioning materials are materials that are distinguished by their electrical, magnetic, optical or chemical properties.

Ways to categorize materials are countless depending on what you are interested in; in my case this is how well the material conducts electricity. Based on this materials can be divided into nonconducting insulators, well conducting metals and semiconductors, where the conductivity can be manipulated more easily than in the former two. The semiconductors have been the focus of the research shown here, more specifically the interfaces that the semiconductors form with different oxides. The interface, the region where one material connects to another through the interface structure, is often the primary source of phenomena that degrade the performance of the devices semiconductors are part of. Semiconductors are the basis of most current electronic devices and components from high to low tech applications, for example the transistors, led lights, solar panels and different sensors. The defining features of these materials are their electrical properties, which can be affected and changed drastically through various means. This property originates from the quantum physical nature of the electrons present in these materials.

The history of semiconductors in electronics started with germanium in the earliest applications in 1940s, but between the 1950s and 1970s electronics transitioned into silicon, which has dominated the field to this day apart from special applications. Not only is silicon very abundant in the earth crust, but it also performs well enough for most applications, which is a big part of its success as a material. [3]

The semiconductor materials are however often very sensitive: impurities and the interfaces they form can seriously hamper the performance of the component they are part of. This is due to the way the semiconductors work. In their pure intrinsic state, they do not conduct electricity well, but even a small concentration of impurities can change this.

Interfaces are unavoidable, as there will always exist surfaces that get oxidized

or there is an interface inherent to the device structure between different materials. What really separates silicon from the other materials is how easily, with proper treatment, it can be protected from the harmful oxidation. Not only does the natural oxide of silicon, SiO_2, form a very good protective layer for silicon components, it also forms a good quality interface that does not greatly lower the device performance. While some defects do form, their number can be greatly reduced with the hydrogen passivation method and the right fabrication. Only when the device size becomes very small do these issues become prominent because as the devices get smaller, the share of surface/interface area becomes a major factor in the component. [4; 5]

As said, silicon is a great material for most semiconductor needs. However, when it comes to the very high end applications, materials with even better properties, like higher electron mobility, different sized band gap, direct band gap, are needed. These materials exist and some very good options are being researched and used currently. What has been in the focus at the University of Turku are the compound III-V semiconductors, which are a widely studied alternative to silicon [1; 2]. There are however many problems that still need to be solved and understood. The biggest problem is the lack of quality interface that they can form. To our knowledge, the natural oxide of these materials forms a problematic interface with the semiconductor underneath [6; 7; 8; 9]. Therefore, something else must be used as the protective layer. What we have studied is different interface structures between the III-Vs and HfO_2 oxide, later on mixing different III-V oxides as a small native oxide layer, which can be beneficial for the interface quality [10]. The goal has been to understand the underlying interface dynamics, what kind of interfaces are stable, what kind of bondings and environments cause the birth of the unwanted defect states and trying to understand the cause of experimentally measured core-level shifts through different model structures and calculated equivalents.

These mentioned studies were done using VASP, a DFT based program that calculates the electronic structure and various properties of the material alongside that. One underlying problem with this kind of setup is that the interface structures are often very large, and require matching two, usually fairly different, crystal structures together. Constructing these interfaces requires a lot of trial and error, which means that the calculations can take a very long time if our initial guess is not close enough to a stable structure. Imagination and knowledge is required to construct many suitable interface structures. With this in mind, we started a project with Prof. Ralf Östermark from Åbo Akademi University to produce a program that can search for these interface structures faster by using semi-empirical potentials. These potentials have limited use as they rely heavily on the form of the potential and the quality of the data used to fit the parameters. Still, they can be used for fast structure searches and molecular dynamics, which would take much longer through the use of DFT. We wanted to take advantage of this and Östermark's expertise, therefore we first started

constructing a silicon optimization method and then moving on to more complex silicon-dioxide. Utilizing these results, the silicon/silicon-dioxide interface could eventually be optimized in the future.

We chose these materials, because the monoatomic silicon provides a good starting point that is good for developing the algorithm. From there, we can move to harder problems with the oxide and interface systems which using the experience we gained with the simpler systems. Silicon structures have been widely studied before, which means that we know the structures we should be getting. This is useful as it makes the development of the algorithm easier. We would have liked to study the III-Vs and their oxides, but unfortunately the potentials for III-Vs were not available and producing them is a complicated procedure we did not have time for.

1.1 What is DFT

Most of the interesting properties of materials, for example conductivity, structural strength and thermal properties, can be traced back to the electrons in the material. Especially the outermost electrons, the so called valence electrons that are least bound to their host atom, often play an important part in bonding between atoms and determining different properties of the material.

Mathematically in quantum physics, the state of these electrons can be described through the so called electron wave function $\psi(x, y, z)$. This is often associated with the probability amplitude of the electron, which means that the squared modulus $|\psi(x, y, z)|^2$ can be seen as the probability density of the electron. There are multiple electrons in a material and summing up their densities one gets the electron density of the material $\rho(x, y, z) = \sum |\psi_i(x, y, z)|^2$. An important density is the ground state density, which is related to the lowest energy state of the system. Due to Hohenberg-Kohn theorems [11], we know two important things about the ground state density:

- If two systems have the same ground state density, then the potential affecting their potentials differ only by a constant.
- There is only one density, the ground state density, that gives the ground-state energy, that is the minimum of the total energy.

It has been shown, that all the properties of the material can be derived from its electron density, this only leaves us the problem of obtaining the density for a given system. This has led to the development of density functional theory (DFT), which deals with solving the electron density of atomic systems. [12]

1.1.1 The Schrödinger equation

Before considering the DFT, the wave function method for solving the ground state of an electronic system is discussed. The key to obtaining the electron density is

to solve the full N-electron Schrödinger equation (1) of the material to obtain the electron wave functions. These wave functions are then used to construct the electron density. The Born–Oppenheimer approximation is often applied, which assumes that the wave functions of nuclei and electrons can be treated separately [13]. This means the heavier nuclei can be fixed in place and we need to only solve the electronic wave functions. The Hamiltonian in this approximation is seperated into three parts:

$$E\psi(\mathbf{r}_1...\mathbf{r}_N) = \left[\overline{\mathbf{T}} + \overline{\mathbf{U}} + \overline{\mathbf{V}}\right]\psi(\mathbf{r}_1...\mathbf{r}_N) \tag{1}$$

Here $\overline{\mathbf{T}}$ and $\overline{\mathbf{U}}$ are called universal operators, as they do not depend on the studied system. They represent the kinetic energy of the electron ($\overline{\mathbf{T}}$) and the interaction energy between the electrons ($\overline{\mathbf{U}}$). All the system specific information is within the component $\overline{\mathbf{V}}$, that is the potential energy of the system. For a system with N electrons and M nuclei, we can further open up the operators a little bit:

$$\overline{\mathbf{T}}\psi(\mathbf{r}_1...\mathbf{r}_N) = \sum_i^N \left(-\frac{\hbar^2}{2m_e}\nabla_i^2\right)\psi(\mathbf{r}_1...\mathbf{r}_N) \tag{2}$$

$$\overline{\mathbf{U}}\psi(\mathbf{r}_1...\mathbf{r}_N) = \frac{e^2}{8\pi\epsilon_0}\sum_{i,j,i\neq j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}\psi(\mathbf{r}_1...\mathbf{r}_N) \tag{3}$$

$$\overline{\mathbf{V}}\psi(\mathbf{r}_1...\mathbf{r}_N) = e^2\sum_i^N v(\mathbf{r}_i)\psi(\mathbf{r}_1...\mathbf{r}_N) \tag{4}$$

$$v(\mathbf{r}) = -\sum_i^M \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|} \tag{5}$$

Here the $v(\mathbf{r}_i)$ is the potential energy caused by the nuclei, Z_i is the charge of nuclei i and ∇_i^2 acts on the coordinates $\mathbf{r}_i$.

1.1.2 Kohn-Sham scheme

Solving the Schrödinger equation becomes very difficult as the amount of electrons in a system increases. Therefore often DFT based methods are used. In DFT the properties of interacting electron gas are solved using a set of one-electron equations related to a fictitious non-interacting electron system. This fictitious system is constructed such that it produces the same electron density as the original system. In this fictitious system, the electrons do not interact, which makes the computation of the wave functions much easier. In a system of N electrons, this reduces the dimension of the equation from 3N to just 3. The drawback is that we have to modify the potential in a way that the solution to the Schrödinger-like equations still provides the wanted density. So our Eq. 1 becomes Eq. 6 for this new system.

$$E\psi(\mathbf{r}) = \left[\overline{\mathbf{T}} + \overline{\mathbf{V}}_{eff}\right]\psi(\mathbf{r}) \tag{6}$$

$$\overline{\mathbf{V}}_{eff}\psi(\mathbf{r}) = v(\mathbf{r})\psi(\mathbf{r}) + \int \frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}'\psi(\mathbf{r}) + V_{XC}(\rho(\mathbf{r}))\psi(\mathbf{r}) \tag{7}$$

Equations (6) - (7) are the so called Kohn-Sham equations, that can be solved self-consistently through multiple iterations.

Here we see that we have transferred the problem from the electron interactions into a different part in the potential. The first term $v(\mathbf{r})$ is the external potential energy term from the original system, second term is the classical repulsion caused by the Coulomb interaction of the electrons and the third is the problematic exchange-correlation potential. The density ρ is also included in the equation, which complicates things a little bit. It means that when we start solving this equation, we have to start with some guess for the density ρ. We then solve the equation and obtain a new density ρ' and repeat this process iteratively until the density converges.

Once the density has been solved, the famous Hohenberg–Kohn theorems guarantee it is an unique solution to the given potential, and that all properties of the system can be extracted from this density [11]. In theory, this method should be exact, provided we have the correct form for the V_{XC}.

1.1.3 The exchange-correlation

Unfortunately the mathematical form of the V_{XC} is not known, which means we have to use approximations for the V_{eff}. Otherwise the Kohn-Sham scheme should actually yield exact results if we do not count the numerical accuracy problems in computations. While the contribution of the XC-term is not big compared to the other terms, it is still vital for any modeling as otherwise bonds will be too weak [14]. We can not approximate it as 0 and its exact form is not known, so we have to find a way to present it in some way. The way we approximate the XC-term is the most crucial part in the accuracy of our DFT-calculations.

There are multiple different approximations for the XC-term, where the computational cost and complexity differ along with the accuracy. To name a few groups, there is LDAs, GGAs, meta-GGAs and hybrid functionals, which we will go through briefly.

LDA stands for local density approximation. In these approximations, as the name implies, we assume that the XC-term in point $\mathbf{r}$ depends only on the density at the same point in the following equation [15].

$$E_{xc}^{LDA}[n] = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}))d\mathbf{r} \tag{8}$$

This means that the approximation is only exact for a system with homogeneous electron density, but it also works fairly well for solids in material physics, but less so for molecules and atoms. The exchange term can be written exactly, but the correlation part is only known at the low and high density limits. Some of the midpoints for the function have also been calculated through quantum Monte Carlo simulations. The different versions of LDA vary mostly in how the correlation part is approximated with an analytical function that produces these limits and interpolates the precalculated values. Some examples of different approximations are Vosko-Wilk-Nusair (VWN), Cole-Perdew (CP), Perdew-Wang (PW92) and Pewdew-Zunger [16; 17; 18; 19].

Next in line comes the class of GGA approximations, which stands for generalized gradient approximation and it was proposed alongside with LDA originally [15]. As the name implies, this approximation also depends on the gradient of the density in the exchange-correlation energy density in the next equation.

$$E_{xc}^{GGA}[n] = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}))d\mathbf{r} \tag{9}$$

In practice, most of the GGA-functionals are constructed by adding a correction term to the LDA-functional

$$\epsilon_{xc}^{GGA}[n] = \epsilon_{xc}^{LDA}[n] + \Delta\epsilon_{xc}[n] \tag{10}$$

These potentials were developed to fix the issues with the LDA. One big problem with LDA is that it overestimates the binding energy between atoms [20]. Also due to the homogenous approximation LDA does not apply well into molecular systems, where the density varies more than in solids. However, GGA-functionals are more difficult to construct and can be very different from each others unlike in the case of LDAs.

One can take GGA even further and introduce even higher degree derivatives. These approximations are called meta-GGA functionals and can include the second derivative of the electron density or kinetic energy density. Hybrid functionals are an even more exotic branch of functionals. They mix in a portion of the exact exchange energy obtained through the Hartree-Fock theory into the exchange-correlation energy obtained through other means, often through the previously mentioned methods. In practice the method is computationally more demanding and has problems with hyperparametrization [21]. When we using DFT we have to make a compromise between the accuracy and used computational resources Fig. **??**.

1.2 Used implementation: VASP

The quantum mechanical calculations presented in this work were all done by using a specialized program called the Vienna Ab initio Simulation Package (VASP) [23; 24;

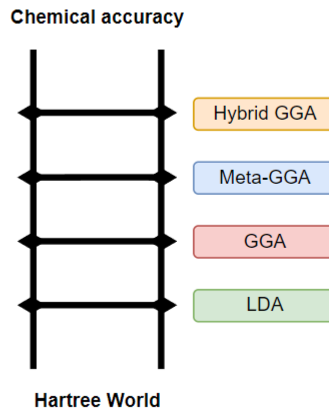


Figure 1. The so called Jacob's ladder of DFT [22]. When we are using DFT we have to make a compromise between the accuracy and the required computational resources. The higher we climb on the ladder, the higher both the accuracy and the requirements will be.

25; 26; 27]. As the name implies, it is an *ab initio* quantum mechanical package that performs the calculations based on the DFT theory. For the exchange-correlation, VASP suppports all of the mentioned, LDA, GGAs, meta-GGAs, Hartree-Fock (HF) method and hybrid functionals, but for my research I used strictly LDA and GGA.

In VASP, the wavefunctions are represented as a series of plane waves, and also as a consequence of this, also the electron charge density and the local potential are affected by this. This is due to many reasons: while a lot of plane waves are needed for accurate presentation, they are very easy to handle computationally. The accuracy of the calculations done with plane waves is also easy to adjust by tuning the basis set size, which is done by changing cut-off energy for the plane-waves. The downsides are that due to the nature of planewaves, they are not good at modeling very sharp features. They for example require the usage of pseudopotentials for the core electrons.

In the following few subsections I will list some of the important outputs and concepts of the VASP calculations, that were relevant during my research.

1.2.1 Pseudopotentials

Pseudopotentials are a natural way to increase the speed of calculations. The pseudopotentials we used and that were supplied with VASP were generated by Kresse with the method described in [28]. The basic idea is that most of the interesting effects in a material arise from the valence electrons. The deeper core electrons on the other hand are very localized around their nucleus and are not really involved in how the material behaves. The pseudopotentials take advantage of this by introducing an effective potential, where the core electrons are included in the description of the

potential of the atom's core.

Another motivation to use pseudopotentials is that some computational approaches use plane waves as the basis set for the computation. The core electrons are very localized, which means that they require more plane waves to model them than the valence electrons usually. The downside of pseudopotentials is that they rely on the assumption that the core and valence electron wavefunctions are not tangled. The wavefunctions of these core electrons will not be present as a result of the computations either naturally, only the valence electron wave functions. Some properties are however influenced by the full wave functions near the nuclei. This is why more advanced methods, like projector augmented wave method (PAW) that is used in VASP, were introduced. PAW divides the space into two regions: the atom centered regions for atom-like electron orbitals and the bonding region between these spheres. These two regions are matched in the boundaries for continuity. The main advantage here is that the full core electron wave functions can still be recovered and used. [27; 29]

1.2.2 The density of states (DOS) and charge density

When it comes to the behavior of a material, the density of states is one of the most important quantities, like a fingerprint of a material. It is a function of energy, that gives us the number of states/energy unit. In ground state at 0 K, the states are filled starting from the lowest energy state until we run out of electrons. Due to the use of the pseudopotentials the core-electron states are not included in the DOS.

The density can further be attributed to different atoms approximately by projecting the wave functions on the orbitals at each atomic site. This is very useful as it helps us understand the source of states at different regions of energy. In our research we used this for identifying the possible sources of problematic states.

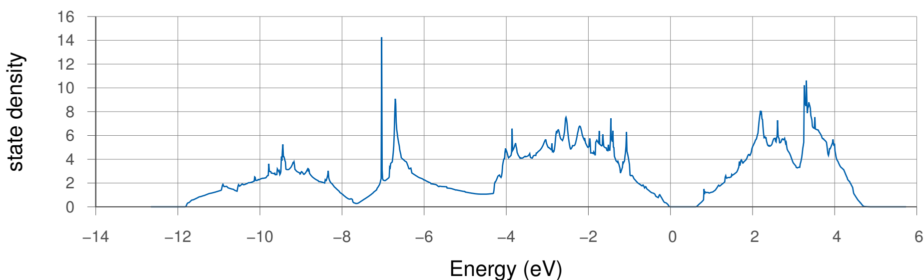


Figure 2. Example of the density of states for silicon.

VASP outputs the total charge density of the system. This helps us visualize the binding and location of the electrons in the system. The partial charge density can also be helpful if we want to visualize only charges corresponding to certain selected bands. It can also be used to approximate the charge on each atom through Bader

charge analysis with the charge density.

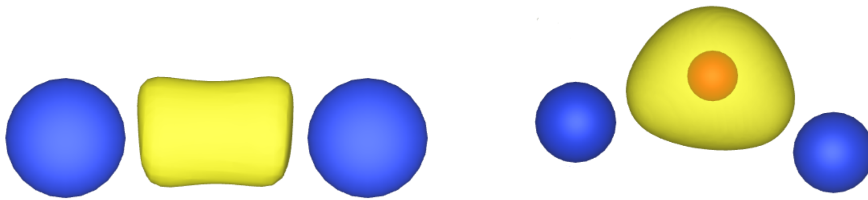


Figure 3. Left: Example of covalent pure silicon system with charge concentrated in the bond between the atoms.
Right: Example of ionic silicon dioxide system with charge concentrated around oxygen that reaches towards the silicon atoms.

1.2.3 Core level shifts

A way to characterize a material is to perform core-level photoelectron spectroscopy on it. In this technique, the binding energies of the core electrons, which are assumed to not take part in the bonding process, are probed to get information about the local area near the atom. Positive shifts mean there has been an increase in the binding energy of the electron, so the electron is more tightly bound into the solid. Negative shifts imply the opposite, decrease in binding energy, meaning the electron can be removed more easily. [30; 31; 32]

We can also estimate these shifts for any atom with a computational approach. However, there is not a feasible way to do it as there are different complex phenomena affecting the experimental shift [31]. The main two ways to calculate the shifts are the initial state approximation and final state approximation. In the former, we assume the energy required to remove the electron is just the binding energy of the electron, which can be calculated from the surrounding electrostatic potential. The final state approximation tries to take into consideration how the electron leaving is affected by the surroundings, mostly the effect of the outer electrons screening the core.

In VASP, the final state is done by calculating the total energy of the cell first in normal state and again with a special potential where the investigated core electron has been moved from the core to the valence. This allows the electronic structure to accommodate the hole left by the electron. The difference between the energies gives us the binding energy of the electron in the final state model. Two calculations are required because to get the relative shift in energies.

Initial state approximation shifts can be calculated in two ways in VASP. First one is the faster way of integrating over the electrostatic potential V with a test charge ρ_{test} at the atom locations $\mathbf{R}_n$. As a result, we get the energy required to remove this

core electron assuming it is centered at this location.

$$\overline{V}_n = \int V(\mathbf{r})\rho_{\text{test}}(|\mathbf{r} - \mathbf{R}_n|)d\mathbf{r} \tag{11}$$

A second way to approximate initial level shifts is to compare the Kohn-Sham energies between different atoms. While the absolute values of the Kohn-Sham energies themselves do not give us the accurate core level energies, the differences between different atoms are often meaningful [33].

These methods can give similar results to experiments, for example with silicon dioxide where the size of the shifts follow the oxidation level linearly[34], but there can be huge differences in the values for other systems; sometimes even different signs for the shifts when using different methods to compute them. What will be truly observed in experimental tests should be some number between these two approximations [32]. This is because these approximations represent the edge cases for the involvement of the system in the process of removing the electron [32].

1.3 Representing the system with supercells

The basis of DFT is the reliance on few assumptions that make the underlying mathematical equations easier to solve, but they are still not simple especially in larger systems. The way the solid material is usually described makes use of the periodic structure that is assumed to exist in solids. These solids, crystals, have atoms arranged in a regular pattern with many internal symmetries. While in real world materials have numerous imperfections, the periodic models often reproduce the properties of these systems very well.

Whenever we do calculations either in the quantum or classical realm, we often want to stick to as small systems as possible. The computational cost increases sharply in both cases. This is one of the reasons the periodic boundary condition is useful, as it allows us to model a massive blocks of material while performing calculations with only few atoms and electrons.

In my studies I have dealt with the following types of systems: bulk systems, slab systems and interface systems.

1.3.1 Bulk crystals

Bulk systems are the smallest of the three types of systems presented here in general. They can be as small as 1 to 2 atom unit cells, for example in the case of iron and silicon and many other monoatomic materials. Larger cells can be required too, especially when more atom types are introduced at various ratios. A 2D example of a bulk system is shown in Fig. 4.

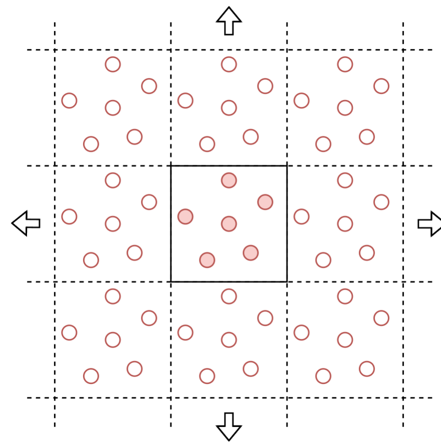


Figure 4. The macroscopic nature of materials is modelled by representing the material with a small unit cells that is assumed to repeat periodically to infinity.

1.3.2 Slab system

In slab calculations a sufficient amount of the bulk structure is included with two open surfaces, with a large void between these surfaces. The plane-wave method requires the system to be periodic in every direction. This means we really are not describing a single infinite slab, but an infinite series of slabs in layers separated by voids. If the space between the slabs is large enough, we get rid off unwanted interaction effects between the slabs. Often the surfaces have to be passivated sufficiently, as otherwise an electric field might form between the surfaces. Passivation also helps with getting rid of surface states. The electric fields can also be removed through dipole corrections. The point of interest in these systems is usually one of the surfaces of the slab and the phenomena around it. Surfaces can have a reconstruction that has larger unit cell than the bulk system. Due to this, slab calculations are typically more computationally demanding than bulk calculations. Especially because when using a planewave basis, the vacuum is as expensive to compute as regions with atoms. An example of a slab system is visualized on the left side of Fig. 5.

1.3.3 Slab and double interface system

Slab interface systems are similar to slab systems, but they also include an interface between two materials inside the slab. Double interface structures are a way to avoid the problem of having a surface in the slab calculation. Essentially, we are introducing a second interface that joins the surfaces that would otherwise exist in a slab calculation. This has some benefits and drawbacks to it. The good part is that we do not have to worry about the surface nor any possible long range interaction hap-

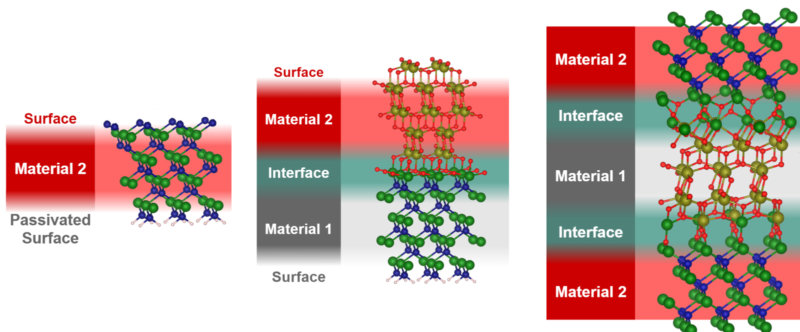


Figure 5. Illustration of the different cells constructed along bulk cells:
Left: A surface slab cell, where one surface is hydrogen passivated. Middle: An interface slab cell, where there is an interface connecting two different materials together. Right: Double interface cell, where the surfaces are replaced with a second interface. No empty void is needed in this cell.

pening between the two surfaces. The downside is that this can lead to cell requiring more atoms. Depending on the structure of the two crystals and the interface, sometimes a double interface cell can not even be made in a way that the two interfaces are identical. An example of these two kinds of interface systems is provided on the right side of Fig. 5.

Constructing interfaces between two materials can be difficult, because one has to match two different crystal structures together. To match the lattices without causing too much strain, a large unit cell is often required if the lattice vectors of the two systems do not match up well. Due to this interface systems especially can lead to fairly heavy calculations. Similarly to surface structures, interfaces can also require increase in the cell size to accommodate the wanted interface structure that is larger than what the bulk system would require.

1.4 The classical semi-empirical model

Semi-empirical potentials are special potentials that try to mimic the interaction and forces between atoms. They have a specific mathematical form that is not usually originating from the quantum mechanical description of the system, but the parameters included in this form are often fitted using data from either experiments or quantum mechanical calculations. Using them howerver does not involve solving of the quantum mechanical equations. The obvious advantage of the potentials is that they allow us to do much faster simulations, like big long timescale molecular dynamics, as there are no hard differential equations to solve. While some charge transfer potentials, like Reaxff[35] and Streitz-Mintmire potential[36], are a bit more computation time consuming, most potentials only require the calculations of interatomic distances in the case of a two-body potential, but also angles if the potential

takes into account interactions including more than two atoms.

The disadvantage is due to the nature of the approximate model. The potential is only as good as the data that was used for the fitting and how well the functional form of the potential can model the data. This means that if the potential is used in a way that is not included in the fit data, the results are not guaranteed to be good. The chosen functional form sets some limitations too, for example all the states of a polymorphic material might not be representable with the same function no matter how much data we have. It is also not guaranteed that a parametrization exists for the elements you want to use, because each elemental interaction requires a separate parametrization for most potentials. Most models can not include many of the quantum mechanical phenomena either, so some things will not be present nor can all the properties be calculated. For example, the electronic wave functions and the density of states can not be obtained. However many structure related features can often be predicted, like atomic positions, forces and stresses. Results should still be confirmed with a more rigorous quantum mechanical calculations, which often are fast to perform as the structure has already been relaxed in the semi-empirical model.

An example of a semi-empirical potential is the Tersoff potential used in the present work. It is a very fast many-body potential originally used for silicon and silicon dioxide, but also extends to carbon and germanium.

The form of the Tersoff potential energy is presented in the equations (12) to (17). The parameters of the Tersoff potential are A, B, c, d, m, n, R, D, β, θ_0, λ_1, λ_2, λ_3. The appearing variables in the sums are r_{ij} and θ_{ijk}, which represent the distance between atoms i and j and the angle between atoms i, j and k. There are the initial sum consisting of factors of two body interactions. Each factor is scaled by a trigonometrical smooth function that goes from 1 to 0 when we go from atomic distance of $R - D$ to $R + D$, both R and D being constants of the potential. For example in silicon dioxide we used, for Si-Si interaction $R - D = 2.5$ Å and $R + D = 2.8$ Å with Tersoff's own potential [37]. The values are similarly low for Si-O interaction, which ensures that the potential fades away quickly after bonds with closest neighbouring atoms.

The factor itself consists of two terms, one repulsive exponential term $f_R(r) = Ae^{-\lambda_1 r}$ and an attraction term $b_{ij} f_A(r) = -b_{ij} B e^{-\lambda_2 r}$, with $f_A(r)$ being another simple exponential term and b_{ij} being the most complex part of the potential, taking into account the three-body interactions of the system that is a bit harder to interpret. The exponential term is often 1 due to λ_3 being 0 in many potentials. The g(θ) is the angle dependent term of the potential, and for a three atom-pairing g(θ) gets its maximum value 1 when the angle between the atoms is a potential predefined value θ_0, as seen in the Fig. 6 for silicon.

Most importantly one can see that a simple energy evaluation requires only going through the three nested sums over all atoms, with each term being easy to compute.

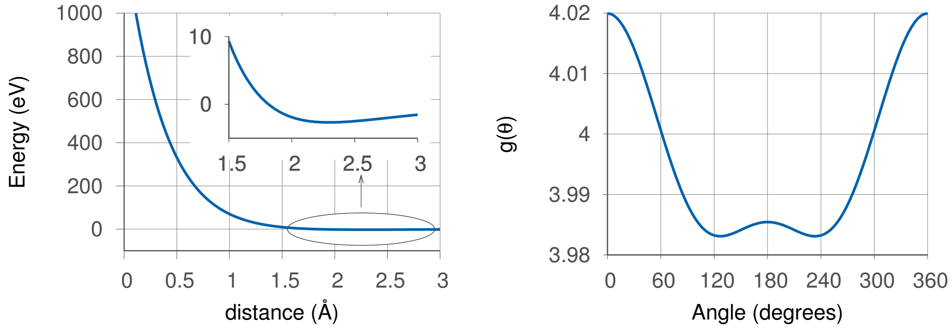


Figure 6. These figures have been made using the parameters for silicon
Left: The form of the purely range dependant factor in the Tersoff potential when there is only two atoms
Right: The form of the purely angle dependant g(θ) factor in the Tersoff potential

For dynamics the forces acting on the atoms can be derived too by taking the gradient of the potential energy. This makes obtaining the forces several orders of magnitude faster than solving the differential equations related to the quantum mechanical approach.

$$E = \frac{1}{2}\sum_{i}\sum_{i\neq j} f_C(r_{ij})[f_R(r_{ij}) + b_{ij} f_A(r_{ij})] \quad (12)$$

$$f_R(r) = Ae^{-\lambda_1 r} \qquad f_A(r) = -Be^{-\lambda_2 r} \quad (13)$$

$$f_C(r) = \begin{cases} 1 & : r < R - D \\ \frac{1}{2} - \frac{1}{2}\sin(\frac{\pi}{2}\frac{r-R}{D}) & : R - D < r < R + D \\ 0 & : r > R + D \end{cases} \quad (14)$$

$$b_{ij} = (1 + \beta^n \zeta_{ij}^n)^{-\frac{1}{2n}} \quad (15)$$

$$\zeta_{ij} = \sum_{k\neq i,j} f_C(r_{ik}) g(\theta_{ijk}) e^{\lambda_3^m (r_{ij} - r_{ik})^m} \quad (16)$$

$$g(\theta) = \gamma\left(1 + \frac{c^2}{d^2} - \frac{c^2}{d^2 + (cos\theta - cos\theta_0)^2}\right), \quad (17)$$

2 Studied materials

The present research was centered around a very limited specific section of the periodic table. First, we focused on the group III-V semiconductors and their interfaces with HfO_2 thin films. Our second branch focused on the structure optimization of silicon and SiO_2. While silicon is widely used for many electronic applications, III-V semiconductors are interesting due to their high performance potential compared to silicon [1; 2]. They have higher electron mobility and a direct band gap, which is useful for many devices. Compared to silicon however, it is harder to make a defect free interfaces between III-Vs and their native oxides [38; 39; 40; 41; 8]. The native oxide, that forms over the semiconductor, in general has a defect density high enough that it hampers device performance significantly, though some studies have shown that a thin native oxide film can be beneficial in some applications [6; 7; 8; 9]. With silicon however, these kinds of defects can be avoided or passivated [4; 5]. HfO_2 is used because it is a high-κ dielectric with a suitable bandgap [42]. For example the InP/HfO_2 interface is a potential component for typical semiconductor devices like solar cells[43; 44], transistors[45; 42] and detectors[46; 47].

Figure 7. The studied materials highlighted in the periodic table.

Group III-V atoms form a bulk structure of zincblende. For HfO_2 we used many different structures. While the monoclinic phase has the lowest energy, some of the other crystal phases were more stable depending on the strain caused by the poor mismatch with the semiconductor, for example anatase and tetragonal structure [48; 49]. While this kind of strain is normally to be avoided, it is justified in this case

as HfO_2 film is assumed to be thin, meaning that it will adhere to the dimensions of the underlying semiconductor.

Semiconductors are present in today's electronics-filled life, but this presence in electronics is not completely obvious to naked eye. Semiconductors are required in almost everything from LEDs, screens, solar cells, diodes, transistors to sensors, which are required for almost any electronic device in people's lives nowadays.

A simple way to give insight into what these materials are is to look what are the other classifications along with it. When it comes to conductivity, materials in general are usually divided into three groups: metals, semiconductors and insulators, which can further be divided into more subgroups. The properties of metals, also known as conductors, and insulators are fairly intuitive: metals conduct heat and electricity very well, while insulators are the opposite. So what does that leave for semiconductors? A very short and kind of obvious answer would be that they are somewhere in between metals and insulators, but before answering we need to take a look at the physical reason for these classifications.

2.1 What is a semiconductor

For a better glimpse into what semiconductors are, we need to approach them through the band-structure of crystalline solids. One lone atom has a discrete set of energies that the surrounding electrons can take. These energies are associated with wave functions that define the state of the electrons. The wave functions are the eigenfunction solutions of the system's Hamiltonian with the energies being the eigenvalues. When two of these atoms are brought together these energy levels start to split. This is because when we bring two atoms together the wavefunctions of the of the two atoms start to overlap. In the case of two atoms this results the energy level splitting into two: bonding and anti-bonding states that have different energies. This effect is more prominent on the outer electrons of the atom as their wavefunctions are often more spread out in real space.

The more of the atoms are brought together the more the energy levels split. Any object we see contains a near uncountable amount of electrons, leading to so much splitting that the resulting energies can be seen and visualised as continuous bands instead of discrete energy levels.

With atoms we have a limited amount of electrons to distribute onto the discrete energy levels and the same applies to solids too. The resulting energy bands are filled only until a certain point, called the Fermi energy. This is a very important energy level that determines a lot of the properties of solid along with the surrounding band structure. In a non zero temperature electrons however get excited by the thermal energy. In this context, we usually talk of Fermi level, which is the energy that has 50% chance of the state being occupied. This energy is determined by applying the Fermi function (Eq. 18) to the density of states. Here E_f is the Fermi energy, k is the

Boltzmann constant and T the temperature. The Fermi function gives the probability of an electron state being occupied at a given temperature at different energies in relation to the Fermi energy.

$$f(E) = \frac{1}{e^{(E-E_f)/kT} + 1} \tag{18}$$

This sounds very similar to Fermi energy, but is not infact the same, most notably in the case of intrinsic semiconductors and insulators, where there is an energy gap between the lowest occupied and highest occupied states. Then the Fermi energy is at the bottom of the gap, while the Fermi level is in the middle of the gap, where there is no states.

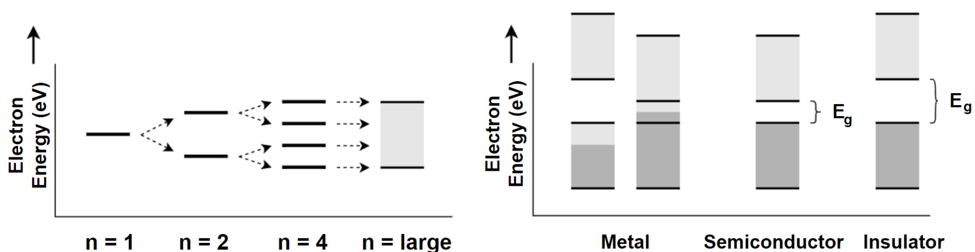


Figure 8. Left: The effect of electrons from different atoms interacting with each others. The energy bands start splitting the more atoms there are, eventually forming bands that are practically continuous.
Right: In materials, the bands formed from splitting energy levels are filled starting from the bottom with electrons present in the system. We can classify materials depending on how the bands are near the highest energy electrons. The darker gray represents filled energy states and the lighter gray represents empty states.

For metals, the Fermi level is somewhere within the band. This explains why most of the metals are so good heat/electricity conductors. The outer electrons in most metals are not heavily bound into the atoms and have many, energy wise, nearby states they can jump into. This allows the electrons to travel more easily in the material, which leads to high electricity and heat conduction.

For insulators, the situation is very different. The Fermi energy is at the top of a fully filled band. There is a huge gap between the highest occupied and highest unoccupied states. The highest occupied band is called the valence band, as it holds the valence electrons of the atoms. On the other side of the gap is the conduction band that holds the lowest unoccupied states that are the way for the material to conduct electrons. This gap causes the features what we associate with insulators like plastics and oxidised materials: very low heat and electricity conduction for example. Because the gap between the states is so large, the energy required to overcome it is big too so that very few electrons will get thermally excited from the valence band to the conduction band, especially in relatively low temperatures like the room temperature.

So where are the semiconductors in this picture? They are in a way closer to insulators, since in semiconductors the energy gap also exist, but it is small enough for an electron in certain conditions to overcome the gap and allow the material to regain some of the metallic properties. This can be due to an external electric field or through thermal excitation for example. The smaller gap also allows the usage of these materials in different optical applications, like sensors, solar cells and LEDs, since the smaller gap corresponds better to visible light wavelengths. The properties of semiconductor devices can be altered heavily by doping the material, which is done by introducing either holes into the valence band or electrons into the conduction band. This can change the properties of a semiconductor by several orders of magnitude. For example the conductivity of a semiconductor can be altered this way heavily, which is not possible in metals. [12]

2.2 Doping semiconductors to alter their properties

There are many specialized uses for semiconductors in very different fields. Some of the most notable ones are transistors and many other electrical components, sensors, solar panels, LEDs and lasers. More broadly, all electronics around us are full of semiconductor materials, where their electrical properties are often one of the key parts for the device to function.

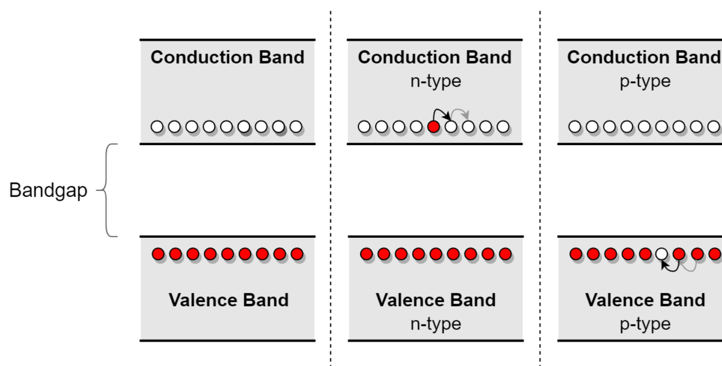


Figure 9. Example of different types of doping and how it introduces states between the bands into the bandgap. On left, we have a intrinsic semiconductor with all states full(empty) in the conduction(valence) band. On right two images, the n-type has been doped with an electron donor element and p-type with an electron acceptor element. This allows easier movement of charges in both situations.

Doping plays a key role in the performance of silicon in these different applications. Doping means introducing more charge carriers to the silicon material by adding impurity atoms. These impurity atoms need to have different valence electron count for this to work and they need to introduce new energy levels around the edges of the bandgap. With silicon that has 4 valence electrons for example, it is common

to use group III atoms to produce p-type semiconductor, or group V atoms to produce n-type semiconductor. These dopant atoms introduce new states into the edges of the energy gap in the semiconductor. This means there will be more charge carrier available in the system. Even small dopant concentrations can affect the electrical properties of the material significantly. In Fig. 9 we illustrate how introducing empty electron states near the top of the valence band will allow the movement of charges in the valence band. Similar thing happens when we introduce filled electron states near the bottom of the conduction band, allowing the movement of charges again.

2.3 The material interface with semiconductors

When two different type of materials connect to each other, the forming bridge region where one atomic structure changes to the other is called the interface. Interface systems are a very important part of research and device development. While bulk materials are important, interfaces appear in everything and can be the source of many interesting effects, both wanted but also often unwanted. This is especially true with semiconductor materials, where in some components the interface is the functional part of the component, for example the MOSFET. For these the interface will not function as wanted, if the structure at the interface is not good enough. A large trap concentration at the interface can for example cause charge accumulation that changes device behaviour.

One of the most famous and well studied interfaces is the SiO_2 interface with silicon. At minimum, the oxide is a protective layer on silicon and it has allowed the quick rise of computing in electronics. Silicon is a rather special case, where with proper treatment it is possible to use the natural oxide as a nearly defect free protective barrier. This means that it is possible to create a Si/SiO_2 interface that does not hamper the semiconductor properties of silicon by introducing additional states inside or near the bandgap. It is not self-evident that the natural oxide of the material will do this, infact, often this is not the case like for example with the III-Vs [38; 39; 40; 41; 8].

This is not only because there exist a quality interface between the bulk-silicon and the silicon dioxide, but quality of the interface can even further be improved by passivating defects, faults in the structure, with hydrogen [50]. Through proper manufacturing process these devices can be constructed in very small sizes. Other materials have potential to be more effective than silicon in devices, but the lack of good quality interface between the new materials and the protective insulator, usually oxide of some sorts, is one of the major obstacles holding them back. Even with silicon and it is natural oxide it took decades to understand the interface well.

We have known about the existance of semiconductors since early 1800s when they were discovered and documented first by Faraday. The history of silicon however begins only from 1950s, when the first silicon transistors were made and the

superior bandgap and native oxide interface were discovered and refined. This is slightly surprising as silicon is one of the most common elements in earth's crust. [3]

In the beginning the understanding of the interface and surface structures was not as important, because the devices were huge so the effect of the interface/surface was small compared to the functioning bulk part. However, as technology has progressed the demand for smaller and smaller devices has increased. The smaller devices have several advantages, like cuts in material costs, energy consumption and space restriction. As the size of the component decreases the portion of the component that forms the interface/surface increases and we have to start taking into consideration the effects from the interfaces and surfaces.

To meet this demand it is important to start studying the new interfaces and materials, especially since it takes a lot of time and requires both experimental and computational contributions. We can not pick just any semiconductor and insulator materials and put them together. The insulator, that also often acts as the protective layer to the semiconductor, has to be chosen carefully so the resulting device is functional. For example the insulator has to have a energy gap around the energy gap in the semiconductor.

After two suitable materials have been chosen, from computational point of view we have to find a way to join them together by making an interface between them. This is where it gets tricky, because the differences in unit cell dimensions make it hard to match the two materials together without distortion. Only a limited amount of orientations are viable from this point of view alone. This is further limited by the computational need for small dimensions, as making these kind of interfaces is faster, they are easier to analyse and faster to compute. These are of course computational limitations, as nature is not limited in the same manner as we are in our computations.

Just matching up the two materials like this is not enough. After that the problematic part of creating the bonds between the two materials needs to be done. This is a very case specific problem, where you try to use up all the electrons in bonds at the interface. The goal here is two fold at this point: find a configuration that both gives a low total energy for the system, so that the interface is realistic and could actually exist outside of calculations, and make sure that interface does not have any harmful effects to the density of states of the system, so that it can still function effectively if a component with this kind of interface was built.

While interfaces are important with almost any material, they are especially so with semiconductors. The function of semiconductors in any device is often reliant on their conductive properties, which in turn are affected by the density of states near the band gap. This is where the interface comes in. In the bulk crystal, we have the periodic structure that houses the desired density of states structure. At the interface, the structure however becomes more problematic. The structure at the interface can be very disordered, with atoms in many different enviroments and

bonding configurations. This kind of system has a very high chance of producing states that are within the critical bandgap region. As discussed before in the Section 2.2 even small amount of new states can be problematic as seen with what happens when even a small amount of doping atoms are introduced. Similarly, states induced from the interface can be seen as a really high level of doping.

In my studies, all of the interfaces are crystallic in a sense, as the unit cells are restricted by atom count and periodic by design. We can still investigate the different configurations of the interface this way by introducing the desired features to the interface. For example, dangling bonds, where an atom has an extra unbounded electron, or dimers, where two atoms are bonded to each others strongly.

2.4 Interface and surface stability and energy

When building different interface models we are often interested in how they compare to each others in terms of stability. We are not only interested in the lowest energy models. To understand the interface we have to also study possible defects at the interface and take into consideration the growth conditions. In the presented research, this was done by approximating the availability of oxygen [51; 52].

In this context, the energy of the interface is

$$\begin{aligned} E_{tot} &= E_{oxide} + E_{interface} + E_{surfaces} + E_{excessO} && (19) \\ \Rightarrow E_{interface} &= E_{tot} - E_{oxide} - E_{surfaces} - E_{excessO} && (20) \\ \Rightarrow A\gamma &= E_{tot} - N_{oxide}\mu_{oxide} - A\gamma_{surf} - N_{O,ex}\mu_{O}, && (21) \end{aligned}$$

where $A\gamma$ is the interface energy, E_{tot} is the total energy of the supercell, N_{oxide} is the number of oxide units with μ_{oxide} being the chemical potential of those units, $N_{O,ex}$ is the amount of extra oxygen atoms in the supercell with μ_O being their chemical potential and $A\gamma_{surf}$ forms the surface energy in the case of a slab calculation, that can be obtained from separate calculations without an interface. In our III-V/HfO_2 studies we considered different structures of HfO_2, which leads to different μ_{oxide} and γ_{surf} terms. The interesting variable here is the chemical potential of oxygen μ_O, which depends on the oxygen availability and affects the relative stability of the different interfaces. Comparison of relative energies for interfaces was done in our publication [53].

3 The basics of optimization

Mathematical optimization is a big field on itself and in terms of mathematical functions, can usually be summarized as finding the best value for the function within some constraints. This often means either maximizing or minimizing the function. If we know the mathematical form of this function, then this can be solved analytically by checking the boundaries and the zero points of the gradient. In more complex cases however, the function might not have a form that can be easily analyzed analytically. In these situations we have to search for the extrema through different algorithms.

Many algorithms are focused on solving very specific types of problems, like simplex[54] for problems where variables have linear relationships and constraints, or its extensions for quadratic problems. These two examples can be solved in a predetermined amount of steps, but with more complex cases we often have to rely on iterative methods where the convergence rate is unknown and not even guaranteed. These methods often evaluate the values, the gradient and the Hessian of the function to converge on a local minimum or maximum of the function.

The process of optimization can be roughly divided into two parts, local and global optimization.

3.1 Local optimization

The goal of local optimization is to take the variables and try to find a local maximum or minimum near them. This can be done using the values of the function or more often using the gradient or even the Hessian of the function, depending on how expensive they are to calculate. These are then used to generate new values for the variables closer to the local extrema. These steps are repeated until we converge on the extrema.

An example of such methods would be the gradient descent, also known as the steepest descent. In gradient descent, the variables represented by vector $\mathbf{x}_n$ are iterated in steps towards the direction of the gradient $\nabla f(\mathbf{x}_n)$, generating us new values $\mathbf{x}_{n+1}$. This is repeated until the search converges, which can take many steps if the extrema is shallow, which is the weakness of the algorithm. The algorithm is very sensitive to the step length variable α. Depending on the complexity of the problem, the step length can be bypassed by doing a line search. In a line search, the

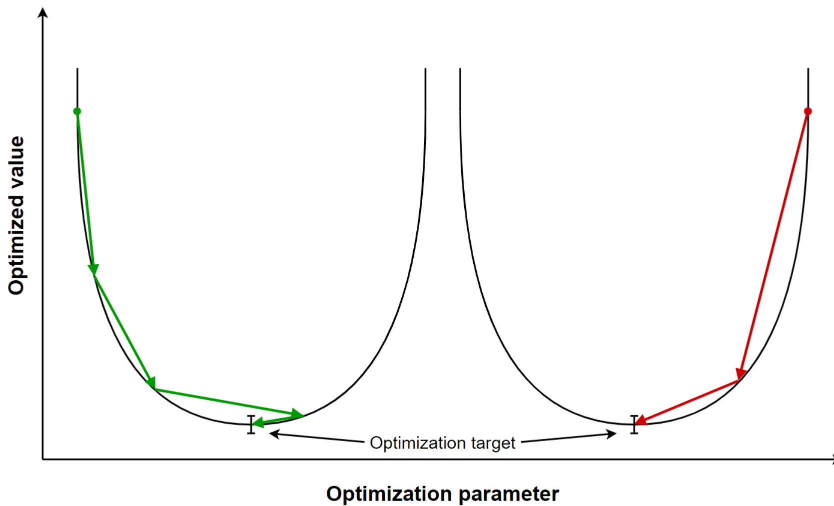


Figure 10. Difference between gradient descent and conjugate gradient descent is the starkest with quadratic functions, where normal gradient descent will often bounce back and forth, while conjugate gradient descent takes only two steps.

step length is optimized in each step [55].

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \alpha \nabla f(\mathbf{x}_n) \tag{22}$$

A more refined group of algorithms are the conjugate gradient methods [56]. Similarly to gradient descent, we take steps where one component is the direction of the gradient, but with an added second component that is a portion of the previous steps taken. The way how the previous steps are mixed in differentiates the conjugate gradient method variants from each others. This allows much faster convergence, especially in the case of quadratic problems like shown in Fig. 10.

Algorithms taking into account the Hessian of the function would be even quicker to converge in general, but especially in high dimension cases, the problem is the calculation of the Hessian, which can be very expensive [55]. The Hessian consists of second-order partial derivatives of the function, which require more evaluation than the first-order partial derivatives. Example of such algorithms would be Newton's method, which takes steps by approximating the optimized function as a parabola during each step, and sequential quadratic programming algorithm which is a more general version of Newton's method, where the approximation is quadratic and constraints can be included too.

As name implies though, these methods are only local. In very simple cases it is possible to find the global extrema through these methods, or sometimes through luck with good starting variables. This becomes apparent if we look at the situation

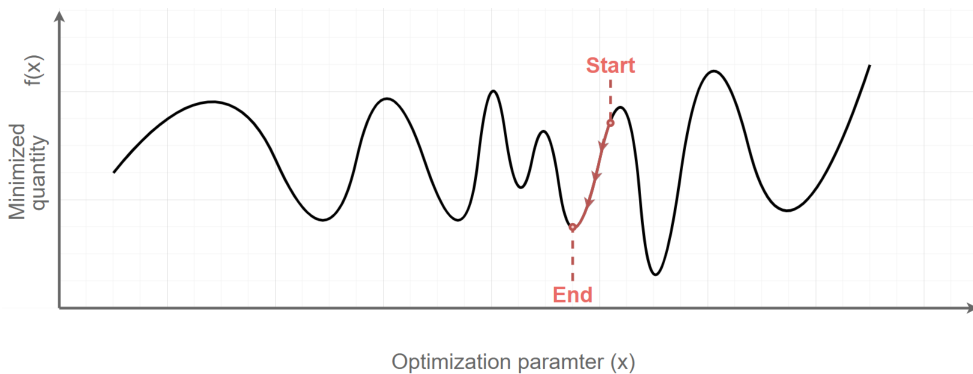


Figure 11. An example of a 1-variable function where it is unlikely to find the global minimum through local optimization. There are multiple local extrema present, which trap the local optimization algorithms and make it difficult to find the global extrema. As shown in the example, success is not guranteed even if our guess locates close to the global minimum initially.

in Fig. 11. In the figure, unless our starting point is right in the vicinity of the global minimum, local optimizing will not converge to the minimum of the function, because the these methods only try to go minimize the function in the immediate vicinity. For complex problems like this, more complicated algorithms are required.

3.2 Global optimization

Global optimization algorithms are often very different compared to the local optimization. They are used when the problem is too complex to solve analytically or through a good guess along with a local optimization. The different algorithms vary in their generality, complexity and applicability. One of the most simple approach would be stochastic random search, where different parameters are generated and locally optimized. This approach can work especially in low dimensional cases, where there is not too many parameters and performing the local optimization is not too expensive.

In annealed optimizing a temperature parameter is introduced, which allows the parameters to move against the gradients [57]. The temperature parameter is then slowly lowered, which allows the system to stabilize on a local extreme point. This kind of heat treatment feels very natural when it comes atomic structures, but it also works well on many other mathematical optimization problems.

Metadynamics is a technique that takes advantage of the history of the search [58]. The algorithm attempts to flood each visited extrema basin. The basin of an extrema is seen as the range of parameters that would, when locally optimized, lead to the this extrema point. Mathematically flooding basins in theory would allow the search to never fall into an already visited extrema, but instead find new ones. The shortcoming of this method is that if the system is too complex, especially in high

dimensional cases, it could take a really long time to flood all the basins. The process of filling the basins is also very delicate.

Evolutionary and genetic algorithms often deal with large populations, that are then evolved with members mixed through different genetic operations to produce new offspring. Throughout the search new population members are generated and mutations are performed on the existing members. This works well when the different parameters of the systems are not deeply entangled with each others. Then the genetic algorithm can identify and combine good features from different members to produce better offspring. This is not viable in every problem however.

In particle swarm algorithm[59], a large number of completely different members are produced in the beginning. These candidates are then subjected to local optimization and other techniques, but on top of that the parameters also feel pull towards other members in the swarm. This way, the swarm eventually converges into a good solution for the optimized problem.

3.3 Optimization in materials physics

In material physics, and especially in the present research, the optimized quantity is often the total energy of the system, the parameters are the locations of the atoms and the gradient of that are the forces acting on the atoms in the system. The low energy structures are inherently interesting as they are the ones most likely to be found in the real world. Finding them through computational means can however, be problematic, if there are a lot of atoms in the systems. The methods mentioned in the previous section have been implemented to different degrees in softwares like USPEX [60], CALYPSO [61] and GASP [62].

While most bulk systems have relatively small unit cells, the problem becomes harder with interfaces. The interface reconstruction can be large area wise and extend multiple layers into the crystal. Especially the oxide part that usually has bigger tendency for amorphic structures has more room to flex into other varying structures.

Many production methods rely on creating the atomic structure through chemical treatment and annealing. We can try recreating these procedures through simulations, but unfortunately even with the big super computers, the time scales and the simulation sizes are still very far from the real situation. Simulated annealing also faces another problem in that the global minimum structure might change depending on the temperature. This applies to many molecules, but also to solids. At lower temperatures the structural changes happen very slowly and the structure can easily get stuck in a high temperature global minimum.

Optimization often faces problems when the amount of parameters exceeds a certain threshold and it becomes impossible to probe the parameter-space fully. This is because increasing the parameter count can lead to exponential growth in complexity, which is exactly what happens in the case of materials physics where every added

atom leads to three new parameters representing the position of this atom. On top of that, due to the nature of atomic potential and bonding, the energy-landscape is riddled with local extrema points that make it even more difficult to try brute forcing the optimization.

If we place atoms down in any given positions in the unit cell, the atoms will follow the forces to the nearby energy minimum. This will not however change the structure massively most of the time. Along with the really high parameter count this means that finding the global minimum through local optimization is very unlikely, unless our guess before the local optimization is very close to the global minimum. This however, would require very intelligent choosing of the initial positions of the atoms, which can be done in some smaller cases. It is worth nothing that real systems might be in one of the near global local minima, so the information about these minima is useful.

4 GHA optimization

Genetic Hybrid Algorithm(GHA) is a platform for optimization algorithms built by Ralf Östermark [63]. It has support for many of the different optimization algorithms and non-linear solvers. These are tools for mixed-integer non-linear programming problems, which help solve the local optimization problem more efficiently.

The GHA platform was also built with parallelization in mind, which is useful for running algorithms on the supercomputers. The platform is designed with genetic algorithms in mind, which means it allows us to run multiple searches concurrently with same or different settings easily and perform genetic operations on the searches. For energy evaluation and annealing treatment we linked and used LAMMPS in library form [64].

Genetic algorithms for structure optimization in material physics have been implemented previously in software packages like USPEX [60], GASP [62] and CALYPSO [61] to name a few.

4.1 The initial phases with the GHA project

A major obstacle in our III-V studies was the construction of different models for the interface, which becomes especially problematic with bigger cell sizes. In 2016 Autumn an opportunity presented itself to do collaboration with Ralf Östermark from Åbo Akademi University. We set out to test how these methods might benefit us in optimizing atomic structures. The initial plan was to start out small, with the first tests done on small FeCr system. From there, the plan was to move onto optimizing bulk silicon, then silicon dioxide and finally try to apply our program to silicon-silicon dioxide interface.

With the initial tests being successful we however quickly found ourselves stuck in the difficult energy landscape of the atomic structures in harder problems. Even in the simple monoatomic case of silicon, finding the diamond structure of silicon proved to be non-trivial when more atoms were included.

In hindsight, we made mistakes, like freezing one layer of atoms in their correct positions. The hope there was that it would promote the correct ordering of atoms through the frozen layer and would allow the crystal structure to form more easily, like crystals grow in nature. This however was not the case, because such growth is not really possible in a small constrained cell of atoms. Instead this frozen layer

actually hindered the optimizing process, as it gives less room for atoms to move around each other. This creates higher potential energy walls and completely cuts off some valleys in the energy landscape that would lead to smaller energies. In the end using a high population approach we were able to find the global minimum reliably. This means processing and altering many structures in parallel on the supercomputers based on our algorithm, that alters and relaxes the structures.

4.2 The algorithm

The layout of our algorithm is presented in Fig. 12. We start out with a pool of generated structures. In the early feature testing phase this was often less than 10, but in in some of the runs done for research presented here the size of the pool was at most 1024. We had different schemes for generating the structures, with the most unbiased method being a random placements of atoms in the unit cell. A more involved and physical method was spreading the atoms around the unit cell randomly and uniformly. While this method was still very unbiased, it also speed up the optimizing process fairly significantly since these structures are more realistic than the completely random ones, where atoms are often clumped up in different parts of the cell. It essentially lets us skip the first iterations of the optimization, while not altering any of the results we get in the end.

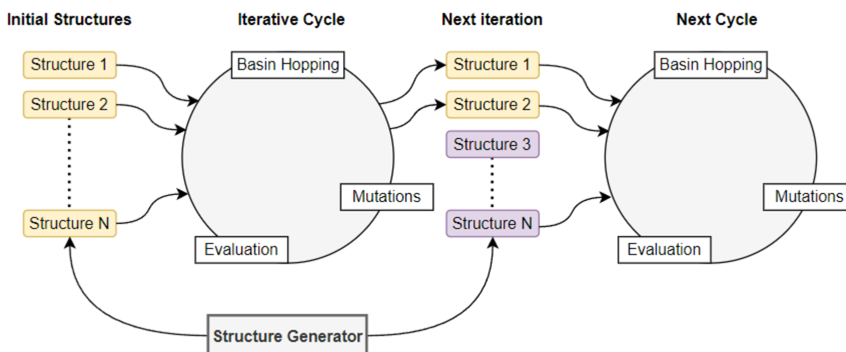


Figure 12. The basic overview of the algorithm. At each time, we are processing N different structures in parallel. Initially we generate a fresh pool of new structures. These then get refined in an iterative loop of different operations, including basin hopping and different targeted mutations. During this process the structures are evaluated constantly and the best ones are saved. After a while, we restart by generating a new pool of structures. Some structures of the previous cycle are retained for the next cycle.

These structures are taken into a loop that performs different actions on them. Due to GHA-platform's parallelization we were able to utilize the supercomputers for this efficiently, which sped up the process. The main way for the algorithm to progress was done using basin hopping [65]. Along this, different targeted mutations

were also applied in various steps. This included moving single atoms intelligently: for example twisting the oxygen bridges in the system and changing which atoms are connected by the oxygen bridges. The structures were evaluated constantly and the best structures were saved.

After this loop was repeated a set amount, we generated a new pool of structures. Most of the previous structures were discarded, but part of them were retained if the algorithm still saw room for improvement with them. These structures were then exposed to the same operations as the first set. The program repeated this cycle for a set amount of cycles before terminating.

Most of the work in this algorithm went into designing the mutations and trying to guide process so that our structures evolved into low energy structures.

4.3 How do we find the global minimum

The local search has usually a guaranteed result and reaching it is just a matter of computational time. However, nothing guarantees that this found minimum is a global one. Behind the surrounding potential barriers there could very well be an even lower minimum, and the local search will not tell us anything about that. This is a very tricky problem in optimization, as there is no efficient foolproof method of finding the global minimum in complex cases. Often the more local minima there are, the trickier it is to find the global minimum as most search methods get trapped or distracted by the other minima.

Basin hopping is a very simple scheme to overcome the problem of barriers in the search of the global minimum. There are many different ways to implement and use it, but the basic principle is usually very similar. The basic idea is to overcome the weakness of gradient based methods and not get trapped in one local minima but instead try to explore the surrounding area too and eventually drift into the global minimum. For this to happen different problem specifi aids may need to be implemented, as is the case with most optimization related tasks. The simplest way is to, once we reach a local minimum, disturb the system enough to overcome the surrounding energy barriers and reach a new local minimum. In Fig. 13 we have illustrated how basin hopping treats a simple 1 variable case.

For example in atomic optimization tasks, these disturbances correspond to moving atoms, but due to geometrical factors there are some ways that are more effective. This includes for example twisting bridge bonds that have a high angle, like the oxygen bonds in SiO_2, where each oxygen has typically only two bonds. Sometimes it is worth going through each atom and just shifting them in one direction slightly to overcome an energy barrier they could not overcome. In the beginning of the search, the way these kind of disturbances are done does not really matter as much. This is because in general, the starting guess is usually pretty bad, sometimes on purpose to ensure wide sample size. When we start influencing the creation of the

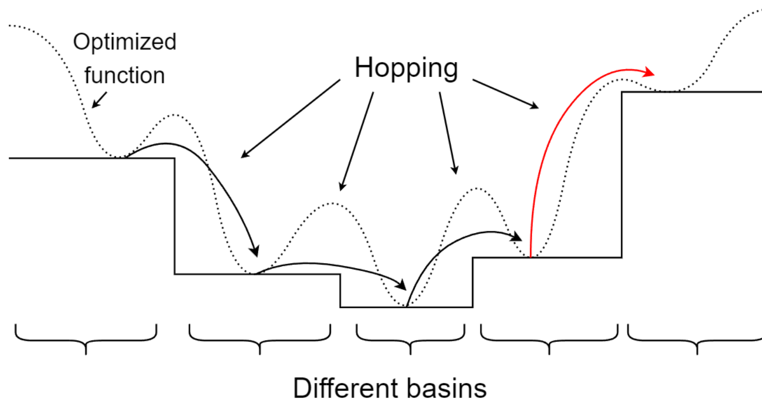


Figure 13. The basic idea of basin hopping is illustrated here. In basin hopping, the optimized function is seen as region of basins. We traverse the function by jumping from basin to basin, with jumps that go uphill being less likely to happen. An unlikely jump is highlighted with red colour here.

starting structure heavily, the risk of producing structures with very similar features increases.

Especially with smaller atomic structures, say around 30 atoms, this does not matter as much, since through few jumps the structure will often very rapidly get to low energy structure with near correct amount of bonds for each atom, like for present examples of bulk silicon and silicon dioxide. That is where the problems begin though. With silicon dioxide it is very hard to guide the search more than that. The amorphous nature of the oxide creates many attractive local minima structures, that will fool the algorithm if it is only looking at the energy of the structures. In reality though these structures are not geometrically close to the wanted global minimum structure. The main difference between the structures is that the global minimum structure is a very highly ordered structure, while most of the other low energy structures are not.

Of course, it all depends on what kind of structures you are looking for in the search. If you just want some low energy structures these amorphous structures might be just what you want. There is usually of course no guarantee that what you get is the true global minimum. Or if you only want some ordered structures, you can run a search and filter out the ordered structures, which in our searches were a very small minority.

In interface calculations however, the lattice constants of the oxide might be forced into something else to accommodate the underlying base structure. In these situations the interface structure itself is unknown, as it usually differs to some degree from the bulk structure of the oxide and substrate, but the close-by layers of the oxide can also be altered.

This kind of structure phase is really hard to study manually. Researchers can create model structures by hand or through some algorithms, but they are always heavily limited by the imagination and creation speed of the researcher. Another problem is that calculations performed on these structures usually relax them to some nearby local minima that might not even be that good. Molecular dynamics can be used to overcome the energy barriers, but there we have very little control over the structure and heat treatments still require a long cooling time which translates to a lot of used simulation time.

5 Publications

Our work has mostly dealt with various interfaces formed by the III-V compound semiconductors and HfO_2. Through different models we have expanded the knowledge about systems with these interfaces: what kind of interfaces are more likely to form, which of them are stable and what are the problems of these structures. The second branch of research is about constructing an optimizing method for atomic structures, with publications made on the optimization of silicon and silicon dioxide bulk. This branch was more about the used method and trying to understand how the optimization of structures works. Our papers highlighted the methods that worked for us and what kind of challenges we had.

5.1 The first interface studies with III-Vs

For the first paper [53] we built our initial interfaces for the HfO_2/III-V using the known models as a rough base [66; 67; 68]. Others were built according to the electron counting rule mentioned before and taking into account the lattice geometries to avoid strain [69; 70; 71]. This included both coherent and semicoherent interfaces.

The main concern was to build stable interfaces that would not contain the harmful interface states near the semiconductor gap. In the paper we presented many models, of which some of the coherent models were gap state free and stable among the studied models under wide range of oxygen conditions during growth. We also show that dangling bonds are responsible for the gap states in different models. Some of the interfaces included dimers which we noted were not the cause of the unwanted gap states.

These models were built for GaAs and GaP compounds, that allowed the coherent models by using the anatase structure of HfO_2. While the coherent model introduces a bit more strain for the oxide, meaning the bulk energy for the oxide is slightly larger, this is offset by the lower interface energies that produce lower total energies for the system in case the oxide part of the system is not too thick.

5.2 The continuation and collaboration with experiments

Second paper in the III-V category had our computational models incorporated into an experimental corelevel study on differently grown interfaces of the InP/HfO_2 sys-

tem [72]. We used some of our previous models for calculating these core level shifts, but we also had to build new systems to include wider array of environments for the core level study.

Through this study we uncovered some unexpected corelevel behavior especially concerning the group III atoms. The magnitude and even the sign of their shift was not explained by the local environment as straightforwardly as we had expected. The expected behaviour would have been something similar to silicon, where the oxidation level of the atom determines the size of the shift. The more oxygens an atom is bonded with, the larger the shift [73]. The group V atoms, phosphorous in this case, seemed to follow this trend. Group III atoms on the other hand varied a lot. The shifts were not as large in general, but they were hard to predict and in certain enviroments they were even negative.

5.3 First steps to optimization

On the second line of the research papers we have been working on the problem of optimizing atomic structures. Using the GHA-platform and various methods like basing hopping and various attempts to guide the search we approached the problem of finding the global minimum structure of a system in two cases: starting from bulk silicon and bulk silicon dioxide[74; 75]. Initially we had expected especially the silicon case to solve easily, but as we started the work we quickly noticed that by increasing the atom count of the system, the cases become very hard to solve quickly. While the small cases with less than 10 atoms solve almost instantly, the difficulty ramps up quickly and after 30 atoms the algorithm start to take noticebly longer to find the solution, especially in the case of silicon dioxide.

In the first paper we were using the silicon bulk as a test case for the optimization. The work was done with a collaboration with an optimization expert from Åbo Academi University and us from materials physics, which meant this problem had new aspects for both of us. Thus starting with a relatively simple and known structure like bulk-silicon was good choice as it allowed us to focus on the understanding and algorithmic side of the problem.

Nevertheless we discovered the difficulties in optimization even in the case of bulk silicon. We focused on a case of 32 silicon atoms in a periodic box. In this box we fixed an atom layer in place to promote the right structure.

While we were able to solve this case eventually, the energy barriers made finding the way to the global minimum, diamond structure, fairly difficult. The study did its job as a step toward more complex systems as we obtained valuable first hand information on our own. While a lot of this was on the algorithmic side, some of the findings were more practical and concrete. Most notably we found that most restrictions during the optimization are fairly detrimental to the optimization. Whether this is the fixed layer we had in our case, or other selectively frozen parts of the structure.

This, we believe, is due to the restrictions removing the structures ability to evolve during the optimization. By this we mean that the fixed atoms create high energy barriers that the surrounding atoms have to accomodiate, as the fixed atoms can not move out of the way.

The diamond structure of silicon was found to be a very dominant minimum, meaning there were not any other nearby stuctures in energy, and the closest structures in energy were also always few iterations away from reaching the diamong structure.

5.4 The difficult task of optimizing SiO_2

From the bulk silicon we moved onto the harder silicon-dioxide bulk, which proved to be an even harder challenge[75]. A 36 atom case of silicon dioxide exhibits a lot of structural variance in the low energy region, which makes the optimization especially difficult. In the silicon case we would eventually naturally drift into the solution, global minimum, but that is not the case here as there is many other attractive solutions. This diversity is born from there being two elemental components. Oxygen is also a smaller atom than silicon, which I believe also complicates the situation. This can been for example with the oxygen bridges, that can lock into multiple different orientations often.

Due to this we tried to find a way to guide the optimization, especially the basin hopping part of the algorithm, as it was one of the major driving parts of exploring the structure-space. Beforehand we suspected that this might difficult, as it appeared that the energy-landscape was very noisy and finding the global minimum was not easy. And this turned out to be true. Guiding the jumping through a selective method that prioritizes low energy structures proved to be detrimental to the global minimum search. This kind of approach would guarantee low energy structures but not the global minimum. As it turns out, in this case, energy is not a good measure for the distance between the global minimum and a given structure. It is likely that many other low-energy structures will trap the search if we use such guidance. This information puts us into a difficult position in the optimization. Energy is the quantity that matters, actually potential energy in this case when not taking into account any other factors. But we can not use energy as a guide during the search as it will not really lead us to the ground-state structure, nor will it serve as a measure on how far we are from the wanted ground-state. There are many amorphous structures having energy near to the global energy minimum even in the smallest studied 18 atom case, and their count explodes even more when we increase the number of atoms.

The starting structure does not seem to be very important in small cases. With for our algorithm centered around basin hopping, doing few iterations of changes and relaxation mixes up the the structure quickly and we end up with structures of very similar energies no matter what kind of starting point we used. Doing these iterations

takes more and more time as the system grows in size, so a good guess does become more important in larger systems.

Since energy was not an ideal guide we tried to use two more structure centered factors as a guide: the silicon-silicon distances and an order factor, that we calculated through a Fourier-transform. The idea of using a criterion derived from silicon-silicon distances arose in a few ways. First the observation that in most ordered and low energy structures the silicon atoms are fairly evenly spread, which in terms of distances means that the atoms are pretty close to maximally distanced from each others. We verified this by mapping the relation between the energy and the silicon-silicon distance. It turned out that the global minimum is not the structure where the silicon atoms are maximally distanced from each others, but pretty close to that situation. Also these structures on average have fairly low energy and have higher concentration of ordered structures or at least a good amorphous structure. In any low energy structures there are no silicon-silicon bonds as they are energetically not as favorable as silicon-oxygen bonds.

6 Conclusions

In this dissertation we have presented our studies on semiconductor materials and structural optimization. The research on semiconductor and oxide interfaces focused on the III-V/HfO_2 systems and was done using DFT implemented in VASP. In this research we started out by building several coherent and semi-coherent models for the interfaces between the two materials. We compared the stability of the different interface models and the electronic structure with the focus being on the states within the energy gap. We identified the source of the trap states to be the dangling bonds present at the interface in these models. The study focused on highly ordered interfaces of 2x2 reconstruction on the 100 surface of the III-Vs. Further studies could expand this research on the gap states by focusing on defects at the interface.

We used the models from the previous study along with additional native oxide models to get computational core-level shifts. These theoretical results were then used to interpret experimental X-ray photoelectron spectroscopy data. In our data the group-V shifts behaved similarly to silicon, where the shifts of an atom increase with higher oxidation. However, the group-III shifts behaved irregularly, and further studies would be needed to fully understand the behavior.

In the silicon and silicon dioxide structure optimization we used a genetic hybrid algorithm to optimize the atomic structure using Tersoff potential. The algorithm relied heavily on basin hopping and different mutations to find a path from a high energy random starting structure to a low energy structure. Especially the silicon dioxide exhibited a large variety of different low energy structures. This included both disordered and ordered structures for both silicon and silicon dioxide. The known ground state structure for both was also reproduced: cubic diamond for silicon and α-quartz for silicon dioxide. With both cases, our research was contained in relatively small structures for the sake of easier analysis and development. In the future, the algorithm could be extended larger structures too, where there would be even larger variety in the low energy structures. Our goal with this line of research was to build different Si/SiO_2 structures, which would be the next logical step in the research and combine the two previous research topics.

List of References

[1] J. Robertson, Y. Guo, and L. Lin. Defect state passivation at iii-v oxide interfaces for complementary metal–oxide–semiconductor devices. *Journal of Applied Physics*, 117(11):112806, 2015. doi: 10.1063/1.4913832.

[2] Jesús Alamo. Nanometre-scale electronics with iii-v compound semiconductors. *Nature*, 479: 317–23, 11 2011. doi: 10.1038/nature10677.

[3] John W. Orton. *The Story of Semiconductors*. Oxford University Press, 1 edition, 2008.

[4] E E Haller. Hydrogen in crystalline semiconductors. *Semiconductor Science and Technology*, 6 (2):73–84, feb 1991. doi: 10.1088/0268-1242/6/2/001. URL `https://doi.org/10.1088/0268-1242/6/2/001`.

[5] M. Stutzmann and J. Chevallier, editors. *Hydrogen in Semiconductors*. Elsevier, Amsterdam, 1991. ISBN 978-0-444-89138-9. doi: https://doi.org/10.1016/B978-0-444-89138-9.50001-2. URL `https://www.sciencedirect.com/science/article/pii/B9780444891389500012`.

[6] Hyunhyub Ko, Kuniharu Takei, Rehan Kapadia, Steven Chuang, Hui Fang, Paul Leu, Kartik Ganapathi, Elena Plis, Ha Kim, Szu-Ying Chen, Morten Madsen, Alexandra Ford, Yu-Lun Chueh, S. Krishna, Sayeef Salahuddin, and Ali Javey. Ultrathin compound semiconductor on insulator layers for high-performance nanoscale transistors. *Nature*, 468:286–9, 11 2010. doi: 10.1038/nature09541.

[7] M. P. J. Punkkinen, P. Laukkanen, J. Lång, M. Kuzmin, M. Tuominen, V. Tuominen, J. Dahl, M. Pessa, M. Guina, K. Kokko, J. Sadowski, B. Johansson, I. J. Väyrynen, and L. Vitos. Oxidized in-containing iii-v(100) surfaces: Formation of crystalline oxide films and semiconductor-oxide interfaces. *Phys. Rev. B*, 83:195329, May 2011. doi: 10.1103/PhysRevB.83.195329. URL `https://link.aps.org/doi/10.1103/PhysRevB.83.195329`.

[8] C. H. Wang, S. W. Wang, G. Doornbos, G. Astromskas, K. Bhuwalka, R. Contreras-Guerrero, M. Edirisooriya, J. S. Rojas-Ramirez, G. Vellianitis, R. Oxland, M. C. Holland, C. H. Hsieh, P. Ramvall, E. Lind, W. C. Hsu, L.-E. Wernersson, R. Droopad, M. Passlack, and C. H. Diaz. Inas hole inversion and bandgap interface state density of 2x1011 at hfo2 inas interfaces. *Applied Physics Letters*, 103(14):143510, 2013. doi: 10.1063/1.4820477.

[9] Lachlan Black, Alessandro Cavalli, Marcel Verheijen, J.E.M. Haverkort, Erik Bakkers, and Wilhelmus Kessels. Effective surface passivation of inp nanowires by atomic-layer-deposited al2o3 with pox interlayer. *Nano letters*, 17, 09 2017. doi: 10.1021/acs.nanolett.7b02972.

[10] M.P.J. Punkkinen, A. Lahti, J. Huhtala, J.-P. Lehtiö, Z.J. Rad, M. Kuzmin, P. Laukkanen, and K. Kokko. Stabilization of unstable and metastable inp native oxide thin films by interface effects. *Applied Surface Science*, 567:150848, 2021. ISSN 0169-4332. doi: https://doi.org/10.1016/j.apsusc.2021.150848. URL `https://www.sciencedirect.com/science/article/pii/S0169433221019097`.

[11] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev. B*, 136:864, 1964.

[12] Joel Gersten and Frederick Smith. *The Physics and Chemistry of Materials*. Wiley, 21 edition, 2001.

[13] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927. doi: https://doi.org/10.1002/andp.19273892002. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19273892002`.

[14] S. Kurth and John Perdew. Role of the exchange-correlation energy: Nature's glue. *International Journal of Quantum Chemistry*, 77:814 – 818, 04 2000. doi: 10.1002/(SICI)1097-461X(2000)77:5⟨814::AID-QUA3⟩3.0.CO;2-F.

[15] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965. doi: 10.1103/PhysRev.140.A1133. URL https://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[16] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58:1200, 1980. doi: https://doi.org/10.1139/p80-159.

[17] Lee A. Cole and J. P. Perdew. Calculated electron affinities of the elements. *Phys. Rev. A*, 25:1265–1271, Mar 1982. doi: 10.1103/PhysRevA.25.1265. URL https://link.aps.org/doi/10.1103/PhysRevA.25.1265.

[18] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45:13244–13249, Jun 1992. doi: 10.1103/PhysRevB.45.13244. URL https://link.aps.org/doi/10.1103/PhysRevB.45.13244.

[19] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, May 1981. doi: 10.1103/PhysRevB.23.5048. URL https://link.aps.org/doi/10.1103/PhysRevB.23.5048.

[20] M. Ernzerhof, J. P. Perdew, and K. Burke. Coupling-constant dependence of atomization energies. *International Journal of Quantum Chemistry*, 64(3):285–295, 1997.

[21] W.R.L. Lambrecht. *Rare Earth and Transition Metal Doping of Semiconductor Materials*. Woodhead Publishing, 2016.

[22] John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, 577(1):1–20, 2001. doi: 10.1063/1.1390175. URL https://aip.scitation.org/doi/abs/10.1063/1.1390175.

[23] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47(1):558 – 561, 1993.

[24] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B*, 49:14251–14269, May 1994. doi: 10.1103/PhysRevB.49.14251. URL https://link.aps.org/doi/10.1103/PhysRevB.49.14251.

[25] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996. ISSN 0927-0256. doi: https://doi.org/10.1016/0927-0256(96)00008-0. URL https://www.sciencedirect.com/science/article/pii/0927025696000080.

[26] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996. doi: 10.1103/PhysRevB.54.11169. URL https://link.aps.org/doi/10.1103/PhysRevB.54.11169.

[27] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994. doi: 10.1103/PhysRevB.50.17953. URL https://link.aps.org/doi/10.1103/PhysRevB.50.17953.

[28] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, Jan 1999. doi: 10.1103/PhysRevB.59.1758. URL https://link.aps.org/doi/10.1103/PhysRevB.59.1758.

[29] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511805769.

[30] J. E. Castle. Practical surface analysis by auger and x-ray photoelectron spectroscopy. d. briggs and m. p. seah (editors). john wiley and sons ltd, chichester, 1983, 533 pp., £44.50. *Surface and Interface Analysis*, 6(6):302–302, 1984. doi: https://doi.org/10.1002/sia.740060611. URL https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/sia.740060611.

[31] M. Weiner and R. E. Watson. Core-level shifts in bulk alloys and surface adlayers. *Phys. Rev. B*, 51(23):17168, 1995.

[32] W. Olovsson, C. Göransson, T. Marten, and I. A. Abrikosov. Core-level shifts in complex metallic systems from first principle. *physica status solidi (b)*, 243(11):2447–2464, 2006.

[33] Noèlia Bellafont, Paul Bagus, and Francesc Illas. Prediction of core level binding energies in density functional theory: Rigorous definition of initial and final state contributions and implications on the physical meaning of kohn-sham energies. *The Journal of chemical physics*, 142:214102, 06 2015. doi: 10.1063/1.4921823.

[34] A. Pasquarello, M. S. Hybertsen, and R. Car. Si2pcore-level shifts at the si(001)-sio2interface: A first-principles study. *Phys. Rev. Letters*, 74(6):1024–1027, 1995.

[35] A.C.T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. Reaxff: A reactive force field for hydrocarbons. *Journal of Physical Chemistry*, A105:9396 – 9409, 2001.

[36] F. H. Streitz and J. W. Mintmire. Electrostatic potentials for metal-oxide surfaces and interfaces. *Phys. Rev. B*, 50:11996, 1994.

[37] Shinji Munetoh, Teruaki Motooka, Koji Moriguchi, and Akira Shintani. Interatomic potential for si-o systems using tersoff parameterization. *Computational Materials Science*, 39(2):334–339, 4 2007. ISSN 0927-0256. doi: 10.1016/j.commatsci.2006.06.010.

[38] M. Hong, J. Kwo, A. R. Kortan, J. P. Mannaerts, and A. M. Sergent. Epitaxial cubic gadolinium oxide as a dielectric for gallium arsenide passivation. *Science*, 283(5409):1897–1900, 1999. doi: 10.1126/science.283.5409.1897. URL https://www.science.org/doi/abs/10.1126/science.283.5409.1897.

[39] Matty Caymax, Guy Brammertz, Annelies Delabie, Sonja Sioncke, Dennis Lin, Marco Scarrozza, Geoffrey Pourtois, Wei-E Wang, Marc Meuris, and Marc Heyns. Interfaces of high-k dielectrics on gaas: Their common features and the relationship with fermi level pinning (invited paper). *Microelectron. Eng.*, 86(7–9):1529–1535, jul 2009. ISSN 0167-9317. doi: 10.1016/j.mee.2009.03.090. URL https://doi.org/10.1016/j.mee.2009.03.090.

[40] É. O'Connor, S. Monaghan, R. D. Long, A. O'Mahony, I. M. Povey, K. Cherkaoui, M. E. Pemble, G. Brammertz, M. Heyns, S. B. Newcomb, V. V. Afanas'ev, and P. K. Hurley. Temperature and frequency dependent electrical characterization of hfo2 ingaas interfaces using capacitance-voltage and conductance methods. *Applied Physics Letters*, 94(10):102902, 2009. doi: 10.1063/1.3089688.

[41] Y. C. Chang, W. H. Chang, C. Merckling, J. Kwo, and M. Hong. Inversion-channel gaas(100) metal-oxide-semiconductor field-effect-transistors using molecular beam deposited al2o3 as a gate dielectric on different reconstructed surfaces. *Applied Physics Letters*, 102(9):093506, 2013. doi: 10.1063/1.4793433.

[42] John Robertson and Robert M. Wallace. High-k materials and metal gates for cmos applications. *Materials Science and Engineering: R: Reports*, 88:1–41, 2015. ISSN 0927-796X. doi: https://doi.org/10.1016/j.mser.2014.11.001. URL https://www.sciencedirect.com/science/article/pii/S0927796X14001168.

[43] J. Wallentin. Inp nanowire array solar cells achieving 13.8% efficiency by exceeding the ray optics limit. *Science*, 339, 2013. doi: 10.1126/science.1230969. URL https://doi.org/10.1126/science.1230969.

[44] S. Z. Oener. Charge carrier-selective contacts for nanowire solar cells. *Nat. Comm.*, 9, 2018. doi: 10.1038/s41467-018-05453-5. URL https://doi.org/10.1038/s41467-018-05453-5.

[45] Oktyabrsky, s. & ye, p. d. fundamentals of iii-v semiconductor mosfets (springer 2010).

[46] C. L. Chen. Wafer-scale 3d integration of ingaas photodiode arrays with si readout circuits by oxide bonding and through-oxide vias, microelectr. *Engineer.*, 88, 2011.

[47] J. Yang. Low leakage of in0.83ga0.17as photodiode with al2o3/sinx stacks, infrar. *Phys. & Techn.*, 71, 2015. doi: 10.1016/j.infrared.2015.04.003. URL https://doi.org/10.1016/j.infrared.2015.04.003.

[48] Alberto Debernardi, Claudia Wiemer, and Marco Fanciulli. Epitaxial phase of hafnium dioxide for ultrascaled electronics. *Phys. Rev. B*, 76:155405, Oct 2007. doi: 10.1103/PhysRevB.76.155405. URL https://link.aps.org/doi/10.1103/PhysRevB.76.155405.

[49] G-M Rignanese. 17(7):R357–R379, feb 2005. doi: 10.1088/0953-8984/17/7/r03. URL https://doi.org/10.1088/0953-8984/17/7/r03.

[50] J Benton, C Doherty, S Ferris, Daniel Flamm, L Kimerling, and H Leamy. Hydrogen passivation of point defects in silicon. *Applied Physics Letters*, 36:670, 04 1980. doi: 10.1063/1.91619.

[51] E. Kaxiras, Y. Bar-Yam, J. D. Joannopoulos, and K. C. Pandey. Ab initio theory of polar semiconductor surfaces. ii. (22) reconstructions and related phase transitions of gaas($\bar{1}\bar{1}\bar{1}$). *Phys. Rev. B*, 35:9636–9643, Jun 1987. doi: 10.1103/PhysRevB.35.9636. URL https://link.aps.org/doi/10.1103/PhysRevB.35.9636.

[52] Guo-Xin Qian, Richard M. Martin, and D. J. Chadi. First-principles study of the atomic reconstructions and energies of ga- and as-stabilized gaas(100) surfaces. *Phys. Rev. B*, 38:7649–7663, Oct 1988. doi: 10.1103/PhysRevB.38.7649. URL https://link.aps.org/doi/10.1103/PhysRevB.38.7649.

[53] A. Lahti, H. Levämäki, J. Mäkelä, M. Tuominen, M. Yasir, J. Dahl, M. Kuzmin, P. Laukkanen, K. Kokko, and M.P.J. Punkkinen. Electronic structure and relative stability of the coherent and semi-coherent hfo2/iii-v interfaces. *Applied Surface Science*, 427:243–252, 2018. ISSN 0169-4332. doi: https://doi.org/10.1016/j.apsusc.2017.08.185. URL https://www.sciencedirect.com/science/article/pii/S0169433217325618.

[54] Jeff Horen. Linear programming, by katta g. murty, john wiley & sons, new york, 1983, 482 pp. *Networks*, 15(2):273–274, 1985. doi: https://doi.org/10.1002/net.3230150211. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/net.3230150211.

[55] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, 2e edition, 2006.

[56] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49:409–436, 1952.

[57] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220 (4598):671–680, 1983. doi: 10.1126/science.220.4598.671. URL https://www.science.org/doi/abs/10.1126/science.220.4598.671.

[58] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002. doi: 10.1073/pnas.202427399.

[59] J. Kennedy and R. Eberhart. Particle swarm optimization. 4:1942–1948 vol.4, 1995. doi: 10.1109/ICNN.1995.488968.

[60] A.R. Oganov and C.W. Glas. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of Chemical Physics*, 124:244704, 2006.

[61] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. Calypso: A method for crystal structure prediction. *Comput. Phys. Commun.*, 183:2063, 2012.

[62] Tipton WW and Hennig RG. A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing empirical potentials. *J Phys Cond Matter*, 25:495401, 2013.

[63] Ralf Östermark. A multipurpose parallel genetic hybrid algorithm for non-linear non-convex programming problems. *European Journal of Operational Research*, 152(1):195 – 214, 2004. doi: https://doi.org/10.1016/S0377-2217(02)00672-0.

[64] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.*, 271:108171, 2022. doi: 10.1016/j.cpc.2021.108171.

[65] D. J. Wales and J. P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *J. Phys. Chem.*, A 101:5111, 1997.

[66] Weichao Wang, Ka Xiong, Geunsik Lee, Min Huang, Robert M. Wallace, and Kyeongjae Cho. Origin of hfo2/gaas interface states and interface passivation: A first principles study. *Applied Surface Science*, 256(22):6569–6573, 2010. ISSN 0169-4332. doi: https://doi.org/10.1016/j.apsusc.2010.04.048. URL https://www.sciencedirect.com/science/article/pii/S016943321000560X.

[67] Weichao Wang, Ka Xiong, Robert M. Wallace, and Kyeongjae Cho. Impact of interfacial oxygen content on bonding, stability, band offsets, and interface states of gaas:hfo2 interfaces. *The Journal of Physical Chemistry C*, 114(51):22610–22618, 2010. doi: 10.1021/jp107880r.

[68] Santosh KC, Hong Dong, Roberto C. Longo, Weichao Wang, Ka Xiong, Robert M. Wallace, and Kyeongjae Cho. Electronic properties of inp (001)/hfo2 (001) interface: Band offsets and oxygen dependence. *Journal of Applied Physics*, 115(2):023703, 2014. doi: 10.1063/1.4861177.

[69] M. D. Pashley. Electron counting model and its application to island structures on molecular-beam epitaxy grown gaas(001) and znse(001). *Phys. Rev. B*, 40:10481–10487, Nov 1989. doi: 10.1103/PhysRevB.40.10481. URL https://link.aps.org/doi/10.1103/PhysRevB.40.10481.

[70] P. W. Peacock and J. Robertson. Bonding, energies, and band offsets of $\mathrm{Si}-\mathrm{zro}_2$ and hfo_2 gate oxide interfaces. *Phys. Rev. Lett.*, 92:057601, Feb 2004. doi: 10.1103/PhysRevLett.92.057601. URL https://link.aps.org/doi/10.1103/PhysRevLett.92.057601.

[71] L. Lin and J. Robertson. Defect states at iii-v semiconductor oxide interfaces. *Applied Physics Letters*, 98(8):082903, 2011. doi: 10.1063/1.3556619.

[72] Jaakko Mäkelä, Antti Lahti, Marjukka Tuominen, Muhammad Yasir, M. Kuzmin, Pekka Laukkanen, Kalevi Kokko, M. Punkkinen, Hong Dong, Barry Brennan, and Robert Wallace. Unusual oxidation-induced core-level shifts at the hfo2/inp interface. *Scientific Reports*, 9, 02 2019. doi: 10.1038/s41598-018-37518-2.

[73] F. J. Himpsel, F. R. McFeely, A. Taleb-Ibrahimi, J. A. Yarmoff, and G. Hollinger. Microscopic structure of the sio2/si interface. *Phys. Rev. B*, 38, 1988. doi: 10.1103/PhysRevB.38.6084. URL https://doi.org/10.1103/PhysRevB.38.6084.

[74] A. Lahti, R. Östermark, and K. Kokko. Optimizing atomic structures through geno-mathematical programming. *Commun. Comput. Phys.*, 25:911–927, 2019.

[75] Antti Lahti, Ralf Östermark, and Kalevi Kokko. Optimization of sio2 with gha and basin hopping. *Computational Materials Science*, page 111011, 2021. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2021.111011. URL https://www.sciencedirect.com/science/article/pii/S0927025621006984.

Original Publications

Antti Lahti & Henrik Levämäki & Jaakko Mäkelä & Marjukka Tuominen & Muhammad Yasir & Mikhail Kuzmin & Pekka Laukkanen & Kalevi Kokko & Marko Punkkinen

Electronic structure and relative stability of the coherent and semi-coherent HfO2/III-V interfaces

Applied Surface Science, 427, 2018, page numbers

I

§

Applied Surface Science 427 (2018) 243–252

Contents lists available at ScienceDirect

Applied Surface Science

journal homepage: www.elsevier.com/locate/apsusc

Full Length Article

Electronic structure and relative stability of the coherent and semi-coherent HfO_2/III-V interfaces

A. Lahti, H. Levämäki, J. Mäkelä, M. Tuominen, M. Yasir, J. Dahl, M. Kuzmin, P. Laukkanen, K. Kokko, M.P.J. Punkkinen*

Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland

ARTICLE INFO

Article history:
Received 31 May 2017
Received in revised form 14 August 2017
Accepted 27 August 2017
Available online 30 August 2017

Keywords:
Semiconductors
Interfaces
DFT
Bonding
Defect
Band gap

ABSTRACT

III-V semiconductors are prominent alternatives to silicon in metal oxide semiconductor devices. Hafnium dioxide (HfO_2) is a promising oxide with a high dielectric constant to replace silicon dioxide (SiO_2). The potentiality of the oxide/III-V semiconductor interfaces is diminished due to high density of defects leading to the Fermi level pinning. The character of the harmful defects has been intensively debated. It is very important to understand thermodynamics and atomic structures of the interfaces to interpret experiments and design methods to reduce the defect density. Various realistic gap defect state free models for the HfO2/III-V(100) interfaces are presented. Relative energies of several coherent and semi-coherent oxide/III-V semiconductor interfaces are determined for the first time. The coherent and semi-coherent interfaces represent the main interface types, based on the Ga-O bridges and As (P) dimers, respectively.

Results: show that interface energy depends sensitively on the type and position of the defects and the atomic structure of the interface. Various coherent interfaces are stable and have band gaps free of defect states in spite of the interfacial structural defects. The semi-coherent interfaces include harmful As dimers and As dangling bonds. If kinetics contributes via the layer by layer oxide growth, the semi-coherent interfaces are formed under the experimentally relevant O-rich growth conditions. This is explained by the basic interfacial structural motifs and the electron counting rule (ECR). An oxidized (3×1) substrate has previously been used to decrease interface defect gap state density. A scenario, which explains why the oxidized substrate leads to a relatively small interface defect density, is presented.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The semiconductor-oxide silicon/silicon dioxide (Si/SiO_2) interface is perhaps the most important interface used in modern electronics. However, alternative interface materials are needed to get more efficient electronic components in future. The HfO_2 is the most promising oxide for semiconductor-oxide interfaces which has been already used for semiconductor devices [1]. Many III–V semiconductors (like gallium arsenide and indium phosphide) have much larger electron mobilities than silicon has. However, it is very difficult to prepare good quality III-V/oxide interfaces, comparable to the Si/SiO_2, *e.g.*, Refs. [2–6]. Especially, the Fermi level is easily pinned in the band gap within the III-V/oxide interfaces due to a high density of interface defects. This phenomenon makes the interfaces useless for microprocessor applications. On the other hand III-V/oxide defects also degrade operation of optoelectronic devices via non-radiative surface recombination.

It is difficult to investigate experimentally buried interfaces. For example, the interpretation of the results obtained by photoemission core-level spectroscopy is quite challenging without model structures. Different kinds of atomic environments can cause similar core-level shifts. On the other hand, the number of *a priori* possible interface structures is practically infinite. Therefore, it is almost impossible to construct realistic III-V/oxide model interface structures using experimental data only. However, proper model interface structures are needed to assess various properties of the interfaces. Computational results are valuable for these purposes.

Knowledge of the atomic structures and interface energies of the oxide/III-V semiconductor interfaces is extremely small compared, *e.g.*, to the vast literature on the III–V semiconductor surfaces. The oxide/III-V interfaces are especially difficult to model. The lattice parameters of III–V semiconductors and oxides generally deviate from each other (lattice mismatch). Furthermore, it is difficult to form III-V/oxide interfaces, which would obey the ECR. The latter

* Corresponding author.
E-mail address: marpunk@utu.fi (M.P.J. Punkkinen).

http://dx.doi.org/10.1016/j.apsusc.2017.08.185
0169-4332/© 2017 Elsevier B.V. All rights reserved.

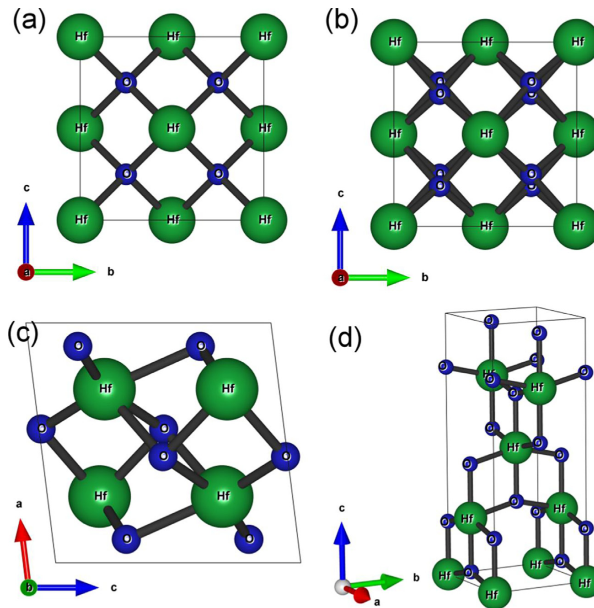


Fig. 1. Illustrations of face-centered cubic (a), tetragonal (b), monoclinic (c) and body-centered tetragonal (d) HfO_2 lattice structures. The Hf and O atoms are shown by green and blue spheres, respectively.

problem originates both from the lattice mismatch and the valence electron numbers of the III–V semiconductors. The same reasons are supposed to make the growth of good quality III-V/oxide interfaces difficult. The oxidation of the III–V semiconductors is challenging also due to the lack of good quality native oxide and the two-component nature of the III-Vs. The oxide can adopt various forms at the interface. The interface structure has to conform to the growing oxide. Otherwise, the interface is quite unstable. This means often that qualitatively different interface structures require qualitatively different oxide structures. This complicates interface modeling significantly. Several qualitatively different interfaces are considered in the present study.

The ideal HfO_2/III-V interface is a prominent example of an interface which does not obey the ECR. The coherent interface is determined to mean here that both the oxide and the semiconductor have the same bulk atomic configuration, disregarding the chemical species, within the interfacial layer. The ideal interface is coherent and does not have defects. Substitutional non-isoelectronic atoms (the number of the valence electrons is not the same), vacancies and dimers are defects.

The ECR is basically the same rule as the octet rule. However, the ECR or the electron counting model (ECM) can also be identified with the rules to count surface or interface bonds. These rules are not trivial, because generally there are surface and interface bonds not found in bulk. The ECR for semiconductor surfaces specify, how the surface or interface dangling bonds are counted [7]. The ECR obeying surface has an energy gap while the ECR breaking surface is metallic. The ECRs have been expressed for ionic oxides and Si [8] and III-Vs [9]. The ECR for interfaces between ionic oxides (like HfO_2) and covalent III–V semiconductors considered in this study are given in other words below.

In this work we concentrate on the HfO_2/III-V(100) interfaces. These interfaces have been modeled using a coherent interface model [10,11] and a semi-coherent interface model [12,13]. Previously used coherent interface model is based on tetragonal HfO_2 and shows a significant lattice mismatch for all III–V semiconductors [10,11] making the model quite unsTable Still, two model structures do not induce defect gap states [10,11]. The semi-coherent interface for HfO_2/GaAs shows negligible lattice mismatch, but leads to defect states within the band gap due to As-As dimers and Ga dangling bonds [12].

The aim of the present investigation is to study the relative stability of several qualitatively different interfaces and find realistic interface models free of defect states. It should be noted that Ga-O-Ga bridges and group V dimers are the basic structural motifs under the relevant relatively O-rich experimental conditions. The relative stability of the main interface types based on these basic structural motifs is unknown so far. These main interface types represent coherent and semi-coherent interfaces. Three different oxides are used in structural models. All structural interface models presented and considered are new. It is also important to find interface models free of gap states which do not have large lattice mismatch, because these energetically favorable interface structures can be grown. It was recently suggested that the existence of the amphoteric defects, group V atom dimers and group V atom dangling bonds, leads to the Fermi level pinning [14]. Therefore, it is also important to understand the relative stabilities of the interfaces with (semi-coherent) and without (coherent) group V dimers with respect to the Fermi level pinning. In both cases interfacial Ga atoms are bonded to oxygen atoms. The Ga oxides are more likely to form than the As oxides at the interface due to their smaller formation energies. Defect free gap interface models

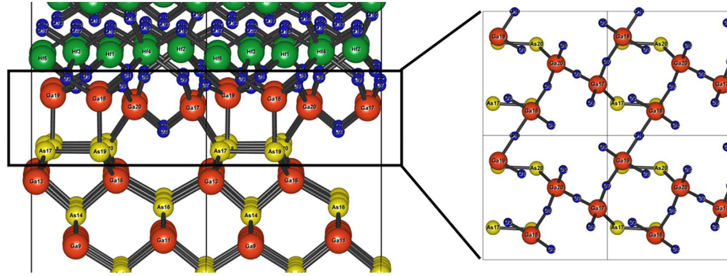


Fig. 2. The semi-coherent O10 HfO_2/GaAs interface structure. There are 10 interfacial O atoms per (2×2) interface area. The positions of the interfacial O, Ga, and As atoms are also shown at the interface plane. The Hf, O, Ga, and As atoms are shown by green, blue, red, and yellow spheres, respectively.

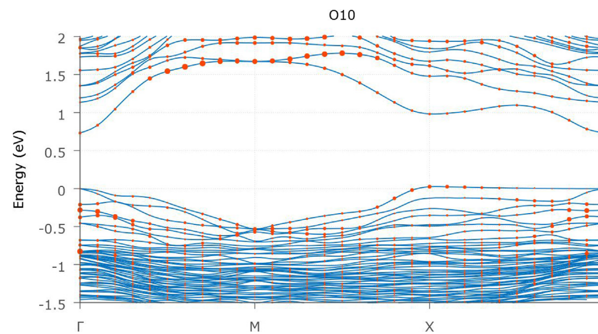


Fig. 3. The band structure of the semi-coherent O10 HfO_2/GaAs interface structure along the high symmetry lines of the 2-dimensional Brillouin zone. Red spheres show the relative weight at the dimer atoms. Zero energy corresponds to the highest occupied state.

have been presented for the three-valent metal oxide/III-V interfaces [2,10,14,15], which obey more easily the ECR. Calculations are performed for the HfO_2/GaAs and HfO_2/GaP interfaces.

2. Computational details

Calculations were performed using an *ab initio* density functional theory (DFT) total energy method within the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) [16] and the local density approximation (LDA) [17,18]. Most of the calculations were performed using the PBE functional. The approach is based on the plane wave basis and projector augmented wave method [19,20] (Vienna *ab initio* simulation package, VASP) [21–24]. The optimization of the atomic structure was performed using the conjugate gradient minimization of the total energy with respect to the atomic coordinates. Atoms were relaxed until the remaining forces were less than 20 meV/Å. The plane wave cutoff energy of 350 eV was used. All test calculations with the cutoff energy of 400 eV showed only marginal differences for the relative interface energies. The As, Ga and P 3*d* as well as Hf 5*p* electrons were treated as core electrons. The *k* point sampling was carried out by the Monkhorst-Pack scheme [25] using a $4 \times 4 \times 1$ mesh. The density of states was calculated using a $8 \times 8 \times 2$ mesh. The origin was shifted to the Γ point.

Both slab unit cells including vacuum and cells with two equal interfaces without vacuum were used. Slab cells including vacuum were used to get total energies and relative interface energies of the systems. The structural optimization is more difficult without vacuum. The interface area is fixed to the substrate (III–V semiconductor) surface area. If vacuum is not used, the cell length perpendicular to the interface has to be optimized manually using a set of different cell lengths. This is laborious, because many different interface systems are considered. Furthermore, the same amount of group III and group V atoms can be used for the coherent and semi-coherent structures, if vacuum is introduced, which makes energetic comparison much easier. On the other hand, experience has shown that total energy is obtained in a good accuracy by the slab unit cells.

Cells with two equal interfaces were used to calculate band structure and density of states (DOS). The magnitude of the energy gap is overestimated with both cell types due to quantum confinement in the direction perpendicular to the interface. Even with 200 Ga and As layers and pseudohydrogen atoms the band gap of the GaAs (0.20 eV), calculated using vacuum, is still larger than the bulk value (0.17 eV) within the PBE. (If the experimental lattice constant is used and the Ga 3*d* states are treated as valence electrons as in the Ref. [26], the magnitude of the band gap within the PBE is 0.56 eV in good agreement with the value of the Ref. [26] which is 0.43 eV.) Therefore, it is not reliable to make conclusions about defect gap states based on the magnitude of the gap. The interface and bulk states have to be identified using the results of the same (interface) calculation. The interpretation of the band structure may be easier, if the vacuum is not used, because then there are two equal interfaces, no pseudohydrogenized surface, and bulk can be identified with the central part of the III–V semiconductor. The calculationally heavier methods, like the hybrid functionals, are beyond the scope

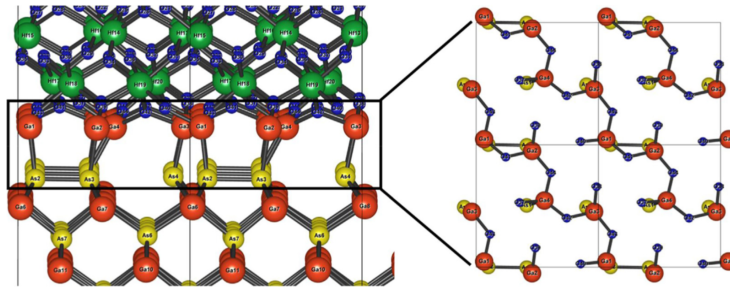


Fig. 4. The semi-coherent O8-strained HfO_2/GaAs interface structure. The positions of the interfacial O, Ga, and As atoms are also shown at the interface plane. The Hf, O, Ga, and As atoms are shown by green, blue, red, and yellow spheres, respectively.

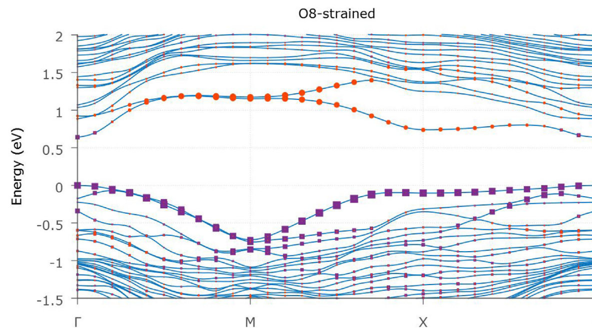


Fig. 5. The band structure of the semi-coherent O8-strained HfO_2/GaAs interface structure along the high symmetry lines of the 2-dimensional Brillouin zone. Red spheres and purple squares show the relative weight at the dimer and dangling bond atoms, respectively. Zero energy corresponds to the highest occupied state.

of the present study. Calculations without vacuum are faster due to smaller cell size and the structural convergence is also obtained with a less amount of iterations.

The unit cells consist of 5-8 layers of group III atoms, 5-7 layers of group V atoms, 5 layers of Hf atoms, and 6 layers of O atoms. The bottom layer group V atoms of the unit cells with vacuum are passivated by pseudohydrogen atoms. The interface area of the III–V semiconductor is (2×2) for most of the unit cells. The HfO_2 part adapts to the surface area of the III–V semiconductor substrate. The total energies are calculated using thinner semiconductor parts.

The most stable form of the HfO_2 has a simple monoclinic structure (space group 14). However, due to the large lattice mismatch monoclinic HfO_2 can not have a coherent interface with III–V semiconductors. This is true also for simple tetragonal HfO_2 (space group 137). [Face-centered cubic HfO_2 (space group 225) is less stable than the monoclinic and tetragonal HfO_2.] Therefore, the monoclinic and tetragonal HfO_2 can have only a semi-coherent interface with the III–V semiconductors. Tetragonal oxide is used to form semi-coherent interfaces, because the tetragonal HfO_2 has a more simple structure than the monoclinic HfO_2 has. Furthermore, larger interface area is needed for the monoclinic HfO_2 which makes calculations much heavier. The anatase HfO_2 (body-centered tetragonal; space group 141) can be taken as distorted monoclinic HfO_2 under strain induced by the III–V semiconductor substrate [27]. Wyckoff positions 4a and 8e ($z = -0.203$) are occupied (origin choice 1). The coherent interface is based on the anatase [27–29]. The various bulk structures of the HfO_2 are shown in Fig. 1.

The c lattice parameter of the tetragonal HfO_2 is directed perpendicular to the interface. The a lattice parameter within the PBE is 4.954 Å. The cutoff energy increase from 350 eV to 400 eV increases, *e.g.*, the a lattice parameter by 1.4%, but affects only slightly the relative stabilities of the different HfO_2 phases. The lattice mismatch [30] for the semi-coherent interface used in the Ref. [12] is 4.1% for GaAs (interface lattice parameter is 4.073 Å) and 0.2% for GaP (interface lattice parameter is 3.915 Å). The lattice mismatch must not be too large, because large lattice mismatch means that the used semi-coherent model is unrealistic. Therefore, we do not consider, *e.g.*, HfO_2/InP interfaces in this study. The lattice mismatch of the coherent interface is 3.3% for GaAs and 0.7% for GaP. The a lattice parameter of the anatase HfO_2 is 3.942 Å within the PBE. The experimental structural parameters for the cubic, tetragonal, and monoclinic HfO_2 phases are given in the Ref. [31].

3. Results and discussion

3.1. Interface states

The semi-coherent interface including tetragonal oxide can have various structures, because the number of Ga atoms per interface area differs from the number of Hf atoms. This means that the interfacial Ga atoms can have different patterns. Different patterns were considered. Fig. 2 shows the structure leading to the smallest total energy. The As dimers form dimer rows. There are not any defect states within band gap of this interface structure, in contrary to the

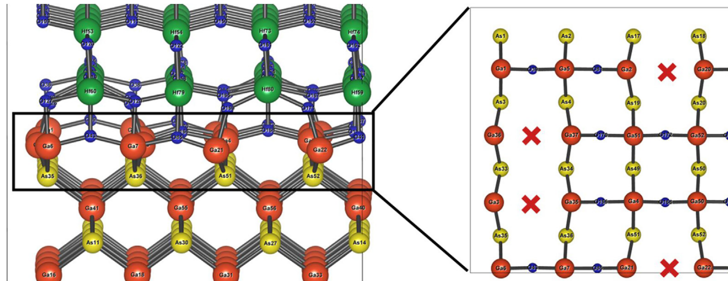


Fig. 6. The coherent OV HfO_2/GaAs interface structure. The positions of the interfacial O, Ga, and As atoms are also shown at the interface plane. The oxygen vacancies which break the Ga-O-Ga chains are denoted by red crosses. The Hf, O, Ga, and As atoms are shown by green, blue, red, and yellow spheres, respectively.

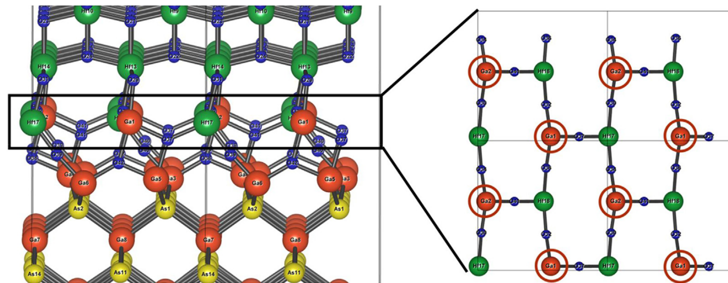


Fig. 7. The coherent OB-GHS HfO_2/GaAs interface structure. The positions of the interfacial Hf, O, and substitutional Ga atoms above the oxygen bridges are also shown at the interface plane. The substitutional Ga atoms are shown by red circles. The Hf, O, Ga, As atoms are shown by green, blue, red, and yellow spheres, respectively.

O10 structure presented in Ref. [12]. The present semi-coherent interface structure is called O10, although it is not the O10 interface presented in Ref. [12]. This interface does not include Ga dangling bonds or any other defects except the group V dimers. The interface presented in the Ref. [12] was reported to have Ga dangling bonds in addition to the As dimers. Fig. 3 shows the band structure of the O10. Some bands especially within the conduction band have significant weight from the dimer atoms. The group V unoccupied dimer bands are located within the conduction band at the Al_2O_3/GaAs interfaces [15].

The present semi-coherent interface satisfies the ECR, which is a mandatory condition to obtain an energy gap. However, the ECR satisfaction does not mean that there are not defect gap states within the bulk band gap. To assess the satisfaction of the ECR it is assumed that the interface region can be divided into a "semiconductor part" and the (ionic) "oxide part". This is true, at least, for the interface models considered in this study. Every atom in the semiconductor part and every oxygen atom has to possess eight valence electrons. The atoms in the semiconductor part have covalent or "shared" bonds which are formed by an electron pair. Interface dangling bond states are also possible. In general, the oxygen atoms in the oxide part receive (and not share) electrons from other atoms. All other atoms in the oxide part donate all valence electrons. These principles express the ECR for the HfO_2/III-V interfaces considered in this study. As usual, they are not intended to describe real charge distributions. This is basically due to the fact that "bond" is not a well-defined quantity, but "one-electron state" is.

The group III atoms occupy Hf sites in the oxide part. The interfacial group III atoms just above the uppermost As atoms can be imagined to be at the boundary of the semiconductor and oxide parts with respect to the ECR. Thus, the group III atoms have both covalent and ionic bonds within the ECR. However, in contrary to the coherent interface, the group III atoms do not occupy bulk positions of the semiconductor part at the semi-coherent interface.

The O8-strained interface is another semi-coherent interface type considered. It has the same kind of bonding as the O10, but the oxide is strained having tetragonal lattice parameters [32,33]. In fact, all oxides are more or less strained on the III–V substrate. However, the O8-strained relaxes to the simple tetragonal (space group 137), if the strain is removed. The space group of the strained bulk oxide is 60 and Wyckoff positions 4c ($y = -0.152$) and 8d ($x = -0.079$, $y = 0.123$, and $z = 0.222$) are occupied. (The internal parameters are calculated for the *a* lattice parameter equal to 5.760 Å). The space group number and the Wyckoff positions were found by the program FINDSYM [34]. Thus 4.073 Å → 5.760 Å [substrate (primitive) lattice parameter was mixed with the oxide lattice parameter]. This interface structure is shown in Fig. 4. The coordination numbers in the O8-strained oxide decrease in partial analogy to the monoclinic structure [32] and the oxide is also called epi(epitaxial)-monoclinic structure [33]. The number of O atoms per (2×2) interface area is 8. It is noted that the O8-strained structure includes one broken As dimer and two As dangling bonds per (2×2) interface area. The As dimer breaking occurs to allow the interface to obey the ECR. Fig. 5 shows the band structure of the O8-strained interface. Some bands show significant weight from the As dimer and dangling bond atoms. The dimer states are within the conduction band and the dangling bond states are at the top of the valence band. Dangling bonds induce defect gap states as shown in Fig. 5.

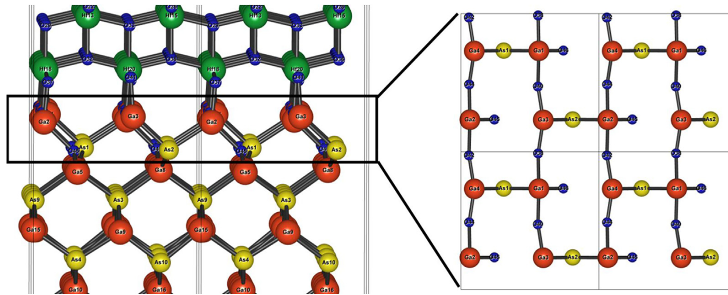


Fig. 8. The coherent OAS HfO_2/GaAs interface structure. The positions of the interfacial O, Ga, and As atoms are also shown at the interface plane. The unbroken Ga-O-Ga chains can be noted. The Hf, O, Ga and As atoms are shown by green, blue, red, and yellow spheres, respectively.

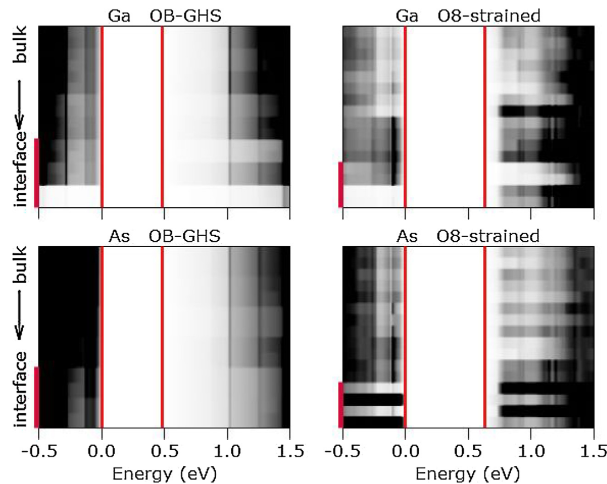


Fig. 9. The local DOS of Ga and As atoms in different layers for the coherent OB-GHS and semi-coherent O8-strained HfO_2/GaAs interfaces. Vertical dimension shows the distance from the interface in blocks of atoms. Horizontal dimension shows energy with respect to the highest occupied state. The red lines indicate the borders of the band gap. Dark red shows the atoms in the interface layers. The Ga atoms closest to the oxide within the OB-GHS interface form a mixed $(Ga,Hf)O_2$ layer. Therefore, the local DOS of these Ga atoms around the Fermi level is very small.

The coherent interface includes the anatase HfO_2. The ideal interface has a deficiency of two electrons per (2×2) interface area with respect to the ECR. One can remove one O atom from the Ga-O interface layer to form an O vacancy (OV interface). There is a certain O atom pattern which leads to the smallest total energy. This interface is shown in Fig. 6. It is the only interface having a (4×4) interface area in this study. It should be noted that the interfacial Ga atoms without interfacial relaxations occupy the lattice sites of both the oxide and the semiconductor bulk. One can also substitute two interfacial Hf atoms by Ga atoms and remove half of the O atoms from the Ga-O interface layer. This interface (OB-GHS), shown in Fig. 7, is characterized by Ga-O-Ga (oxygen) bridges, which include two Ga atoms and one bridging O atom, and not by the Ga-O-Ga chains as the ideal interface. The ECR can also be satisfied by substituting half of the interfacial As atoms by O atoms without removing O atoms. This interface (OAS) is shown in Fig. 8. The ECR is satisfied also by substituting half of the interfacial Ga atoms by Hf atoms (not shown). In spite of the defects, these coherent interface structures do not reveal gap states. However, if half of the Hf atoms are substituted by As atoms to satisfy the ECR, a clear defect state is found around the middle of the band gap. This defect state is originated from the substitutional As atoms within the oxide. The substitutional O atoms in the AOS structure are neighbored by Ga atoms and no As-O bonds are formed. The results obtained in this study show that defect free energy gaps and dimer free interface structures can be obtained for coherent interfaces, although structural defects at the interface are introduced (O vacancies and some substitutional atoms). The relaxation of the interface makes this possible. It should be noted that similar defects within the bulk parts induce defect gap states.

Fig. 9 shows the DOS of the OB-GHS and O8-strained [having both (2×2) interface area] for Ga and As atoms in different layers. It can be noted that the local DOS is not increased at the interface for the coherent OB-GHS indicating that there are no defect gap states. On the other hand, there is increased local DOS for As dangling bond and dimer atoms for the semi-coherent O8-strained indicat-

ing localized defect bands. It can also be noted that the dangling bond atoms induce gap states while the dimer atoms do not. The band gap is larger for the O8-strained interface. This reflects a general behavior found. The band gap is larger for the semi-coherent interfaces than coherent interfaces. It is assumed that the artificial increasing of the band gap due to quantum confinement, compared to the calculated bulk band gap, is influenced by the interface structure. The local DOS of Hf and O atoms around the band gap is very small and is not shown.

3.2. Interface stability

Relative interface energies are calculated using the *ab initio* atomistic thermodynamics [35,36] to assess the relative stability of the interfaces. The relative interface energy γ is calculated for the most of the considered models as follows

$\gamma A = E_{tot} - N_{HfO2}\,\mu_{HfO2} - N_{O,ex}\mu_O - \gamma_{surf}A - C$, (1)

where E_{tot} is the total energy of the interface supercell, A is the interface area, N_{HfO2} is the number of HfO_2 units in the supercell, μ_{HfO2} is the chemical potential of the HfO_2, $N_{O,ex}$ is the number of excess O atoms in the supercell, μ_O is the chemical potential of the O, γ_{surf} is the surface energy of the HfO_2 and C is some constant. Different HfO_2 types are considered. Therefore, μ_{HfO2} and γ_{surf} depend on the interface model structure.

The μ_{HfO2} is equal to the bulk energy of the HfO_2 calculated by the incremental method [37]. The incremental bulk energy is calculated using the same (2×2) area and k point mesh which are used for the interface calculations. The formation energy (heat of formation) H_f of the HfO_2 determines the range of the μ_O. The formation energy is given by

$H_f = E_{HfO2} - E_{Hf} - E_{O2}$, (2)

where E_{HfO2} and E_{Hf} are the total bulk energies of the HfO_2 and Hf and E_{O2} is the total energy of the oxygen molecule (-9.84 eV and −10.46 within the PBE and LDA, respectively). The following values are obtained by normal bulk calculations. The H_f for the tetragonal HfO_2 is −10.75 eV and −12.20 eV within the PBE and LDA, respectively. The experimental value is −11.86 eV [38]. Total energy difference between the tetragonal and anatase phases is 0.26 eV and −0.14 eV within the PBE and LDA, respectively. Thus, the anatase is incorrectly more stable within the PBE. Total energy difference between the monoclinic and anatase is −0.32 eV within the LDA which is in good agreement with the value given in the Ref. [27] (-0.23 eV).

The μ_O varies within the range

$(E_{O2} + H_f)/2 \leq \mu_O \leq E_{O2}/2$. (3)

Almost corresponding range is obtained, if the lower limit is calculated by

$2E_{HfO2} - E_{Hf2O3} \leq \mu_O$, (4)

where E_{Hf2O3} is the total bulk energy of the Hf_2O_3. The value of the left side of Eq. (4) is −10.08 eV (used in figures) and −10.80 within the PBE and the LDA, respectively. The condition given by Eq. (4) expresses relatively O-poor growth conditions in which Hf_2O_3 [39] and Hf are formed. The upper limit of Eq. (3) expresses relatively oxygen-rich growth conditions. The more O-rich HfO_3 is not stable at ambient conditions [40]. All considered interface model structures are O-rich and have, therefore, excess O atoms. The energy difference between two variations of the OV structure is calculated using the (4×4) cell. The less stable modification includes also Ga-O-Ga chains without O vacancies.

The relative stability of various HfO_2/GaAs interfaces are shown as a function of the chemical potential of the oxygen in Fig. 10. The OV and OAS are coherent interface structures and the O10, O9 and O8-strained are semi-coherent interface structures. The O10 structure was constructed by attaching the As dimer atoms to interfacial Ga atoms in a way which preserves the As dimers as well as As bonds to Ga atoms above and below the As atoms. A different in-

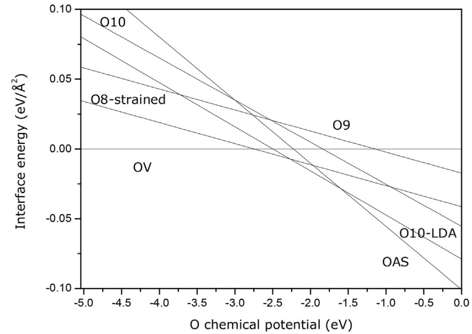


Fig. 10. The relative interface energies of the HfO_2/GaAs coherent OV and OAS and semi-coherent O10, O9 and O8-strained interface structures as a function of the O chemical potential in relative to the O chemical potential in O_2 molecule.

plane orientation between the HfO_2 and III–V parts (no rotation) was also tested, but it leads to slightly higher interface energies as well as non-symmetric distortions within the Hf-O layer above the interfacial layer which induces defect gap states resembling bulk defects. There are two arrangements of the interfacial Ga atoms which lead to relatively small interface energies within the set of different O10 structures. One of the slightly less stable O10 interfaces includes “alternating” As dimers which do not form As dimer rows. Another one includes a two-fold coordinated O atom within the As dimer rows and not between them as in Fig. 2. The interface energy of the latter O10 is increased by about 7 meV/Å^2 probably due to the repulsion between the O atom and As dimers. If the two-fold coordinated O atom is removed from this structure, the most stable O9 interface, having 9O atoms per (2×2) area (not shown), is formed. One As dimer is broken to obey the ECR. One of the separated As atoms has one dangling bond and the other As atom forms two covalent bonds with interfacial Ga atoms. If the two-fold coordinated O atom is removed from the most stable O10 (shown below the other O atoms in Fig. 2) to form the O9 structure, a larger interface energy is obtained (by about 8 meV/Å^2). This O9 interface includes one Ga dimer but no As dangling bonds. Other O9 interfaces were considered as well. All structures shown in Fig. 10, except the OAS, include the same number of Ga and As (as well as pseudohydrogen) atoms. The total energy of the bulk As was calculated using a hexagonal cell [41] to estimate the relative interface energy of the OAS structure. The chemical potential of the As was estimated under extremely Ga-rich growth conditions for Fig. 10, *i.e.*, the formation energy of the GaAs is added to the total energy of the bulk As. The OAS interface is less stable under less Ga-rich growth conditions. The number of excess O atoms is 3, 6, 5, 4 and 4 for OV, OAS, O10, O9 and O8-strained interfaces, respectively.

The OV, OAS and O8-strained are stable for different ranges of the O chemical potential. This suggests that both coherent and semi-coherent interfaces based on different structural motifs could be found in experiments. Obviously the O-rich interfaces (many excess O atoms) tend to be stabilized under the O-rich growth conditions (relatively large O chemical potential). However, there are also many other factors contributing to the interface energy like bond number (coordination), bond types, bond distortions, ECR and defects. The O8-strained interface is quite stable with respect to the other semi-coherent interfaces O10 and O9, although it includes two As dangling bonds. The lateral densities of the Ga and Hf atoms are (not) equal for the O8-strained (O10 and O9) structure(s) which probably destabilizes O10 and O9 with respect to the O8-

Table 1
The amount of interfacial atoms having coordination number not equal to the corresponding bulk coordination number within the HfO_2/GaAs interfaces per (2 × 2) area. The "n-f" denotes n-fold coordination. The bulk coordination numbers of the Hf and O atoms are 6 and 3 for the OV, AOS, and O8-strained interfaces, and 8 and 4 for the O10 and O9 interfaces, respectively. In practice, interatomic distances below 3 Å characterize a bond here (either covalent or ionic) excluding O-O distances. The O9* denotes a non-stable interface described in text.

	Hf	O	Ga	As
OV	1 5-f		2 5-f	
AOS			2 5-f	
O10	4 7-f	8 3-f 1 2-f	2 5-f	
O9	3 7-f	6 3-f	2 5-f	1 3-f
O9*	4 7-f	8 3-f	2 5-f	
O8-strained				2 3-f

strained interface. This is reflected in the coordination numbers of the interfacial atoms which are shown in Table 1. It is noted that the interfacial Hf, O, and Ga atoms can keep their bulk coordination within the O8-strained structure. The Ga atoms occupy the Hf atom positions at the semi-coherent interfaces, but one of the interfacial Hf positions is not occupied within the O9 and O10 interfaces. It is found that the LDA makes semi-coherent interfaces more stable than the PBE (only O10-LDA shown).

The relative stability of some HfO_2/GaP interfaces are shown as a function of the chemical potential of the oxygen in Fig. 11. The FOL (full oxygen layer) and OB (oxygen bridge) models are variations of the OV model corresponding to different O concentrations within the interface layer. The number of excess O atoms is 4 and 2 for FOL and OB interfaces, respectively. The FOL is an ideal interface.

It is noted that the stability intervals and energy differences are very similar to those obtained for the HfO_2/GaAs. The ECR breaking OB and FOL structures are not stable. The ECR breaking should lead to occupation of antibonding states or deoccupation of the bonding states which increases the total energy. The interface energy of the O9 structure is practically identical to the interface energy of the FOL in Fig. 11. The LDA makes semi-coherent interfaces less stable than the PBE in contrary to the HfO_2/GaAs (not shown). The total energy of the bulk P was calculated for the orthorhombic black phosphorus [42,43]. The chemical potential of the P was estimated under extremely Ga-rich growth conditions for the OAS interface.

It is assumed that no stable interface is obtained by removing O atoms from the O10 structure. The assumption is intuitive, because removed O atoms mean energetically very unfavorable holes in the ionic oxide part. This is due to the stiffness of the ionic bond. On the other hand, As dangling bonds are formed. The interface energy increases rapidly with O removals (Fig. 3 in the Ref. [12]). The less O-rich interfaces should have some less O-rich non-defective oxide interface planes. However, this is very difficult, because the interface should sustain relatively small lattice mismatch and allow Ga atoms keep approximatively their positions to maintain the original bonding through the interface. The strained O8 structure avoids oxygen vacancies due to the smaller lateral density of the oxide. The O vacancy might be energetically more favorable within the coherent OV structure (unoccupied Ga dangling bond) than within the semi-coherent interfaces. On the other hand, the O vacancy within the OV structure breaks only three ionic bonds (Table I). It should be noted that vacancies, Ga dangling bonds, and substitutional atoms within the bulk parts increase total energy significantly. The interface energy depends sensitively on the atomic structure of the interface. This means that small lattice mismatch does not necessarily mean small interface energy, because the arrangement of bonds is important.

The relative stability of the OB-GHS structure is not shown in Figs. 10 and 11, because the OB-GHS interface has different amount of Hf and Ga atoms compared to the other interfaces. Two Hf atoms were substituted by Ga atoms in the Hf layer closest to the interface in the O10 and O8-strained interfaces to allow comparison. The results are similar to those shown in Figs. 10 and 11 (not shown). The ECR can also be satisfied by substituting two interfacial Ga atoms by Hf atoms or two interfacial Hf atoms by As atoms within the coherent interface as mentioned (the latter one induces defect gap states). The relative stabilities of these interfaces were not assessed. However, the formation energy of the Ga_2O_3 is considerably larger than that of the As_2O_3, which should destabilize the latter interface.

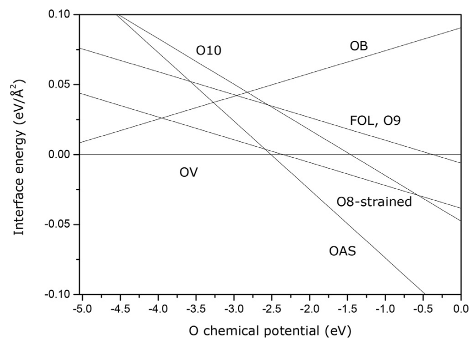


Fig. 11. The relative interface energies of the HfO_2/GaP coherent OAS, OB, OV, and FOL interface structures and semi-coherent O10, O9 and O8-strained interface structures as a function of the O chemical potential in relative to the O chemical potential in O_2 molecule.

The interface energies show that both coherent and semi-coherent interfaces have a stability interval within the range of the O chemical potential. It is of course in principle possible that there is some semi-coherent interface with more O atoms which has a larger stability interval than those considered in this study. However, that kind of interface was not found in this study. It is difficult to imagine energetically favorable positions for additional O atoms in the O10 and O8-strained structures. The same mixed As-O layer as in the AOS interface structure breaks dimers. Substitutional O atoms also lead to the violation of the ECR through excess electrons within the semi-coherent interfaces. These factors contribute to the destabilization of the interfaces. It should also be noted that the broken dimers in the O8-strained structure can not be fixed by additional O atoms. Non-substitutional additional O atoms have to occupy some two-coordinated atom positions, which does not increase effectively the amount of the bonds, because no extra electrons are needed for the additional bonds.

Based on the results shown in Figs. 10 and 11, it should be relatively easy to grow coherent interfaces without dimers, if only interface thermodynamics is considered. However, the stability of the oxide also affects the structure of the interface. The formation energies of the considered oxides vary approximately within 0.2-0.3 eV per formula unit. This energy difference per monolayer of the HfO_2 equals to about 0.013–0.020 eV/Å^2. The experimental difference is not known, however, because the anatase is not stable in the bulk form. It is estimated that the formation energies of the considered oxides are increased by about 0.3 eV per formula unit, if the oxides are grown on the InP substrate (the O8-strained is broken on the InP). The lattice parameter of the InP is about 4% larger than that of the GaAs. It is noted that considering symmetry, at least the tetragonal oxide and the anatase can form easily (100) interface. With some particular initial interface structures part of the tetragonal oxide is transformed into anatase within the PBE (these structures are excluded from interface energy calculations). Therefore, the total energies of the O9 HfO_2/GaAs interfaces within the

PBE are determined by stopping the geometrical relaxation when the tetragonal oxide starts to break. It is possible to form an interface using the monoclinic HfO_2 (the *c* axis perpendicular to the interface) similar to the O10 interface. This interface has probably an interface energy similar to the O10. It is assumed that the interface energy calculated for the tetragonal HfO_2 is a lower bound for the similar interface of the monoclinic structure, because the interface formed with the monoclinic structure looks less symmetric and the oxygen atoms do not form smooth planes. However, the calculations to justify this assumption are beyond the scope of the present study due to the significantly increased calculational burden.

The particular direction of the monoclinic *c* axis found in the experiments might also be due to the fact that the dimer rows in the O10 can not be formed otherwise [27].

It was suggested that the ECR acts during oxide growth layer by layer [44]. This means that the growing oxide layer can not change atoms with deeper layers. The OV model is the only coherent interface satisfying this condition. However, this interface might be found only in the quite extreme growth conditions. Using the temperature-dependent O chemical potential [45,46] it is obtained that the relative O chemical potential is approximately -1.17 eV and -2.47 eV at 600 K and 1200 K, respectively. The O_2 pressure is assumed to be 10^{-10} in relative to the atmospheric pressure. If the corresponding relative O_2 pressure is 1, the relative chemical potentials are dropped to -0.58 eV and -1.28 eV. Therefore, high temperature and low O_2 pressure is needed to stabilize the OV interface according to the interface energies. These numbers also explain, why the relatively O-rich interfaces were considered in this study.

The relative stability of the semi-coherent interfaces with respect to the OV interface under the O-rich growth conditions can be explained as follows. The considered interfaces are O-rich, which means that there are excess O atoms (*i.e.*, the O coverage is more than 0.5 monolayers at the interface). If the lateral density of the Hf and Ga atoms are the same and the lateral density of the Ga atoms and excess O atoms are the same (as in the ideal coherent and semi-coherent O8-strained structures), a Ga atom donates two electrons to an excess O atom. Alternatively, it can be said that two covalent bonds are formed. Therefore, a Ga atom has one electron for covalent bonds with the underlying As atoms. There is typically one covalent bond per one Ga atom at the semi-coherent interface. However, an As atom has 1.5 electrons for a covalent bond due to the dimer bond. Therefore, there are 0.5 excess electrons per a Ga (or As) atom. On the other hand, there are two covalent bonds per a Ga atom at the coherent interface. Therefore, there is a deficiency of 0.5 electrons per a Ga atom (an As atom donates 1.25 electrons to one covalent bond). The excess O atoms consume electrons. This means that the semi-coherent interfaces favor more O-rich growth conditions than the coherent interfaces.

Only the coherent OV model satisfies the ECR layer by layer. In fact, the Ga-O layer within the OV interface can be taken as an ultrathin Ga_2O_3 oxide. However, the OV interface is not stable under usual growth conditions as shown above. It was recently shown that a specific (3×1) oxidized InAs reconstruction [47] can be used to decrease interface defect state density by a factor of 40 for HfO_2/InAs [6]. A similar effect was found also for HfO_2/(In,Ga)As [48]. It is noted that the arsenic oxide adopts usually the As_2O_3 stoichiometry which is more covalent than the HfO_2. The As atoms are three-fold coordinated in the various As_2O_3 oxides [49] and have an occupied dangling bond. This explains why there are less O atoms per As atom in the As_2O_3 than Hf atom in the HfO_2, although the valence electron number of the As atom is larger than that of the Hf atom. However, an As atom can also be four-fold coordinated in thin As oxide at the III–V semiconductor surface. This was tested for some model oxide/GaAs interfaces. In these models two uppermost layers of the ideal As-terminated GaAs surface are oxidized. The second layer is a Ga-O layer similar to the coherent interfaces. The three-fold (four-fold) coordinated As atom donates zero (one) electrons to the crystal. Therefore, with equal lateral As atom densities at both sides of the III–V semiconductor interface the As oxide can (only) donate electrons. Electron donation from the oxide to the semiconductor favors interface bonding similar to the coherent interfaces. It is important to note that the coherent interfaces include a Ga-O layer which inhibits the dimer formation due to the bonding geometry. The tight Ga-O layer, which does not relax much, may also prevent the oxidation of the deeper semiconductor layers and amorphization of the interface. When the HfO_2 is grown on the As oxide/III-V substrate system, the Ga-O layer is possibly not destroyed due to the kinetics. The As oxide layer may be vanished or segregated to the surface of the growing HfO_2. It is supposed that a thick ordered As oxide on the III–V substrates is not possible. Therefore, it is possible that the ordered (3×1) reconstruction does not reveal a coherent interface between the As oxide and the III–V semiconductor. However, the oxidized surface is ordered and may include a Ga-O layer similar to the coherent interfaces. It is noted that the As oxide can accommodate possible electronic mismatch at the interface by different coordinations of the As atoms. When two three-fold coordinated As atoms form connecting bonds between them by an oxygen bridge, two electrons are donated to the crystal. Electronically this is equivalent to the dimer formation from dangling bonds. However, in this case the oxide, instead of the semiconductor, adjusts the valence electron number to match the bond number. It is possible that the formation of the (3×1) reconstruction is due to kinetics rather than thermodynamics. Still, the As oxide is capable to balance electronically the system due to the dangling bonds. The Ga-O layer may make the growing oxide to adopt the coherent interface structure.

4. Conclusions

Several prominent new model HfO_2/GaAs(100) and HfO_2/GaP(100) interfaces were presented. Both coherent and semi-coherent HfO_2/GaAs and HfO_2/GaP interfaces, which do not have defect gap states, were found. It was shown that various interfacial defects, which can not be avoided within the HfO_2/III-V interfaces, do not necessarily cause defect gap states. The As dangling bond, which induces gap states, seems to be a typical defect at semi-coherent interfaces. Relative stabilities of the two main interface types, based on the Ga-O-Ga bridges and group V dimers, were determined using various structures. Interface energy depends sensitively on the type and position of the defects and the atomic structure of the interface. Coherent interfaces are desirable, because they do not include harmful As dimers. Interface thermodynamics shows that a coherent interface is quite stable also under the relevant O-rich growth conditions. However, the semi-coherent interfaces should be found, if kinetics prohibits atom changes between growing oxide layers. This is explained by the interface bonding (which is different at the coherent and semi-coherent interfaces) and the ECR. The same reasoning is used to explain why an ordered interface without group V dimers can be formed at the interface of the As oxide and the III–V substrate. Interface structures and results presented can be used to develop new interface models.

Acknowledgments

The Magnus Ehrnrooth Foundation is acknowledged for financial support (A. L.). The computer resources of the Finnish IT Center for Science (CSC) and the FGI project (Finland) are acknowledged.

Supplementary material

The structures for the HfO_2/GaAs O10 (Fig. 2), O8-strained (Fig. 4), OV [both (4×4) and (2×2)] (Fig. 6), OB-GHS (Fig. 7), OAS (Fig. 8), O9, FOL, and OB interfaces are given in VASP POSCAR/CONTCAR format as supplementary data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apsusc.2017.08.185.

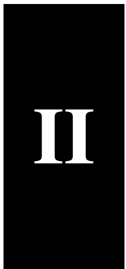
References

[1] J. Robertson, R.M. Wallace, High-K materials and metal gates for CMOS applications, Mat. Sci. Eng. R 88 (2015) 1.
[2] M. Hong, J. Kwo, A.R. Kortan, J.P. Mannaerts, A.M. Sergent, Epitaxial Cubic Gadolinium Oxide as a Dielectric for Gallium Arsenide Passivation, Science 283 (1999) 1897.
[3] M. Caymax, G. Brammertz, A. Delabie, S. Sioncke, D. Lin, M. Scarrozza, G. Pourtois, W.-E. Wang, M. Meuris, M. Heyns, Interfaces of high-*k* dielectrics on GaAs: Their common features and the relationship with Fermi level pinning, Microelectron. Eng. 86 (2009) 1529.
[4] E. O'Connor, S. Monaghan, R.D. Long, A. O'Mahony, I.M. Povey, K. Cherkaoui, M.E. Pemble, G. Brammertz, M. Heyns, S.B. Newcomb, V.V. Afanas'ev, P.K. Hurley, Temperature and frequency dependent electrical characterization of $HfO_2/In_xGa_{1-x}As$ interfaces using capacitance-voltage and conductance methods, Appl. Phys. Lett. 94 (2009) 102902.
[5] Y.C. Chang, W.H. Chang, C. Merckling, J. Kwo, M. Hong, Inversion-channel GaAs(100) metal-oxide-semiconductor field-effect-transistors using molecular beam deposited Al_2O_3 as a gate dielectric on different reconstructed surfaces, Appl. Phys. Lett. 102 (2013) 093506.
[6] C.H. Wang, S.W. Wang, G. Doornbos, G. Astromskas, K. Bhuwalka, R. Contreras-Guerrero, M. Edirisooriya, J.S. Rojas-Ramirez, G. Vellianitis, R. Oxland, M.C. Holland, C.H. Hsieh, P. Ramvall, E. Lind, W.C. Hsu, L.-E. Wernersson, R. Droopad, M. Passlack, C.H. Diaz, InAs hole inversion and bandgap interface state density of 2×10^{11} cm^{-2} eV^{-1} at HfO_2/InAs interfaces, Appl. Phys. Lett. 103 (2013) 143510.
[7] M.D. Pashley, Electron counting model and its application to island structures on molecular-beam epitaxy grown GaAs(001) and ZnSe(001), Phys. Rev. B 40 (1989) 10481.
[8] P.W. Peacock, J. Robertson, Bonding Energies, and Band Offsets of Si-ZrO_2 and HfO_2 Gate Oxide Interfaces, Phys. Rev. Lett. 92 (2004) 057601.
[9] L. Lin, J. Robertson, Defect states at III-V semiconductor oxide interfaces, Appl. Phys. Lett. 98 (2011) 082903.
[10] J. Robertson, L. Lin, Defect gap states on III-V-semiconductor-oxide interfaces, Microelectron. Eng. 88 (2011) 1440.
[11] W. Wang, K. Xiong, G. Lee, M. Huang, R.M. Wallace, K. Cho, Origin of HfO_2/GaAs interface states and interface passivation: A first principles study, Appl. Surf. Sci. 256 (2010) 6569.
[12] W. Wang, K. Xiong, R.M. Wallace, K. Cho, Impact of Interfacial Oxygen Content on Bonding Stability, Band Offsets, and Interface States of GaAs:HfO_2 Interfaces, J. Phys. Chem. C 114 (2010) 22610.
[13] K.C. Santosh, H. Dong, R.C. Longo, W. Wang, K. Xiong, R.M. Wallace, K. Cho, Electronic properties of InP(001)/HfO_2 interface: Band offsets and oxygen dependence, J. Appl. Phys. 115 (2014) 023703.
[14] D. Colleoni, G. Miceli, A. Pasquarello, Origin of Fermi-level pinning at GaAs surfaces and interfaces, J. Phys.: Condens. Matter 26 (2014) 492202.
[15] G. Miceli, A. Pasquarello, Accurate determination of charge transition levels of the As-As dimer defect at GaAs/oxide interfaces through hybrid functional, Appl. Phys. Lett. 103 (2013) 041602.
[16] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. 77 (1996) 3865.
[17] D.M. Ceperley, B.J. Alder, Ground State of the Electron Gas by a Stochastic Method, Phys. Rev. Lett. 45 (1980) 566.
[18] J.P. Perdew, A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B 23 (1981) 5048.
[19] P.E. Blöchl, Projector augmented-wave method, Phys. Rev. B 50 (1994) 17953.
[20] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B 59 (1999) 1758.
[21] G. Kresse, J. Hafner, *Ab initio* molecular dynamics for liquid metals, Phys. Rev. B 47 (1993) 558.
[22] G. Kresse, J. Hafner, *Ab initio* molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium, Phys. Rev. B 49 (1994) 14251.
[23] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mat. Sci. 6 (1996) 15.
[24] G. Kresse, J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169.
[25] H.J. Monkhorst, J.D. Pack, Special points for Brillouin-zone integrations, Phys. Rev. B 13 (1976) 5188.
[26] Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, P. Blaha, Towards efficient band structure and effective mass calculations for III-V direct band-gap semiconductors, Phys. Rev. B 82 (2010) 205212.
[27] A. Debernardi, C. Wiemer, M. Fanciulli, Epitaxial phase of hafnium dioxide for ultrascaled electronics, Phys. Rev. B 76 (2007) 155405.
[28] G. Giorgi, A. Korkin, K. Yamashita, Zirconium and hafnium oxide interface with silicon: Computational study of stress and strain effects, Comp. Mat. Sci. 43 (2008) 930.
[29] G. Giorgi, L.R.C. Fonseca, A. Korkin, K. Yamashita, Impact of the crystal structure of HfO_2 on the transport properties of model HfO_2/Si/HfO_2 silicon-on-insulator field-effect transistors: A combined DFT-scattering theory approach, Phys. Rev. B 79 (2009) 235308.
[30] M.P.J. Punkkinen, K. Kokko, H. Levämäki, M. Ropo, S. Lu, L. Delczeg, H.L. Zhang, E.K. Delczeg-Czirjak, B. Johansson, L. Vitos, Adhesion of the iron-chromium oxide interface from first-principles theory, J. Phys.: Condens. Matter 25 (2013) 495501.
[31] G.-M. Rignanese, Dielectric properties of crystalline and amorphous transition metal oxides and silicates as potential high-κ candidates: the contribution of density-functional theory, J. Phys.: Condens, Matter 17 (2005) R357.
[32] V. Fiorentini, G. Gulleri, Theoretical Evaluation of Zirconia and Hafnia as Gate Oxides for Si Microelectronic, Phys. Rev. Lett. 89 (2002) 266101.
[33] B.R. Tuttle, C. Tang, R. Ramprasad, First-principles study of the valence band offset between silicon and hafnia, Phys. Rev. B 75 (2007) 235324.
[34] H.T. Stokes, D.M. Hatch, FINDSYM: program for identifying the Space Group Symmetry of a Crystal, J. Appl. Cryst. 38 (2005) 237.
[35] E. Kaxiras, Y. Bar-Yam, J.D. Joannopoulos, K.C. Pandey, *Ab initio* theory of polar semiconductor surfaces. I. Methodology and the (2×2) reconstructions of GaAs(111), Phys. Rev. B 35 (1987) 9625.
[36] G.-X. Qian, R.M. Martin, D.J. Chadi, First-principles study of the atomic reconstructions and energies of Ga- and As-stabilized GaAs(100) surfaces, Phys. Rev. B 38 (1988) 7649.
[37] J.C. Boettger, Nonconvergence of surface energies obtained from thin-film calculations, Phys. Rev. B 49 (1994) 16798.
[38] J.X. Zheng, G. Ceder, T. Maxisch, W.K. Chim, W.K. Choi, First-principles study of native point defects in hafnia and zirconia, Phys. Rev. B 75 (2007) 104112.
[39] K.-H. Xue, P. Blaise, L.R.C. Fonseca, Y. Nishi, Prediction of Semimetallic Tetragonal Hf_2O_3 and Zr_2O_3 from First Principles, Phys. Rev. Lett. 110 (2013) 065502.
[40] J. Zhang, A.R. Oganov, X. Li, K.-H. Xue, Z. Wang, H. Dong, Pressure-induced novel compounds in the Hf-O system from first-principles calculations, Phys. Rev. B 92 (2015) 184104.
[41] D. Schiferl, C.S. Barrett, The crystal Structure of Arsenic at 4·2, 78 and 299 °K, J. Appl. Cryst. 2 (1968) 30.
[42] Y. Takao, Electronic structure of black phosphorus: Tight binding approach, Physica (Amsterdam) 105 B (1981) 93.
[43] Y. Du, C. Ouyang, S. Shi, M. Lei, Ab initio studies on atomic and electronic structure of black phosphorus, J. Appl. Phys. 107 (2010) 093718.
[44] J. Robertson, L. Lin, Bonding principles of passivation mechanism at III-V-oxide interfaces, Appl. Phys. Lett. 99 (2011) 222906.
[45] N.I. Al-Zoubi, M.P.J. Punkkinen, B. Johansson, L. Vitos, Completeness of the exact muffin-tin orbitals: Application to hydrogenated alloys, Phys. Rev. B 81 (2010) 045122.
[46] M.P.J. Punkkinen, P. Laukkanen, J. Lång, M. Kuzmin, J. Dahl, H.L. Zhang, M. Pessa, M. Guina, L. Vitos, K. Kokko, Structure of ordered oxide on InAs(100) surface, Surf. Sci. 606 (2012) 1837.
[47] M.P.J. Punkkinen, P. Laukkanen, J. Lång, M. Kuzmin, M. Tuominen, V. Tuominen, J. Dahl, M. Pessa, M. Guina, K. Kokko, J. Sadowski, B. Johansson, I.J. Väyrynen, L. Vitos, Oxidized In-containing III-V(100) surfaces: Formation of crystalline oxide films and semiconductor-oxide interfaces, Phys. Rev. B 83 (2011) 245401.
[48] X.Y. Qin, W.E. Wang, R. Droopad, M.S. Rodder, R.M. Wallace, A crystalline oxide passivation on $In_{0.53}Ga_{0.47}As(100)$, J. Appl. Phys. 121 (2017) 125302.
[49] A. Matsumoto, Y. Koyama, A. Togo, M. Choi, I. Tanaka, Electronic structures of dynamically stable As_2O_3, Sb_2O_3, and Bi_2O_3 crystal polymorphs, Phys. Rev. B 83 (2011) 214110.

Jaakko Mäkelä & Antti Lahti & Marjukka Tuominen & Muhammad Yasir & Mikhail Kuzmin & Pekka Laukkanen & Kalevi Kokko & Marko Punkkinen & Hong Dong & Barry Brennan & Robert Wallace

Unusual oxidation-induced core-level shifts at the HfO2/InP interface

Scientific Reports, 9, 2019



www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Unusual oxidation-induced core-level shifts at the HfO_2/InP interface

Jaakko Mäkelä[1], Antti Lahti[1], Marjukka Tuominen[1], Muhammad Yasir[1], Mikhail Kuzmin[1,2], Pekka Laukkanen[1], Kalevi Kokko[1], Marko P. J. Punkkinen[1], Hong Dong[3,4], Barry Brennan[3,5] & Robert M. Wallace[3]

Received: 1 October 2018
Accepted: 5 December 2018
Published online: 06 February 2019

X-ray photoelectron spectroscopy (XPS) is one of the most used methods in a diverse field of materials science and engineering. The elemental core-level binding energies (BE) and core-level shifts (CLS) are determined and interpreted in the XPS. Oxidation is commonly considered to increase the BE of the core electrons of metal and semiconductor elements (*i.e.*, positive BE shift due to O bonds), because valence electron charge density moves toward electronegative O atoms in the intuitive charge-transfer model. Here we demonstrate that this BE hypothesis is not generally valid by presenting XPS spectra and a consistent model of atomic processes occurring at HfO_2/InP interface including negative In CLSs. It is shown theoretically for abrupt HfO_2/InP model structures that there is no correlation between the In CLSs and the number of oxygen neighbors. However, the P CLSs can be estimated using the number of close O neighbors. First native oxide model interfaces for III-V semiconductors are introduced. The results obtained from *ab initio* calculations and synchrotron XPS measurements emphasize the importance of complementary analyses in various academic and industrial investigations where CLSs are at the heart of advancing knowledge.

The x-ray photoelectron spectroscopy (XPS) is widely utilized not only in the characterization of the chemical composition of materials but also to understand and control various scientifically and industrially interesting phenomena such as atomic layer deposition, catalysis, materials protection, operation of electronic devices, and photoelectrochemical reaction (*e.g.*, refs[1–15]). In research and development of these phenomena, the main XPS objective is typically determination and interpretation of the CLS, which are further combined with results of other measurements to obtain interrelationships between important properties. The CLSs are commonly interpreted in terms of electronegativity differences between elements. Excess (deficit) charge in the valence shell of an atom decreases (increases) the BE of a core electron according to the classical electrostatic case of the potential inside a uniformly charged spherical surface. Charge transfer in oxides is often expressed in terms of the oxidation state. This interpretation applies nicely to silicon oxidation, because a Si atom has four valence electrons. Therefore, the oxidation number of a silicon atom (0, +1, +2, +3, +4) is equal to the number of oxygen neighbors. Coincidentally, even the numerical values of the CLSs of the Si atoms (in eVs) equal roughly to the oxidation numbers[6]. The CLSs of other oxides are interpreted often in the same way. The BE is increased with the number of oxygen neighbors. However, it is much less clear, to what extent this model can be applied to other, especially more complex systems like oxide/III-V semiconductor interfaces. In general, the CLSs depend on several factors, not just on the atomic on-site charge and different complex environments can induce similar CLSs.

In this work, we report that the semiconductor oxidation can surprisingly cause negative CLSs by presenting theoretical and experimental results for the HfO_2/InP junction. In more general terms, the presented results reveal how one should be cautious when analyzing the XPS spectra solely in terms of the electronegativities of elements and number of oxygen neighbors. Furthermore, the oxidation-induced CLSs of a semiconductor are interpreted, which is further essential to understand phenomena like the ALD mechanisms[4] and the formation of surface defects harmful to electronics and photonics devices[10,16–19]. The HfO_2/InP interface is a prototypical insulator/semiconductor junction and also a potential component for devices like transistors[10,16–24], nanowire solar

[1]Department of Physics and Astronomy, University of Turku, FI-20014, Turku, Finland. [2]Ioffe Physical-Technical Institute, Russian Academy of Sciences, St. Petersburg, 194021, Russian Federation. [3]Department of Materials Science and Engineering, The University of Texas at Dallas, Richardson, Texas, 75080, USA. [4]Present address: Department of Electronics and Tianjin Key Laboratory of Photo-Electronic Thin Film Device and Technology, Nankai University, Tianjin, 300071, China. [5]Present address: National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, United Kingdom. Correspondence and requests for materials should be addressed to J.M. (email: jaakko.m.makela@utu.fi) or M.P.J.P. (email: marpunk@utu.fi)

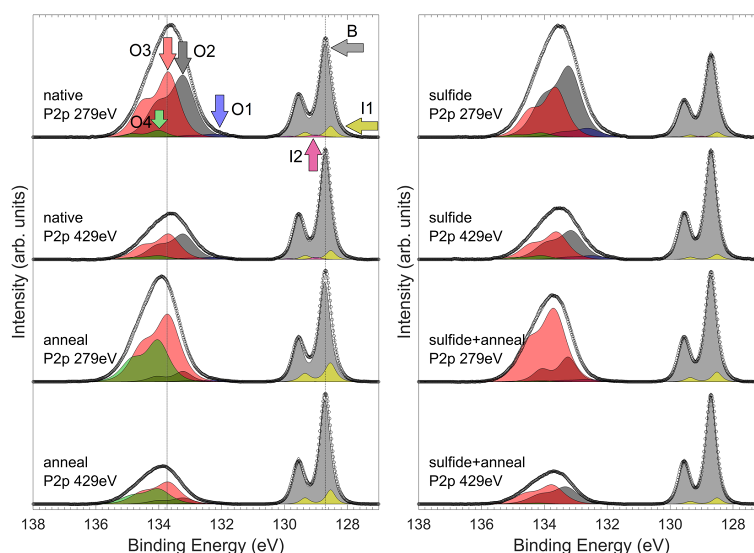


Figure 1. P *2p* spectra with fitted peaks. Vertical lines have been placed to illustrate the clear shift of the envelope of the O components even though B has been calibrated to 128.7 eV in these figures. On the left side the measurements of the corresponding native oxide experiments are shown, and sulfide treated on the right side. The energy label shows the chosen *hν* of the photons.

cells[25–27], and infrared detectors[28,29]. In these applications, HfO_2 on InP typically acts as a dielectric and/or passivates the semiconductor crystal against environment-induced changes. It is essential to minimize the amount of interfacial defects, which can cause for example the Fermi-level pinning, non-radiative recombination, and leakage currents via the defect-induced electron states. Significant progress has been made in reducing the densities of such harmful band gap states (e.g., refs[10,16–24]). The development of atomic layer deposition (ALD) of insulator films has significantly aided this progress. Still, HfO_2/III-V junctions contain too many defects as compared to the strict industrial reference of HfO_2/Si. To reduce the defect concentration and to improve device performance, it is crucial to understand and control the chemical and physical properties of the HfO_2/III-V interfaces, where III-V crystals become oxidized. XPS has been widely utilized in the studies to find interrelationships between the chemical composition and electrical properties of the HfO_2/III-V interface[10,16–19].

In this work, calculated CLSs based on the *ab initio* models for HfO_2/InP have been combined with synchrotron-radiation XPS measurements of the HfO_2/InP junctions grown by ALD. We focus on the CLSs of In *3d* and P *2p*, which are obtained with high enough resolution and surface sensitivity concerning the analysis made here, and yield well distinguishable changes as a function of photon energy and different sample treatments.

Results

We first discuss the differences in the spectra as a function of probing depth and sample treatments, starting from P 2*p* that exhibits the most systematic differences, then move on to In 3*d* and see how In bonding is changed with respect to P. The interpretations made are also supported by supplementary information with Hf 4*f* and S 2*p* spectra (see Fig. S1). These effects are then related to, and compiled consistently with the complementary data from computational results.

P 2*p* measurements. From Fig. 1 it can be seen that the P 2*p* emission around the InP bulk peak (about 129 eV, used as reference) is very narrow and exhibits the well-defined 2*p* doublet, in contrast to a broad oxide-related emission at 134 eV. The deconvolution of the P 2*p* spectra around the InP bulk emission is however complicated by the fact that the branching ratio varies from 0.4 to 0.47, instead of the theoretically predicted 0.5, when only single component (*i.e.*, both $2p_{3/2}$ and $2p_{1/2}$ peaks) is included in the fitting of the emission at 129 eV. Thus additional components, I1 and I2, are introduced for consistency. It should be noted that particularly I2 is not very reliable due to only slight CLS which results in large changes in intensity ratios when the shifts are varied by even +/− 0.05 eV.

The other P 2*p* emission components: O1, O2, O3 and O4 are necessary to reproduce the characteristics of the emission around 134 eV. It has been commonly considered that native oxide of InP causes features at 134 eV due to P containing oxides at P^{+5} oxidation state, such as $InPO_4$. Components at this BE have also been attributed

	Signal intensity (%)				Surface (150 eV) to bulk (300 eV) signal ratio (each peak referred to bulk peak intensity)		
	native 150 eV	native 300 eV	anneal 150 eV	anneal 300 eV	native	anneal	(anneal/native)
B (0 eV)	28,6	49,8	29,1	48,9	1,00	1,00	1,00
I1 (−0.18 eV)	1,9	2,0	5,2	6,7	1,66	1,30	0,79
I2 (+0.30 eV)	0,6	0,2	0,7	2,9	4,35	0,39	0,09
O1 (+3.51 eV)	2,3	2,9	1,4	0,5	1,36	4,45	3,27
O2 (+4.50 eV)	32,4	21,2	8,5	8,7	2,66	1,64	0,61
O3 (+4.97 eV)	33,2	21,7	34,9	25,7	2,67	2,27	0,85
O4 (+5.30 eV)	1,1	2,2	20,2	6,6	0,83	5,16	6,19
	Signal intensity (%)				Surface (150 eV) to bulk (300 eV) signal ratio (each peak referred to bulk peak intensity)		
	sulfide 150 eV	sulfide 300 eV	anneal 150 eV	anneal 300 eV	sulfide	anneal	(anneal/sulfide)
B (0 eV)	25,9	48,3	35,7	56,8	1,00	1,00	1,00
I1 (−0.18 eV)	1,3	0,6	2,7	2,9	3,74	1,49	0,40
I2 (+0.30 eV)	0,3	0,0	0,0	0,0	-	-	-
O1 (+3.51 eV)	5,1	2,4	1,6	0,2	3,94	12,31	3,12
O2 (+4.50 eV)	39,7	23,5	12,1	18,5	3,15	1,04	0,33
O3 (+4.97 eV)	25,7	21,6	47,3	20,5	2,21	3,67	1,66
O4 (+5.30 eV)	2,1	3,6	0,6	1,0	1,10	0,97	0,88

Table 1. First columns (under "Signal intensity") give the proportional intensities of the fitted peaks for P $2p_{3/2}$ for each measurement. Next two ("native" or "sulfide" and "anneal") express the relative average proximity of the state to the surface, higher number indicating closer to the surface, and the last column ("native/anneal" or "sulfide/anneal") the change in average distribution due to annealing, >1 indicating shift towards the surface and <1 towards the bulk. Upper panel represents the values for the native oxide sample and bottom panel for the sulfide treated sample.

to $In(PO_3)_3$, or could be related to InPHfO species. Here we have not observed P^0 –type emission around +1 eV, which has been usually associated with pure P-P bonding or P clusters (*e.g.*[30,31]).

To suggest justified interpretations about the origin of each fitted component, we will discuss the depth distribution from which each component arises. The relative intensities of the components in the P 2*p* spectra have been listed in Table 1. Furthermore, to understand the relative depth from which the component arises, the peak intensities have been scaled with the bulk peak intensity of the corresponding measurement. Then the obtained values for the 150 eV kinetic energy (KE) measurement before and after annealing are divided by the corresponding values of the 300 eV KE measurement. This analysis provides the proportional increase of a component in the topmost layers; the higher value, the nearer to the outer surface a component arises. All of the obtained values for significant components other than B are >1, meaning that they are closer to the surface than bulk, as expected. I2 makes an exception, but its intensity is so low that reliable intensity analysis is not possible for I2. Furthermore, when the values for the annealed sample obtained this way are divided by the corresponding ones before the annealing, it can be deduced how much the average depth distribution of each signal changed due to the annealing treatment. If the average distribution would stay in place without movement of the average position, the last ratio would remain unity, since the distance from bulk would not change. This approach is especially sensitive to the changes in the topmost surface layers due to exponential dependence of the signal on emission depth. It is to be noted that, due to the exponential dependence, this ratio will remain unity, if there is slight broadening or movement of the specific state towards the surface, and simultaneously significant broadening or movement towards the bulk.

Next we present the effects of annealing observed in P 2*p* on the native oxide sample. From the left panel of Fig. 1 as a function of annealing, one can see an increase in the I1 intensity, suggesting that some of the O-P bonds are reduced into interface related chemical state such as P dimers. This is consistent with the changes in the depth distribution of I1 and O1: the depth distribution of I1 stays relatively constant; *i.e.*, close to the interface, while the average depth of O1 moves closer to the outer surface because of loss of such P species near the bulk InP (note that when we talk about movement, it could mean either diffusion, or, a reconfiguration of chemical bonds differently on the surface and on the bulk side, resulting in a movement of the average depth of a given state). Another prominent difference is that the intensity of O4 is tremendously increased, and its depth distribution moves toward the surface while O2 has moved deeper and decreased, and O3 remained in its average depth distribution and increased. We note that if the effects observed were due to oxide growth as 'thickening' with no chemically induced redistribution, each of the components should be observed moving towards the outermost surface, since they are referred to bulk signal depth. However, we see the effects in both directions without suppression in the bulk peak signal intensity, indicating that the effects are indeed due to local and, eventually, extended conversion of one compound into another at different depths, or possibly also cross-diffusion of different species. To recapitulate, the amount of the highest BE component in the oxide film increases, and its presence as well as increase is more pronounced close to the surface. The O1 component most likely represents an unstable phase at the

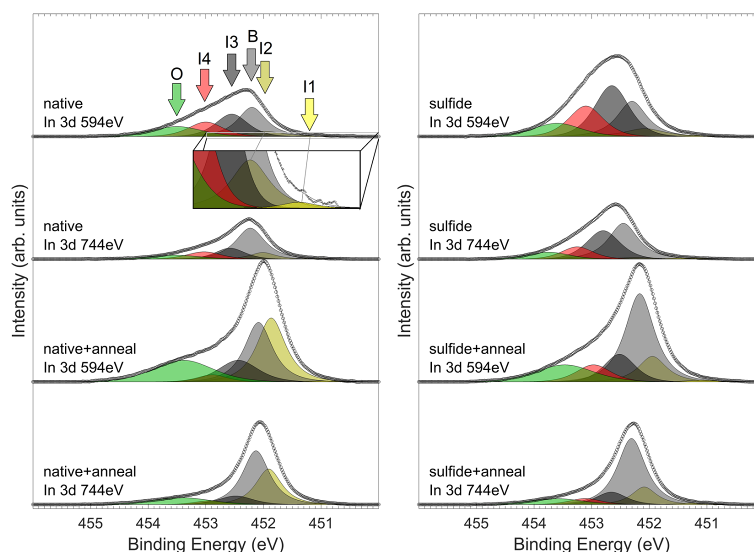


Figure 2. In 3*d* spectra with the fitted peaks. On the left side, sulfide treated sample with corresponding experiments are shown. On the right side, corresponding spectra of sulfide treated sample are shown. The energy label shows the chosen *hν* of the photons. It is noteworthy that I2 emission is observed for all measurements, yet increased dramatically for native + anneal sample.

interface dissociating into P and O. These atoms can further form P-P bonds at the interface as well as more highly oxidized P-O above this region.

The effects of sulfide treatment are observed as increased bulk emission after anneal, less decomposition of O1 into I1 tentatively assigned to P-P bonding at the interface, and distinct conversion of all oxidation states more uniformly into O3 state (+4.97 eV) instead of O4 (+5.30 eV), especially near the surface. A consistent model explaining this effect will be discussed after computational analysis regarding CLSs of interfacial native oxide models.

In 3*d* measurement. In Fig. 2 it can be seen that the intensity of component B in the In $3d_{3/2}$ spectra increases in relation to the other components when the bulk sensitivity of the measurements is increased. Thus, B is straightforwardly interpreted as the component arising from the emission of the bulk crystal chemical state, and its intensity is bound to vary similarly as for P 2*p* B, as described previously. Further justification for the exact position of the bulk component is gained from only slight variation in the BE position (<0.3 eV for a given sample). Table 2 shows the proportional integrated signal intensities of each peak for all of the measurements.

It is clear from Figs 1 and 2 that the In emission changes much more than the P emission due to the annealing. This is consistent with the bond formation energetics[32]: P-O-P and In-O-P bonding configurations are stronger than pure In-O-In. Thus, InP appears to be an exception among various III-V crystals because often the oxidation of group-III elements (*e.g.*, In) leads to a more stable oxide phase than the group-V (*e.g.*, As) oxidation[33].

The other components, I1, I2, I3, I4 and O were introduced to accommodate all the spectral features observed as a function of surface sensitivity and/or different treatments. These same components are fitted to all of the spectra even though we note that some components might be attributed to totally different chemical states or compounds due to different treatments. The variation has not been fitted as separate components, due to a finite resolution, but is rather taken into consideration as inhomogeneous broadening (FWHM) of chemical states (*i.e.*, there is no separate peak for e.g. In-S as compared to the native oxide sample due to the close proximity of existing peaks). A highly noteworthy observation is, that even though the B peak BE is carefully considered and adjusted, I2 with a negative CLS persists for each of the results.

To study the depth distribution of each emission component, the intensity values of the In 3*d* components are listed in Table 2, similarly to P 2*p* in Table 1.

Before the annealing the origin of I2 is close to the bulk boundary while I4, I3 and O lie increasingly closer to the outer surface. However, during the annealing the depth distribution changes, so that the average depth of the I2 signal moves closer to the surface. All the other interface related components stay fairly still in terms of depth distribution. The proportional bulk intensity remains similar before and after annealing, meaning that there is no significant net segregation towards the surface (that would result in lower proportional B signal intensity),

	Signal intensity (%)				Surface (150 eV) to bulk (300 eV) signal ratio (each peak referred to bulk peak intensity)		
	native 150 eV	native 300 eV	anneal 150 eV	anneal 300 eV	native	anneal	(anneal/native)
B (0 eV)	31,8	50,7	29,5	46,6	1,00	1,00	1,00
I1 (−1.10 eV)	0,7	0,9	0,5	0,8	1,24	0,90	0,73
I2 (−0.22 eV)	7,7	9,3	32,6	29,4	1,32	1,75	1,33
I3 (+0.35 eV)	25,2	16,4	13,7	8,4	2,45	2,58	1,06
I4 (+0.80 eV)	17,4	12,7	3,4	3,3	2,19	1,63	0,75
O (+1.30 eV)	17,2	9,9	20,3	11,6	2,76	2,76	1,00
	Signal intensity (%)				Surface (150 eV) to bulk (300 eV) signal ratio (each peak referred to bulk peak intensity)		
	sulfide 150 eV	sulfide 300 eV	anneal 150 eV	anneal 300 eV	sulfide	anneal	(anneal/sulfide)
B (0 eV)	23,1	40,2	46,9	59,8	1,00	1,00	1,00
I1 (−1.10 eV)	0,6	1,3	0,9	1,1	0,74	0,96	1,30
I2 (−0.22 eV)	6,9	2,5	12,9	13,8	4,81	1,19	0,25
I3 (+0.35 eV)	35,4	32,7	14,0	10,7	1,88	1,67	0,89
I4 (+0.80 eV)	21,3	13,4	9,6	5,6	2,77	2,19	0,79
O (+1.30 eV)	12,7	9,8	15,8	8,9	2,25	2,25	1,00

Table 2. First columns (under "Signal Intensity") give the proportional intensities of the fitted peaks for In $3d_{3/2}$ for each measurement. Next two ("native" or "sulfide" and "anneal") express the relative average proximity of the state to the surface, higher number indicating closer to the surface, and the last column ("native/anneal" or "sulfide/anneal") the change in average distribution due to annealing, >1 indicating shift towards the surface and <1 towards the bulk. Upper panel represents the values for the native oxide sample and bottom panel for the sulfide treated sample.

or alternatively that there is significant concomitant evaporation of In species from the surface. The I3 and I4 components quite well retain their depth distribution close to the bulk, meaning that they are likely related to the bulk-native oxide interface. This can also be true for the origins of I2 if there are two overlapping components with a very small difference in BE around the I2 position. I3 and O closely match with the BE shifts reported previously for chemical states of oxide In^{+3} and $InPO_4$, respectively [*e.g.*[30,31]]. The oxidation state In^{+3} has been commonly fitted with a relatively large peak width to take into account of the commonly observed inhomogeneity in bonding environment of an amorphous native oxide[30,31]. However, in this work we have used quite narrow peak width due to a clear variation in the spectral shape that could not be described using a single peak with the Gaussian emphasis. In contrast, the two separate components, I3 and I4, have been introduced, while the peak O has been fitted with a broad shape. The observation of three separate components reflects the fact (discussed further in the next section) that the same oxidation state, sometimes coarsely attributed as In^{3+} can be contained within several markedly different compounds, and the exact BE shift is dictated by the specific constituents and bonding environment. The large shift of component O from the bulk peak offers an additional support for the bulk peak position by fixing its shift to 1.2 eV so that the shoulder-like feature is well fitted and bulk peak intensity variation is still well explained as a function of surface sensitivity and similar behavior as in P 2*p*.

It is interesting that the emission at the negative BE side (i.e., I2 component) greatly increases during the annealing treatment. Negative CLSs in the group-III spectra of insulator/III-V junctions are typically interpreted as metallic group-III atoms or clusters/droplets and/or filled dangling-bond states. Intuitively, the I2 component's signal depth and its variation due to annealing (Table 2) suggests that the I2 origin is In atoms detached from the native oxide and diffused into the HfO_2 film toward the surface. I3 and I4 seem like native oxide components, decomposing somewhat during annealing, and possibly reconfiguring into states corresponding to either I2, or O component that is interpreted to be found at the boundary of native oxide and HfO_2 according to our straightforward analysis.

Very similar trends are found for the sulfide treated sample. However, increase in the proportion of B emission is significantly higher (as dictated by the similar trend in P 2*p*), and there is a more significant decrease in the initially higher I3 and I4 components; these components are likely related to In-S bonding sites at the interface area, and overlapping with oxide peaks; both In-S and native oxide peaks are tentatively assigned to I3 and I4. Moreover, as they are significantly reduced due to annealing while bulk-emission increases, it is suggested that S-containing interface transforms into a more abrupt barrier between InP and HfO_2, leaving less dangling bonds as described below, and seen also here as more intense B signal due to more ideal reconfiguration beneath the oxide. Lower increase in I2 and O suggests also less detachment of In to diffuse and/or reconfigure in the HfO_2 film, consistent with earlier literature[19]. Some In could still diffuse to the surface from the interface, but a more limited supply will result in much less observed In on the surface, especially after prolonged annealing as In will most likely also evaporate when reaching the surface. This effect is also consistent with a significantly higher proportional increase of B signal after annealing.

	In 3*d*	P 2*p*
In_2O_3		
Exp.	0.1–0.3	
CS (IS)	−1.23 (−1.35)	
CS (IS) 4*d*	−1.08 (−1.71)	
$InPO_4$		
Exp.	1.0–1.3	5.2–5.3
CS (IS)	0.18 (0.24)	5.21 (2.13)
CS (IS) 4*d*	0.17 (−0.27)	4.78 (2.13)
$In(PO_3)_3$		
Exp.	1.8	6.2
CS (IS)	0.57 (0.48)	6.91 (3.19)
CS (IS) 4*d*	0.58 (−0.12)	6.46 (3.20)
P_2O_5		
Exp.		6.8–7.5
CS (IS)		7.53 (4.23)

Table 3. The experimental and calculated In 3*d* and P 2*p* relative binding energies in In_2O_3, $InPO_4$, $In(PO_3)_3$ and P_2O_5. The calculations were done within the complete screening (CS) and initial state (IS) models. The In 4d states were core electrons or valence electrons (4d). The experimental values are from refs[58–60]. The experimental In and P binding energies in InP are equal to 444.4 eV and 128.8 eV[58,60].

Calculated bulk oxide and interface core-level shifts. The In 3*d* and P 2*p* relative core-level binding energies of several common bulk oxides of In and P were calculated, and the results are presented in Table 3. The calculational results reveal interesting trends and set important reference values for oxidized semiconductor systems. The relative binding energies are represented with respect to the Fermi level (not the vacuum level), which is the common practice in experiments. The In 3*d* and P 2*p* core-level binding energies in InP are set to zero. Therefore, positive (negative) CLS means increased (decreased) BE.

The P 2*p* relative BEs or CLSs within the complete screening (CS) model are in a relatively good agreement with the experimental ones. However, the disagreement is larger for the In 3*d* CLSs, which are underestimated by about 1.0–1.2 eV, if the In 4*d* electrons are treated as valence electrons. Still, the experimental trend in the In 3*d* CLS shown in the Table 3 is reproduced by the calculations. Obviously, there are many potential sources of errors both in the calculations and experiments. Concerning calculations, in particular, the hybrid Heyd-Scuseria-Ernzerhof (HSE) density functional[34] and homogeneous background charge as a replacement for the additional neutralizing electron in the complete screening calculation[35] were tested. These methods did not change the In 3*d* or P 2*p* CLS significantly. It should be noted that the CLS are usually calculated for systems, like surfaces or impurities, which can be modeled by a single geometrical construction. This increases accuracy significantly.

Several remarks can be made. The P CLSs are much larger than the In CLSs. The P valence charge is strongly bound due to the increased nuclear charge (which is not compensated by the increased electronic repulsion), and therefore, the electronic charge transfer in the ionic bond leads to larger CLS. Experimental core-level shifts are indeed often interpreted intuitively in terms of the transferred valence charge, see *e.g.* ref.[36]. It is commonly assumed that charge transfer increases with the ionicity of the bond. Therefore, ionic bonds should induce larger charge transfer than covalent bonds do. The In and P atoms lose electronic charge in oxides which increases the binding energies of the In and P core states. However, this is obviously not the whole story, because the In 3*d* initial state model CLSs (calculated with In 4*d* states in valence) are negative especially in the In_2O_3 ionic oxide. The CLSs are often interpreted more quantitatively in terms of the oxidation state which is occasionally even identified with the number of nearest neighbor O atoms, because this interpretation is valid for the SiO_2/Si interfaces[6]. However, In and P have oxidation states of +3 and +5 in all compounds considered in the Table 3. Furthermore, the In and P atoms occupy octahedral and tetrahedral positions, respectively, in all considered oxides. Still, the In or P CLSs are significantly different in various compounds. It can be noted that the CLSs become larger as the oxygen concentration increases. The second nearest neighbor configuration is also changed with the composition. Ionization generally increases the CLS in relative to the initial state model CLS, and this effect is much stronger for the P CLSs than the In CLSs. The discrepancies between the experimental and calculated complete screening CLSs might be tentatively contributed to the non-complete screening in the experiments.

The CLSs of the In and P impurities in the HfO_2 are −0.18 eV and 6.17 eV, respectively. The corresponding initial state CLSs are 0.17 eV and 4.19 eV. An interfacial atom can have a different (nearest) neighbor atomic configuration than the bulk atoms have. Furthermore, the interface dipole may affect the CLS. Different HfO_2/InP interface models were constructed to investigate, how the CLSs depend on the atomic environment. A semi-coherent model (O10), which has a relatively small lattice mismatch and which does not show interface states in the band gap, was introduced for the HfO_2/GaP and HfO_2/GaAs interfaces[37,38]. The lattice mismatch of this model can be kept relatively small for the HfO_2/InP interface by replacing the simple tetragonal HfO_2 (space group 137) used in the Ref.[37] in the O10 model with the anatase HfO_2 (body-centered tetragonal; space group 141)[37,39]. It should be noted that the CLSs of the In and P impurity atoms do not depend on the chosen HfO_2 phase.

N_O	CS P	IS P	CS In	IS In
0	0.17–0.49	0.28–0.42	−0.41–0.20	−0.16–0.33
1	0.48–2.03 (1.10–2.03)	0.18–1.05 (0.71–1.05)	−0.58–0.52	−0.39–0.95
2	2.10–3.06	1.50–2.87	−0.39–0.47	−0.31–0.67
3	3.05–4.48	1.00–3.15	−0.38–0.46	−0.19–0.24
4	5.56	3.00	−0.63–0.66	−0.38–0.39
5	5.77	3.64	−0.84–−0.33	−0.16–0.25
6			−0.10	0.12

Table 4. The P 2*p* and In 3*d* complete screening (CS) and initial state (IS) model CLSs of several compositionally different semi-coherent HfO_2/InP interfaces grouped in terms of the close O neighbors. A P (In) atom has a close O neighbor, if the interatomic distance is smaller than 2.0 Å (2.7 Å). The chosen cutoffs are somewhat arbitrary (as the concept of bond), but the found trends are not affected by this slight arbitrariness. One P + 1 configuration is considered unlikely (having two relatively distant Hf neighbors in addition to one O neighbor). The parenthesis show values without this configuration. There is only one value for the P N_O (number of close O neighbors) equal to four and five. These P atoms are above the interface layer and substitute Hf atoms. Similarly there is only one value for the In N_O equal to six.

The impurity CLSs calculated for the bulk ground state monoclinic structure are practically identical to those calculated for the anatase structure. Nine different HfO_2/InP interface models based on the O10 model were constructed to investigate the CLSs of the impurity and interfacial atoms. The InP part can be either In or P terminated whereas the In and P concentrations vary at the first layer of the HfO_2 part. There are many different kinds of In and P atom environments at the interface due to these variations. The first layer of the InP part is dimerized and oxygen atoms can be inserted into the dimers. All considered interfaces have an energy gap which points out that semi-coherent interfaces are electronically flexible. The electron counting rule (ECR) can be satisfied by In unoccupied and P occupied dangling bonds. The complete screening CLSs of the P atoms can be classified in terms of the close O neighbors (*i.e.*, O bonds) which is shown in Table 4 (only complete screening CLSs are considered below). This is very simple and interesting taking into account that some of the interface atoms have an unusual atomic neighbor configuration. Furthermore, the relative valence band offsets vary within an interval of 1.5 eV. The robustness of the classification suggests that the same principle might be applied also to different kind of interfaces (*e.g.*, more diffuse interfaces). It should be noted that the concept of "close O neighbor" is convenient in practice. Taking into account the calculated As 2*p* CLSs at the Al_2O_3/GaAs interfaces[40] (the only previous first-principles study of oxide/III-V interface CLSs found) it is possible that similar interpretation might be valid for III-V semiconductors more generally. It should also be noted that the P CLSs are not increased significantly by increasing the number of close O neighbors from four.

However, it is shown in the Table 4 that there is no correlation between the CLSs and the number of close neighbors of the In atoms. This shows that the charge transfer model expressed in terms of the number of oxygen neighbors is not generally valid assuming that the calculated results reproduce the experimental trends. It has been shown that the local charge or *d* charge of transition metal atoms is approximately constant in different "charge states" in several oxides[41,42]. However, the phenomena expressed in terms of the "oxidation state" or "charge state" are real in these cases[41,42].

An interesting correlation was found between the composition of the interface layer at the boundary of the HfO_2 oxide part and the band offset. The valence band maximum of the oxide part seems to decrease with respect to the InP part with the concentration of the P atoms within the oxide interface layer. This might be attributed to the occupied P dangling bonds which cause gap states. The band offsets reported by Santosh *et al.* for a different O10 model (In terminated interfaces were not considered) seem to follow roughly this trend[22]. The band offsets are reflected also in the CLSs of the substitutional/impurity In and P atoms within deeper layers of the HfO_2. The CLSs for the substitutional In and P in the centre of the HfO_2 part within the O10 model are −0.89 eV and 5.49 eV. The magnitudes of these CLSs are smaller than those calculated for pure separate HfO_2. The decreased CLSs imply that due to the interface there is a band offset which decreases the CLS within the HfO_2/InP interface system relative to the separate bulk HfO_2 and InP phases. The band offset is larger for other interface models within an interval of 1.5 eV increasing the CLS.

It is possible that there is some native oxide between the HfO_2 and InP parts, and this is the most realistic scenario for the investigated structures especially since HfO_2 was specifically grown on a native oxide on one of the samples. Two $InPO_4$ model oxides, constrained to the InP interface area, are introduced to calculate CLSs for two coherent HfO_2/$InPO_4$/InP double interface systems. The first model oxide ($InPO_4$-a) is based on the HfO_2 anatase structure in which every second atomic layer in direction of the longest lattice parameter is substituted with either In or P atoms. However, the doubling of the length of the *c* lattice parameter decreases total energy as the PO_4 tetrahedra can be oriented in different ways. The second oxide model structure ($InPO_4$-b) can be relaxed from an orthorhombic initial structure (by non-symmetric atomic displacements). The space group number of this initial structure is 80. There are two O atom Wyckoff positions (8a) ($x = 0.232$; $y = 0.301$; $z = 0.297$; $x = 0.778$; $y = 0.160$; $z = 0.160$). The Wyckoff *z* parameters for the In and P (4*a*) positions are 0.697 and 0.098, respectively. The resulting $InPO_4$-a and $InPO_4$-b oxide structures may depend on the size of the chosen cell due to disorder. The structures used are calculated for an (2×2) interface area. The unrelaxed model oxides are composed of similar structural motifs with different stackings. Total energies of these oxides are larger than the total energy of the

ground state $InPO_4$ structure. The total energies of the $InPO_4$-a and $InPO_4$-b are 0.17 eV/atom and 0.09 eV/atom with respect to the total energy of the ground state $InPO_4$ (orthorhombic lattice, ref.[43]). For comparison, the only bulk $InPO_4$ phase with a square face, scheelite[43], has a relative total energy of 0.10 eV/atom. Therefore, the total energies of the considered model oxides are not unrealistically high. Furthermore, the decreasing of the interface area increases the total energy of these model $InPO_4$ oxides in opposition to the HfO_2. Therefore, it is not relevant to consider semi-coherent $InPO_4$/InP interfaces based on the O10 model. The bulk complete screening P 2*p* CLSs (In 4*d* states in core) for the $InPO_4$-a and $InPO_4$-b are 5.02 eV and 4.55 eV, respectively. The corresponding values for the ground state and scheelite $InPO_4$ are 5.21 eV and 5.27 eV.

The $InPO_4$-a and $InPO_4$-b oxides may grow in thin films, if the interface energy for the ground state $InPO_4$ structure is relatively high. The total energy of the double interface system with an $InPO_4$-a thin film (two In-P oxide double layers) is slightly lower than that based on the $InPO_4$-b [0.26 eV per (1×1) interface area], which points out that the interface energy is relatively low for the $InPO_4$-a structure. The result also shows that a relatively significant bulk $InPO_4$ total energy difference (~0.1 eV/atom) may be compensated by the interface energy difference in thin films. It is supposed that the chosen terminations of the $InPO_4$ models are energetically favorable due to the characteristic form of the PO_4 tetrahedron found in all $InPO_4$ phases (no broken PO_4 tetrahedra).

The CLSs for the double interface systems were calculated using from two to four In-P oxide double layers. The highest P CLSs originate just below the HfO_2, while the other P layers in the $InPO_4$ show quite similar CLSs among each other. The P CLSs for the interface layer just below the HfO_2 and other $InPO_4$ layers are 5.03–5.06 eV and 4.63–4.66 eV, and 5.23–5.35 eV and 4.45–4.65 eV for the $InPO_4$-a and $InPO_4$-b, respectively. The native oxide and interface regions of this unit cell structure are shown in the Fig. 3. The CLSs are quite similar, although the difference between the bulk $InPO_4$-a and $InPO_4$-b CLSs is slightly larger. The valence band offset is smaller for the $InPO_4$-a, which decreases the CLSs for the $InPO_4$-a. The orientation of the PO_4 tetrahedra are different in the $InPO_4$-a and $InPO_4$-b, which may contribute to the relative band offset. Thus, band bending is possible also without composition changes within the interfacial layers at the oxide and semiconductor parts. The results point out that different In and P CLSs may be originated from a chemically uniform interface system. Thus, the different experimental CLSs do not originate necessarily from different oxide films having, *e.g.*, In_2O_3 and $InPO_4$ compositions.

Discussion

Finally, the experimental CLSs are analyzed using the calculated CLSs. The experimental P CLSs (I1, I2, O1, O2, O3, O4) are −0.18, 0.30, 3.51, 4.50, 4.97, 5.30 eV. It is noted first that the experimental P CLSs are in good agreement with the calculated ones for the model $InPO_4$/InP interfaces. The results show that P oxidation states +1 and +2 are missing. This suggests that the first interface layer in the InP part is composed of In atoms, because the P dimers probably tend to be oxidized. However, if the interface includes P-P dimers, they cause small positive shift 0.2–0.5 eV for P 2*p* according to the calculations. The O1 peak (3.51 eV) vanishes with annealing which means that the broken PO_4 tetrahedra disappear (Table 4). The relative intensity of the O2 (4.50 eV) is decreased whereas the relative intensities of the O3 (4.97 eV) and O4 (5.30 eV) are increased by annealing which could reflect thinning of the $InPO_4$ part (because then the relative weight of the layer just below the HfO_2 increases), but the depth analysis gives reason to suspect other effects than just thinning. On the other hand, the relative amount of different $InPO_4$ phases could be changed. Alternatively, when considering the effect of previously observed indium out-diffusion[18], it is likely that composition also changes. A noteworthy observation about $InPO_4$ CLSs is that all of the native oxide stacks considered produce smaller shifts in the mid-layer of the native oxide than the corresponding bulk oxide (about 4.6 eV for $InPO_4$-a and $InPO_4$-b vs. 5.2 eV for $InPO_4$ bulk). Thus, it is possible that out-diffusion could cause In-deficient phases in mid-layers of the native oxide (originally mainly composed of $InPO_4$) similar to $In(PO_3)_3$ (6.2–6.9 eV in bulk), that would match the O4 BE (5.3 eV). This is consistent with the higher stability of P-O bonding as compared to In-O[18]. Since there is out-diffusion of In in the HfO_2, the depth analysis is well reasoned: O4 is observed an increase especially further away from bulk than other components, probably because the rate of out-diffusion is likely higher closer to the native oxide/HfO_2 interface.

Furthermore, the sulfide treatment has been observed to suppress the indium out-diffusion[18]. In our experiments and based on the above analysis, this is observed in P 2*p* as the lack of O4 signal, or In-deficient bonding, consistently with the amount of In staying relatively constant in the sulfide/native oxide film. The O3 component intensity increases, which is likely related to the increased relative weight of the layer just below the HfO_2 as described previously. Here, also thinning of the sulfide/native oxide film is plausible, since the proportional emission of B signal increases after annealing.

In order to justify the analysis above, similar effects need to be observed also for In. However, it is to be noted that our computational results underscore the difficulty in making well justified interpretations about the In 3*d* XPS results for our samples, as the CLSs are found with only slight offset from the bulk core-level. Furthermore, the shifts are not consistent with the amount of nearest-neighbor O, or straightforwardly with valence charge, as opposed to P 2*p*. However, the relative differences between BEs of different In-P oxide bulk phases are close to the ones reported in literature. In_2O_3 is however typically associated with positive shifts, contrary to the computational results. On the other hand, reference data tables suggest very similar BEs for bulk In_2O_3 and InP[44], which is why a small negative shift for In_2O_3 in the structure for any particular oxide/InP systems is not beyond reasoning, but on the contrary, suggested also by the calculations. Without taking this into consideration, there is a considerable chance of misinterpretation since, as mentioned, elemental/metallic In can cause very similar shifts.

The out-diffusion of indium being the established culprit of device performance degradation on HfO_2/InP interfaces, it is of necessity to consider this effect as has been done above. The defective sites accountable for the diffusion (interstitial defects containing In)[45] are not, however taken into account in the spectral analysis. A concentration of these defects that would be detectable in XPS (0.1–1%) would also significantly alter the oxide characteristics and cause much higher amount of trap states that has been observed[3]. However, despite a small

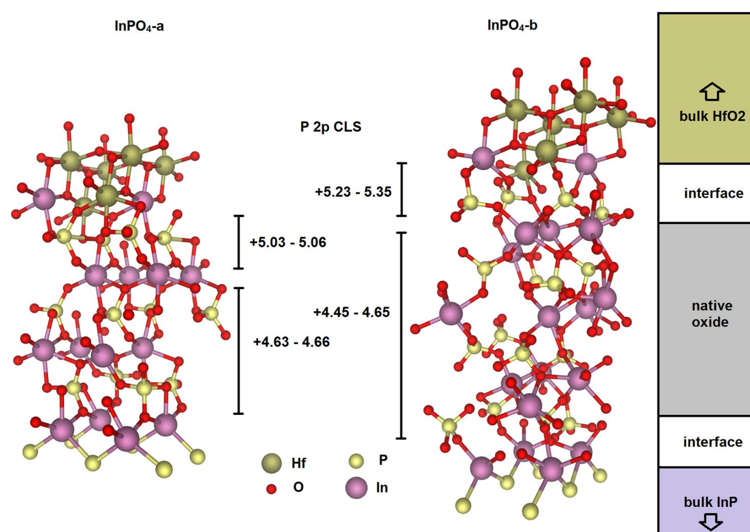


Figure 3. $InPO_4$-a and $InPO_4$-b structures between bulk InP and HfO_2 shown with the calculated native oxide CLSs of P 2p of the corresponding structures (eV).

concentration, a significant amount of In can diffuse to the surface if the flow is continuous as can be assumed during the annealing. This has been observed previously with LEIS[4], and thus, surface segregated In needs to be considered as a chemical state observable especially in the surface sensitive XPS setup such as has been utilized here.

Based on the analysis for P 2*p* above, and the fact that stoichiometrically identical compound can cause markedly different CLSs, we assign the In 3*d* peaks I3, I4 and O to $InPO_4$ and $In(PO_3)_3$ contained within the native oxide or at the interface of HfO_2. The concentration of $InPO_4$ [I3 and I4] decreases as $In(PO_3)_3$ [O] bonding increases near the interface boundary after annealing, consistent with the analysis above. The depth distributions stay fairly constant, which is consistent apart from $In(PO_3)_3$ which changes much more towards the surface for P 2*p*. This could be because, the stoichiometric factor of In in $In(PO_3)_3$ is smaller than that of P, so that similar change in concentration has only a third of an effect on the intensity of the corresponding component, and because there is likely overlapping of states related to either interface bonding or other oxide phases, that do not change their depth distribution, and thus diminish the effect of changes in $In(PO_3)_3$ seen in depth distribution. The I2 is increased dramatically and brought closer to the surface after the annealing. However, the depth distribution is not changed as dramatically, indicating that there is increase in the concentration of this component both near the bulk and the surface. Here the previous reports and our complementary computational results bring insight onto identification of the state; I2 actually likely consists of two overlapping peaks: In_2O_3 type emission that is observed near the bulk, and elemental/metallic In at the surface due to out-diffusion. Some of the segregated In is evaporated, which is observed as an increase in the B signal relative to the others.

Noting the formation of In_2O_3, we suggest a following model which accounts for all of the effects discussed above. Initially the native oxide film or the thin oxide present also in sulfide-treated sample consists mainly of $InPO_4$. During annealing, the P atoms tend to bond with O, producing $In(PO_3)_3$, or other In-deficient phases. For P, the most prominent differences are at the topmost layers of the native oxide, indicating either relatively more significant proportion of P atoms having multiple bonds to the O in HfO_2 layer, or more prominent formation of $In(PO_3)_3$. This relieves In, that can form In_2O_3 or other In-rich phases locally, causing some phase separation. Some of the O and/or In atoms could be provided by phase separation at the oxide/InP interface into P dimers. On the other hand, atomic In in the upper layers of native oxide is able to diffuse to the surface through vacancy sites in HfO_2. This makes sense, since as noted above, the most marked differences are at the topmost layers of the native oxide, and the differences are due to more P-O bonds. It is possible that during annealing in the HfO_2 film, extra O vacancies are formed, through which In has been reported to diffuse. Thus, the degradation occurs due to the inherent chemistry of the native oxide film as a result of these synergistic effects, and thus, it is readily prevented by saturation of InP surface dangling bonds with e.g. S after the sulfide treatment. However, being an *ex-situ* method, the sulfide passivation is not able to prevent the formation of bonding observed in the native oxide altogether, which is why similar effects are seen on the sulfide-treated sample, but to a lesser extent.

	Shape	FWHM (eV)	BE position (eV)
B	GL(80)	0.39–0.45	128.55–128.85
I1	GL(80)	0.35–0.5	B-0.18
I2	GL(80)	0.35–0.5	B + 0.3
O1	GL(65)	0.65–0.85	B + 3.51
O2	GL(65)	0.65–0.85	B + 4.5
O3	GL(65)	0.65–0.9	B + 4.97
O4	GL(65)	0.65–1.0	B + 5.3

Table 5. Peak fitting parameters for P $2p_{3/2}$ components before and after annealing as well as for as-grown and S-treated samples. BE position of components O1–O4 was allowed to vary 0.1 eV from their fixed position.

	Shape	FWHM (eV)	BE position (eV)
B	GL(87)	0.5–0.6	452.05–452.45
I1	GL(60)	0.5–0.7	B-1.1
I2	GL(87)	0.5–0.8	B-0.22
I3	GL(60)	0.6–0.8	B + 0.35
I4	GL(60)	0.6–0.7	B + 0.8
O	GL(50)	0.8–1.2	B + 1.3

Table 6. Peak fitting parameters for the In $3d_{3/2}$ components.

Conclusions

The presented results for HfO_2/InP junctions demonstrate that the semiconductor oxidation can cause negative CLSs (*i.e*., a decrease in core-level BE as compared to the clean semiconductor), in contrast to the common hypothesis that the material oxidation causes positive CLSs, which is based on the charge-transfer model and the well-understood SiO_2/Si system. The P CLSs can be estimated robustly at the abrupt HfO_2/InP interfaces considering the number of close O neighbors irrespective of the other atomic neighbors, resembling the SiO_2/Si system, but no similar correlation was found for the In CLSs. The In CLSs cannot be explained by the number of close O neighbors. The results emphasize that the special care needs to put on determining the reference BE (e.g., bulk emission peak position) by changing the surface-sensitivity of the measurements.

To strengthen the XPS analysis and to utilize full potential of the method, we have combined *ab initio* calculations and synchrotron XPS in the study of the example case of HfO_2/InP. Two model structures for the $InPO_4$/InP were introduced. These are the first model interfaces structures for native oxides of III-V semiconductors which can be used, *e.g*., to estimate, whether coherent or semi-coherent interface growth is preferred. A correlation was found between the number of P atoms in the interface oxide layer and the band offset at the semi-coherent HfO_2/InP interfaces. A model consistent with our experiments and calculations as well as previous reports concerning annealing effects on HfO_2/InP system has been presented. We suggest that annealing can induce effects at the oxide/semiconductor interface that result in CLSs without necessarily changing the chemical stoichiometry, but rather the bonding configuration. Furthermore, markedly different chemical states can be observed at the same BE. These effects complicate XPS analyses, and the results underline the importance of complementary studies and high resolution XPS data. Here, we have been able to identify the atomic origins of CLSs that can remain totally hidden in the traditional laboratory XPS spectra. Our findings may pave the way for systematic improvement of the interpretation of CLS in relation to characterization of materials at the atomic scale both in academic and industrial investigations where CLS are at the heart of advancing knowledge.

Methods

Sample and measurement setup.

Our XPS experiments were carried out in the synchrotron radiation centre MAX-lab, Lund, Sweden, at beamline I311. The base pressure of the experimental station was in 10^{-10} mbar range. The photon energy, $h\nu$, was varied to measure the P 2*p* and In 3*d* peaks with two different kinetic energies (KE, i.e., surface sensitivities): 150 eV and 300 eV ($h\nu$ of 279 eV and 429 eV for P 2*p*, and 594 eV and 744 eV for In 3*d*). Gaussian broadening of the signal arising from the instrumentation is estimated to be less than 0.15 eV. Two samples were investigated. An InP(100) crystal with a native oxide film on top of which a HfO_2 film was grown by ALD. Another InP(100) sample was treated by 10% $(NH_4)_2S$ aqueous solution diluted from 20% aqueous solution. TDMA-Hf was used as the metal precursor, and H_2O vapor as the oxidant precursor with ultrahigh purity N_2 gas as the carrier gas. The temperature of ALD was at 250 °C, and 20 cycles of ALD of HfO_2 were grown on the InP wafer by a pulse sequence of Hf/purge/H_2O/purge for 0.1 s/10 s/0.1 s/10 s, respectively. The growth corresponds to a uniform film thickness of approximately 1.6 nm with an established growth rate of 0.08 nm per cycle[18]. After the ALD growth the samples were transferred to the *ex-situ* synchrotron radiation centre. The samples were measured before and after post-growth annealing at 400–450 °C in the UHV system, to investigate temperature dependent compositional changes in the oxide film and at the oxide-semiconductor interface.

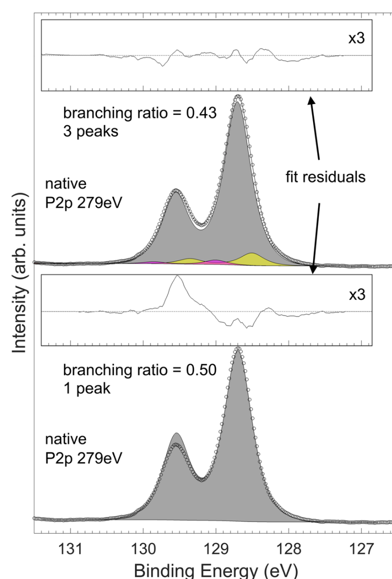


Figure 4. An example P 2*p* spectrum with separate fittings showing the effect of branching ratio parameter and additional components to the fit residual. The applied parameters (top spectrum) result in a low residual especially at branching point and systematically reduced around both P $2p_{3/2}$ and $2p_{1/2}$ without overemphasizing any features.

Spectral analysis. A fitting procedure similar to one used for describing surface CLSs for In containing semiconductors was applied with CasaXPS software version 2.3.16[46]. However, due to overlapping of In $3d_{5/2}$ with Hf $4p_{1/2}$, In $3d_{3/2}$ was used in the fitting because this peak still gives a high intensity. This approach is convenient due to the relatively high spin-orbit splitting of In 3*d* (~7.5 eV) so that overlapping of the spin-orbit peaks can be avoided. Fitting was carried out to deconstruct XPS spectra into a minimum number of individual components that were required to reproduce the spectral features observed. An essential requirement in the fitting was that higher $h\nu$, which provides a more bulk sensitive measurement, caused higher relative intensity for the bulk peaks. Even though bulk peak intensity ratio of P $2p_{3/2}$ to In $3d_{3/2}$ is not 1:1 (due to different photoionization cross-section and photon flux), the relative difference in absolute bulk signal intensity needs to vary identically between treatments for these peaks with a given KE, since bulk bonding consists of In-P bonds only and this bonding environment starts from beneath the exact same depth. Thus, the bulk peak intensity ratio of In $3d_{3/2}$ at a given KE before and after annealing treatment was bound to vary in similar ratio as the bulk P $2p_{3/2}$ at the same KE, by adjusting the bulk peak BE position. 1:1 stoichiometry condition in InP bonding environment was more reliably satisfied this way, since P $2p_{3/2}$ bulk signal was observed without significant overlappings. Variation in the photon flux from the synchrotron ring was taken into account by scaling the intensity of each measurement with the ring current during the corresponding measurement, so that absolute intensity values of different measurements could be reliably compared this way. The actual photon flux from the beamline optics to the sample could not be taken into consideration, but should be similar for two measurements with the same $h\nu$ after scaling with the ring current.

The fitting parameters which reproduced the spectral envelopes are shown in Tables 5 and 6. The spin-orbit splitting was 0.85 eV and branching ratio 0.40–0.43 between the P $2p_{3/2}$ and $2p_{1/2}$ peaks. A P 2*p* peak sum with 0.5 branching ratio, that would account for the envelope spectrum could not be introduced, which is obvious from the absence of any significant tail features at higher or lower BE. No other photoelectron peak should be observed at this BE for Hf, O, P, In or C either. Thus, we expect this discrepancy to be caused by some other external effect, such as diffraction or multiplet splitting. Figure 4 illustrates an example of the significant reduction in asymmetric fit residual for P 2*p* bulk peak area when changing the branching ratio and adding two adjacent components. The justifications for the introduced components are further discussed in the Results section.

A systematic BE increase of 0.1–0.2 eV was observed in all of the S-treated sample's components. We note that this could be related to charge redistribution near the interface, but such analysis is omitted since the Fermi-energy was not calibrated separately for the measurements of the two different samples.

Calculations

Calculations were performed using an *ab initio* density functional theory (DFT) total energy method within the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA)[47]. The approach is based on the plane wave basis and projector augmented wave method[48,49] (Vienna *ab initio* simulation package, VASP)[50–53]. The interfaces were modeled using unit cells with two equal (single or double) interfaces without vacuum. The optimization of the atomic structure was performed using the conjugate gradient minimization of the total energy with respect to the atomic coordinates. Atoms were relaxed until the remaining forces were less than 20 meV/Å. The plane wave cutoff energies of 350 eV and 500 eV were used for interface and bulk calculations, respectively. All test calculations with the cutoff energy of 500 eV showed only marginal differences for the interface CLSs and total energies. The In 4*d* and P 3*d* as well as Hf 5*p* electrons were treated as core electrons within the interface systems. The In 4*d* electrons were treated both as valence electrons and core electrons in the bulk calculations. The interface *k* point sampling was carried out by the Monkhorst-Pack scheme[54] using a $4 \times 4 \times 1$ mesh for (2×2) interface area. The origin was shifted to the Γ point.

The HfO_2/InP interface unit cells consist of 8–9 layers of group III atoms, 8–9 layers of group V atoms, 5 layers of Hf atoms, and 6 layers of O atoms. The HfO_2/$InPO_4$/InP double interface unit cells consist of 10–14 layers of group III atoms, 10–14 layers of group V atoms, 4–6 layers of Hf atoms, and 11–19 layers of O atoms. The In_2O_3, $InPO_4$, $In(PO_3)_3$ and P_2O_5 initial structures are obtained from the refs[43,55–57]. The initial state CLSs were determined by calculating the electrostatic potential at each ion core. The Fermi level is set to the middle of the band gap. The complete screening calculations (core hole and an extra screening valence electron) were calculated using large supercells (about 100 atoms for bulk calculations) to minimize artificial interaction of the core-ionized atoms. Some test calculations were performed using even larger cells (*e.g.*, 640 atoms for In_2O_3). The atoms in the central layers of the InP part represent bulk atoms in the interface calculations. The inaccuracy of the CLSs with respect to the length of the InP part is assessed to be smaller than 0.1 eV. The interface area is (2×2) except for the In and P impurity calculations in which an (4×4) interface area was used.

Data Availability

The photoelectron spectra and data used for the computational studies are available from the corresponding authors on reasonable request.

References

1. Sokolowski, E., Nordling, C. & Siegbahn, K. Chemical Shift Effect in Inner Electronic Levels of Cu Due to Oxidation. *Phys. Rev.* **110**, 776 (1958).
2. Egelhoff, W. F. Jr. Core-level binding-energy shifts at surfaces and in solids. *Surf. Sci. Rep.* **6**, 253 (1987).
3. Briggs, D. & Seah, M. P. Practical Surface Analysis, Volume 1: Auger and X-ray Photoelectron Spectroscopy (John Wiley & Sons Inc., West Sussex 1990).
4. Pehlke, E. & Scheffler, M. Evidence for site-sensitive screening of core holes at the Si and Ge (001) surface. *Phys. Rev. Lett.* **71**, 2338 (1993).
5. Alden, M., Skriver, H. L. & Johansson, B. Ab initio surface core-level shifts and surface segregation energies. *Phys. Rev. Lett.* **71**, 2449 (1993).
6. Himpsel, F. J., McFeely, F. R., Taleb-Ibrahimi, A., Yarmoff, J. A. & Hollinger, G. Microscopic structure of the SiO2/Si interface. *Phys. Rev. B* **38**, 6084 (1988).
7. Pasquarello, A., Hybertsen, M. S. & Car, R. Si 2p Core-Level Shifts at the Si(001)-SiO2 Interface: A First-Principles Study. *Phys. Rev. Lett.* **74**, 1024 (1995).
8. Tu, Y. & Tersoff, J. Structure of the silicon-oxide interface, Thin Sol. *Films* **400**, 95 (2000).
9. Bongiorno, A., Pasquarello, A., Hybertsen, M. S. & Feldman, L. C. Transition Structure at the Si(100)-SiO2 Interface. *Phys. Rev. Lett.* **90**, 186101 (2003).
10. Oktyabrsky, S. & Ye, P. D. Fundamentals of III-V Semiconductor MOSFETs (Springer 2010).
11. Walter, A. L. *et al*. X-ray photoemission analysis of clean and carbon monoxide-chemisorbed platinum(111) stepped surfaces using a curved crystal. *Nat. Comm.* **6**, 8903 (2015).
12. Zheng, Z. *et al*. Semiconductor SERS enhancement enabled by oxygen incorporation. *Nat. Comm.* **8**, 1993 (2017).
13. Gopal, C. B. *et al*. Equilibrium oxygen storage capacity of ultrathin CeO2-δ depends non-monotonically on large biaxial strain. *Nat. Comm.* **8**, 15360 (2017).
14. Timm, R. *et al*. Self-cleaning and surface chemical reactions during hafnium dioxide atomic layer deposition on indium arsenide. *Nat. Comm.* **9**, 1412 (2018).
15. Tian, B. *et al*. Supported black phosphorus nanosheets as hydrogen-evolving photocatalyst achieving 5.4% energy conversion efficiency at 353K. *Nat. Comm.* **9**, 1397 (2018).
16. Robertson, J. & Wallace, R. M. High-K materials and metal gates for CMOS applications, Mater. *Sci. & Engin.* **88**, 1 (2015).
17. Galatage, R. V. *et al*. Effect of post deposition anneal on the characteristics of HfO2/InP metal-oxide-semiconductor capacitors. *Appl. Phys. Lett.* **99**, 172901 (2011).
18. Dong, H. *et al*. Indium diffusion through high-k dielectrics in high-k/InP stacks. *Appl. Phys. Lett.* **103**, 061601 (2013).
19. Xu, M. *et al*. New insights in the passivation of high-k/InP through interface characterization and metal–oxide–semiconductor field effect transistor demonstration: Impact of crystal orientation. *J. Appl. Phys.* **113**, 013711 (2013).
20. Dong, H. *et al*. *In situ* study of the role of substrate temperature during atomic layer deposition of HfO2 on InP. *J. Appl. Phys.* **114**, 154105 (2013).
21. Dong, H. *et al*. *In situ* study of e-beam Al and Hf metal deposition on native oxide InP (100). *J. Appl. Phys.* **114**, 203505 (2013).
22. Santosh, K. C. *et al*. Electronic properties of InP (001)/HfO2 (001) interface: Band offsets and oxygen dependence. *J. Appl. Phys.* **115**, 023703 (2014).
23. Galatage, R. V. *et al*. Accumulation capacitance frequency dispersion of III-V metal-insulator-semiconductor devices due to disorder induced gap states. *J. Appl. Phys.* **116**, 014504 (2014).
24. Dong, H. *et al*. Silicon Interfacial Passivation Layer Chemistry for High-k/InP Interfaces, ACS Appl. *Mat. & Interf.* **6**, 7340 (2014).
25. Wallentin, J. *et al*. InP Nanowire Array Solar Cells Achieving 13.8% Efficiency by Exceeding the Ray Optics Limit. *Science* **339**, 1057 (2013).
26. Oener, S. Z. *et al*. Charge carrier-selective contacts for nanowire solar cells. *Nat. Comm.* **9**, 3248 (2018).
27. May, M. M., Lewerenz, H.-J., Lackner, D., Dimroth, F. & Hannappel, T. Efficient direct solar-to-hydrogen conversion by *in situ* interface transformation of a tandem structure. *Nat. Comm.* **6**, 8286 (2015).

28. Chen, C. L. *et al*. Wafer-scale 3D integration of InGaAs photodiode arrays with Si readout circuits by oxide bonding and through-oxide vias, Microelectr. *Engineer.* **88**, 131 (2011).
29. Yang, J. *et al*. Low leakage of In0.83Ga0.17As photodiode with Al2O3/SiNx stacks, Infrar. *Phys. & Techn.* **71**, 272 (2015).
30. Cuypers, D. *et al*. Study of InP Surfaces after Wet Chemical Treatments. *ECS Journal of Solid State Science and Technology* **3**, N3016–N3022 (2014).
31. Adelmann, C. *et al*. Surface Chemistry and Interface Formation during the Atomic Layer Deposition of Alumina from Trimethylaluminum and Water on Indium Phosphide. *Chem. Mater.* **25**, 1078 (2013).
32. Chen, G., Visbeck, S. B., Law, D. C. & Hicks, R. F. Structure-sensitive oxidation of the indium phosphide (001) surface. *J. Appl. Phys.* **91**, 9362 (2002).
33. Kaspari, C., Pristovsek, M. & Richter, W. Deoxidation of (001) III–V semiconductors in metal-organic vapour phase epitaxy. *J. Appl. Phys.* **120**, 085701 (2016).
34. Heyd, J. & Scuseria, G. E. Efficient hybrid density functional calculations in solids: Assessment of the Heyd-Scuseria-Ernzerhof screened Coulomb hybrid functional. *J. Chem. Phys.* **121**, 1187 (2004).
35. Van den Bossche, M. *et al*. Effects of non-local exchange on core level shifts for gas-phase and adsorbed molecules. *J. Chem. Phys.* **141**, 034706 (2014).
36. Bagus, P. S., Illas, F., Pacchioni, G. & Parmigiani, F. Mechanisms responsible for chemical shifts of core-level binding energies and their relationship to chemical bonding. *J. Electron Spectrosc. Relat. Phenom.* **100**, 215 (1999).
37. Lahti, A. *et al*. Electronic structure and relative stability of the coherent and semi-coherent HfO2/III-V interfaces. *Appl. Surf. Sci.* **427**, 243 (2018).
38. Wang, W., Xiong, K., Wallace, R. M. & Cho, K. Impact of Interfacial Oxygen Content on Bonding, Stability, Band Offsets, and Interface States of GaAs:HfO2 Interfaces. *J. Phys. Chem. C* **114**, 22610 (2010).
39. Debernardi, A., Wiemer, C. & Fanciulli, M. Epitaxial phase of hafnium dioxide for ultrascaled electronics. *Phys. Rev. B* **76**, 155405 (2007).
40. Miceli, G. & Pasquarello, A. First principles study of As 2p core-level shifts at GaAs/Al2O3 interfaces. *Appl. Phys. Lett.* **102**, 201607 (2013).
41. Quan, Y., Pardo, V. & Pickett, W. E. Formal Valence, 3d-Electron Occupation, and Charge-Order Transitions. *Phys. Rev. Lett.* **109**, 216401 (2012).
42. Raebiger, H., Lany, S. & Zunger, A. Charge self-regulation upon changing the oxidation state of transition metals in insulators. *Nature (London)* **453**, 763 (2008).
43. López-Moreno, S. & Errandonea, D. Ab initio prediction of pressure-induced structural phase transitions of CrVO4-type orthophosphates. *Phys. Rev. B* **86**, 104112 (2012).
44. Moulder, J. F., Stickle, W. F., Sobol, P. E. & Bomben, K. D. Handbook of X-ray photoelectron spectroscopy, Vol. 40 (Perkin Elmer Eden Prairie, MN, 1992)
45. Hu, Y. *et al*. Origin of Indium Diffusion in High-k Oxide HfO2. *ACS Appl. Mater. Interfaces* **8**, 7595 (2016).
46. Mäkelä, J. *et al*. Line shape and composition of the In 3d5/2 core-level photoemission for the interface analysis of In-containing III-V semiconductors. *Appl. Surf. Sci.* **329**, 371 (2015).
47. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
48. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
49. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
50. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558 (1993).
51. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251 (1994).
52. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. *Mat. Sci.* **6**, 15 (1996).
53. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
54. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).
55. Karazhanov, S. Z. *et al*. Phase stability, electronic structure, and optical properties of indium oxide polytypes. *Phys. Rev. B* **76**, 075129 (2007).
56. Pauling, L. & Sherman, J. The Crystal Structure of Aluminum Metaphosphate, Al(PO3)3. *Z. Kristallogr.* **96**, 481 (1937).
57. Abarenkov, I., Tupitsyn, I., Kuznetsov, V. & Payne, M. The electronic structure of crystalline phosphorus pentoxide and the effect of Ag impurity. *Phosph. Res. Bull.* **10**, 123 (1999).
58. Hollinger, G., Bergignat, E., Joseph, J. & Robach, Y. On the nature of oxides on InP surfaces. *J. Vac. Sci. Technol. A* **3**, 2082 (1985).
59. Hollinger, G. *et al*. On the chemistry of passivated oxide-InP interfaces. *J. Vac. Sci. Technol. B* **5**, 1108 (1987).
60. Hoekje, S. J. & Hoflund, G. B. Surface characterization study of InP(100) substrates using ion-scattering spectroscopy, Auger electron spectroscopy and electron spectroscopy for chemical analysis I: Comparison of substrate-cleaning techniques. *Thin Solid Films* **197**, 367 (1991).

Acknowledgements

This work has been supported by University of Turku Graduate School (UTUGS) and the Academy of Finland (project no. 296469). The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

Author Contributions

J.M. carried out fitting analysis of the XPS spectra. B.B., R.M.W., and P.L. planned the samples and measurements. H.D., B.B., and R.M.W. prepared the samples. M.T., M.Y., M.K., and P.L. executed the experiments at synchrotron facility. A.L. and M.P.J.P. constructed the atomic models and performed the computational part of the study. J.M., M.P.J.P., H.D., B.B., R.M.W., P.L., and K.K. analyzed the results. All of the authors participated in writing of the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-37518-2.

Competing Interests: The authors declare no competing interests.

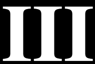
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

Antti Lahti & Ralf Östermark & Kalevi Kokko
Optimizing Atomic Structures through Geno-Mathematical Programming

Communications in Computational Physics, 25(3), 2019, 17



Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2017-0253

Vol. **25**, No. 3, pp. 911-927
March 2019

Optimizing Atomic Structures through Geno-Mathematical Programming

Antti Lahti[1,2,*], Ralf Östermark[3] and Kalevi Kokko[1,2]

[1] *Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland.*
[2] *Turku University Centre for Materials and Surfaces (MatSurf), Turku, Finland.*
[3] *Åbo Akademi University, School of Business and Economics, FIN-20500 Turku, Finland.*

Received 15 December 2017; Accepted (in revised version) 18 April 2018

Abstract. In this paper, we describe our initiative to utilize a modern well-tested numerical platform in the field of material physics: the Genetic Hybrid Algorithm (GHA). Our aim is to develop a powerful special-purpose tool for finding ground state structures. Our task is to find the diamond bulk atomic structure of a silicon supercell through optimization. We are using the semi-empirical Tersoff potential. We focus on a 2x2x1 supercell of cubic silicon unit cells; of the 32 atoms present, we have fixed 12 atoms at their correct positions, leaving 20 atoms for optimization. We have been able to find the known global minimum of the system in different 19-, 43- and 60-parameter cases. We compare the results obtained with our algorithm to traditional methods of steepest descent, simulated annealing and basin hopping. The difficulties of the optimization task arise from the local minimum dense energy landscape of materials and a large amount of parameters. We need to navigate our way efficiently through these minima without being stuck in some unfavorable area of the parameter space. We employ different techniques and optimization algorithms to do this.

AMS subject classifications: 82-08

PACS: 81.05.Cy

Key words: Optimization, geno-mathematical programming, bulk silicon, semi-empirical potential.

1 Introduction

Our interest in especially semiconductor-oxide interface and surface structures is due to their prevalence in modern electronics and devices; these structures heavily affect the

*Corresponding author. *Email addresses:* `ailaht@utu.fi` (A. Lahti), `ralf.ostermark@abo.fi` (R. Östermark), `kokko@utu.fi` (K. Kokko)

 ©2019 Global-Science Press

performance of the different components present in the devices. On the interface there are two different interconnected crystal structures often leading to structures hard to predict. This is what makes them a difficult research object. Finding these structures by hand, i.e. designing the interfaces by trial and error needs special knowledge and takes a lot of time, which is why a flexible, powerful optimizing tool would be beneficial for many research problems across the field. Measuring the interface structures experimentally is also difficult as they are buried in the material. That is why simulations and calculations conducted on many different interface models are indispensable in understanding the nature and behavior of these structures. The scale of these structures is often measured in Ångström's (Å) which is short for 10^{-10}m.

We started a collaboration between the School of Business and Economics at Åbo Akademi University and University of Turku's Materials Research Laboratory in order to develop a tool for optimizing atomic structures, with the algorithm especially tuned for interface structures. In geno-mathematical programming artificial intelligence is connected to mathematical programming methodology on parallel supercomputers. The approach provides a powerful basis for coping with difficult irregular optimization problems and solving them concurrently. We were also interested in the performance of our special-purpose algorithm in the physics based problem of optimizing atomic structures of materials, designed using a modern numerical platform as a base. The optimization is done by minimizing the potential energy, measured in electronvolts (eV), of the structure. The difficulty does not lie in finding a nearby local minimum from a given starting structure, as this can usually be achieved through the steepest descent method in a small amount of steps, but in navigating past all these minima to the global minimum. The energy landscape is filled with these local traps that do not reveal much if any indication on where the true global minimum lies.

Exploring the whole landscape is only doable in very small cases. This is because along with the increasing parameter count the number of local minima of the task rises exponentially with the number of atoms: for example with Lennard-Jones clusters it was shown that the number of minima multiplies by around 2-3 per atom added [23]. This makes almost any interesting interface or surface system hard to study. In this article, we show that even our small silicon case can be problematic if not treated properly.

In theory, good molecular dynamics (MD) simulation should be able to find the global minimum given enough time and proper annealing. In practice, the required time is often very large and might require a lot of parameter fine-tuning while still leaving defects at the end of the simulation. Of course, even then we can never be sure that the result is the true global minimum, unless we know the answer beforehand. Large scale computing is a crucial part of bigger MD simulations and advances in that field continue to be made even today, for example speeding up node and core communication [7, 26], reducing memory usage [7], improving threading [14] and creating faster algorithms for force computations [8]. In our work presented in this paper, all the cores work fairly independently, but in the future the algorithm could be expanded to include more communication between the cores. We did a series of MD simulations of our silicon test case

shown in Section 2, where we find that while we cannot guarantee the global minimum, it becomes very likely to be found as the simulation time increases.

Advanced techniques like basin hopping [24], meta dynamics [11], swarm simulation [2,3], genetic and evolution based methods have been developed and used over the years to find the global minimum of a system. They have been implemented to various degrees in software like USPEX [15], GASP [27] and CALYPSO [25] to name a few.

2 Silicon: 20-atoms case

Because finding the one global minimum amongst all the multiple local minima is a challenging task, we chose a well-known case of bulk silicon for our first study. The more difficult case of interfaces is left for future research. The structure of silicon we are trying to find is the well-known diamond structure, which can be represented as a cubic box with a periodic boundary condition, also known as a unit cell. We chose our optimization task to be a $2\times2\times1$ a supercell of these cubic diamond cells. The dimensions of this cell were not part of the optimization. The supercell contains 32 atoms in total, but we fixed all of the border atoms in place as shown with grey atoms in Fig. 1. This leaves 20 atoms for optimization, each having spatial coordinates x, y and z that gives us 60 parameters in total. Unlike in some smaller optimization tasks, we could not solve a problem of this size by brute force. The parameters have only simple box constraints determined by the supercell we are using. The cell dimensions are 10.86 Å, 10.86 Å and 5.43 Å, which leads to atom i's coordinates (x_i, y_i, z_i) having the box constraints given in Eq. (2.1):

$$0\leq x_i\leq 10.86,\quad 0\leq y_i\leq 10.86,\quad 0\leq z_i\leq 5.43. \tag{2.1}$$

We have fixed the border atoms for the following reasons:

- The end goal is to produce a package that searches interface and surface structures, which usually have a fixed known bulk structure surrounding the optimized region.
- It provided a starting point for the optimizer while also making it easier for us to analyze if the produced structure was correct or what kind of techniques would be needed to correct it.
- At the start we did not want to concern our optimizer with the periodic boundary conditions.

This leaves three of the four layers to be optimized. These layers have 8, 4 and 8 atoms. The middle layer has 8 atoms in it too, but four of those are frozen in place on the border of the supercell.

We further divide this optimization task into three different cases with 60, 43 and 19 parameters with the latter two having assumptions that reduce the parameter count. In

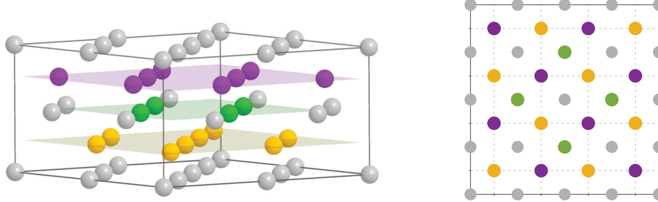


Figure 1: Presentation of our test case target solution from an angled and top-down perspective. The optimized atoms are colored while the fixed atoms are grey and located on the borders of the supercell. This is the ideal diamond structure for silicon we are trying to achieve.

the 60-parameter case all atoms have their positions optimized independently. The 43-parameter case assumes that the atoms reside in 3 layers. All the atoms share 3 height coordinates in total, reducing the parameter count by 17. The final 19-parameter case assumes, in addition that the two 8 atom layers have the atoms arranged in squares with 4 rows and 4 columns. This drops the parameter count from 43 to 19, as the new 8 row/column parameters replace the 32 x- and y-parameters of 16 atoms. In the structure of Fig. 1 these 8 parameters correspond to columns and rows of the yellow and purple atoms. In the global minimum structure these rows and columns are evenly spaced.

Especially the layer assumption is reasonable in many interface and surface models as long as we give some leeway for the atoms within the layer. In the 19-parameter case the row/column assumptions for atoms on a layer can also be useful as the low energy structures often have some form of symmetry that just needs to be found. In general using symmetries can be very advantageous as the reduction in parameters yields a smaller dimensional problem to be explored. The structures produced through these high symmetry cases can also be used as a starting point for runs where the symmetry assumptions are relaxed.

2.1 The Tersoff potential

We tried searching for the diamond structure through traditional molecular dynamics to give us a reference point for comparison. We used the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software for our simulations [17]. In this test case, we are using the Tersoff-potential [21], which is a fast many body bond-order potential[†]. The Tersoff potential allows fast computations and still describes silicon fairly well in different environments. Comparative testing with alternative potentials and pseudopotential methods (e.g. the slower but more accurate ReaxFF-potential [22] and VASP [10]) is left for future research.

The potential energy formula consists of a sum of two- and three-body interactions

[†]Potential energy refers to the Tersoff potential throughout the paper

and is implemented in LAMMPS in the following form:

$$
\begin{aligned}
&E=\frac{1}{2}\sum_i\sum_{i\neq j}f_C(r_{ij})[f_R(r_{ij})+b_{ij}f_A(r_{ij})],\\
&f_R(r)=Ae^{-\lambda_1 r},\quad f_A(r)=-Be^{-\lambda_2 r},\\
&f_C(r)=\begin{cases}1, & r<R-D,\\ \frac{1}{2}-\frac{1}{2}\sin(\frac{\pi}{2}\frac{r-R}{D}), & R-D<r<R+D,\\ 0, & r>R+D,\end{cases}\\
&b_{ij}=(1+\beta^n\zeta_{ij}^n)^{-\frac{1}{2n}},\\
&\zeta_{ij}=\sum_{k\neq i,j}f_C(r_{ik})g(\theta_{ijk})e^{\lambda_3^m(r_{ij}-r_{ik})^m},\\
&g(\theta)=\gamma\left(1+\frac{c^2}{d^2}-\frac{c^2}{d^2+(\cos\theta-\cos\theta_0)^2}\right),
\end{aligned}
$$

where the sums go over all atoms, r_{ij} is the distance between atoms i and j, θ_{ijk} is the bond angle between bonds ij and ik, f_C is a smooth cutoff function, f_R is a repulsive two-body interaction and term $b_{ij}f_A$ is the three-body interaction. The parameters A, B, D, R, β, λ_1, λ_2, λ_3, γ, m, n, c, d and θ_0, are material specific constants. We used the values specified in the Ref. [21] for these parameters.

2.2 Steepest descent and annealing simulations

We wanted to see how good the traditional molecular dynamics methods are at solving this problem. Starting with the steepest descent method, we generated 100000 random initial structures and applied the steepest descent to each of them. The optimization was terminated when a local minimum was reached within the desired energy tolerance; we found out that increasing the tolerance from 10^{-8} eV improved the results only marginally. The energy distribution of the found minimum in these optimization runs is presented in Fig. 2 along with the distribution of optimization steps required to find the minimum. From this figure we see that none of these runs could even come close to finding the global minimum around -148 eV. Indeed, most of the runs stop in the region -135 eV to -130 eV, which is the same region our algorithm easily is stuck into, but more on this later in Section 4.

It is worth noting that if we free the border atoms, the steepest descent method actually becomes noticeably more effective, shifting the minima peak and allowing us to reach lower energies (Fig. 2 dashed line). This must be due to the cell being less rigid, allowing the atoms to move around more freely. However, we chose to keep these atoms frozen, as interface structure calculations do have a more rigid bulk part surrounding the interface. In any future cases, however one should consider keeping the surrounding bulk relatively flexible, in order not to hamper the optimization process.

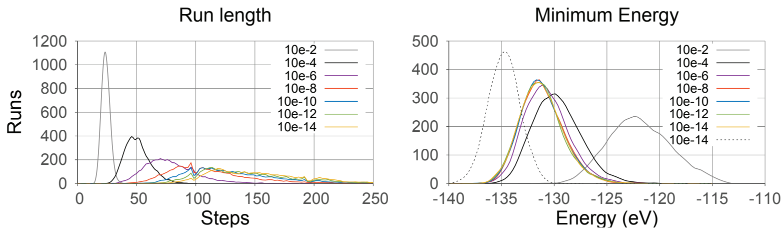


Figure 2: These are the distributions of lowest energy and steps required to achieve that energy from 100000 steepest descent simulations. Different curves correspond to a different energy tolerance used by LAMMPS to end the search. The solid lines correspond to runs with the border atoms fixed in place. The dashed line is from a similar simulation with only one fixed atom. Fixing only one atom made the simulations slightly faster than freeing all of the atoms. The graphs have been smoothed. We see that around the tolerance of 10^{-8} the results start converging.

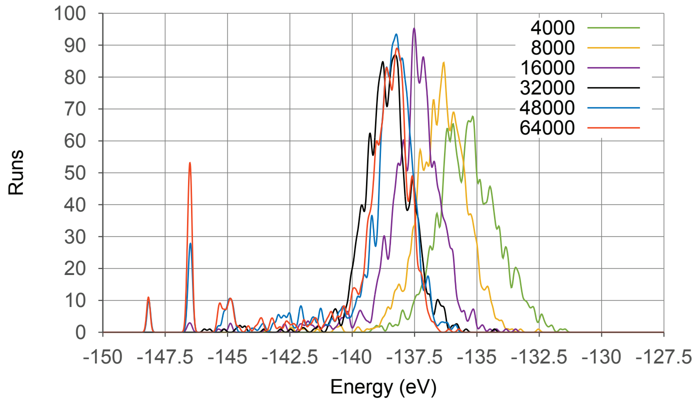


Figure 3: LAMMPS annealing done from random starting positions. At the end of each simulation, we searched the nearest minimum through the steepest descent method. Each line corresponds to a different amount of applied annealing. The graph legend gives the amount of annealing steps done by LAMMPS, which also equals to simulated time in femtoseconds.

After these simulations, we switched to considerably slower annealing simulations [9], where we were able to obtain better results (Fig. 3). In these annealing simulations the system is in a heat bath, which gives the atoms enough kinetic energy to overcome the potential energy barriers. The temperature of the system is slowly reduced to a very low temperature allowing the structure to settle in a local minimum.

In these simulations one step done by LAMMPS corresponds to 1 femtosecond(fs) of simulated time. The calculations now absorb considerably more CPU time, as we set

a single annealing to take 4000, 8000, 16000, 32000 48000 and 64000 steps. In contrast most of the steepest descent calculations took between 50 and 300 steps, with the average being around 100. Because the simulations take that much longer, we only did 1000 simulations/annealing time. In Fig. 3, we can see that most runs end with an energy above -140 eV, even if we increase the simulation time.

We see that even this simple case is hard to solve quickly through traditional methods of annealing and steepest descent. Even with 64000 steps, annealing had a success rate of only 1.1%. If we increase the duration of the simulation by tenfold to 640000 steps, the success rate rises noticeably to 41.9%. This run is not shown in Fig. 3 as it would cause a very tall spike at the global minimum energy and render the graph harder to interpret.

We believe this is a good first test case for our optimizer providing enough challenge and a straightforward way to progress into more challenging SiO_2/Si-interfaces and surface structures in the future.

3 Algorithm: methods and techniques

3.1 Introduction to genetic hybrid algorithms

Genetic hybrid algorithms are a combination of two parts, genetic and local search, that complement each others. Genetic algorithms are known for being population based search algorithms, that try to copy the process of natural evolution through natural selection and genetic dynamics. They were first described by Holland in the 70s [6].

In most cases the genetic algorithms are good at exploring the landscape of the whole parameter space and discovering the regions which possibly have the wanted global minimum [4,18]. They do this by trying to use the information gained from the known good solutions and exploring the parameter space widely. The genetic algorithms however sometimes have trouble actually pin pointing the global minimum after they have found the region it belongs to. This is often caused by the algorithm's inability to make the necessary small changes to the system [19].

Different local search methods can cover for this weakness of genetic algorithms [5, 12]. From a given starting point the local search methods are usually very efficient at exploring the region and finding the nearby local minima. The local search methods use the information available from the surrounding area to find a good nearby minimum and continue from there on. The quality of starting point is critical for the method. If the starting point is not in the funnel of the global minimum, then it is very unlikely that the global minimum will be found.

Genetic hybrid algorithm's power comes from the combination of these two methods, where you take advantage of their individual strengths [5,12]. By using the genetics to explore the vast parameter space and using the different local search methods to explore the interesting areas for the global minimum. In structure optimization the genetic methods include operations like exchanging layers between two structures, using parts of known low energy structures to produce very different possible solution structures

and different kinds of smaller mutations, like just moving atoms around and switching coordinates/layers in a given structure. The most basic local search methods are nearest minimum locating algorithms like steepest descent and different sequential quadratic programming (SQP) algorithms. When these are combined with methods like basin hopping with its various forms and annealing, we get local search methods that are good at exploring the nearby minima neighborhood.

3.2 GHA platform and our solver

We are using the Genetic Hybrid Algorithm (GHA) as a platform for algorithmic development [16]. It is a computational platform for designing special purpose algorithms for difficult numerical problems, extensively tested in economics and engineering. Through GHA we can access different algorithms and non-linear solvers, like nlpqlp, snopt, fsqp, dncong, cplex, kkt_ql and gurobi, which are powerful tools for mixed-integer non-linear programming problems. We have linked LAMMPS to GHA as a library for easy and fast potential energy evaluation. We can also extract forces from LAMMPS for fast gradient evaluation. For the local search, in addition to the algorithms implemented through GHA, LAMMPS offers annealing and steepest descent. Later in Table 3 we present a comparison done between these algorithms and a sequential quadratic programming (SQP) implementation.

Next we introduce our program structure, parallelization and the essential low-level logic (Fig. 4). The searches we have done were multi-core jobs, but there is no search boosting communication between the cores during the search. All cores are prepared independently from the same parameters in a preprocessor-function. From there each core runs the search for a preset length. The exception to this is when we at times choose to stop the search when one of the cores has found the solution, sending out an interrupt signal that will force all the other cores to stop the search. At the end each core will do their core specific post processing involving mostly clean up of the memory and some result processing. The root level I/O is done last, including most importantly information of the search.

More specifically each core performs the global minimum search in a series of runs. The general structure of the search has been illustrated in Fig. 4. Each of these runs starts with the generation of a large pool of random structures. These are then ranked by their energy and only a population of POP members is retained; typically this is between 4 and 64 in our calculations. With this population we will then perform genetic manipulation, in this case arithmetic cross-over and non-uniform mutations, and continue processing them in series of iteration loops.

Each member of the generated population launches into a series of iteration loops. One loop consisting of a series of mixed evaluator() and accelerator() calls. The standard progression is done through box-constrained optimization(BFGS) using evaluator() calls while the accelerator() is responsible for more radical changes, like different kind of mutations to the structure.

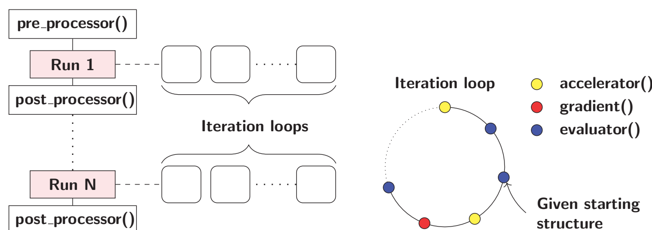


Figure 4: The general structure of the main loops and working of GHA in our case. The main program consists of a series of runs, which launch into multiple iterative loops, one for each population member, that try to solve the problem from different starting positions. The figure is only illustrative as the accelerator()-function is called when it is needed.

After the set amount of iteration cycles only the best structure with lowest energy will continue to the next run cycle; others are discarded after the run. The search ends after the chosen number of completed runs is reached. Since we know the global minimum structure, we also choose to end the search early incase a core reaches the solution.

A very important genetic part of our problem specific accelerator() code are the different implemented mutations that are used to generate new structures from a given structure. The most efficient type of mutations to the structure proved to be the simplest ones where we move one or two atoms simultaneously to a better position. Simultaneous moves of two atoms are especially helpful in breaking the atoms out of strong local minimum potential wells.

Also in the layered 19- and 43-parameter cases it was pretty straight forward to implement a method that crosses the layers from two models with each other. Yet, we can use this type of mutation to nudge the structure from a local minimum to the funnel of another minimum. This is much harder to implement in the 60-parameter case, as the layers are much harder to identify. One of the important reasons for mutations is their ability to produce vastly different structures without starting from scratch so we do not end up exploring only a small portion of the parameter space.

4 Results and comparisons

4.1 Basin hopping

Basin hopping is a well-known algorithm for structure optimization [24]. We implemented a simple form of basin hopping for comparison purposes. To put it shortly, this method tries to jump from a local minimum well to another, eventually hoping to funnel into the global minimum.

We start from a local minimum, chosen by randomly placing the atoms and determining the minimum through steepest descent. Then we choose a nearby structure as

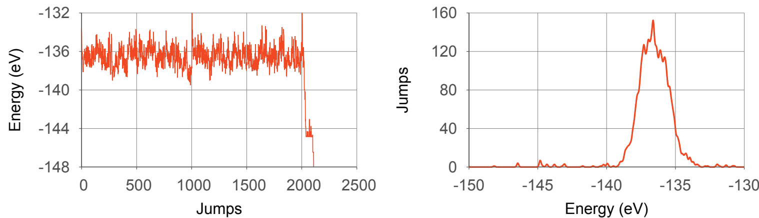


Figure 5: On the left, the energy progression of a typical 2108 jump basin hopping run. The energy spikes shortly at every 1000 jumps because we randomize our current location. We see that after we break the energy barrier of -140 eV we reach the ultimate diamond structure at -148 eV quickly. On the right we present the energy distribution of the terminal jump location, showing where we spend most of the time in the run.

the jump location, coarsely evaluating its energy by assigning it the value of the nearest known local minimum energy. The jump is carried out using the Metropolis algorithm: if the estimated energy is lower than the energy of our current location, we jump. Otherwise the jump is done if a random number is lower than $\exp(-\frac{\Delta E}{kT})$, where ΔE is the energy difference of these two locations, k is Boltzmann's constant and T is a temperature constant affecting the jump chance. If the random number is higher, we try again with a different jump location until a jump is successful. After the jump, we use steepest descent to reach the local minimum and prepare for another jump.

The method has its problems as it too can get trapped into a region of local minima and never reach the funnel of the global minimum. We noticed that in our simple case too, but we were able to avoid the trapping by randomizing our current location every 1000 jumps. The same effect could be achieved by doing a large enough jump – given that the required step length can be set correctly. We did a series of 1000 search runs and were able to find the global minimum consistently, although the time it takes varies a lot. All searches found the solution: on average it would take around 2237 jumps and 38 steps to find the local minimum after the jump. The spread is very high as sometimes we could find the minimum almost right away, in less than 50 jumps, and sometimes it could even take over 10000 jumps. This task as a whole also took a bit over 86 million steps for LAMMPS to complete. Each subsequent jump takes slightly longer on average due to each jump using the known local minima in the jumping process. This list of minima grows as the search progresses leading to an approximately quadratic relation between the search time and the number of jumps made. This could become problematic in a more complex task, requiring more advanced methods for processing and sorting the stored minima. The average 2237 jump search would roughly take around 80 seconds of CPU time. In Fig. 5 we have illustrated the energy distribution of the found minima and the progression of one basin hopping search that took 2018 jumps. We see that most of the jumps land on local minima in the range from -138 eV to -135 eV. Once we get past the apparent barrier around -139 eV the search quickly finds the global minimum.

4.2 Geno-mathematical search

Having implemented the GHA-LAMMPS interface and introduced the problem specific code described above, we conducted a series of independent parallel runs on CrayXC40 at CSC (Helsinki). The parameter cases are shown in Table 1. The simulations with only 1 solution were done using an early mesh interrupt broadcasted when a core found the known optimum solution at -148.17 eV. This was determined by the evaluated potential energy of the structure. Since the number of local solutions close to the global optimum is small and the distance between the global minimum and the nearest local solution is tangible, the advancement to the global solution from a favorable initial point is relatively fast. The nearest local minima are single-atom dislocations at around -146 eV.

Table 1: Results obtained with 1024/4096 parallel cores, where each search found the known global energy optimum $f^*=-1.4817e+02$ eV. The runs with only 1 solution had an interruption signal broadcasted to stop the search once the solution was found. Each core conducts its own search. Each row corresponds to one search done with the given number of cores in parallel. The success rate for the interrupt searches is left out, because only one core finds the solution in each of the cases before the search is forced to stop by the interrupt signal.

n	Cores	Success rate	f^* energy (eV)	Average time (CPU s/core)
(i)		*Success rate in massive search*		
19	1024	98.9%	-1.4817e+02	1920
19	1024	89.2%	-1.4817e+02	1080
19	1024	99.0%	-1.4817e+02	1740
(ii)		*Solution speed with early mesh interrupt*		
19	1024	-	-1.4817e+02	25.2
19	4096	-	-1.4817e+02	12.6
43	1024	-	-1.4817e+02	1740
43	4096	-	-1.4817e+02	995
60	1024	-	-1.4817e+02	43.2
60	4096	-	-1.4817e+02	63.6

The message of Table 1 is that, when early mesh interrupt is activated, the silicon structure optimization problem is solved to optimum ($f^*=-1.4817e+02$) in at most 30 CPU-minutes from an arbitrary starting point using concurrent search with 1024 parallel cores and different parametrizations. With early mesh interrupt and $n=19$ and $n=60$ the processing time is less than 64 CPU-seconds. Studying the sensitivity of CPU-time to mesh size and corroborating the evidence with massively parallel Monte Carlo simulation are left for future research. We also note that the efficiency of our algorithm is critically dependent on the starting point for the local search. Whereas we have applied simple genetic manipulation, more advanced initialization techniques applied in future development efforts, such as variants of e.g. Basin hopping may influence performance significantly.

The $n=43$ search takes significantly longer to complete than the other cases. We believe this must be due to the rigidity problems mentioned earlier in Section 2.2. The constraints on the system cause high-energy barriers that are hard to overcome for the system. We are not sure why this does not affect the $n=19$ case. The significant drop in parameter count may overweight the problems caused by the energy barriers.

We did another series of searches, shown in Table 2, where we fixed the random number seed to core specific values to give us replicable starting structures for the runs. The purpose of these runs is to show how short high population runs compare to longer runs with a smaller population. From these searches we see that using higher population in a short run is much more beneficial than just letting a low population search go on for longer. The results with the interrupt signal can be quite misleading due to how the parallelization behaves in very short runs. On the operating system level different cores start the search at different times. This becomes apparent, when we notice that the similar run with no interrupt signal actually has a core find the solution over twice as fast as with the interrupt signal. This is why we left out the mean time for these runs as it would be heavily influenced by the cores that have not even started yet. If the search time is measured in minutes the differences in core start up aren't significant anymore.

In the long run, the 60-parameter case is slower than the 19-parameter case as expected. The completion time of the cores – the time needed to obtain the global optimum from a unique random starting point – in the long and short 60-parameter run is presented in Fig. 6. The graph suggests that the high population runs are more suitable for short search runs.

Table 2: Runs using the same random variable seed as in Table 3. The purpose of these calculations was to pit long search, with low population count (POP) and many iterations, to short search, with less iterations and high population count. The short search, while having much lower success rate, finds the global minimum faster. The highest reached runtime was nearing 14 hours in the 60-parameter case. The mean time for short runs with interrupt is left out, because it doesn't give useful information due to all searches being stopped when one of the cores finds the solution. Success rate is left out for the same reason, as only one solution is found.

Search Job	n	POP	Cores	Success rate	Time (s) Mean	Time (s) Fastest	Energy (eV) Best	Energy (eV) Mean	Energy (eV) Variance
Short with interrupt	19	64	4096	-	-	18.1	-1.4817e+02	-119.19	88.39
Short with interrupt	60	64	4096	-	-	116	-1.4817e+02	-122.16	20.22
Short no interrupt	60	64	1024	3.71%	95.3	42.6	-1.4817e+02	-135.95	22.72
Short no interrupt	60	64	4096	3.03%	94.4	41.8	-1.4817e+02	-135.86	21.77
Long no interrupt	19	4	1024	97.9%	1940	12.7	-1.4817e+02	-148.06	1.208
Long no interrupt	60	4	1024	99.5%	5090	66.4	-1.4817e+02	-148.13	0.3406

We did three searches with mesh size 1024 where we fixed the seed for the random number generation like in Table 2 and tried to optimize the structures using only one of three methods: in the first search we only used the steepest descent method from LAMMPS, in the second we used the BFGS/SQP-algorithm [1] and in the third we used only LAMMPS annealing. The purpose of these runs was to see how these methods

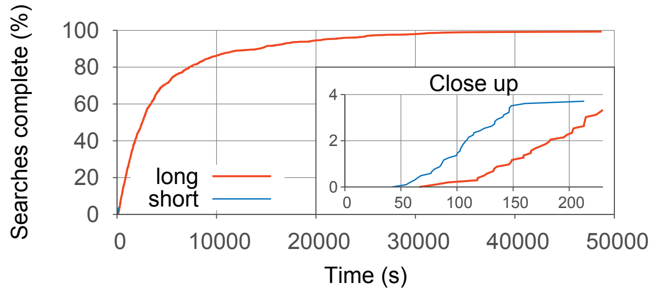


Figure 6: The search completion rate of different cores as a function of time in the case of the long and short 60-parameter run in Table 2. In the close up we see that the short run(blue line) initially has faster solution rate, but the solution rate starts to stagnate towards the end of the run and the long run with a smaller population catches up to it.

perform in isolation compared to our results in Table 2 and each others, given the same starting structures. There was no big difference between the first two searches neither in the obtained energy values nor in the solution time. The results are presented in Table 3.

Table 3: Comparing the steepest descent and BFGS/SQP-algorithms from the same starting position, with mesh size 1024. We also listed a similar simulation that used annealing only.

Search Job	n	Mean time (ms)	Best energy (eV)	Mean energy (eV)	Variance
Steepest descent	60	1.09	-135.945	-129.774	5.74455
BFGS/SQP	60	1.11	-135.397	-129.306	5.39542
LAMMPS annealing	60	11.6	-131.15	-123.974	6.2765

5 Discussion

It is worth noting that often we are not only interested in the global minimum but also in the local minima nearby as they may be more stable under different circumstances. Some of them can also be useful when comparing simulations to data obtained from experiments, as crystals are not perfect in reality. They can sometimes also give useful information about different reaction paths, especially when studying a system with defects, complex interfaces or surfaces. This kind of information can be stored during the simulation and analyzed afterwards. Through replicable core specific random numbers, we can study the solution process of any interesting crystal structure in detail on a single core by specifying the output level of GHA correspondingly.

There is still the issue of identifying the interesting minima, as most of the found local solutions are uninteresting amorphic structures. There are measures for differenti-

ating structure, like the fingerprint function used in USPEX and geometric measures like neighbor counts and the radial distribution function [13].

As we found out with the LAMMPS simulations, fixing some of the atoms actually does not help at all; instead, it hinders the optimization by making some energy barriers harder to bypass. Further testing is required, but it would seem that fixing atoms in place, even if those spots are known correct locations, is not an efficient choice. Restricting their movement to a small area surrounding this location might allow the structure to flex around the energy barriers.

We have probed different methods for generating the initial raw pool of structures. As a word of caution, it is easy to make a structure generation algorithm that leaves out a big portion of the possible structures, but it is harder to control what is left out. We have to be careful not to rule out some possibly successful structures. At the same time, some methods produce mostly unwanted high-energy structures or, for example identical inferior structures. There is also the question of how much we want to refine the initial structures. When two atoms are close to each other, the associated two-body term gives an exponentially growing positive contribution to the total potential energy of the structure. This makes energy-based ranking of the structures slightly problematic, so we have either discarded or modified these converging structures to get a proper evaluation of their energy.

One way to circumvent energy spiking is to design an initial structure having every atom close to the typical silicon bond distance from its neighbor. The procedure is not too time consuming and ensures we are not wasting potentially good structures, since atom pair proximity related energy spikes are avoided. Another way is to develop the structures with unusually high energies a bit, either through applying some steepest descent type of optimization or identifying and moving the problematic atoms.

We tried a method of inserting atoms into the cell one by one. However, we quickly noticed that this was not an effective strategy, as it leads to the atoms clustering in some areas of the cell. We believe this is because the silicon atoms want to form quadruple bonds, which is not possible in their natural location when the simulation area is still half-empty in the beginning of the optimization. As we start inserting the atoms, the only place where they can form these bonds to achieve lower energy is in the region where there already are atoms. This leads to unwanted clustering in some region of the cell. Perhaps this could be circumvented by restricting the placement of an atom to areas far from the previously placed atoms.

Another issue is how good the initial structures should be. With the future in mind, creating an algorithm that generates as good initial structures as possible is desirable. However, we did not want to have too good structures right now, as one of the points of this test case was to assess the performance of the GHA optimizer and develop methods to advance through complex energy landscapes to the global minimum.

Even after we implemented more structure optimization specific techniques the major energy obstacle was between -140 eV and -130 eV. Most of the calculations would end up there easily, but getting forward proved to be challenging. We tried analyzing

the structures around the problematic energy area above −140 eV, but the local minima found there did not really resemble the desired diamond structure. It is a known problem in optimization, that the structures have a tendency to turn out fairly disordered and amorphic. That was the case here too. We tried developing some metric to differentiate these structures and find the ones that could potentially lead to the global minimum but we did not find such metric.

If a structure could break the barrier at 140 eV, the optimizer often ended up progressing to the perfect diamond structure around −148.17 eV. We believe that the small number of minima between −140 eV and the global minimum of −148.17 eV and shallowness of the minima in this region make it easy for the system to jump to the funnel of the global minimum. An eyeball test indicates a clear ordered structure with only one or two atomic dislocations and rest of the diamond structure intact around this energy region. That means that most of the minima between −148 eV and −140 eV correspond to structures, which can be corrected with just a few well-done atom displacements. Our calculations confirm this, as passing the barrier around −140 eV usually leads to a rapid advancement to the global minimum.

In the annealing Fig. 3, we see that the graph becomes spiky around this region from −148 eV to −140 eV, with small gaps between the spikes. The spikes are not very sharp though, which we believe is because of the frozen atoms in the cell. Most of the spikes correspond to some 1-2 atom dislocations. For example, ideally, the single atom dislocations should correspond to a unique structure, but that is not the case here. When we introduce a dislocation defect, its location affects the energy of the resulting structure. If the dislocation is near the corner or the edge of the cell, the environment reacts differently as some atoms are unable to move. We tested this by doing simple dislocation defects to the system and noticed a 0.5 eV difference depending on where the dislocation was. With an analogous two-atom dislocation, we noticed differences up to 2 eV.

It is clear that the solution for the type of task considered in this study is very dependent on the initial structure. As we saw in our long searches, a single core could take anywhere from ten seconds to fourteen hours to reach the known global minimum. At the same time, concurrent processing of our algorithm with a large mesh from core specific random numbers with early mesh interrupt activated guarantees – beyond reasonable doubt – the global solution in seconds, when modelling the structure as a full 60-parameter or restricted 19-parameter problem. This is indeed an encouraging and exceptional result in a difficult global optimization problem.

6 Conclusions

We have successfully used the optimizing platform GHA in our case of silicon structure optimization and presented the results. GHA has inbuilt support for the necessary large scale computing and has the required versatility to solve the case, producing results that are comparable to other methods like basin hopping and annealing. With the future

in mind, we hope to refine and extend this approach to more relevant material physics problems related to oxides, interfaces and defects.

In order to solve more complicated atomic structures than the one probed in this study, we would have to be able to reduce the processing time significantly and guarantee the optimal solution using only a single processor. This would allow, e.g. massively parallel search of interesting crystal candidates using modern Geno-mathematical techniques, with potential for new discoveries in atomic interface and surface structures.

At the same time, there is a lot of room for improvement of the search algorithm. Incorporating the basin hopping and other popular methods is one way we could try to achieve this. Another possibility is to use elaborate statistical distributions for the starting positions. In the searches presented in this paper, we used the uniform distribution, but distributions generated through, e.g. Copula-theory [20] might prove to be better suited for this case.

Acknowledgments

The Magnus Ehrnrooth Foundation is acknowledged for financial support (A. L.). The computer resources of the Finnish IT Center for Science (CSC) and the FGI project (Finland) are acknowledged. Advice of the experts at CSC on installing LAMMPS on Cray XC40 is gratefully acknowledged.

References

[1] X. Chen. Convergence of the bfgs method for lc1 convex constrained optimization. *J. Control Optim.*, 34:20512063, 1996.

[2] R. C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. *Proceedings of the Sixth International Symposium on Micromachine and Human Science, Nagoya, Japan*, pages 39–43, 1995.

[3] R. C. Eberhart and J. Kennedy. Particle swarm optimization. *Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ*, pages 1942–1948, 1995.

[4] D. E. Goldberg. Genetic algorithms in search, optimization, and machine learnin. *Addison-Wesley*, 1989.

[5] W. E. Hart. Adaptive global optimization with local search. *Doctoral Dissertation. San Diego: University of California*, 1994.

[6] J. Holland. Adaptation in natural and artificial systems. *The University of Michigan*, 1975.

[7] C. Hu, H. Bai, X. He, B. Zhang, N. Nie, X. Wang, and Y. Ren. Crystal md: The massively parallel molecular dynamics software for metal with bcc structure. *Computer Physics Communications*, 211(Supplement C):73–78, 2017. High Performance Computing for Advanced Modeling and Simulation of Materials.

[8] C. Hu, X. Wang, J. Li, X. He, S. Li, Y. Feng, S. Yang, and H. Bai. Kernel optimization for short-range molecular dynamics. *Computer Physics Communications*, 211(Supplement C):31–40, 2017. High Performance Computing for Advanced Modeling and Simulation of Materials.

[9] S. Kirkpatrick, G. J. C. D., and M. P. Vecchi. Optimization by simulated annealing. *J Phys Cond Matter*, 220:671680, 1983.

[10] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47(1):558–561, 1993.

[11] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[12] F. G. Lobo and D. E. Goldberg. Decision making in a hybrid genetic algorithm. *IEEE International Conference on evolutionary Computation, USA: IEEE Press*, pages 122–125, 1997.

[13] A. O. Lyakhov, A. R. Oganov, and M. Valle. How to predict very large and complex crystal structures. *Computer Physics Communications*, 181:1623–1632, 2010.

[14] C. M. Mangiardi and R. Meyer. A hybrid algorithm for parallel molecular dynamics simulations. *Computer Physics Communications*, 219(Supplement C):196–208, 2017.

[15] A. Oganov and C. Glas. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of Chemical Physics*, 124:244704, 2006.

[16] R. Östermark. A multipurpose parallel genetic hybrid algorithm for non-linear non-convex programming problems. *European Journal of Operational Research*, 152(1):195–214, 2004.

[17] S. Plimpton. Bayesian method for global optimization. *J. Comp. Phys.*, 117:1–19, 1995.

[18] P. Preux and E.-G. Talbi. Towards hybrid evolutionary algorithms. *International Transactions in Operational Research*, 6:557–570, 1999.

[19] C. Reeves. Genetic algorithms and neighbourhood search. *Evolutionary Computing, AISB Workshop*, 865, 1975.

[20] A. Sklar. Fonctions de repartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

[21] J. Tersoff. New empirical approach for the structure and energy of covalent systems. *Physical Review B*, 37(12):6991–6999, 1988.

[22] A. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. Reaxff: A reactive force field for hydrocarbons. *Journal of Physical Chemistry*, A105:9396–9409, 2001.

[23] P. K. Venkatesh, M. H. Cohen, R. W. Carr, and A. M. Dean. Bayesian method for global optimization. *Phys. Rev. E*, 55(5):6219–6232, 1997.

[24] D. J. Wales and J. P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *J. Phys. Chem.*, A 101:5111, 1997.

[25] Y. Wang, J. Lv, L. Zhu, and Y. Ma. Calypso: A method for crystal structure prediction. *Comput. Phys. Commun.*, 183:2063, 2012.

[26] B. Wu, S. Li, Y. Zhang, and N. Nie. Hybrid-optimization strategy for the communication of large-scale kinetic monte carlo simulation. *Computer Physics Communications*, 211(Supplement C):113–123, 2017. High Performance Computing for Advanced Modeling and Simulation of Materials.

[27] T. WW and H. RG. A grand canonical genetic algorithm for the prediction of multicomponent phase diagrams and testing empirical potentials. *J Phys Cond Matter*, 25:495401, 2013.

Antti Lahti & Ralf Östermark & Kalevi Kokko
Optimization of SiO2 with GHA and basin hopping

Computational Materials Science, 0927-0256, 2021, 10

IV

ARTICLE IN PRESS

Computational Materials Science xxx (xxxx) xxx

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Optimization of SiO_2 with GHA and basin hopping

Antti Lahti [a,b,*], Ralf Östermark [c], Kalevi Kokko [a,b]

[a] *Department of Physics and Astronomy, University of Turku, FI 20014, Turku, Finland*
[b] *Turku University Centre for Materials and Surfaces (MatSurf), University of Turku, FI 20014, Turku, Finland*
[c] *School of Business and Economics, Åbo Akademi University, FI 20014, Turku, Finland*

ARTICLE INFO

Keywords:
Optimization
Bulk SiO_2
Semi empirical potential
Basin hopping

ABSTRACT

In this paper we continue to develop our structural optimization algorithm built earlier on a numerical platform, the Genetic Hybrid Algorithm (GHA). Our goal now is to extend our algorithm to oxides, find an effective way to search for the known global minimum, alpha-quartz as a test case, and report our results and findings for this system. We studied unit cells of different sizes: 18, 36 and 72 atoms, but most of the presented results are for cases with 18 and 36 atoms. The algorithm makes heavy use of the basin hopping method for searching for the global minimum of the system. We show how we were able to apply basin hopping most effectively in this case and which variables were of importance.

We identify three other low energy structures near the global minimum structure, that trap the search. We show that the energy guided basin hopping can be detrimental to the search and structure-based guiding works more reliably. Two different structure based guides were used, one that tries to maximize the shortest silicon–silicon bond in the cell, while the other tries to maximize the calculated order parameter. The guiding was implemented by generating multiple different options for the basin hopping jumps, and doing weighted choosing on those options based on their properties.

1. Introduction

Atomic structures pose a very difficult optimization problem that is still relevant especially when looking at complex interfaces, surfaces and compound materials [1,2]. The difficulty comes from the large amount of local minima that hinder any global minimum search. Only a very small range of structures will converge to the global minimum through traditional gradient descent optimization. Stumbling into the global minimum by scattering atoms around the unit cell randomly is extremely unlikely. The local minima obtained this way do not give us easily usable information about the global minimum either. This means that many local relaxations are required with most conventional methods if we want to find the global minimum [1,3,4]. Especially different minima and basin hopping algorithms have proved successful with clusters and molecules, but also solids as well [3,5,6].

Our motivation in developing this algorithm is to get a tool for creating interface structures for oxide thin films on semiconductors. These interface structures are difficult to construct and are computationally demanding as the required unit cells quickly become large depending on the reconstruction at the interface and how well the oxide and semiconductor lattice vectors match. This is why we believe it would be advantageous to get a tool that targets this problem specifically. An effective tool would allow us to rely less on our skill, intuition and pure guessing when trying to discover the low energy interfaces.

One of the crucial parts of the algorithm for these thin film interfaces will be optimizing the oxide film, which is why we are focusing on the oxide optimization in this paper before attempting to optimize the actual entire interface system. When the oxide film is thin, it is reasonable to assume that the oxide film retains the dimensions of the semiconductor base. This is why we have chosen to not include the volume dimensions to the optimization, as it would increase the computational cost while not being required for our end goal. In the future, if we want the algorithm to have a more general application, the dimensions would have to be included in the optimization.

Because the used potential is based on classical physics, it is also import to check the results with *ab initio* calculations afterwards. This is why it is important to find other low energy structures near global minimum, as the order of the structures might change when the energies are recalculated.

The true global minimum, also known as ground state in material physics, is important as it is the default configuration for the material at 0 K, but other low energy structures are interesting too as they might be present at other conditions or due to the way the material was formed.

* Corresponding author at: Department of Physics and Astronomy, University of Turku, FI 20014, Turku, Finland.
E-mail addresses: ailaht@utu.fi (A. Lahti), Ralf.Ostermark@abo.fi (R. Östermark), kokko@utu.fi (K. Kokko).

https://doi.org/10.1016/j.commatsci.2021.111011
Received 6 June 2021; Received in revised form 25 October 2021; Accepted 27 October 2021
0927-0256/© 2021 Published by Elsevier B.V.

Please cite this article as: Antti Lahti, *Computational Materials Science*, https://doi.org/10.1016/j.commatsci.2021.111011

ARTICLE IN PRESS

Historically the computational search for the global minimum structures really kicked off in the 80s, when the computational power was enough to search through large quantities of structures of difficult materials, like silicates for example [7–9]. The search was often done using case specific cost functions, that would be designed with the wanted bond structures in mind using bond valence model and Coulomb interaction [10]. These kind of special potentials and cost functions are still useful today, as doing quantum mechanical simulations is several magnitudes slower in comparison. Combining this kind of fast evaluation of structures with genetic algorithms later on also proved beneficial [10].

A different geometrical approach can also be effective, if enough information of the system is known. For example Foster et al. [11] identified 152 distinctly different crystalline structures for SiO_2 by exploring geometrically the possible structures where each node has 4 connections, similar to how many bonds silicon atoms have in many SiO_2 structures. Then, by inserting silicon atoms to the network sites and bridging the sites with oxygen atoms, they were able to find the global minimum α-quartz [12,13] and many other ordered SiO_2 structures. This approach of course heavily limits the structures that can be found, as there is a built in constraint in the algorithm.

From the perspective of potential energy landscape theory, the energy surface is seen to contain basins of attraction, that lead to different local minima structures [14,15]. In this article we define the distance of a structure from another structure as being proportional to the number of optimization steps required to reach the minimum. We use this to express if a structure is closer to the global minimum than the false attractors in the system from the optimization point of view. These false attractors are stable local minima, that have a low energy close to the global minimum. They are problematic as the search often gets trapped into their vicinity, requiring us to either change the structure radically or start the search again from another structure.

1.1. Our previous case and eventual goal

In this paper we tackle this atomic structure optimization problem through modern well-tested numerical platform: the Genetic Hybrid Algorithm (GHA) [16]. It is a computational platform for designing special purpose algorithms for difficult numerical problems, extensively tested in economics and engineering. It was developed by Ralf Östermark from the School of Business and Economics at Åbo Akademi University. In geno-mathematical programming artificial intelligence is connected to mathematical programming methodology on parallel supercomputers. The approach provides a powerful basis for coping with difficult irregular optimization problems and solving them concurrently. In GHA the genetic population solutions are also subjected to heavy local alterations in frequent passes within the algorithm. We have linked LAMMPS [17] to GHA as a library for easy and fast potential energy evaluation. We can also extract forces from LAMMPS for fast gradient evaluation. For the local search, in addition to the algorithms implemented through GHA, LAMMPS offers annealing and Polak–Ribiere version of the conjugate gradient descent.

We are interested in seeing how far our approach with this platform can take us in the structure optimization problem. We are also looking to extend our knowledge on the problematic oxide structures. Previously we published an article on the optimization of the atomic structure of bulk silicon [18]. This paper continues the work expanding to the system of two component silicon oxide. The more complex compound leads to a case that has a more problematic energy landscape; it is harder to navigate to the global minimum. We chose silicon oxide as it is a logical next step from silicon to a more complex system while also providing a bridge to the interface structures between silicon and thin silicon dioxide films, that are the next goal of our research after this topic.

Silicon oxide is more problematic than pure silicon due to the oxygen atoms making the structure more flexible. Especially in the low energy structures the oxygen atoms form bridges between silicon atoms, which can twist and rotate creating a larger variety of structures. This leads to a high amount of disordered and amorphous low energy structures, which makes it hard to find the more ordered global minimum structure. These structures are problematic, as they are hard to differentiate and it is hard to avoid them and find the real global minimum in the midst of them. Geometrically many of these structures near the global minimum in energy are very different. Additionally most of them also lack the same degree of order present in the global minimum structure.

1.2. The studied SiO_2 system

We used the Tersoff potential [19] for calculating the potential energy of the system with the parametrization found in the paper [20]. The Tersoff potential is a many body bond-order potential originally designed for silicon. The potential allows us to evaluate the energy through fast computation while still being fairly accurate when it comes to replicating bonding in silicon dioxide. This is a quality that was beneficial for us as the study required a lot of iterative testing of sometimes very small changes in the algorithm. We used the timestep of 1 fs in all of our experiments. The used cutoff distances are defined in the used potential ranging from 2.8 Å for Si–Si–Si interaction to 2.0 Å for O–O–O interaction [20].

For SiO_2 in the case of our fixed volume, the known ground state structure has an unit cell of 18 atoms, which consists of silicon atoms surrounded by roughly a tetrahedron of oxygen (Fig. 1). This way each silicon has 4 oxygen atom neighbors and each oxygen has 2 silicon neighbors. We have chosen to omit the box dimensions from the optimization, because our goal for the algorithm are the interface systems with oxide thin films on semiconductors, where the dimensions of the semiconductor substrate determine the lattice constant of the oxide too. To make the algorithm a general structural optimization tool the dimensions of the box should be taken as variables too.

Our goal was to make an optimization algorithm that could find the ground state structure using GHA as a platform. We looked at different sized super cells of 18, 36 and 72 atoms. Even a smaller 9 atom unit cell exists, but this case converges to the solution too easily to be used for improving the algorithm. Larger cells increase in complexity and computation time, which makes studying them undesirable for now as even the 72 atom case proved to be challenging.

Working with these three different sizes helps us in making the resulting algorithm more general: in the beginning we focused on getting the 18 atom case working well, but noticed we had to implement new methods and modify the existing ones to get results in the other cases with more atoms. For example basin hopping behaved differently when we increase the size of the super cell: converging takes longer but the initial energies are slightly better, the average energies are more tightly grouped and there is more variety in structures near the global minimum.

2. Algorithm

We use Ralf Östermark's GHA as the platform for out optimization [16]. The main benefits are easy parallelization, processing many structures at the same time and access to the algorithms and solvers within the platform.

The biggest building block of the algorithm is a so called run (Fig. 2). Each run starts with a preprocessor(), that generates the starting population of structures and initializes the program. The population consists of mostly newly generated structures, but we often include few structures from the previous run if the structures still show room for improvement. A structure that had seen improvement in energy within the last 20 jumps would be included into the new run. Depending on the type of search, the population size would be typically between 4 and 64. These structures are then iteratively worked on by the algorithm,

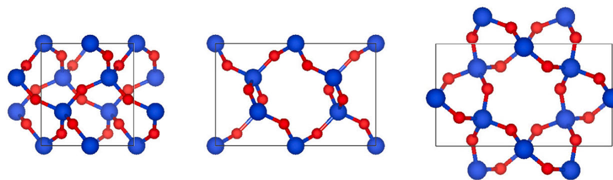


Fig. 1. A rendition of the smallest studied cell with 18 atoms from 3 cartesian angles with the black lines marking the edges of the cell and bonds being visualized through blue and red bridges between the atoms. This is the known global minimum of the case. The minimums for the subsequent 36/72 atom cases are 2x1x1 and 2x2x1 supercells of this cell, with the latter being fairly close to a cube when it comes to proportions.

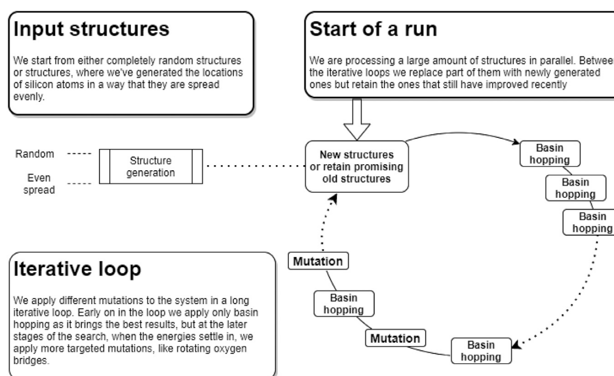


Fig. 2. A general rendition of a single run within the algorithm. We process many structures in parallel in iterative loops. Within these loops we apply different operations and mutations to the structures at hand. At the end of each loop, we discard most structures and generate new ones to replace them. With this population we then start a new run.

and periodically exposed to basin hopping and other mutations. We apply mutations only at the later stages of the search, because early on basin hopping is more effective. We terminate these runs periodically and start anew, because they often get trapped in different low energy regions that our algorithm cannot escape well from.

2.1. Basin hopping

Basin hopping is a structure exploration method we make extensive use in our studies in this algorithm. It is a general global minimization technique presented by Wales Doye [21], that aims to explore the potential energy surface by repeated random perturbations to known minima. After the perturbation a local optimization is done and the new minima is accepted based on the selected criterion, for example if the energy is lower than the energy of the minima we started the perturbation from. This is a way to avoid the energy barriers within the system and iteratively explore and search for structures that minimize the criterion value [21].

2.1.1. The effect of the jump length on the search

The jumping is done completely randomly. We use the optimum jump distance we found out to be roughly $d = 0.46$ Å. In a jump each of the i atoms is slightly moved by a random vector (r_{i1}, r_{i2}, r_{i3}), where the variables r are random numbers between $-d$ and d. The solver is sensitive to changes in the jump distance and it depends on the cell size and shape. This was easiest to test with the 18 atom case, which can still be solved with the pure basin hopping pretty reliably. In Fig. 3 we see the effect of jump distance to the solution rate. Solution rate here is the success rate of us finding the global minimum with the given

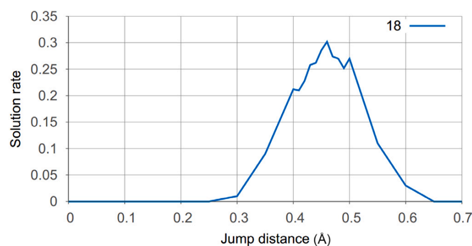


Fig. 3. The system is sensitive to the jump distance. Here we have plotted the solution rate of 18 atom case with pure basin hopping with different jump distances. We had a set of 10000 structures and the solution rate is the rate at which the basin hopping algorithm managed to find the global minimum from these 10000 initial structures. The sharp peak makes it apparent that using the correct jump distance is very important. The bigger cells exhibited similar behavior.

jump distance. There is a peak in solution rate around 0.46 Å, which drops fast if we go much higher or lower. The number was found by looking at two numbers, the jump distance that on average produces the lowest energy structures and the jump distance which has the best chance of finding the global minimum. These numbers were very close to each others (±0.02 Å), which is not much compared to the inherit randomness in the jumping itself.

It is especially interesting to see the sharp drop happen after the peak too. The system is effectively jumping at random, but this still has

some drift towards the global minimum. Increasing the jump distance too much seems to break this process, as we break the existing structure too much.

We used the same jump distance for both silicon and oxygen. We experimented with different jump distances for the atom types and found to our surprise that the optimal jump distances were very close to each others, with silicon preferring slightly higher numbers and being more sensitive to the jump distance: wrong length for silicon jumps had a much higher impact on the range of energies found than the length of oxygen jumps. This seems understandable, as silicon atoms have more mass, which probably means moving them around determines the change in the structure more than moving oxygen atoms that slip into new places more easily. The effect is easiest to see when we choose to not move Si/O atoms during jumps. When silicon atoms are not moved in the jumps the oxygen atoms require larger jumps than before and the search barely ever manages to find the global minimum. Vice versa when we do not move oxygen atoms in the jumps at all, the optimal silicon jump length is slightly raised but the success rate of the search is only marginally different. All the atoms were allowed to move in the gradient descent relaxation following each jump.

Increasing the cell size from 18 to 36 and 72 atoms did not seem to have impact on the optimal jump distance. The only effect was that the range of viable jump distances became slightly wider. By this we mean that the peak in Fig. 3 is wider: it is possible to find the global minimum using both slightly smaller and larger jump distances.

2.2. Initial structures

We generate structures through two different methods. The first one is a completely random placement of atoms in the unit cell. The problem with this method is, that the generated structure is almost always unphysical: the atoms are usually clustered in many places around the cell and the bonding of atoms is not realistic with many atoms having too many or too few bonds of incorrect lengths and types. The only benefit is that it guarantees an unbiased starting locations and it is very unlikely we repeat a structure.

Our second scheme for generating the initial structures is more physical and should still be usable for most other materials too. The goal of this algorithm is to still place the atoms around the unit cell randomly, but with an additional constraint that they are evenly spread. We want there to be some inherent randomness still so that we can use the algorithm to generate many different starting points for our optimization, but we also want them to be more physically realistic, which will save us time in the optimization and hopefully direct the optimization towards more promising direction. We do this by first generating a set of random positions into the unit cell, but instead of stopping at the atom count, we generate way more than we need. We then start discarding positions one by one, determined by the neighborhood of the position. If a position is in a dense cluster with other generated points close to it, we discard it and re-evaluate the remaining positions again. We repeat this until we are back down to our desired atom count, which in our case would be 18/36/72 depending on the case. The amount of initially generated positions affects the variety of the structure we get from this algorithm, the less positions we generate the more random the resulting structure will be and less evenly spread the points will be. On the other hand, if we generate a massive amount of positions initially, all the structures we get will be very similar to each others as the spread of the atoms is close to optimal.

2.3. Constraints and trying to promote crystal formation

During the optimization none of the atoms have any restrictions on their location. In the previous silicon case we noticed that trying to keep any part of the structure static always led to worse results, which we also noticed in this case too through testing. We believe that the static atoms introduce inflexibility to the system, which closes some relaxation paths hindering the optimization process significantly. As the result, we would notice higher average energies during the search and a smaller chance for the search to find the global minimum.

With the bulk silicon we kept a layer frozen in place but with silicon oxide we tried freezing promising SiO clusters and noticed quickly that this was not useful. Even if we cheated and used the information of the known correct global minimum structure to freeze some clusters into correct geometry, it was never beneficial to do so. Giving the atoms small wiggle room could be one possibility to the inflexibility, but we believe this still would not solve the problem completely. Also we would still have to face the problem of identifying the correct parts of the cell to freeze. The main motivation behind this kind of attempt to fix the position of atoms in a cluster was to promote crystal growth towards the global minimum structure. However, it turned out the flexibility of the structure is more beneficial for the search than forcing a cluster or a layer of atoms on their correct positions.

3. Energy and smallest silicon–silicon distance as guides

During the testing we noticed that energy is not a very helpful guide in this problem. Instead we looked for a better, more structure related, quantity to be used as a vague guide during our search. One good shared quality between the low energy structures is that they exhibit almost always no silicon–silicon bonds. Also in the global minimum structure the silicon atoms are distributed in a way that they are nearly as far away from each others as possible. Of course this would not work if our cell size was also a variable instead of being fixed.

This inspired us to calculate the shortest silicon–silicon distance (SSSD) for each structure and use it as a soft guide in parts of our search: for example when deciding to backtrack to some already visited structure or choose a new structure from several generated options. Using the average silicon–silicon distance did not perform as well in our tests as the shortest distance. Similarly, using oxygen–oxygen or oxygen–silicon distances did not yield any good results.

3.0.1. The gap between 2.55–2.75 Å in silicon–silicon distances

Interestingly, if you gather a lot of structures along the search and plot the SSSD in relation to energy, you get left side of Fig. 6. Most of these runs were successful at finding the global minimum so there is a high concentration of structures around it in the top left corner of the figure. More interestingly, there is a noticeable gap in the SSSD distances.

With this gap SSSD does give us a nice divider on if we have a structure without any silicon–silicon bonds. In our algorithms we determined there to be a bond with two silicon atoms, if their distance was less than 2.65 Å. So in Fig. 6 the lower cluster represents structures with one or more silicon bonds.

We get the right side of Fig. 6 if we split these structures with silicon bonds based on how many bonds there are. On the right side figure starting from the bottom, we have 0 bonds in the first big cluster, 1 Si–Si bond in the second cluster, 2 Si–Si bonds in the third cluster etc. From this figure we see that in general, the lower energies correlate fairly well with the structures with few silicon bonds. Curiously the clusters seem to form a staircase type of formation.

Going through our data we inspected some of these structures with silicon–silicon bonds of lengths between 2.6 Å and 2.8 Å, and found that they manifested roughly in two categories shown in Fig. 4. First is a bond between 4 oxygen bonded and 3 oxygen bonded silicons. Second one is not a bond but a small cluster caused by a tri-coordinated oxygen atom, that is surrounded by a triangle of silicons atoms. Two of the silicon atoms surrounding this oxygen then have this rarer inter-atomic distance. This explains why the structures inside the gap are so rare: the tri-coordinated oxygen clusters are not very stable and break easily. On the other hand the first case is altered easily by either the silicons forming a closer bond, which is encouraged by the potential,

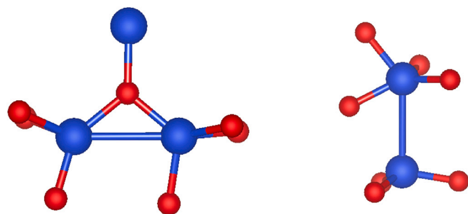


Fig. 4. Rare unstable silicon distances (2.55–2.75 Å):
On the left: a tri-coordinated oxygen bringing silicons atoms closer
On the right: a possible bond 2.7 Å between with two silicon atoms with 3 and 4 oxygens.

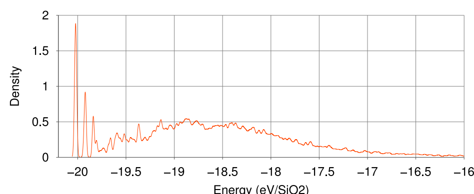


Fig. 5. The energy distributions of found minima in the 18 atom cell case. The distribution is from short searches that do managed to find the global minimum. The distribution of found minima is nearly continuous until we get close to the global minimum, roughly −19.80 eV/SiO_2 with global minimum being the peak at −20.21 eV/SiO_2. The following peaks are the global minimum rotated 90 degrees, the β-quartz structure and an unnamed ordered structure.

or broken by an oxygen bridge between the silicons, which separates them.

The energy of these structures varies quite a lot: typically they are not low energy structures, ranging from −10 eV/SiO_2 to −17 eV/SiO_2, but there are exceptions near −19 eV/SiO_2 with global minimum being at −20 eV/SiO_2.

3.0.2. *Going forward with SSSD as a guide*

We tried to use this Si–Si-bonding derived criterion to our advantage and see if it could be used to guide the search. We want to favor structures with no Si–Si-bonds, where the silicons are as far apart from each others as possible. An additional guide for the search would be useful as the energy itself is not a good indicator of the closeness of given structure to the global minimum. This is because even if we are geometrically close to the global minimum, we will still find a lot of high energy structures (Fig. 7) and we can find low energy structures far away (structurally) from the global minimum.

This leads to energy guide alone most of the time doing more harm than good. While it does often quickly converge to some local minimum, it also traps the search very easily to a minima neighborhood. Comparably random Monte Carlo jumping accomplishes the same in a slightly larger amount jumps, while having less risk of getting trapped. Accompanied with smart mutations, back pedaling on the search path and frequent restarts the results are much better than trying to reinforce an energy guide. We were pleasantly surprised that SSSD guiding during the hopping actually was beneficial in many instances. While the problems of energy guidance are still present with the SSSD guide too, the results seems to be noticeably better as we show later.

3.0.3. *On the vicinity of the global minimum*

It is hard to quantify when a structure is close to the global minimum structurally. We have tried this by picking some of these low energy structures and observing them more closely. Just based on visual inspection most of them do not look ordered, but they do have somewhat evenly distributed silicon atoms connected by oxygen bridges. We also tried a more rigorous method of exploring the nearby minima through different mutations and jumps. If the structure was close to the global minimum you would expect us to find it this way eventually, but this is not the case for most of the low energy structures. This has lead us to the conclusion that there are multiple low energy minima funnels that attract the search away from true global minimum.

3.1. *Short look into the energy landscape*

Previously [18] with the pure silicon case we noticed that optimization task in a 32 atom case often got stuck around −4.3 eV/Si, that is roughly 0.3 eV/Si away from the global minimum −4.64 eV/Si structure. There were few structures between this area and the global minimum, but reaching those structures always meant we were on our way converging to the global minimum. With this new case of SiO_2 it became interesting to ask if and how this behavior changes.

3.1.1. *Setup*

We tried to map the nature of the structure of the energy landscape in two ways. First we made small changes to the known ground level structure by moving all the atoms slightly. We increased the size of these changes and compared the resulting distribution of structural energies. This gave us an idea of the nature of the local minima near the global minimum.

Secondly we did basin hopping for a set amount of 1000 jumps and looked at the distribution of the found minima to get a broader view of the case. We chose these two cases because we have the luxury of approaching this problem from the beginning and the end. By doing perpetuations we get a glimpse on the structures near the global minimum from the energy point of view. By doing basin hopping from a random starting structure we see how the energy varies when we are not near the global minimum. It is important to have understanding on both of these cases, as the search should be approached differently depending on the stage of the search. In the beginning it is easy to get to lower energies by randomly disturbing the current structure, but this becomes less viable as we get to lower energy structures and more specific mutations to the known problems in the structures become more efficient.

We were also interested in seeing how the change in cell size affects this. We expected that the energies would draw closer to the global minimum as the cell size was increased, since the amount of structures with low energies should grow larger.

3.1.2. *Results: the structures near the global minimum*

The longer searches had a high chance of success in the 18 atom case. The distribution of visited energies is presented in Fig. 5. We see that the energy landscape looks very different this time, as the uncertain region is much wider. The funnel to the global minimum seems to be much harder to find this time, as the search spends more time much further away from the global minimum energywise, or at least it is not apparent when looking at the energy of the structures only. It is also shadowed by other low energy minima that were not present in the previous silicon case [18].

If we look at the Fig. 5, we can separate 4 peaks at the lower end of the energy scale. First one is the global minimum α-quartz. Second one is α-quartz, but rotated 90° so the 4.914 Å and 5.406 Å lattice vectors switch places and the structure is deformed slightly. Third one is another known ordered stable structure, β-quartz, which is stable at higher temperatures [22]. Fourth one is slightly more interesting, as it shows some similar features to the α-quartz, but there are few key differences in bond geometry. We tried restarting the stimulation from the latter three of these structures a thousand times and running the optimization algorithm for a lengthy time, but never managed to find the global minimum. We took this to mean that they are indeed deep false attractors that really hinder our attempts to find the global minimum, as the algorithm gets easily trapped in their neighborhood. In

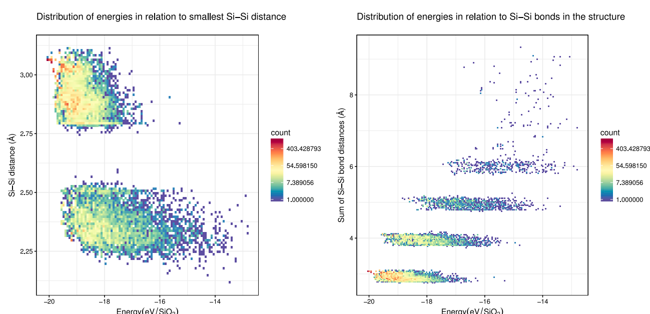


Fig. 6. Example of energies found in short basin hopping runs in the 18 atom case. The visited minima are represented in this heatmap. On the left we see that by plotting the energy against the smallest Si–Si distance, we get two clusters of structures: one with Si–Si bonds (bottom) and one without (top). On the right we have done same kind of classification, but further divided the block with Si–Si bonds to structures with 1,2,3 etc. Si–Si bonds. So starting from bottom, the first cluster has structures that have no Si–Si bonds, second cluster has structures with one Si–Si bond, third cluster has structures with two Si–Si bond etc. A distance below 2.65 Å was defined to be a bond in this instance.

comparison, a similar amount of optimizing from randomly generated structures would have yielded the global minimum at a decent rate in the 18 atom case.

We did not analyze the structures after the fourth one, as that one already was showing signs of disorder which only increase in the case of most structures as we increase the energy. Also the structures were not separated into clear peaks anymore after this point which lessened their value. Analyzing these kind of structures is also very arduous and not very fruitful usually.

3.1.3. Results: general look and stability of global minimum

There is an interesting trend when we look at the different local minima from bond length perspective, especially if we look at silicon–silicon bonds. As we know, the global minimum structure exhibits only silicon–oxygen bonds. On the other hand the various other more amorphous local minima structures, they may or may not have silicon–silicon bonds. We collected a sample of 300000 local minima structures and paired the smallest silicon–silicon distance with the energy of the structure. These structures were randomly generated by using basin hopping to explore the area. The heatmap of these structures can be seen in Fig. 6 on the left. This kind of behavior is of course somewhat expected, as silicon–silicon bonds are not energetically as preferable as silicon–oxygen bonds. It is still interesting to note that having no silicon bonds almost guarantees a fairly low energy structure, and having two or more guarantees a high energy structure.

In Fig. 7 we are showing data of the minima surrounding the global minimum. This data was produced by us trying to jump from the global minimum by moving every atom randomly 0 Å to 0.8 Å in the direction of the cell sides. Smaller distances like 0.6 Å or 0.7 Å would barely ever get out of the global minimum, or would not at all. With 0.8 Å roughly 4% of the jump attempts would still fail to escape the minimum. For reference, the optimal jump distance to find the global minimum is roughly 0.46 Å, as shown in Fig. 3.

There is a large variety of structures around the global minimum energy and SSSD wise, that is similar to the range we would obtained from random exploration in Fig. 6. The concentration of structures is slightly different, but the overall shape is similar. Fortunately the global minimum is at least more stable than the surrounding, or most likely any other, minimum: on average it requires much bigger jumps to get out of its influence than it does for any other structure we have observed. The jumps were required to be roughly twice as long to acquire similar escape chance compared to an average local minima during a search. This should also mean it is easier to fall into it, which makes it easier to find. Unfortunately other highly ordered low energy structures exhibit similar behavior.

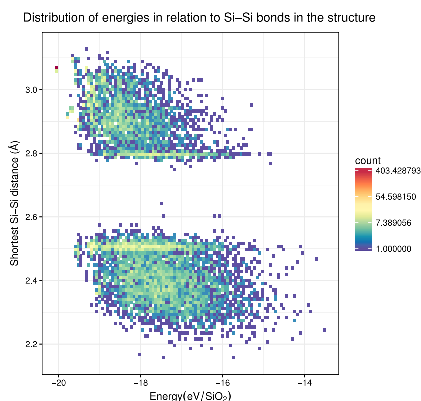


Fig. 7. Distribution of found minima when disturbing the global minimum structure of alphaquartz. We see a wide array of structures energywise, some of which include silicon–silicon bonds (lower half) and some of which do not (upper half). This image was produced from 10000 jump attempts in the 18 atom case.

3.1.4. Conclusions

As said before, locating the global minimum is a difficult task in this case. It is made difficult by our inability to recognize when we are close this desired structure. Energy itself will not really guide as to the minimum. This can be pretty easily seen from Fig. 7 that showcases the large range of energies just surrounding the global minimum, which is really similar to the range we get just by exploring the local minima randomly through basin hopping. On the other hand, We also know of several ordered structures that are close in energy to the global minimum, but structurally clearly completely different. The silicon distances give us a promising guide to lower energies structures, which we made use of in our algorithm.

This is the crux of this optimization problem. It is pretty easy to find good structures, that by many metrics, like energy or silicon–silicon distances, should be close the global minimum. But breaking through the last steps to find the wanted global minimum is challenging.

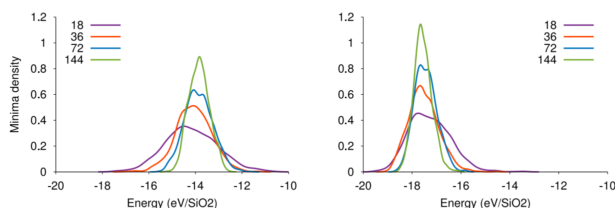


Fig. 8. Each curve represents 10000 runs of a case of 18/36/72/144 atoms, where we used gradient descent to relax a system from different initial structure.
On the left: Completely randomly positioned atoms
On the right: Silicons and oxygens are distributed uniformly in a random manner.

4. Approaching the problem through different methods

As in the previous article with silicon-bulk [18], it is interesting to see how the problem behaves under some basic solver methods. This kind of initial approach gives us some understanding on how this test case behaves. We applied the gradient descent version implemented in LAMMPS. We also tried to solve the problem through the LAMMPS molecular dynamics and pure basin hopping. While possible in theory statistically it is very unlikely to ever find the global minimum through gradient descent as our start guess has to be almost the correct structure already. Through molecular dynamics and basin hopping we would expect to find solution however in a reasonable computational time, at least with the smaller cases.

4.1. Gradient descent

First we checked how this problem behaves under simple gradient descent optimization. This type of optimization is highly dependent on the initial starting structure as it will just follow the gradients to a local minimum. We generated starting points using two different methods and applied gradient optimization. For each type we generated 10000 initial structures and optimized them with gradient descent. The resulting distribution of energies can be seen in Fig. 8.

The first way to generate the starting structures was distributing the atoms completely randomly to the system, which in practice never results in a good structure. In this type of structure the different types of atom are rare evenly distributed and the structure often includes very sparse and dense regions when looking at the atom density. Second method was a bit more involved: we distributed the silicon atoms roughly uniformly, but randomly otherwise, to the unit cell and then followed by adding the oxygen atoms, either to form bridge spots between two silicon atoms that are close enough or to fill empty areas in the cell.

Gradient descent alone we cannot hope to reach the global minimum near -20 eV/SiO_2. The difference between of the energies between the two starting point schemes is massive, roughly 4 eV/SiO_2, but this evens out fast after just few jumps and additional gradient descents. In a test, the completely random sample caught up in energies after we performed around 15–30 jumps on the structures with the number increasing along with the atom number. However generating the evenly distributed start structures takes longer, which cancels out some of the speed benefits.

Interestingly increasing the size of the of simulation cell sharpens the peak, but does not shift it massively. We believe the changes in width is explained by two things: first the small cell shape and size is hindering the gradient descent by blocking out some paths of relaxation, which is why some runs get stuck in high energy structures when there is less atoms. On the other side, with more atoms it is less likely that we get lucky and have a structure that relaxes to a low energy structure.

In our test, this type of single gradient descent method has never managed to find the global minimum, or even gotten close in energywise. This highlights the complexity of this problem even in the relatively small 18 atom case.

4.2. Molecular dynamics

To give perspective to the problem it would be interesting to find how fast this kind of problem gets solved by traditional molecular dynamics and annealing. In our simple approach we did this by assigning a random structure as a starting point and running the system in a heat bath for 25000 timesteps, then slowly cooling it down. Longer heat bath did not seem to have noticeable affects nor did the used starting structure as the heat bath would rearrange the atoms completely. This treatment was done with system sizes of 18, 36 and 72 atoms. Collected data from these calculations is presented in Table 1. The table gives us some perspective to the problem. The 18 atom case solves in a reasonable time, but we were unable to solve the other two cases using this method. The used criterion was that the energy of the found structure was below -20 eV/SiO_2, which in inspection turns out to be the global minimum always. What is interesting is that we see the mean energy of found structures is not really affected by the size of the simulation cell. On the other hand the deviation indicates that the structures are clustered energywise. These effects were noticed with gradient descent too.

While the runtime of this kind of run increases fairly linearly with the atom count, the success rate decreases extremely fast. The small case of 18 atoms gets solved nicely with longer simulations, but at 36 atoms we are already struggling and it gets worse at 72 atoms. This is probably due to amorphous nature of the SiO_2 coming to the effect with the increasing unit cell size. With more atoms it is also a lot more likely to still have few atoms out of place still at the end, even though we might be really close to the desired solution. We tested feeding some of these structures from the longer annealing runs into our optimization algorithm and many of them did indeed find the global minimum quickly.

If we look at the energy distribution of the found structures in the 18 atom case (Fig. 9), we see that the energy/atom has a similar structure to the one in Fig. 5, except the peak for β-quartz is missing and the unnamed structure are missing. With the bigger cells these peaks are however missing, and the distributions drops to zero below the -19.875 eV/SiO_2 mark. Because SiO_2 can be amorphous, the larger cell allows the structure to adjust much better than in the smaller cells, making the amorphous structures more viable energy wise. This in turn causes problems for the annealing process, leading to much smaller success rate at finding the real global minimum among the increased number of amorphous structures, even though the average energy is really low and comparable in different sized unit cells.

Table 1
Gathered simulation statistics on molecular dynamic runs of different lengths, with annealing time referring to the amount of taken timesteps. We see that the mean energy of the runs is not majorly affected by the size of the simulation cell, but the energy deviation and success rate of finding the global minimum are.

	Anneal time	25000	50000	100000	200000	400000	800000	1600000	3200000	6400000
	Mean (eV/SiO_2)	−19.25	−19.33	−19.46	−19.54	−19.61	−19.66	−19.72	−19.76	−19.80
18	Dev (10^{-2} eV/SiO_2)	9.14	7.35	5.59	4.82	5.06	5.56	4.73	3.98	4.32
	Success %	1%	2%	3.5%	4.5%	6.5%	16%	21.5%	20%	33%
	Mean (eV/SiO_2)	−19.25	−19.37	−19.45	−19.53	−19.59	−19.64	−19.68	−19.72	−19.75
36	Dev (10^{-2} eV/SiO_2)	4.81	3.39	2.45	1.69	1.83	1.17	0.93	1.07	0.57
	Success %	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Mean (eV/SiO_2)	−19.23	−19.34	−19.43	−19.51	−19.57	−19.61	−19.65	−19.70	−19.72
72	Dev (10^{-2} eV/SiO_2)	1.55	1.52	1.05	0.94	0.76	0.66	0.49	0.46	0.45
	Success %	0%	0%	0%	0%	0%	0%	0%	0%	0%

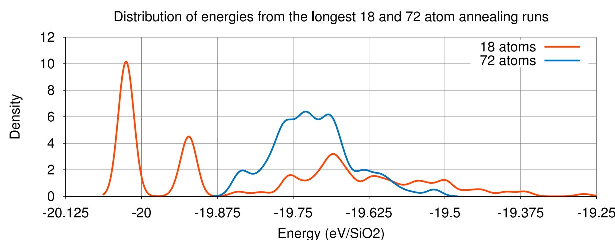


Fig. 9. Presented are the distributions of energies from the longest 18 and 72 atom annealing runs. The 18 atom case is similar to the one presented along with the short basin hopping runs in Fig. 5, but with the 3rd peak of β-quartz missing. The 72 atom case also follows the similar trend from gradient descent runs: the energies form a more clustered hill around the average energy, meanwhile the 18 case stretches further into lower and higher energies.

4.3. Our searches with GHA coupled with basin hopping

Our genetic-hybrid algorithm ran multiple structure searches in parallel iteratively, applying basin hopping in the mutation part of the loop. We tried several different implementations of the basin hopping method. What is interesting is that the completely free jumping seems to work the best, instead of accepting all the jumps going down in energy and only part of the jumps going up in energy. A variation where this chance is altered based on the history of previous jumps to skew the acceptance chance towards 50% is called minima hopping, and has been successful especially with molecules [23]. This kind of behavior however was not beneficial in our tests, instead it is better to accept new jumps always if we are looking only at energy. It is possible that our searches were not long enough for minima hopping method to climb out of the potential well it was trapped in.

We believe this is because of the shape of the energy landscape; it does not have a clear path in decreasing energy to the global minimum. This was apparent in couple of ways. The first sign of this we saw when we found multiple local minima that were close to the global minimum energy, both ordered and amorphous, with all silicon atoms having four oxygen bonds and no silicon–silicon bonds. These minima would easily trap the search. The second sign of this we saw previously in Section 3.1.3 when disturbing the global minimum and observing the minima found around it. We saw that the distribution of these minima is broad in energy terms. Likewise if we are interested in the closest Si–Si distance we see that it too exhibits the same behavior.

A more successful implementation was a variation of basin hopping, where we generate J jumps from the original structure, and choose one of them based on some metric. This kind of method actually turns out to be slightly faster than normal basin hopping, but only if you take into consideration the numbers of jumps made. Because each jump requires roughly J times the processing the method is overall slower. That is, if you use a good metric, as a bad one can prove to be much more detrimental to the search. The metrics we compared were energy, SSSD, a combination of these two and an order based metric, that was calculated from the Fourier transformation. The combination metric uses mainly SSSD, which is to say it favors the structures with less silicon bonds, but solves the ties with the energy metric instead of comparing the bond lengths.

We calculated an order value from the Fourier transformation by trying to evaluate the spikiness of the Fourier transform made on the atomic positions. Doing the transform was a slow process and it was sensitive to the many parameters in the algorithm, probably because of the fairly low resolution we had to use due to costly speed of a three dimensional Fourier transformation. Nevertheless we did find settings that seemed to rank the found structures sufficiently based on their order. The most ordered structures were structures like the previously mentioned α- and β-quartz. We also gathered a big amount of amorphous structures that we arranged based on their order. By eyeball testing we could not see anything wrong in the way the structures were ordered and decided to include this in our guided basin hopping comparison presented in the next section. In theory, this kind of guiding could be effective at the later stages of optimization when we have reached low energies. It could also face the same problems we have had with energy guides, if the order landscape is even more erratic than the energy landscape. We note that our order parameter will not generalize well for other systems; instead new order parameter has to be introduced. We were more interested in seeing how this kind of guidance would affect the search compared to our other methods.

4.3.1. Results

In this section we present how well our algorithm does with this problem using the different guides for the basin hopping. For the data presented in Fig. 10 and Table 2 we ran the algorithm 1000 times with each one performing 300 basin hopping jumps. We restricted the jumps to 300 because in our experience, a search would have found the global minimum often by the 250th jump if it was going to find it at all. The amount of times the algorithm was ran was restricted to 1000 due to the slow speed of the Fourier transformation done for the order guide.

Very counter intuitively, using energy as metric does not work well at all, as seen from Fig. 10 and Table 2. We think the increasing number of options makes it really difficult for the search to get out of a deep

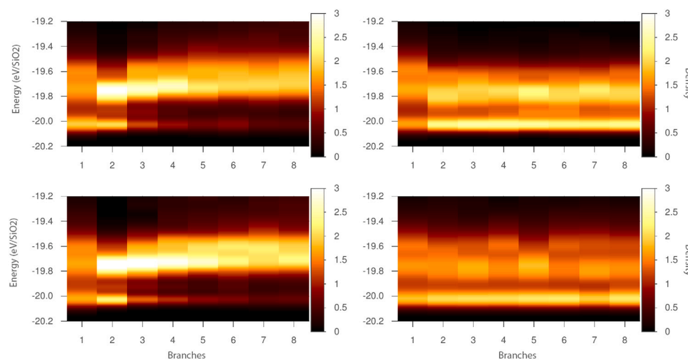


Fig. 10. The heatmaps of the distribution of resulting energies from 1000 samples of guiding basin hopping through either energy (top left), shortest Si–Si bond distance (top right), a combination of the previous two (bottom left) or an order based quantity (bottom right). The *x*-axis represents the amount of jump choices generated and vertical intensity shows the distribution of best found minima energies. We see the harmful effect of energy guidance in both of the two energy guided cases. With bond distance guidance we see a slight improvement when increasing the amount of jump options from 1 to 2, but further increases do not give more benefits. The order guided search finds the global minimum well at similar rates as the Si–Si bond distance based search.

Table 2
Statistics of GHA basin hopping simulations, where at each jump we made the algorithm choose the jump from different number of options. Here the algorithm tried to minimize the amount of Si–Si bonds and maximize the distances between silicons. We see that creating two options is beneficial, but creating more less so. These statistics were obtained from a small sample size of 1000 where the algorithm was ran for 300 jumps each.

Options	Energy guide only							
	1	2	3	4	5	6	7	8
Success rate (%)	20.2	18.3	11.6	8.8	7.9	9.2	6.6	6.6
Mean E (eV)	−19.75	−19.78	−19.69	−19.63	−19.60	−19.59	−19.57	−19.56
Mean jumps for solution	167	153	121	113	128	119	126	120
Mean jumps for best	182	193	196	201	198	194	192	195
Options	Si–Si bond length guide only							
	1	2	3	4	5	6	7	8
Success rate (%)	19.7	24.2	25.5	25.7	23.7	23.5	23.5	24.2
Mean E (eV)	−19.73	−19.79	−19.80	−19.80	−19.80	−19.80	−19.79	−19.79
Mean jumps for solution	169	167	158	163	156	160	164	155
Mean jumps for best	181	180	176	178	182	179	178	177
Options	Combination of previous two							
	1	2	3	4	5	6	7	8
Success rate (%)	19.9	20.6	11.7	9.3	7.6	7.2	6.4	6.6
Mean E (eV)	−19.73	−19.79	−19.72	−19.67	−19.63	−19.63	−19.60	−19.59
Mean jumps for solution	172	142	130	136	117	102	101	104
Mean jumps for best	182	187	196	196	195	191	191	194
Options	Order guide only							
	1	2	3	4	5	6	7	8
Success rate (%)	20.4	23.9	24.4	24.4	25.2	25.5	22.9	25.1
Mean E (eV)	−19.74	−19.76	−19.76	−19.76	−19.77	−19.75	−19.74	−19.76
Mean jumps for solution	167	168	164	168	156	155	154	162
Mean jumps for best	183	185	187	188	187	182	181	187

local minimum. With the increasing amount of options, there is a high chance at least one of them will jump back to this deep minimum, as based on our findings they seem to have wider area of attraction than other surrounding minima with higher energy.

On the other hand, using the number of silicon–silicon bonds as a metric appears to have a positive effect on the system, which starts to decline after there are more than 3 to 4 options. This is a vague metric as the compared value is an integer, which is probably why it is a better metric as it makes it harder for the system to get stuck. We thought we could further improve this by adding either energy, or smallest silicon–silicon distance as a secondary measure, used in case two candidates have the same amount of silicon bonds, but this only had a negative influence on the results. Oddly the order based jumping gives similar results to the SSSD one. The more successful order and Si–Si bond length guides reach their best success rate of finding the global minimum at around 3–4 options per jump.

Interestingly from the contents of Table 2 we see that while the success rate of energy guided hopping is much less likely to succeed, it is more likely to succeed early. We think this is due to the energy focus driving the search structure to the closest deep attraction faster, but after reaching it the search will struggle to break free. With a system with not as deep false attractors this might work really well, like our previous silicon case, but with silicon dioxide and its numerous low energy amorphous structures and few extremely low energy ordered structures this is not the case.

ARTICLE IN PRESS

Using the order parameter as a guide worked actually well. The downside is that the algorithm is of course really slow when using this guide as the transform is computationally expensive. What is interesting is that this guide seems to be the least affected by the amount of generated jump alternatives and the success rate almost steadily increases along with the amount of jump alternatives.

The algorithm works better than the molecular dynamics method, as it takes computationally less time and scales better into the 36 atom case. We did not produce statistically enough data to show for the 36 atom case, but tests showed that the amount of generated jump alternatives had a similar affect to the success rate as in the 18 atom case, with the success rates being noticeably lower of course.

5. Conclusions

We approached the optimization of 18, 36 and 72 atom fixed volume cells of SiO_2 by combining genetic algorithms in Ralf Östermark's GHA-platform with basin hopping and other more system specific mutations to help with exploring the difficult energy-structure landscape. Basin hopping was found to be sensitive in this system, as the range of viable distance done for the perpetuations in basin hopping jumps was found to be narrow. Addition restrictions on the jumps were hard to apply too, because these restrictions would often trap the search into some local area and prevent us from finding the global minimum.

As expected, we found out that the SiO_2-bulk was a much harder system to optimize than the Si-bulk we optimized in our previous study [18]. The oxide exhibits much more structural variety than the Si-bulk. An indication of this were the three identified structures that were close to the global minimum in energy and wider spectrum of energies encountered during the search, both near and far from the global minimum. Of the three mentioned structures, the lowest in energy was a stretched out version of the global minimum, the second lowest was a known β-quartz structure and third an unnamed amorphous structure. There were other low energy structures present in our searches, which were almost as low in energy, but these three stood out when we analyzed the frequency of the structures that came up during the searches.

We found that directing the search towards the global minimum was difficult. Most of the time forcing restrictions lead to worse results than letting the search wander freely, as the repeated local optimizations still direct the search towards low energy structures effectively. If the goal is to just find low energy structures, the energy guided basin hopping worked very well for that purpose. It descents quickly to a low energy structure that usually is not the global minimum.

If we on the other hand are only interested in finding the global minimum, then structure based guiding is more beneficial. When mapping the different SiO_2 structures based on different geometrical features, we discovered that there is a gap between 2.55 Å and 2.75 Å when looking at the shortest silicon–silicon distance in the whole structure, with the global minimum and other low energy structures having near maximal value for this variable. Therefore, an algorithms that aims to maximize the shortest silicon–silicon distance in the system seemed to work really well and improved our chances of finding the global minimum, especially when we increase the atom count. An algorithm that aimed to maximize the order value through the basin hopping performed similarly well. Our order parameter has the downside that it is tied to this system specifically and for different systems we would have to implement a new one.

With the future in mind, we hope to refine the algorithm more and extend to optimizing more relevant systems in material physics: for example the interfaces between silicon and silicon dioxide thin films with defects included in the structure. For more general bulk material optimization we would also have to include the dimensions of the volume to the optimization algorithm as the current study relies on knowing them beforehand.

CRediT authorship contribution statement

Antti Lahti: Software, Writing, Visualization, Investigation. **Ralf Östermark:** Methodology, Software. **Kalevi Kokko:** Conceptualization, Methodology, Supervision, Reviewing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The Magnus Ehrnrooth Foundation, Finland is acknowledged for financial support (A. L.). The computer resources of the Finnish IT Center for Science (CSC) and the FGI project (Finland) are acknowledged. Advice of the experts at CSC on installing LAMMPS on Cray XC40 is gratefully acknowledged.

References

[1] S. Woodley, R. Catlow, Crystal structure prediction from first principles, Nature Mater. 7 (2008) 937.
[2] Z. Chen, W. Jia, X. Jiang, S.-S. Li, L.-W. Wang, SGO: A fast engine for ab initio atomic structure global optimization by differential evolution, Comput. Phys. Comm. 219 (2017) 35–44.
[3] C. Glass, A. Oganov, N. Hansen, USPEX—Evolutionary crystal structure prediction, Comput. Phys. Comm. 175 (2006) 713.
[4] D. Wales, H. Scheraga, Global optimization of clusters, crystals, and biomolecules, Science 285 (1999) 1368.
[5] S. Goedecker, W. Hellmann, T. Lenosky, Global minimum determination of the Born-Oppenheimer surface within density functional theory, Phys. Rev. Lett. 95 (2005) 055501.
[6] S. De, B. Schaefer, A. Sadeghi, M. Sicher, D. Kanhere, S. Goedecker, Relation between the dynamics of glassy clusters and characteristic features of their energy landscape, Phys. Rev. Lett. 112 (2014) 083401.
[7] J. Maddox, Crystals from first principles, Nature 335 (1988) 201.
[8] S.C. Parker, Prediction of mineral crystal structures, Solid State Ion. 8 (1983) 179–186.
[9] S. Kickpatrick, J.C.D. Vecchi, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[10] S.M. Woodley, P.D. Battle, J.D. Gale, A. Richard, C. Catlow, The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation, Phys. Chem. Chem. Phys. 1 (10) (1999) 2535–2542.
[11] M.D. Foster, A. Simperler, R.G. Bell, O.D. Friedrichs, F.A.A. Paz, J. Klinowski, Chemically feasible hypothetical crystalline networks, Nature Mater. 3 (4) (2004) 234–238.
[12] W.L. Bragg, R.E. Gibbs, The structure of α- and β-quartz, Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character 109 (751) (1925) 405–427.
[13] G.A. Lager, J.D. Jorgensen, F.J. Rotella, Crystal structure and thermal expansion of α-quartz SiO_2 at low temperatures, J. Appl. Phys. 53 (10) (1982) 6751–6756, http://dx.doi.org/10.1063/1.330062.
[14] T.A. Weber, F.H. Stillinger, Inherent structures in polyatomic liquids: Simulation for Si2F6, J. Chem. Phys. 95 (5) (1991) 3614–3626.
[15] M. Goldstein, Viscous liquids and the glass transition: A potential energy barrier picture, J. Chem. Phys. 51 (1969) 3728.
[16] R. Östermark, A multipurpose parallel genetic hybrid algorithm for non-linear non-convex programming problems, European J. Oper. Res. 152 (1) (2004) 195–214.
[17] S. Plimpton, Bayesian method for global optimization, J. Comput. Phys. 117 (1995) 1–19.
[18] A. Lahti, R. Östermark, K. Kokko, Optimizing atomic structures through geno-mathematical programming, Commun. Comput. Phys. 25 (2019) 911–927.
[19] J. Tersoff, New empirical approach for the structure and energy of covalent systems, Phys. Rev. B 37 (12) (1988) 6991–6999.
[20] S. Munetoh, T. Motooka, K. Moriguchi, A. Shintani, Interatomic potential for Si-O systems using tersoff parameterization, Comput. Mater. Sci. 39 (2) (2007) 334–339, http://dx.doi.org/10.1016/j.commatsci.2006.06.010.
[21] D.J. Wales, J.P.K. Doye, Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms, J. Phys. Chem. A 101 (1997) 5111.
[22] A. Wright, M. Lehmann, The structure of quartz at 25 and 590C determined by neutron diffraction, J. Solid State Chemistry 36 (3) (1981) 371–380.
[23] S. Goedecker, Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems, J. Chem. Phys. 120 (2004) 9911–9917, http://dx.doi.org/10.1063/1.1724816.