Ab initio doorzoeken van geschikte wolfraamlegeringen
voor de eerste wand in kernfusiereactoren

Ab Initio Screening of Suitable Tungsten Alloys
as First-Wall Material in Nuclear Fusion Reactors

Kurt Lejaeghere

Promotoren: prof. dr. ir. V. Van Speybroeck, prof. dr. S. Cottenier
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Toegepaste Natuurkunde

Vakgroep Toegepaste Fysica
Voorzitter: prof. dr. ir. C. Leys
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2013 - 2014

UNIVERSITEIT
GENT

ISBN 978-90-8578-690-0
NUR 928, 971
Wettelijk depot: D/2014/10.500/36

This research was conducted within the Center for Molecular Modeling.

Leden van de examencommissie

Voorzitter

prof. dr. ir. Rik Van de Walle (Universiteit Gent)

Leescommissie

prof. dr. Patrick Bultinck (Universiteit Gent)
prof. dr. Stefaan Cottenier (Universiteit Gent, *copromotor*)
prof. dr. ir. Xavier Gonze (Université Catholique de Louvain)
prof. em. dr. ir. Guido Van Oost (Universiteit Gent)

Overige leden

prof. dr. ir. Leo Kestens (Universiteit Gent)
dr. Lorenzo Malerba (Studiecentrum voor Kernenergie)
prof. dr. Jörg Neugebauer (Max-Planck-Institut für Eisenforschung)
prof. dr. Dimitri Van Neck (Universiteit Gent, *secretaris*)
prof. dr. ir. Veronique Van Speybroeck (Universiteit Gent, *promotor*)
prof. dr. ir. Toon Verstraelen (Universiteit Gent)

Contents

Voorwoord

Dreigend. Omineus zelfs. Ik vermoed dat deze termen het beste beeld bieden van de waarschuwingen die aan het schrijven van deze doctoraatsthesis voorafgingen. Het merendeel van mijn illustere voorgangers was het er immers over eens: eenmaal de laatste fase van het doctoraat intreedt, herleidt de vrolijke doctoraatsstudent zich tot een zwoegende letterzetter, die tot in het holst van de nacht over zijn kluwen van nota's gebogen zit. Ik heb het eerlijk gezegd niet gemerkt. De laatste maanden zijn als in een waas voorbijgevlogen. Nu de proefdruk me van over mijn bureau ligt aan te staren, durf ik me in een onbewaakt moment af te vragen of het schrijven een soort trance met zich meebrengt.

De rest van mijn doctoraat staat me nochtans helder voor de geest. De pure voldoening wanneer je een probleemstelling tot zijn meest essentiële componenten hebt kunnen herleiden en oplossen. De intense discussies, die aanleiding geven tot wildste ideeën. En bovenal, de schoonheid van de wetenschap. Als het plaatje klopt, dan is het Af. Met een hoofdletter.

En nu is deze periode bijna achter de rug. Het voelt als de eerste kleine stapjes van een kind dat de grotemensenwereld ontdekt. Op zo'n moment kijk je achterom, en zie je dat dit niet mogelijk was zonder enkele mensen die je eerst bij de hand hielden, om langzamerhand, haast ongemerkt, de steun weg te nemen. Om plots te beseffen dat je al op eigen benen staat.

Mijn twee toeverlaten verschillen misschien wel als dag en nacht van elkaar, en bieden net daarom de perfecte mix voor de startende onderzoeker. Veronique en Stefaan, zonder jullie was dit avontuur niet mogelijk, en waren de eerste aarzelende stapjes misschien snel in een valpartij ontaard. Veronique, je leerde me de knepen van het vak en hoe je als groentje in de vraatzuchtige academische wereld stand kan houden. Ook je visie op lange termijn kunnen weinig je nadoen: je hebt een ongeëvenaard besef van wat nodig is om succesvol de eindmeet te halen, en ik hoop daar ook iets van opgestoken te hebben. Stefaan, jij leerde me dan weer stil te staan en na te denken. Resultaten produceren is relatief gemakkelijk, maar om daar ten volle de implicaties van in te zien... Bovendien had ik vaak de indruk dat we elkaars klankbord waren, meer nog dan begeleider en doctoraatsstudent. Ik hoop dat we dat in de toekomst mogen blijven doen.

Hoewel slechts twee personen je daadwerkelijk kunnen helpen bij die eerste stappen, zijn er altijd de vele supporters, die langs de zijlijn aanmoedigen of onderweg een dankbaar steuntje in de rug geven. Na drieënhalf jaar zijn jullie met zo velen, dat ik jullie onmogelijk allemaal kan opnoemen. Daarom alvast een welgemeend woord van dank voor iedereen die me hielp zo ver te raken, zowel binnen het Centrum voor Moleculaire Modellering – mijn thuis de voorbije jaren en podium voor de onvolprezen FML's – als in Gent, België, en het buitenland.

Ik wil toch enkele personen nog speciaal vermelden. Zo is er Michel, toch wel wat de peetvader van het CMM. Je wijze raad en gourmandise worden sterk gesmaakt. Ook mijn bureaugenoten, Kim en Louis, verdienen een woordje, om dag in dag uit mijn geraaskal te moeten doorstaan – en ik het hunne. Ward, die een schitterende website op poten hielp te zetten om ons Delta-project met de wereld te delen. De minimeeters, die week na week de vrijdag weten te kruiden met DFT-miserie en -successen. Steven, die ten koste van zijn eigen tijd heldhaftig voor me in de bres sprong en tijdens het schrijven van deze thesis mijn oefeningenlessen deels overnam. Dimitri en An, voor de intrigerende samenwerking bij het vak Kwantummechanica I, nu ze waardig in Michels voetstappen getreden zijn. Mijn thesisstudenten, Francis en Craig, met wie de samenwerking steevast interessante resultaten leverde. En natuurlijk Wim: altijd behulpzaam en een grafisch genie. Vergaap u maar aan de uitnodigingen, of aan Figuur 1.8. Die zijn beide zijn werk.

Ik wil natuurlijk ook enkele niet-CMM'ers bedanken. In eerste plaats de mensen van HPC, die altijd klaar stonden om de problemen met de supercomputer aan te pakken. Jilt en Leo, voor de nuchtere materiaalkundige kijk. Serge en Nele bij OCAS, die me bij mijn stage en masterthesis warm maakten voor toepassingsgericht materiaalonderzoek. Guido, Lorenzo, Dmitry and Alexander, who provided feedback from the fusion community. Also beyond Flanders, there were many supporting people, who enthusiastically stepped into our Delta project. Marcin, Chandler, Xavier, ... there are just too many of you, but I really appreciate the trouble you have already taken to make this initiative work.

Ten slotte rest me nog mijn gezin te bedanken. Waar het voorgaande mijn academische familie betrof, zijn jullie natuurlijk al veel langer bij project 'Kurt' betrokken. Jullie onverdroten steun heeft me al door veel woelige waters geloodst, en ook de voorbij drieënhalf jaar waren niet zonder jullie mogelijk. Laat staan hoe de kleurenschema's van deze thesis eruitgezien zouden hebben.

Kurt Lejaeghere
14 mei 2014

Samenvatting

Materiaalonderzoek speelt een belangrijke rol in de menselijke geschiedenis. De vooruitgang van onze samenleving wordt immers in grote mate bepaald door de grondstoffen waarover we beschikken. Door nieuwe en meer efficiënte materialen te ontwikkelen, leerden onze voorouders betere werktuigen, stevigere gebouwen en nieuwe producten te vervaardigen. Op die manier leidde de overgang van het stenen naar het bronzen tijdperk en uiteindelijk het gebruik van ijzer tot verschillende maatschappelijke revoluties. Ook vandaag nog zijn materialen doorslaggevend. Zo zou het huidige succes van de elektronica niet mogelijk zijn zonder gespecialiseerde siliciumtechnologie.

Er is ook een keerzijde aan het grote belang van materialen. Het zorgt ervoor dat veel innoverende ontwikkelingen staan of vallen met de beschikbaarheid van het juiste materiaal. Dit is vooral problematisch door de complexiteit van veel hedendaagse vraagstukken. Kernfusie is hét voorbeeld bij uitstek. Om voor de toenemende wereldbevolking een menswaardige levensstandaard te garanderen, zijn immense hoeveelheden energie nodig. Toch moet dit gebeuren zonder onze ecologische voetafdruk nog verder te vergroten. Kernfusie is voor deze doeleinden uitermate geschikt. Fusie is een CO_2-neutrale en vrijwel onuitputtelijke energiebron en laat bijna geen langlevend radioactief afval achter. Het is bovendien volkomen veilig, doordat in fusiereactoren geen kettingreacties optreden: bij het stilvallen van de centrale houden alle processen onmiddellijk op. Toch is kernfusie nog niet voor morgen. Na verschillende decennia van wetenschappelijk onderzoek en politiek getouwtrek is pas recent de constructie van de testreactor ITER begonnen, en het is nog niet duidelijk hoe we die naar een commercieel niveau kunnen opschalen. Ook hier is het gebrek aan geschikte materialen een van de belangrijkste beperkingen.

De omstandigheden waaraan fusiematerialen blootgesteld worden zijn dan ook niet te onderschatten. Aan de binnenwand van het reactorvat kan de temperatuur oplopen tot 900°C, waarbij we niet eens de instabiliteiten van het fusieplasma in rekening brengen. Daarnaast wordt die 'eerste wand' ook gebombardeerd met grote hoeveelheden neutronen, die als restproduct van de fusiereacties optreden. Momenteel lijkt slechts één materiaal in staat

die stroom van warmte en deeltjes ook op de schaal van commerciële reactoren te weerstaan: wolfraam. Dit metaal smelt pas bij 3400°C en ondervindt relatief weinig schade door de neutronenimpact. Daar staat tegenover dat het bij kamertemperatuur bros is, wat het erg moeilijk maakt wolfraam te verwerken tot reactoronderdelen en bovendien ook de levensduur op lange termijn niet ten gunste komt. Daarom zoeken wetenschappers naar een beter vervormbaar alternatief.

Een mogelijke oplossing ligt in het legeren van wolfraam. Door andere elementen aan het zuivere metaal toe te voegen, kunnen de materiaaleigenschappen gewijzigd worden zonder de voordelen van puur wolfraam te verliezen. Dit bleek alvast mogelijk met rhenium, hoewel de prijs een struikelblok is: rhenium kost slechts twee keer minder dan goud. Er is daarom nood aan een goedkopere optie, maar hoe kunnen we alle mogelijkheden voldoende grondig bestuderen? Experimenten zijn erg tijdintensief, zodat het noodzakelijk is om al een idee te hebben welke samenstellingen best werken. Gelukkig is de kracht van kwantumfysische computersimulaties zodanig toegenomen dat veel legeringen eerst computermatig gescreend kunnen worden en slechts een kleine selectie aan dure laboratoriumproeven onderworpen hoeft te worden. Dit proefschrift heeft dan ook tot doel *op grote schaal de eigenschappen van wolfraamlegeringen te berekenen en op zoek te gaan naar een goedkope samenstelling die de kwaliteiten van puur wolfraam bij hoge temperaturen combineert met een goede vervormbaarheid.* We gaan daarbij te werk in drie stappen.

In een eerste fase is het belangrijk zicht te krijgen op de nauwkeurigheid van onze computersimulaties. Hoewel we van een kwantummechanische methode gebruik maken, dichtheidsfunctionaaltheorie of DFT, krijgen we op verscheidene niveaus met onzekerheden te maken. Zo hangen de DFT-resultaten af van de manier waarop de wiskundige vergelijkingen in computermethoden verwerkt werden (*numerieke fouten*). Aangezien praktische procedures vaak vereenvoudigingen vereisen, leidt dit tot numerieke ruis. Bovendien geeft DFT ook zonder deze praktische problemen geen perfecte weergave van de werkelijkheid. De theorie zelf is correct, maar voor één ingrediënt (de zogenaamde 'functionaal') is geen exacte vorm gekend. Men moet daarom zijn toevlucht nemen tot beredeneerde gokken, wat voor kleine maar niet-verwaarloosbare afwijkingen zorgt (*intrinsieke fouten*). Ten slotte zijn sommige materiaaleigenschappen te complex om met de huidige computerkracht rechtstreeks te berekenen. We maken dan gebruik van eenvoudige relaties tussen microscopische grootheden, die we wel kunnen simuleren, en het materiaalgedrag waarin we geïnteresseerd zijn. Ook dit houdt een benadering in (*semi-empirische fouten*), zodat we bij onze uiteindelijke voorspellingen met een drievoudige foutenmarge rekening moeten houden.

Merkwaardig genoeg zijn, ondanks verschillende decennia van vastestof-

simulaties, nog geen algemene en kwantitatieve onzekerheden op de voorspelde grootheden beschikbaar. In een experimentele context is het achterwege laten van foutenmarges nochtans *not done*. We stellen daarom een ijkingsprocedure voor waarbij we elk van de drie fouten op basis van een verzameling testmaterialen afschatten. Voor numerieke fouten definiëren we een Δ-criterium, dat ons toelaat verschillende DFT-codes met elkaar te vergelijken. Dit gebeurt door met beide codes te voorspellen hoe de testmaterialen op volumeveranderingen reageren, en het verschil over alle materialen heen uit te middelen. Voor de bepaling van de intrinsieke en semi-empirische fouten maken we daarentegen gebruik van een statistische procedure. Op die manier kunnen we systematische afwijkingen onderscheiden en de resterende fout in algemeen toepasbare foutenvlaggen gieten. Bovendien is een vergelijking met meer geavanceerde technieken mogelijk. Zo vinden we dat de mate waarin materialen bij hogere temperaturen uitzetten even goed voorspeld wordt door benaderende formules als door intensieve berekeningen van de roostertrillingen.

Deze bevindingen laten toe de eigenschappen van wolfraamlegeringen te bepalen en afwijkingen op onze DFT-berekeningen in kaart te brengen. In een tweede stap is het echter ook nodig tussen de vele resultaten naar het beste materiaal op zoek te gaan. Wanneer we slechts één doel voor ogen hebben (een hoge smelttemperatuur, bijvoorbeeld), kan dat door het metaal met de beste waarde te selecteren. In het geval van kernfusie moeten we daarentegen aan meerdere vereisten voldoen. Niet alleen moet de legering de gunstige eigenschappen van zuiver wolfraam behouden, maar het dient ook minder bros te zijn zonder de prijs al te veel te doen stijgen. Die vereisten zijn vaak tegenstrijdig, en het gaat er in de praktijk om voor het beste compromis te kiezen. Dat is trouwens niet alleen het geval voor fusie: ook bij andere hoogtechnologische toepassingen zijn zo'n afwegingen onontbeerlijk.

Wanneer meerdere criteria tegelijk geoptimaliseerd dienen te worden, is een ruwe selectie alvast mogelijk door enkel de Pareto-efficiënte materialen te beschouwen. Voor de meeste legeringen is er altijd een alternatief te vinden dat op alle punten beter scoort, maar dat is niet het geval voor legeringen in de Paretoverzameling. Toch blijven er dan nog heel wat mogelijkheden over, vooral wanneer van een grote hoeveelheid samenstellingen vertrokken werd. We voeren daarom een nieuw criterium in, dat we de minimale winfractie noemen. Het drukt uit hoeveel we winnen of verliezen wanneer we één Paretomateriaal boven de andere kiezen. De wolfraamlegeringen met de hoogste minimale winfracties zijn dan het meest geschikt voor kernfusie.

De ontwikkeling van de minimale winfractie en de ijking van DFT-berekeningen voor vaste stoffen bieden een kader voor de derde en laatste stap van dit proefschrift: de grootschalige simulatie van wolfraamlegeringen voor fusiedoeleinden. In de praktijk bestaan er een bijna onbegrensd aantal mo-

gelijkheden: zo kan het combineren van slechts drie niet-radioactieve stoffen met wolfraam al op 88 560 verschillende manieren, en die legeringselementen kunnen dan nog eens in verschillende hoeveelheden aanwezig zijn. We beperken ons daarom tot een grote maar haalbare deelverzameling van materialen, waarbij tot twee metalen uit twintig kandidaten aan wolfraam toegevoegd worden in één vastgelegde verhouding. Dat komt neer op 210 mogelijke samenstellingen of 9 170 DFT-berekeningen. De studie van zo'n groot aantal wolfraamlegeringen vereist een verregaande automatisering van de simulaties en gebruik van enkele van de grootste supercomputers van België. Op basis van deze berekeningen merken we een grote stabiliteit wanneer de legeringselementen erg gelijkaardig zijn aan wolfraam zelf. Om ook een gunstige vervormbaarheid en een lage prijs te bekomen, lijkt het dan weer interessanter om twee stoffen toe te voegen die erg van elkaar verschillen, en waarvan enkel hun *gemiddelde* op wolfraam lijkt. Verder onderzoek is aangewezen om na te gaan of deze conclusies ook in experimenten opgaan.

Naast het mogelijke belang voor kernfusie, zijn de technieken die in dit proefschrift voorgesteld worden ook ruimer toepasbaar. Zo heeft de bepaling van onzekerheden in vastestofsimulaties ertoe geleid dat verschillen tussen DFT-methoden voor het eerst systematisch bestudeerd worden. Momenteel worden verschillende initiatieven opgestart waarin ontwikkelaars van DFT-codes en atomaire potentialen de grenzen van de numerieke nauwkeurigheid opzoeken. De minimale winfractie is dan weer interdisciplinair inzetbaar voor verschillende baanbrekende technologieën. Bijna elke complexe probleemstelling, gaande van bouwprojecten tot ruimtevaart, gaat immers gepaard met meerdere (tegenstrijdige) vereisten, waarin ons nieuwe criterium klaarheid kan scheppen. Ten slotte is ook onze geautomatiseerde procedure voor de berekening van ternaire wolfraamlegeringen eenvoudig uit te breiden. De beschikbaarheid van nauwkeurige en grootschalige DFT-berekeningen bevestigt zo de waardevolle rol van computersimulaties in het ontwerp van innovatieve materialen.

Summary

There are little research areas that shape society as materials design does. Indeed, our society advances as the materials it uses improve. By developing new and more efficient compounds, mankind was able to craft better tools, construct sturdier buildings and introduce novel commodities. As such, the transition from the Stone Age to the Bronze Age and the eventual discovery of iron catalyzed several technological revolutions. Even now, materials are still a determining factor in scientific breakthroughs. Without the development of silicon, for example, the thriving electronics industry would be all but non-existing.

The prominent role of materials is not always helpful. Because of it, many innovating developments critically depend on the availability of the right material. Most present-day problems are quite complex, however, so suitable compounds are increasingly hard to find. Nuclear fusion demonstrates this particularly well. To ensure an acceptable standard of living for our growing world population, huge amounts of energy are required, yet we need to take care not to strain our ecological reserves any further. Nuclear fusion addresses both issues. Fusion does not emit CO_2 nor leave any substantial amount of long-living radioactive waste, while still providing a virtually unlimited supply of energy. It is moreover perfectly safe as it does not involve a chain reaction: shutting down the plant brings all processes to an immediate halt. Nevertheless, the road to nuclear fusion is a long and painstaking one. Several decades of scientific research and political debates have only recently resulted in the construction of the ITER test reactor, and it is still unclear how to scale up these technologies to a commercial level. Here too, one of the principal obstacles is the lack of appropriate materials.

Indeed, the conditions inside working fusion reactors are extremely harsh. Even without taking plasma instabilities into account, temperatures at the inner surface of the reactor vessel lie between 500 and 900°C (900 to 1700°F). In addition, this 'first wall' is constantly bombarded with massive numbers of neutrons, a side product of the fusion reactions. At this time, only one material seems to be able to withstand these impinging fluxes of heat and particles at the scale of commercial reactors: tungsten. Tungsten only melts at 3400°C (6200°F) and is relatively insensitive to neutron damage. It is brit-

tle at room temperature, on the other hand, making it hard to process into reactor components and compromising its long-term stability. Scientists are therefore searching for a more ductile, and hence more malleable, alternative.

One possible solution is alloying. By adding other elements, tungsten's brittleness can be addressed without affecting its favourable properties too much. Such a ductilization has already been achieved with rhenium, but its price prevents large-scale applications: rhenium is half the price of gold. A cheaper option is therefore required, but how can we investigate all possibilities in sufficient detail? Experiments take much time and are best limited to a few compositions that are suspected to work well. Fortunately, the power of quantum physical computer simulations has increased to such an extent that they may be used to computationally screen large numbers of alloys before subjecting a small selection to expensive lab tests. This dissertation hence aims to *perform high-throughput simulations of tungsten alloys and search for an inexpensive compound that combines tungsten's favourable high-temperature properties with an acceptable ductility*. We proceed in three stages.

In a first stage, we need to know how accurate our computer simulations are. Even for a quantum mechanical method like density-functional theory (DFT), there are uncertainties at multiple levels. On the one hand, DFT results depend on how the mathematical equations were implemented into computer methods (*numerical errors*), as practical procedures require simplifications and give rise to numerical noise. Moreover, even without these practical issues, DFT does not represent reality perfectly. The theory itself is correct, but one ingredient (the so-called 'functional') is not known exactly. Educated guesses are available, but they lead to additional deviations (*intrinsic errors*). Finally, the currently available computing power is not sufficient to compute some complex materials properties directly. We instead apply simple relations between microscopic quantities, which we can simulate, and the materials behaviour we are interested in. The corresponding approximation (*semi-empirical errors*) leaves our overall predictions with a three-fold error bar.

Remarkably, several decades of solid-state simulations have not led to general and quantitative uncertainties for predicted properties yet. In an experimental context, error bars are indispensable, though. We therefore propose a benchmark procedure in which each of the three errors is estimated from a set of test materials. For numerical errors, we define a Δ criterion, which allows us to compare different DFT codes. It predicts the response of the test materials to volume changes and averages the difference between the two codes over the entire test set. To determine intrinsic and semi-empirical errors, on the other hand, we use a statistical procedure. It enables us to distinguish systematic deviations and express the remaining errors as generally

applicable error bars. A comparison to more advanced techniques is then also possible. We, for example, find the expansion of materials at high temperatures to be equally well predicted by approximate formulas as by intensive calculations of the lattice vibrations.

These considerations allow determining the properties of tungsten alloys and mapping the expected deviations from DFT. In a second stage, it is also necessary to screen many results for the best material. When we only aim to optimize a single objective (a high melting temperature, for example), we only need to select the metal with the best value. In the case of nuclear fusion, however, we need to deal with multiple requirements. Not only does the alloy need to retain the favourable properties of pure tungsten, but it should also be more ductile without raising the price too much. These requirements often do not match, and in practice, we are looking for the best compromise. Such issues are moreover not unique to fusion: for other cutting-edge technologies, similar considerations are essential as well.

When multiple criteria need to be optimized simultaneously, a preliminary selection is already possible by only taking the Pareto efficient materials into account. For most alloys, there is always a way to simultaneously improve all objectives, but for compounds in the Pareto set, that is not the case. Nevertheless, the remaining number of possibilities is sometimes still too large, especially when many options were considered to begin with. We therefore introduce a new criterion, which we call the minimum win fraction. It expresses how much we win or loose by choosing one Pareto material over another. The tungsten alloys with the highest minimum win fractions are then most suited for nuclear fusion.

The development of the minimum win fraction and the quantification of solid-state DFT errors offer the necessary background to proceed with the third and last stage of this dissertation: the large-scale search for fusion-related tungsten alloys. In practice, the number of possible alloys is nearly unlimited: only three non-radioactive elements, for example, can already be combined with tungsten in 88 650 different ways, and this can be done at different concentrations. We therefore restrict ourselves to a large but manageable subset of materials, in which up to two dopants (out of twenty candidates) are added to tungsten in one predetermined composition. This boils down to 210 potential compounds or 9 170 DFT calculations. Studying this many tungsten alloys requires thoroughly automating our simulations and using some of the largest supercomputers in Belgium. Based on these calculations, we notice the tungsten alloys to be the most stable when the two alloying elements closely resemble tungsten itself. Then again, to also obtain a favourable ductility and low price, it appears more interesting to combine two strongly different dopants, of which only their *average* resembles tungsten. Further research may show whether these conclusions are also valid

experimentally.

The techniques proposed in this dissertation are not only relevant for nuclear fusion. By determining the uncertainties in solid-state simulations, for example, we try to convince the modelling community how important it is to systematically study the effect of different DFT approaches. At this point, several initiatives are being set up in which the developers of DFT codes and atomic potentials address these numerical aspects. The minimum win fraction, on the other hand, is even useful over a wide range of scientific disciplines and technological applications. Indeed, almost every complex problem, varying from construction projects to space navigation, comprises multiple (conflicting) requirements, on which our novel criterion may shed light. Finally, the automated procedure for the calculation of tungsten ternaries is also straightforwardly extensible. The availability of accurate and large-scale DFT calculations thus confirms the valuable role of computer simulations for the design of innovating materials.

1 General introduction

1.1 Nuclear fusion

1.1.1 Free energy for all?

We live in a golden age. Although the economic crisis has taken its toll, most Western families still have a roof over their head, food to sustain them and energy to heat them. There have never been more cars on the road and a flight to the other side of the world has never been easier. In the meantime, high-tech appliances are flooding our homes, provided by an ever revolutionizing industry. The same standard of living is now also becoming available to the rest of the world. Life in many emerging economies has already improved significantly and one can only hope developing countries to follow in their footsteps soon. Ideally, each individual should have access to the same services and the same high level of welfare.

Such an endeavour not only requires social and economic changes. The energy-related challenges are considerable too. The world population is already at 6.8 billion and is expected to rise to 8.6 billion by 2035 [1]. Combined with an increase in income, especially in emerging economies, this puts a heavy load on the supply of electricity. By 2035, the demand is estimated to increase by one third and the price to go up by 15 % [2]. A major question is therefore how to address the booming power consumption. The issue is aggravated by the need for sustainable energy sources. CO_2 emissions already amount to 31.2 gigatonnes a year and would grow to 37.0 by 2035. According to International Energy Agency, this corresponds to a long-term global

warming of 3.6°C [2]. Moreover, fossil fuel reserves are dwindling, with only 53 years of oil, 69 years of natural gas and 129 years of coal left [2]. Renewable energy sources are slowly taking their place, at already one third of the total electricity output, but their intermittent character make them less reliable. Nuclear fission is an option too, but its safety is being questioned since the Fukushima incident. Several European countries are even abandoning nuclear power entirely. There is therefore a pressing need for a new and sustainable source of baseline energy production.

1.1.2 A potential solution

A promising candidate is nuclear fusion [3–5]. Contrary to nuclear fission, it gains energy from joining two light atomic nuclei, rather than breaking up a heavy nucleus into smaller parts (Figure 1.1 (top)). In both reactions, a more stable nucleus is created, as can be seen from a plot of the binding energy per nucleon (Figure 1.1 (bottom)). In nuclear fusion, however, the stability difference and energy gain are much higher, enabling a higher return.

Fusion is a common phenomenon in the interior of the sun. Each second, hydrogen isotopes crash into each other, giving rise to the production of helium and a massive amount of energy. This so-called proton-proton chain reaction is not suited for use on Earth, however. The only reason solar fusion works, is because of the sheer size of the sun, as solar conditions take a year for even one in ten billion protons to react. Researchers instead aim to harness the fusion power of deuterons and tritons. These are the nuclei of two isotopes of hydrogen (with one and two neutrons respectively), which yield a helium nucleus, a neutron and 17.6 MeV of energy upon reacting. Contrary to the fusion of two protons, the so-called D-T reaction has a larger cross section (5 barn) at much lower impacting energies (110 keV). It therefore occurs more frequently than its H-H counterpart, rendering it more suitable for applications on Earth. Deuterium (the atom based on a deuteron) constitutes 0.015 % of all hydrogen and is therefore abundantly present. Tritium is much rarer, but can be bred from widely available lithium through neutron impact.

Nuclear fusion outperforms nuclear fission in many respects. The energy gain per reaction is much higher, for example, as was mentioned before. Lighter nuclei are moreover not radioactive or have a fast decay rate, giving fusion an additional edge. Finally, there is no chain reaction involved, allowing the operators to shut down the reactor at any point and avoiding scenarios like in Chernobyl or Fukushima. Compared to other energy sources, the raw materials for fusion energy production are geographically dispersed and abundantly available (hydrogen and lithium are two of the most common elements on Earth). Because of its lack of highly radioactive waste or CO_2 emission, it is also an environmentally friendly power source.

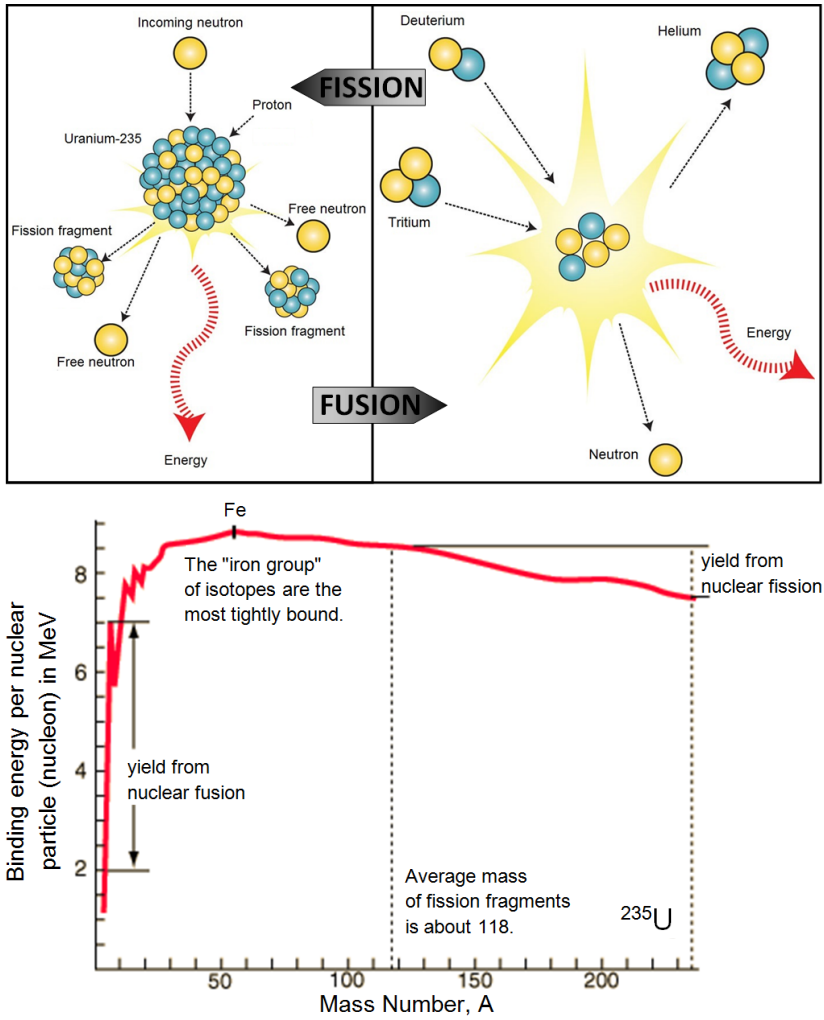


Figure 1.1: Nuclear fusion and nuclear fission, compared in terms of the nuclear reactions [6] (top) and the binding energy per nucleon [7] (bottom).

1.1.3 Containing the power

Achieving nuclear fusion is not trivial, however. To expedite the reaction, minimum temperatures of the order of 100 million K are needed. A working fusion reactor therefore requires some means of containing this tremendous heat. In practice, the fusion particles are often confined magnetically [3–5][1].

[1] We will not elaborate on less well developed alternatives like inertial fusion, despite having already achieved a few successes as well [8].

This is possible because at such high temperatures, all matter ionizes, forming a plasma. The charged nature of plasmas allows exploiting the Lorenz force, which causes the ions and electrons to follow helical trajectories around magnetic field lines. Tamm and Sakharov already realized in 1951 that in a toroidal configuration, where the magnetic field lines are closed, this effect enables capturing all charged particles and effectively containing the fusion heat. This set-up was called a tokamak, which stands for тороидальная камера с магнитными катушками (**to**roidal'naya **ka**mera s **ma**gnitnymi **k**atushkami) or 'toroidal chamber with magnetic coils'.

A tokamak is driven by both toroidal and poloidal magnetic fields (see Figure 1.2). The toroidal fields are generated by poloidal coils and give rise to the basic plasma confinement. Without modifications, however, such a plasma exhibits a pressure gradient, which causes it to expand outwards. Poloidal fields can counteract this effect. These fields are primarily produced by the motion of the plasma itself, which induces a toroidal current around a magnetic yoke. By using some additional toroidal magnets, the plasma can be shaped in more detail. The overall effect is a toroidal reactor where the

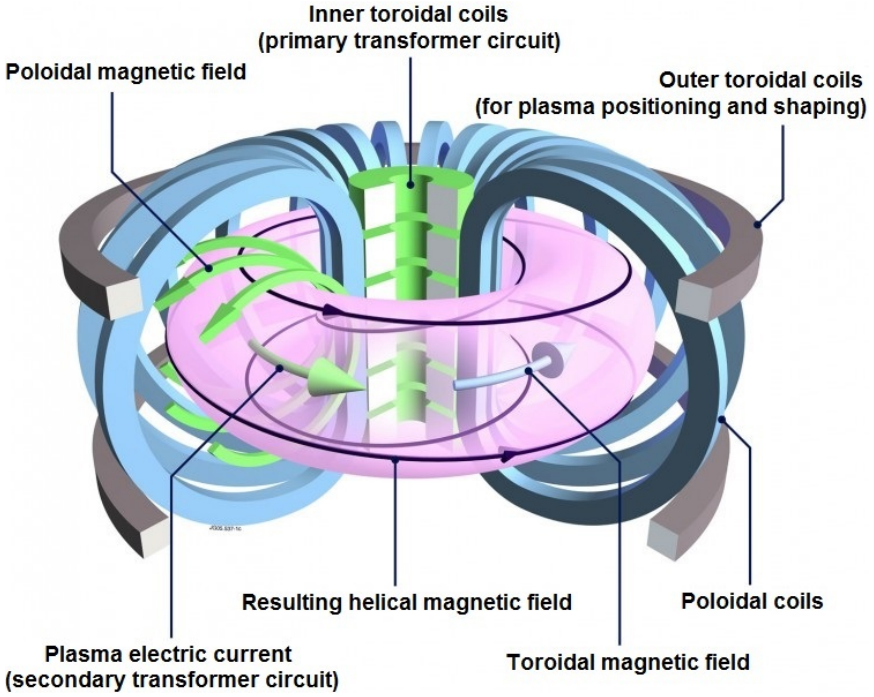


Figure 1.2: Toroidal and poloidal fields in a tokamak configuration for fusion plasmas [9].

plasma is confined by helical field lines. Actual fusion is then achieved using several heating mechanisms (ohmic, particle injection or RF heating).

Magnetic confinement allows containing the most extreme conditions that correspond with nuclear fusion. Not all effects can be shielded, however. Temperatures at the edge of the plasma still amount to more than 800 K [10]. In addition, large amounts of helium nuclei and neutrons are generated, which need to be dealt with. Since neutrons are not charged, they can penetrate the magnetic field lines and damage the reactor vessel. The helium nuclei are contained, but without some extraction mechanism, they will hinder the reactive species and suffocate the fusion reaction. Finally, tokamak plasmas also suffer from inherent instabilities. Even in the parts of operating space where they can be controlled, typical instabilities like edge-localized modes (ELMs) are unavoidable and may still damage plasma-facing materials [3].

These considerations demonstrate how important the inner components of a tokamak are. They have to withstand high neutron and heat fluxes (for normal operation and for off-normal events), without compromising the lifetime of the fusion reactor. There should moreover be a possibility for removing waste particles. Figure 1.3 shows how a typical D-T plasma vessel addresses these issues. The major part of the inner surface is called the first wall, and absorbs most of the incoming heat. It covers water- or helium-cooled blanket modules, which contain the necessary lithium to breed tritium. At the bottom of the reactor vessel, a divertor is added to clean the plasma from exhaust gases, such as helium. Waste particles are guided towards it by means of open magnetic field lines (separated from the closed field lines by a magnetic separatrix). At the divertor, ions are then neutralized and evacuated using vacuum pumps. Of course, this also means that the divertor experiences even worse particle and heat fluxes than the first wall.

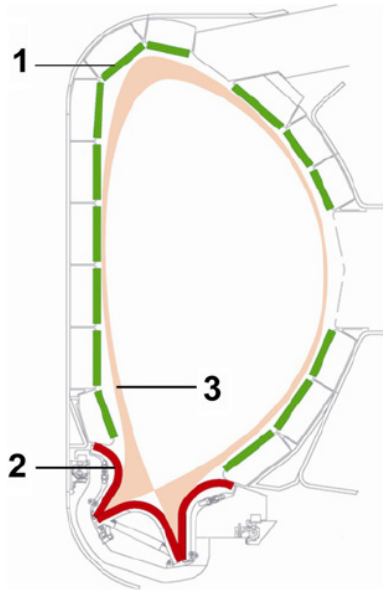


Figure 1.3: Cross-section of a tokamak [10] with (1) the first wall, (2) the divertor and (3) the separatrix. Magnetic field lines beyond the separatrix are open and direct particles towards the divertor.

1.1.4 A global strategy

Fusion reactors have come a long way since their pioneering days in the 1950s (Figure 1.4). Nevertheless, constantly scaling up has rendered several issues even more critical and has produced additional, unexpected challenges. International collaborations have therefore been set up to support the transition of nuclear fusion to a commercially viable power source. Instigated on a European level, a fast-track procedure was proposed in 2001 to expedite the fusion development [11, 12]. It entails the parallel development of several test reactors and facilities and should yield the first commercial fusion plant by 2050. In this way, the European Fusion Development Agreement (EFDA) aims to generate 30 % of the world electricity production from nuclear fusion by 2100.

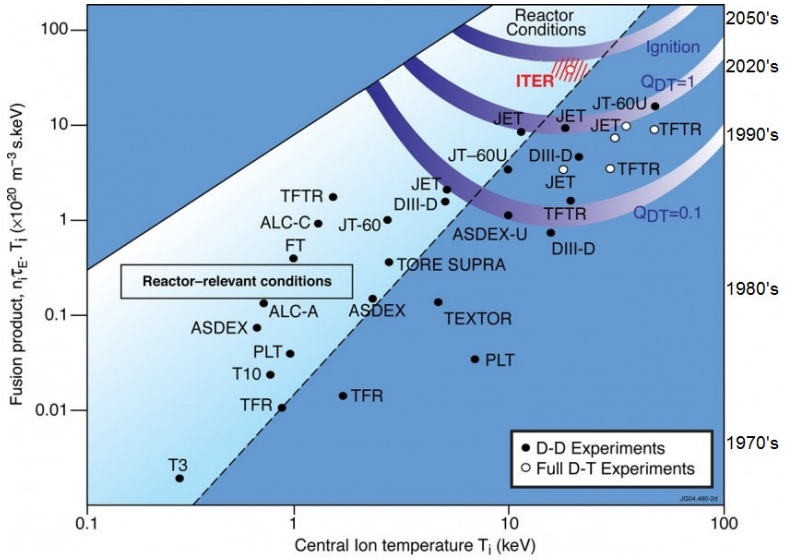


Figure 1.4: Progress of nuclear fusion over the last decades (right axis) and expected performance of ITER, expressed in terms of the central ion temperature and the fusion triple product (a figure of merit for confined plasmas) [9]. Q_{DT} is the energy efficiency of a D-T fusion reaction, with 1 corresponding to break-even conditions.

The fast-track strategy primarily focuses on the ITER and DEMO reactors to bridge the gap between science-driven tokamaks and industry- and technology-driven plants [12]. ITER is currently being built in Cadarache, in the south of France. As it paves the way for large commercial-scale machines, its name is suitably chosen as the Latin word for 'road'. ITER is a key facility in the global fusion strategy [12, 13]. Not only will it be the first reactor

that produces significantly more power than it consumes, it will also investigate several design questions for future machines. Within its 30-year operation time, it aims to address several physical, materials-related and technical questions that have remained unresolved until now. In doing so, it will provide the necessary information to build DEMO, the first demonstration power plant. DEMO [12, 14], for which construction is foreseen in the early 2030s, will generate hundreds of megawatts of power. It represents the last step before the first real commercial plant and will therefore deal with several related aspects, such as net electricity output to the grid and tritium breeding in a closed fuel cycle.

ITER cannot prepare for all challenges in DEMO, however. DEMO's increased reactor efficiency comprises higher neutron fluxes, which the ITER materials are not able to withstand. A separate facility will therefore be built, specifically aimed at investigating the influence of intense neutron fluxes on candidate materials [12, 14]. Construction of this International Fusion Material Irradiation Facility (IFMIF) is planned to start between 2017 and 2021 in Rokkasho, Japan. By assessing the effects of neutron damage in parallel to the development of more efficient tokamaks, it enables a significant speed-up of the fusion program. The experimental work will moreover be combined with theoretical and modelling efforts, which can help extrapolating the known physics and guide the investigations by IFMIF.

1.2 Fusion materials

1.2.1 Stringent requirements

Despite the promises of nuclear fusion, there is still much work to be done before it can be considered as a valid alternative for conventional energy sources. One of the most critical issues is the development of materials that are able to survive the harsh conditions in the plasma vessel. Indeed, the EFDA road map identifies seven challenges for fusion energy, three of which are directly related to the reactor materials [12]. Although many high-strength compounds are already available, the requirements for fusion materials are of a scale that has not been encountered before. The identification and development of more suitable compounds is therefore indispensable.

As mentioned previously, the most apparent forms of plasma-material interaction are due to the huge heat and neutron fluxes. Although the particular numbers depend on the reactor design, minimum temperatures are expected to vary between 800 K (first wall [10]) and 1100-1200 K (divertor [3]). Heat fluxes can be as high as 2 MW/m^2 for the first wall under normal operation, and increase up to peak loads of 10-20 MW/m^2 at the divertor [10]. The effect of the neutrons is substantial as well. The damage they do is measured

in dpa or displacements per atom. In a material with 1 dpa, neutron collisions have displaced each atom on average 1 time during its lifetime. While the neutron power fluxes will 'only' be 0.5 MW/m^2 in ITER, corresponding to 5 dpa in steel [13], up to 100 dpa neutron damage is expected in DEMO's tungsten components [15].

These heat and neutron fluxes drastically affect the plasma-facing materials [10, 15, 16]. When inappropriately dealt with, the incident heat causes metals to melt and recrystallize, substantially degrading their properties. This is especially the case when the thermal conductivity is low, leading to large temperature gradients and considerable surface heating. The incoming neutrons, on the other hand, generate collision cascades within the reactor materials, damaging the crystalline structure and causing severe embrittlement. Moreover, by transmuting the constituent elements, they not only render the plasma-facing materials radioactive, but change their composition and properties as well [17, 18]. The combined effect of heat and neutron fluxes also modifies surfaces and sputters away several millimeters during a component's lifetime. These sputtered atoms can be detrimental to the plasma stability if they reach the reactor interior.

Not all effects are as clear-cut. There are also more subtle issues which need to be taken into account [15, 16]. It is important, for example, to ensure a good chemical compatibility between the structural materials and the coolant. Most designs for DEMO propose a tungsten divertor, cooled with oxygen-containing helium. Since tungsten is known to be prone to oxidation, this might lead to the release of volatile tungsten oxides. Another problem is the generation of thermomechanical stresses when the thermal expansion of different materials or different areas within the same material does not match. Finally, the large amount of hydrogen and helium in the plasma can also affect the surrounding components, forming cavities and bubbles and causing embrittlement. Each of these issues becomes even more critical in view of the long lifetime and system availability that will be required of future fusion reactors.

1.2.2 The fusion materials family

Several classes of materials have been proposed to address the harsh conditions for plasma-facing components [19–21]. Their suitability depends on the reactor design, which imposes certain materials requirements, safety margins and production and processing limitations. Conventional metals, for example, such as austenitic stainless steels or alloys based on copper, titanium, chromium or nickel, may serve to strengthen the structural parts of the reactor. When the exposure to neutron fluxes becomes too critical, however, low-activation materials offer a better alternative. Typical examples are

ferritic-martensitic steels, vanadium alloys and ceramic composites. In the high-flux areas of the reactor vessel, such as at the edge of the first wall and at the divertor, more specialized choices are required. In ITER, the first-wall armour plates will be made out of beryllium. Its low atomic number guarantees a low energy loss – Be particles are fully ionized in the centre of the plasma, so they do not emit radiation by electronic excitation [22] – and beryllium is good at gettering oxygen from the vessel vacuum. For the divertor strike point tiles, on the other hand, carbon fibre composite is foreseen because of its high thermal shock resistance. However, both beryllium and carbon fibre composite are ill-suited for the most intensely radiated components. In these circumstances, beryllium melts, while the carbon-based material suffers from massive erosion and tritium retention. In ITER, these conditions only apply to certain parts of the divertor, but in DEMO, even the first wall will be subject to too high heat and neutron fluxes.

Refractory metals are less susceptible to these issues. Tungsten in particular is often used as a first barrier between the plasma and the rest of the reactor vessel. Although its application will be limited to the dome and baffle region of the divertor in ITER [20], DEMO's plasma-facing components will almost entirely consist of tungsten, with tungsten armour plates in the first wall and structurally used tungsten in the divertor [14]. Tungsten's advantages are numerous [14, 16, 19]. Not only does it have the highest melting point of all pure metals, it also expands little at high temperatures and possesses a good thermal conductivity and surface heat capability. Tungsten is moreover not easily activated or sputtered and has a low vapour pressure, mitigating the problems because of its high atomic number. As a structural material, it has a high Young's modulus and creep resistance. Of course, there are also some drawbacks when using tungsten. Its low fracture toughness, high ductile-to-brittle transition temperature and neutron-induced embrittlement complicate its use in structural components, both during production and operation. As mentioned before, tungsten is also very prone to oxidation, which could lead to the release of highly volatile WO_2 particles. Finally, tungsten is quite expensive, especially in comparison to steels: there is a 40-fold price difference between iron-chromium and tungsten sheets (prices in 2000 [23]). Table 1.1 quantitatively illustrates some of the properties of tungsten in the context of nuclear fusion and compares to a ferritic-martensitic steel.

1.2.3 Tungsten – State of the art

Despite these unresolved problems when using tungsten, it remains the best candidate for application in plasma-facing components. It is therefore critical to address the shortcomings of tungsten without compromising its fa-

Table 1.1: Some fusion-relevant properties of pure tungsten and F82H ferritic-martensitic steel: price for a 2.3 mm sheet, ultimate tensile strength (uts), linear thermal expansion coefficient (α_l), thermal conductivity (κ), fracture toughness (K_Q), thermal figure of merit for the surface heat capability (M), ductile-to-brittle transition temperature after irradiation (DBTT), maximum tensile elongation after irradiation (ϵ_{max}) and neutronic response after 15 MW y/m^2.

	W		F/M	
price [USD/kg]	200	[23]	5.5	[23]
uts (300 K) [MPa]	380	[23]	660	[23]
(1000 K) [MPa]	320	[23]	200	[23]
α_l [10^{-6}/K]	4	[19]	10	[24]
κ [W/K m]	85	[19]	33	[19]
K_Q [MPa $\sqrt{\text{m}}$]	30	[19]	50-500	[25]
M [–]	11.3†	[23]	2.7‡	[19]
DBTT [K]	1200	[23]	240-320	[24]
ϵ_{max} [%]	0	[23]	10-15	[24]
neutron damage [dpa]	45	[16]	170	[16]

† at 1300 K; ‡ at 900 K

vourable properties. Especially its brittleness urgently requires an answer. It makes tungsten hard to process into structural components and imposes a threshold service temperature. The issue is moreover aggravated by neutron irradiation, as this increases the ductile-to-brittle transition temperature even further (Figure 1.5). Brittleness can be characterized by a material's deformation properties (a low maximum elongation) or its fracture behaviour (a low fracture toughness), and both phenomena are related to microstructural features, which are strongly affected by plasma exposure [15]. When defect clusters accumulate, for example, they pin dislocations, which hardens the material and decreases its plasticity. Dislocation channelling can occur as well. In that case, the dislocations pick up defects along their way through the crystal, leaving behind almost defect-free channels and exposing the grain boundaries to higher stresses. Even helium can affect the ductility, since helium cavities may promote intergranular fracture. Although this list is certainly not exhaustive, it demonstrates that reliable fusion materials require controlling both the intrinsic ductility or brittleness and the interaction with the microstructure.

Given the importance of ductility, it has been the focus of much fusion research over the last decades. So far, several processing techniques have

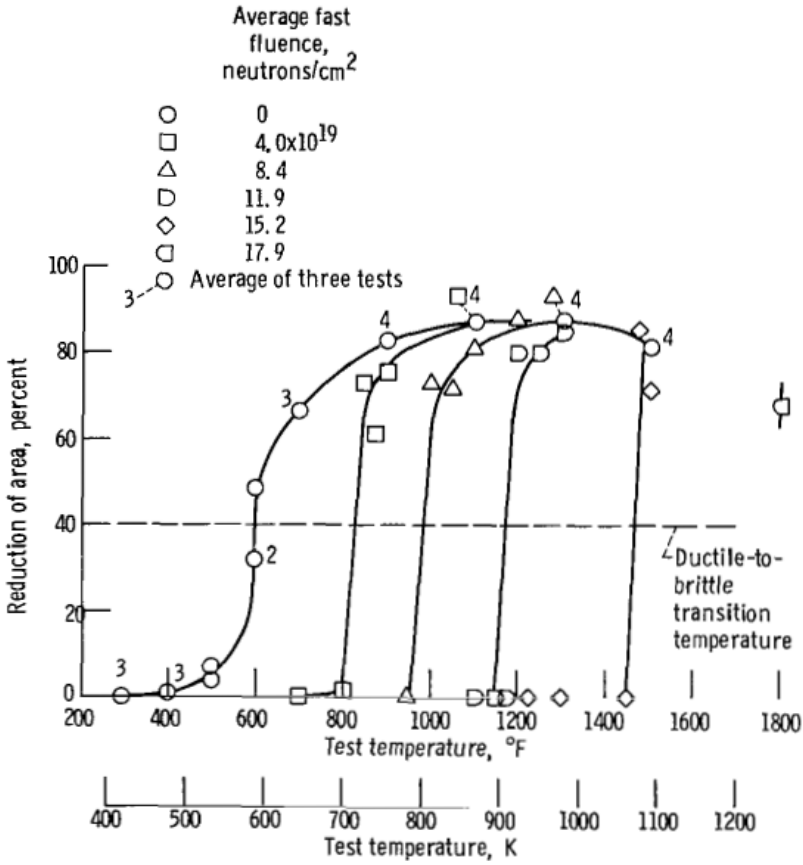


Figure 1.5: Shift of the ductile-to-brittle transition temperature of tungsten under irradiation with neutron energies above 1 MeV [26]. A considerable amount of necking (cross-section area reduction) in a tensile test is indicative for ductile samples.

been found to mitigate tungsten's brittleness, but they often give rise to other issues. A common method is the use of nanostructuring to achieve ultrafine-grained tungsten compounds [14]. This leads to a redistribution of the impurities and a higher dislocation density. Such a reduction in grain size is observed when dispersing oxides (ThO_2, La_2O_3, Y_2O_3) or TiC into the tungsten matrix [10,14,19,23], or when adding small amounts of potassium to stabilize the grain boundaries (WVM) [14]. This increases the fracture toughness at room temperature by a factor of 5 [14] and shifts the recrystallization temperature upwards [27]. Unfortunately, the lower brittleness is not maintained under neutron irradiation [19]. The ductility improves in a similar way when applying severe plastic deformations. On the other hand, it also causes the

recrystallization temperature to deteriorate. Not all approaches focus on a reduction in grain size. The low toughness of tungsten can, for example, also be addressed by using metal composites, such as Ta- or W-fibre-reinforced tungsten [10].

One of the most straightforward options in the search for more ductile tungsten compounds is alloying, however. Although alloys mostly become harder and more brittle due to solid solution strengthening, some tungsten-based solid solutions or intermetallics possess properties superior to the pure metal. The best-known example is that of rhenium. Already at concentrations of a few percents, rhenium is observed to improve the fracture toughness of tungsten and decrease the ductile-to-brittle transition temperature [23, 27, 28]. Alloying with iridium is thought to have similar effects [10]. On the other hand, both rhenium and iridium are extremely expensive. A less costly alternative is to add 3 to 6 % of Fe and Ni, which raises the fracture toughness to 60-200 MPa $\sqrt{\text{m}}$ [23], but does deteriorate the high-temperature properties (Plansee's DENSIMET alloy, for example [27]). There is therefore much room for improvement, especially since the large range of possible tungsten alloys remains all but uncharted territory. With all intuitive options exhausted by now, *the next stage in the search for ductile tungsten alloys entails a systematic screening of all alloying elements and concentrations.*

1.3 Computational materials design

1.3.1 Down the rabbit hole: from real size to atomic scale

Which alloying elements improve the properties of tungsten most? So far, this question has remained unanswered. Indeed, screening all possibilities is a daunting task. Of all chemical elements, 83 have stable isotopes [29], and each of them can be introduced into the tungsten host material. While 'only' 82 binary alloy systems exist, there are already 3 321 ternaries and 88 560 quaternaries. When combined with a 5 %-interval concentration sampling, for example, this means that 820 binary, 149 445 ternary and 10 627 200 quaternary tungsten-rich compounds need to be evaluated. The characterization of this many candidates is nearly impossible. Each alloy would have to be manufactured with a sufficiently high purity and be submitted to a considerable number of measurements. To date, only one approach has the potential to address such a search within a reasonable time[2]: a computational one.

[2] 'Reasonable' is to be interpreted loosely: if we assume that we can characterize all basic properties of a single compound in 3 000 hours on one core, a reasonable assumption for the currently available resources, it would take 12 days to assess all binaries, 74 months for the ternaries and 431 years for the quaternaries on the 528 16-core nodes of the largest Flemish supercomputer muk (see Appendix A).

Unfortunately, it is not straightforward to investigate the macroscopic behaviour of a material on a purely theoretical basis (also called first principles or *ab initio*). Phenomena like ductility originate at multiple length and time scales, and not all aspects can be simulated within the same formalism. A classic approach is then to proceed in a hierarchical way: at each computational level, key parameters can be retrieved from increasingly more expensive methods at increasingly smaller length scales (Figure 1.6) [34–36]. Such

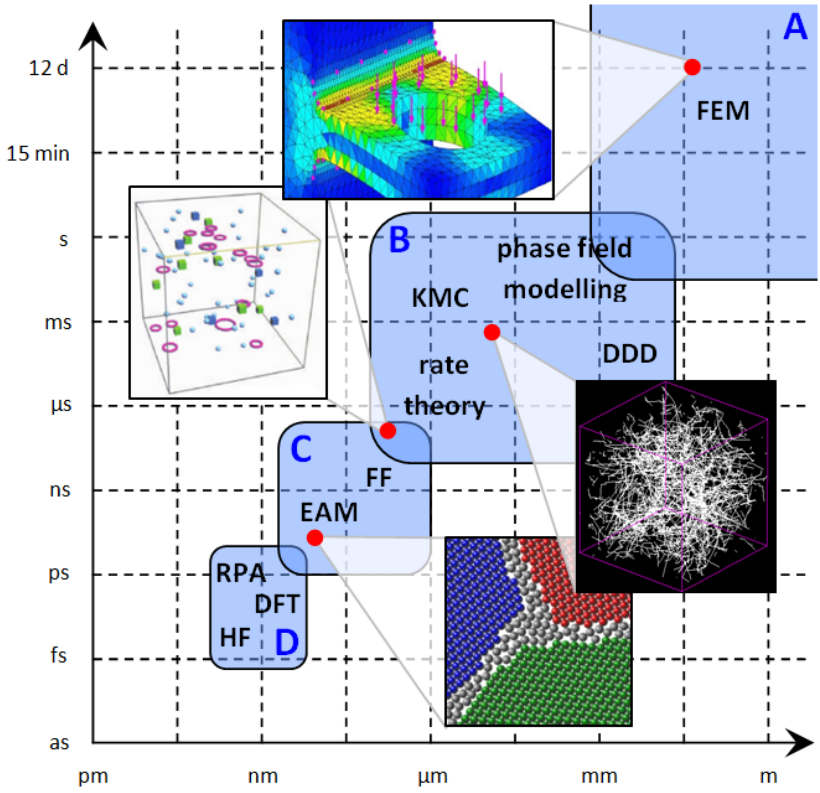


Figure 1.6: Hierarchical approach to materials modelling, using macroscopic (A), microstructural (B), semi-empirical (C) and quantum mechanical methods (D). Each level is illustrated by a few example techniques: the finite-elements method (FEM), phase field modelling, kinetic Monte Carlo (KMC), discrete dislocation dynamics (DDD), rate theory, force fields (FF), the embedded-atom method (EAM), density-functional theory (DFT), Hartree-Fock (HF) and the random-phase approximation (RPA). The inset pictures represent, from large to small length scales, real-size components [30] (FEM), dislocation networks [31] (DDD), defect clusters [32] (KMC) and grain boundaries [33] (EAM).

a multiscale strategy has already proven successful in the past [32,37,38] and provides much insight in physical mechanisms that are not easily identified experimentally.

At the macroscopic scale (Figure 1.6, set A), one of the most common engineering techniques is the finite-elements method (FEM) [39]. It divides a full-scale object into polyhedral microcells, on which the governing constitutive laws of the material are imposed and numerically solved. The FEM approach allows applying localized stresses, deformations and thermal fluxes, and investigates how they propagate through the material. It requires accurate information about the material's behaviour, which is obtained experimentally or, when accurate experimental information is unavailable, computationally. Indeed, the constitutive laws can also be obtained from continuum mechanical theory or microstructural mesoscale models (Figure 1.6, set B). Kinetic Monte Carlo methods (KMC) [40] or rate theory [41], for example, may be used to describe the mobility and reactions between defects, taking into account the probabilities and rates with which these structures are interacting. The former method is object-, event- or lattice-based, while the latter employs a mean-field formalism. Phase field modelling [42], on the other hand, investigates the microstructural evolution via differential equations in terms of phase-dependent order parameters. Finally, discrete dislocation dynamics models (DDD) [43] allow simulating complicated dislocation networks and the interaction with surrounding defect structures. From this, the plastic deformation can be evaluated. Each of these approaches again critically depends on further input from smaller scales (Figure 1.6, set C), as it is necessary to know the interaction energy of a dislocation and an interstitial cluster, for example, or the stress and strain field generated by a dislocation loop. Stress and strain fields follow from anisotropic continuum theory [44], which is based on the assumptions of a continuous medium and linear elasticity. In the case of dislocations, it relates the characteristics of infinite straight dislocations to their orientation and the material's elastic properties. The interaction of large defect clusters, or even the behaviour of a single dislocation in a defected material, require considering the material from an atomic point of view, using molecular dynamics, for example. While several approximate methods exist to deal with large numbers of atoms, such as force fields (FF), embedded-atom potentials (EAM) or tight-binding approaches [34], these in turn depend on accurate knowledge about atomic interactions (Figure 1.6, set D). Such information can only be reliably obtained when employing state-of-the-art quantum mechanical techniques – post-Hartree-Fock (post-HF), density-functional theory (DFT) or the random-phase approximation (RPA), for example [45] – which become computationally more intensive as their accuracy grows. Quantum mechanical calculations can moreover provide us with local materials properties, such as elastic constants, defect trapping en-

ergies or reaction enthalpies.

This descent into the microscopic world involves many steps. As a result, a good correspondence between the separate scales needs to be ensured [34]. This is not only the case when a purely hierarchical approach is followed, as described above, but also when the different length scales are combined interactively. In concurrent calculations, for example, all scales are solved simultaneously and coupled with each other. Another technique is to impose multiscale boundary conditions. In that case, macroscale information is fed to atomistic simulations at the border of the simulation box, through lattice impedances. The present work will focus on only one computational level, however, the quantum mechanical one (see further), and couple to larger scales via semi-empirical approximations (i.e. QSPRs, see Section 1.3.3).

1.3.2 It's a quantum world

For existing materials, a top-down multiscale approach as outlined above provides the best way to understand the underlying physics. In the design of new materials, on the other hand, it is more interesting to vary the atomic structure and assess the macroscopic repercussions. Computational materials design hence corresponds to a bottom-up approach, in which the properties of the pure materials are first characterized using quantum mechanical simulations. This information is then employed as input for larger-scale models, gradually including microstructural effects and macroscopic component shapes and conditions. Such an approach shows that, no matter how complicated the real-life material becomes, its behaviour is already determined at the smallest length scales, although we might not know a closed expression that incorporates all factors of influence. A multiscale approach replaces that unknown equation by a more elaborate, but feasible scheme. Nevertheless, all required information for the macroscopic material behaviour is inherently present in atomic phenomena. A systematic screening at that level may therefore provide a glimpse of the bigger picture, in addition to providing the necessary information for the next link in the computational chain.

Density-functional theory At the atomic scale, a solid is nothing more than electrons and nuclei. Their energy E_{tot} is described by the Hamiltonian H_{tot}, which contains both kinetic (T) and Coulombic potential energy parts (V). This gives rise to a many-body Schrödinger equation:

$$\hat{H}_{tot}\,|\Psi_{tot}\rangle = \left(\hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn}\right)|\Psi_{tot}\rangle = E_{tot}\,|\Psi_{tot}\rangle \qquad (1.1)$$

with e standing for electrons and n for nuclei. When assuming that the nuclear and electronic motion can be decoupled (the so-called clamped nuclei

assumption in the Born-Oppenheimer approximation), this equation simplifies to:

$$\hat{H}\,|\Psi\rangle = \left(\hat{T}_e + \hat{V}_{ee} + \hat{V}_{en}\right)|\Psi\rangle = E\,|\Psi\rangle \tag{1.2}$$

where the Hamiltonian H, the wave function $|\Psi\rangle$ and the energy E are purely electronic, and only parametrically depend on the nuclear positions. Nevertheless, $|\Psi\rangle$ does still depend on all N electrons. Solving this multidimensional differential equation is therefore almost impossible, and in practice, several approximations are available. We will focus on density-functional theory (DFT) [46, 47], however, as it also includes some part of the electron correlation, and thus combines a relatively high speed for high-throughput applications (see Section 1.3.3) with a good accuracy.

The breakthrough of DFT already dates back to the 1960s with the works of Pierre Hohenberg, Walter Kohn and Lu Jeu Sham [48, 49]. Because of its flexibility and accuracy, the theory gained in popularity over the years, first for modelling solids and later for molecular systems. In 1998, this culminated in a Nobel prize in chemistry for Walter Kohn to honour his work in this field [50]. Today, the availability of increasingly better approximations and increasingly larger computing power have made it one of the most widely used quantum mechanical approaches.

DFT replaces the search for the ground-state wave function by the search for the ground-state electron density. Indeed, the Hohenberg-Kohn theorems [48] prove that in the ground state, the electron density $\rho(\boldsymbol{r})$ contains sufficient information to determine the energy and that it fulfils the variational principle as well. It is therefore possible to fully express the energy as a functional of $\rho(\boldsymbol{r})$:

$$T_e[\rho] + V_{ee}[\rho] + V_{en}[\rho] = E[\rho] \tag{1.3}$$

This functional equation can be solved using the Kohn-Sham scheme [49], which states that an independent-particle model (IPM) exists that gives rise to the same ground-state electron density:

$$\hat{H}^{KS}\,|\psi_i\rangle = \left(\hat{T}^{IPM} + \hat{V}_{eff}\right)|\psi_i\rangle = \epsilon_i\,|\psi_i\rangle \tag{1.4}$$

or in coordinate space and using atomic units

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(\boldsymbol{r})\right]\psi_i(\boldsymbol{r}) = \epsilon_i\psi_i(\boldsymbol{r}) \tag{1.5}$$

with $\sum_i |\psi_i(\boldsymbol{r})|^2 = \rho(\boldsymbol{r})$, and the total wave function of the non-interacting system a single Slater determinant of the one-particle orbitals $|\psi_i\rangle$ ($i = 1, \ldots, N$). The effective potential energy V_{eff} and potential $v_{eff}(\boldsymbol{r})$ (its functional derivative with respect to ρ) now only operate on one particle at a time,

which allows solving this equation much more easily. Of course, these quantities depend on the density that equations (1.4) or (1.5) are solved for, so an iterative scheme is needed, resulting in a self-consistent field (SCF).

By requiring the Kohn-Sham equations (1.4) to yield the same ground-state density as (1.3), it follows that

$$\frac{\delta T_e[\rho]}{\delta\rho(\boldsymbol{r})} + \frac{\delta V_{ee}[\rho]}{\delta\rho(\boldsymbol{r})} + \frac{\delta V_{en}[\rho]}{\delta\rho(\boldsymbol{r})} = \frac{\delta T^{IPM}[\rho]}{\delta\rho(\boldsymbol{r})} + v_{eff}(\boldsymbol{r}) \tag{1.6}$$

or

$$v_{eff}(\boldsymbol{r}) = -\sum_{\alpha} \frac{Z_\alpha}{|\boldsymbol{r} - \boldsymbol{R}_\alpha|} + \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}' + v_{xc}(\boldsymbol{r}) \tag{1.7}$$

v_{eff} is hence equal to the sum of the ionic potential v_{en} (with Z_α and $\boldsymbol{R}_\alpha$ the atomic numbers and positions of the ions, respectively), the Hartree potential $v_H(\boldsymbol{r})$ (a mean-field Coulombic term) and the exchange-correlation potential $v_{xc}(\boldsymbol{r})$. For the last contribution, no analytical expression is known.

As the Kohn-Sham ground-state density $\sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2$ is equal to the real $\rho(\boldsymbol{r})$, it can be plugged into (1.3) to calculate the corresponding energy. Using (1.6) and (1.7), we obtain:

$$E[\rho] = T^{IPM}[\rho] + \int \rho(\boldsymbol{r}) v_{en}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} + \frac{1}{2}\int \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}' + E_{xc}[\rho] \tag{1.8}$$

Note that the energy cannot be calculated as the sum of the orbital energies ϵ_i, since the Kohn-Sham orbitals $|\psi_i\rangle$ do not possess any physical meaning.

To determine the total energy (1.8) and the Kohn-Sham potential (1.7), we require an expression for the exchange-correlation energy $E_{xc}[\rho]$ and its functional derivative $v_{xc}[\rho; \boldsymbol{r}]$. These are the most critical ingredients of Kohn-Sham DFT, as everything else does not entail any major approximations. For the exchange-correlation energy, which is mostly denoted as 'functional', only guesses exist, and this forms the main source of errors with respect to the exact quantum mechanical solution.

A simple yet surprisingly efficient guess of the exchange-correlation functional assumes each infinitesimal volume element to behave as a uniform electron gas, the so-called local-density approximation or LDA. More advanced strategies [51] also include a dependence on the gradient of the electron density (the generalized-gradient approximation or GGA), the kinetic-energy density (meta-GGA) or even some exact exchange between the Kohn-Sham one-particle orbitals, similar to Hartree-Fock (hybrid functionals). The reliability of these functionals has provided DFT with a good predictive value,

making it one of the most popular choices for the quantum mechanical treatment of materials.

For solids, especially the GGA functional of Perdew, Burke and Ernzerhof [52] (PBE) has proven to perform well over a wide class of materials [53]. It fulfils the correct bounds and recovers the LDA energy in the limit of a uniform electron gas. It moreover closely resembles the PW91 functional [54,55] (except for surfaces [56]), although the latter was based on an analytic fit. PBE is therefore able to produce reliable energetic and structural information at an acceptable computational cost, making it one of the most popular functionals to describe condensed matter [57]. Unfortunately, PBE's preference for large gradients of the electron density carries too much weight, leading to overly large volumes (as opposed to LDA, which makes them too small) [58]. Several GGAs have therefore revised the original PBE form to resolve this issue (e.g., PBEsol [59] and WC [60]). Other GGA functionals try to improve the description of molecules (such as revPBE [61] and RPBE [62]) or even molecules and solids simultaneously (HTBS [63]). Nevertheless, none of these GGAs currently seems to be able to outperform regular LDA for vibrational and thermodynamical properties [64], or PBE for cohesive energies [53, 65]. As a matter of fact, not even common meta-GGA [66, 67] or hybrid functionals [68] are able to improve the PBE cohesive energies.

Plane-wave basis sets When a suitable functional has been chosen, density-functional theory generally reduces to solving the Kohn-Sham equations (1.4) and determining the energy (1.8). Although this can be done with grid-based methods [69] or Green's functions [70], the most common approach decomposes the one-particle orbitals using a basis set $|\phi_n\rangle$. Equations (1.4) then take a matrix form, and this generalized eigenvalue problem can be solved more easily [71]:

$$HC = SCE \tag{1.9}$$

The corresponding matrix elements are given by $H_{mn} = \langle\phi_m|\hat{H}^{KS}|\phi_n\rangle$, $C_{ni} = \langle\phi_n|\psi_i\rangle$, $S_{mn} = \langle\phi_m|\phi_n\rangle$ and $E_{ij} = \epsilon_i\delta_{ij}$.

The exact treatment of the Kohn-Sham formalism requires a complete (and hence infinite) basis set, but in practice, only a finite number of basis functions can be taken into account. If the used basis set is large enough, the results will nevertheless converge to the correct values. An efficient choice of the shape of the basis functions may moreover improve this convergence, reducing the computational effort drastically[3]. On the other hand, efficient basis sets are less transferable, so a compromise is generally required.

[3]Indeed, solving the eigenvalue equation requires a diagonalization of the Hamiltonian matrix, which has a complexity of $\mathcal{O}(N^3)$ for N-component basis sets [72].

An efficient choice of the basis set implies anticipating on the expected shape of the single-particle orbitals. Molecular problems therefore often employ atom-like basis functions, such as Slater-type orbitals and contracted Gaussian functions. For condensed-matter simulations, on the other hand, it is best to already incorporate the periodicity of the studied material. Plane waves are particularly suited to this end. They are moreover inherently orthogonal, reducing the overlap matrix S in the Kohn-Sham matrix equation (1.9) to the identity matrix. Finally, the use of plane waves is also compatible with the quantum mechanical treatment of solids, which says that all eigenstates can be written as Bloch functions:

$$\psi_i(\boldsymbol{r}) = \psi_{\boldsymbol{k},n}(\boldsymbol{r}) = u_{\boldsymbol{k},n}(\boldsymbol{r})\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{1.10}$$

in which $u_{\boldsymbol{k},n}(\boldsymbol{r})$ are periodic. Because of the periodicity, the quantum number i separates into a periodic contribution (the wave vector $\boldsymbol{k}$) and a band number n. It can be proven that only solutions with $\boldsymbol{k}$ in the first Brillouin zone (BZ) are unique.

Since the functions $u_{\boldsymbol{k},n}(\boldsymbol{r})$ follow the same periodicity as the real-space lattice, we can write

$$\psi_{\boldsymbol{k},n}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{G}} c_{\boldsymbol{k},n}(\boldsymbol{G})\, \mathrm{e}^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}} \tag{1.11}$$

where $\boldsymbol{G}$ stands for the reciprocal lattice vectors and the factor $1/\sqrt{\Omega}$ (with Ω the unit cell volume) is used for normalization purposes. The wave function can therefore be described by an infinite yet discrete set of plane-wave basis functions. This basis set can be limited in a natural way by noting that larger wave vectors $\boldsymbol{k}+\boldsymbol{G}$ correspond to increasingly smaller wavelengths. Imposing an upper limit for $|\boldsymbol{k}+\boldsymbol{G}|$ hence introduces a minimum resolution. This maximal wave number G_{max} is often expressed as an energy, the cutoff energy

$$E_{cut} = \frac{G_{max}^2}{2} \tag{1.12}$$

Using a plane-wave basis set allows expressing some parts of the DFT formalism in reciprocal space. This substantially simplifies the Kohn-Sham equations (1.4) and (1.7) [73, 74]:

$$\left\langle \boldsymbol{k}+\boldsymbol{G} \,\middle|\, \hat{T}^{IPM} + \hat{V}_H + \left(\hat{V}_{en} + \hat{V}_{xc}\right) \,\middle|\, \psi_{\boldsymbol{k},n} \right\rangle = \epsilon_{\boldsymbol{k},n} \left\langle \boldsymbol{k}+\boldsymbol{G} \,\middle|\, \psi_{\boldsymbol{k},n} \right\rangle \tag{1.13}$$

or

$$\frac{1}{2}(\boldsymbol{k}+\boldsymbol{G})^2 c_{\boldsymbol{k},n}(\boldsymbol{G}) + \frac{4\pi}{\Omega}\sum_{\boldsymbol{G}'}\frac{\rho(\boldsymbol{G}-\boldsymbol{G}')}{|\boldsymbol{G}-\boldsymbol{G}'|^2}c_{\boldsymbol{k},n}(\boldsymbol{G}') + \frac{1}{\sqrt{\Omega}}\int_\Omega \mathrm{e}^{-i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}}\left(v_{en}(\boldsymbol{r})+v_{xc}(\boldsymbol{r})\right)\psi_{\boldsymbol{k},n}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = \epsilon_{\boldsymbol{k},n}c_{\boldsymbol{k},n}(\boldsymbol{G}) \quad (1.14)$$

These equations need to be solved iteratively, since some terms depend on the Bloch functions they are solved for. The reciprocal components of the electron density, for example, are calculated as

$$\rho(\boldsymbol{G}) = \int_\Omega \rho(\boldsymbol{r})\,\mathrm{e}^{-i\boldsymbol{G}\cdot\boldsymbol{r}}\,\mathrm{d}\boldsymbol{r} \quad (1.15)$$

with

$$\rho(\boldsymbol{r}) = \frac{1}{V_{BZ}}\int_{BZ}\sum_n w_{\boldsymbol{k},n}\left(\frac{1}{\Omega}\sum_{\boldsymbol{G},\boldsymbol{G}'}c^*_{\boldsymbol{k},n}(\boldsymbol{G}'-\boldsymbol{G})c_{\boldsymbol{k},n}(\boldsymbol{G}')\mathrm{e}^{i\boldsymbol{G}\cdot\boldsymbol{r}}\right)\mathrm{d}\boldsymbol{k} \quad (1.16)$$

and $w_{\boldsymbol{k},n}$ standing for the orbital occupation numbers. In their plane-wave shape, the Kohn-Sham equations therefore involve multiple integrals, which need to be solved numerically. Grids are defined in real and reciprocal space to replace the integrals over $\boldsymbol{r}$ and $\boldsymbol{k}$ by summations. The discretization of reciprocal space is referred to as k-point sampling, for which several schemes are known to yield accurate integrated values [75, 76]. The integrations over $\boldsymbol{r}$ correspond to Fourier transforms, and are calculated using FFT algorithms on a so-called FFT mesh. The inverse transformation does not require any additional approximations, as it is calculated at the already discrete $\boldsymbol{G}$-points, which only need to fulfil the condition imposed by the cutoff energy.

Pseudopotentials and the projector augmented-wave method As mentioned in the previous paragraph, the size of a plane-wave basis set is determined by the cutoff energy, which defines a maximum wave number G_{max}. As this quantity grows, the basis set becomes more accurate and is able to describe more detailed features (up to $\lambda_{min} = 2\pi/G_{max}$). Unfortunately, the computational cost increases accordingly. This is especially a problem when the electron density contains high-frequency components, which is the case near the nuclei because of the strong Coulomb potential. Accurately representing these characteristics would therefore require the use of a very large G_{max}.

In practice, the basis set requirements are often mitigated by freezing the core electrons, using a one-time single-atom calculation. Although the environment does influence the core electrons, especially their kinetic energy, the frozen-core approximation only changes the total energy up to second-order effects [78]. For the valence electrons, on the other hand, a smoother potential is introduced. This surrogate potential must reduce the required number of plane waves for the description of the wave functions near the nuclei, while ensuring that the correct wave function is obtained beyond a certain threshold radius, to accurately represent the interatomic binding (Figure 1.7).

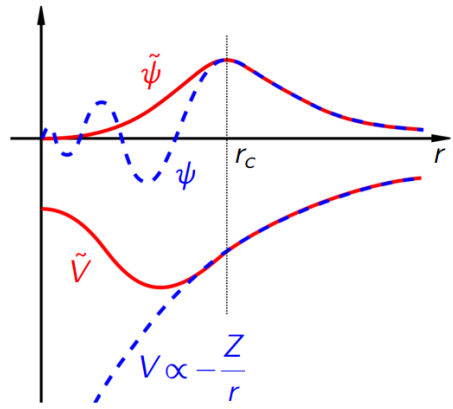


Figure 1.7: Schematic illustration of a pseudization scheme [77], in which the real potential V and wave function ψ are smoothed to $\tilde{V}$ and $\tilde{\psi}$ within a cutoff radius r_c to remove high-frequency components.

When directly incorporated into the Kohn-Sham formalism, this smooth potential is referred to as a pseudopotential [79]. In this approach, soft radial wave functions $\tilde{\psi}$ and pseudopotentials $\tilde{v}$ are first sought to fulfil the necessary conditions for single atoms. These then allow constructing a projector-based, pseudized Hamiltonian that has the pseudized wave functions as solutions (e.g. [80]). There is still much flexibility, however: pseudizing the atomic wave function at another energy, choosing another type of projector or imposing a different core-valence splitting all yield different pseudopotentials. Moreover, $\tilde{\psi}$ can be constructed using different constraints, leading to either norm-conserving [81] or ultrasoft [82] pseudopotentials. Norm-conserving pseudopotentials are defined as to maintain the core-region charge of a single atom, while this condition is relaxed for ultrasoft pseudopotentials by adding a so-called augmentation charge to the valence electrons. This additional freedom allows choosing softer potentials and wave functions, which hence require smaller basis sets.

The projector augmented-wave method (PAW) [83] was inspired by the same idea, but proceeds in a different way. It also proposes a new Hamiltonian with pseudized wave functions for eigenstates, but this $\hat{\tilde{H}}$ is defined by a transformation $\hat{\mathcal{T}}$ between the pseudized and the all-electron solutions. The self-consistent procedure therefore implicitly takes into account the full wave functions. Afterwards, the transformation $\hat{\mathcal{T}}$ allows determining the

exact wave function again. In addition, working with the full wave functions allows writing the local parts (centred around a nucleus and hence highly oscillatory) in a different basis set, the so-called partial waves. In most cases, these partial waves correspond to two or three single-atom orbitals per quantum number of the orbital momentum (i.e. two or three electronic shells are taken into account per value of ℓ). Such a basis set is much more efficient than a plane-wave one, and the PAW method is therefore in principle softer and more accurate than pure pseudopotential approaches.

The augmented plane-wave method A pseudization scheme is not the only way to reduce the required number of plane-wave basis functions. The highly oscillatory behaviour near the nuclei can be maintained by describing it with a different basis set [84]. Such an explicit treatment of all electrons makes it possible to access information in the region around the nuclei, for example. Augmented plane-wave methods (APW) [85,86] employ atom-type orbitals for this, which are located in so-called muffin tin spheres S_α around the ions α and which match plane waves in the interstitial space I:

$$\phi_{\boldsymbol{k},\boldsymbol{G}}(\boldsymbol{r},E) = \begin{cases} \frac{1}{\sqrt{\Omega}} \mathrm{e}^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}} & \boldsymbol{r} \in I \\ \sum\limits_{\ell m} A_{\ell m}^{\alpha,\boldsymbol{k}+\boldsymbol{G}}\, u_\ell^\alpha(\boldsymbol{r}-\boldsymbol{R}_\alpha,E)\, Y_\ell^m(\boldsymbol{r}-\boldsymbol{R}_\alpha) & \boldsymbol{r} \in S_\alpha \end{cases} \tag{1.17}$$

Here Y_ℓ^m are spherical harmonics and u_ℓ represent the radial solutions of the Schrödinger equation for an isolated atom. Unfortunately, the latter depend on the energy for which the Kohn-Sham equations are solved. Rather than following a cumbersome iterative procedure, a possible solution is to expand the wave functions u_ℓ around a so-called linearization energy E_0. These linearized augmented plane-wave (LAPW) basis functions are [87,88]:

$$\phi_{\boldsymbol{k},\boldsymbol{G}}(\boldsymbol{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} \mathrm{e}^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}} & \boldsymbol{r} \in I \\ \sum\limits_{\ell m} \Big[A_{\ell m}^{\alpha,\boldsymbol{k}+\boldsymbol{G}}\, u_\ell^\alpha(\boldsymbol{r}-\boldsymbol{R}_\alpha,E_{0,\ell}^\alpha) & \\ \quad + B_{\ell m}^{\alpha,\boldsymbol{k}+\boldsymbol{G}}\, \dot{u}_\ell^\alpha(\boldsymbol{r}-\boldsymbol{R}_\alpha,E_{0,\ell}^\alpha) \Big]\, Y_\ell^m(\boldsymbol{r}-\boldsymbol{R}_\alpha) & \boldsymbol{r} \in S_\alpha \end{cases} \tag{1.18}$$

and the additional parameters B are obtained by imposing a continuous derivative at the boundary of the muffin tin sphere. To also take into account multiple shells of semi-core electrons (which share the same ℓ with the valence electrons), the basis set can be extended with local orbitals (LO),

which are zero in the interstitial space:

$$\phi^{\alpha}_{LO,\ell m}(\boldsymbol{r}) = \begin{cases} 0 & \boldsymbol{r} \in I \\ \Big[A^{\alpha}_{LO,\ell m}\, u^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{0,\ell}) \\ \quad + B^{\alpha}_{LO,\ell m}\, \dot{u}^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{0,\ell}) \\ \quad + C^{\alpha}_{LO,\ell m}\, u^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{1,\ell})\Big]\, Y^{m}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}) & \boldsymbol{r} \in S_{\alpha} \end{cases} \tag{1.19}$$

The constants C follow by normalizing. Unfortunately, the use of LAPW and LO wave functions requires larger basis sets than APW methods. A speed-up of the calculations is possible by using a third kind of basis set, which combines augmented plane waves at fixed energies E_0 with two kinds of local orbitals (APW+lo) [89, 90]:

$$\phi_{\boldsymbol{k},\boldsymbol{G}}(\boldsymbol{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}} & \boldsymbol{r} \in I \\ \sum\limits_{\ell m} A^{\alpha,\boldsymbol{k}+\boldsymbol{G}}_{\ell m}\, u^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E_{0,\ell})\, Y^{m}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}) & \boldsymbol{r} \in S_{\alpha} \end{cases} \tag{1.20}$$

$$\phi^{\alpha}_{lo,\ell m}(\boldsymbol{r}) = \begin{cases} 0 & \boldsymbol{r} \in I \\ \Big[A^{\alpha}_{lo,\ell m}\, u^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{0,\ell}) \\ \quad + B^{\alpha}_{lo,\ell m}\, \dot{u}^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{0,\ell})\Big]\, Y^{m}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}) & \boldsymbol{r} \in S_{\alpha} \end{cases} \tag{1.21}$$

$$\phi^{\alpha}_{LO,\ell m}(\boldsymbol{r}) = \begin{cases} 0 & \boldsymbol{r} \in I \\ \Big[A^{\alpha}_{LO,\ell m}\, u^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{0,\ell}) \\ \quad + C^{\alpha}_{LO,\ell m}\, u^{\alpha}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}, E^{\alpha}_{1,\ell})\Big]\, Y^{m}_{\ell}(\boldsymbol{r}-\boldsymbol{R}_{\alpha}) & \boldsymbol{r} \in S_{\alpha} \end{cases} \tag{1.22}$$

The constants in the lo and LO basis functions are in that case determined by normalization and continuity at the sphere boundaries.

Because no pseudization is involved, (L)APW methods are often considered to provide a golden standard for DFT accuracy. Indeed, numerical parameters can be adjusted systematically and the converged result will not depend on atomic potentials. Nevertheless, achieving numerical convergence is not always straightforward. Increasing the number of basis functions and k points is helpful, but it is also important to take into account sufficient valence electrons and not to truncate the spherical harmonic expansion too soon. The muffin tin radii need to be examined as well, as smaller radii generally yield better results (only the deepest core states are truely atom-like), but require larger numbers of plane waves. Finally, the particular choices for the linearization energies matter as well.

1.3.3 A computational conveyer belt

The methods described above played a principal role in the success of solid-state DFT for computational materials design. They allowed the highly accu-

rate description of quantum physics by density-functional theory to be combined with fast and efficient algorithms. By the turn of the millennium, the available computing power had moreover boomed, enabling the generation of so far unimaginable quantities of first-principles predictions. It therefore became clear that a new research domain needed to be investigated: that of high-throughput computational materials design.

High-throughput studies are not new. In experimental environments, the automated search for new compounds dates back centuries. Edison's search for a suitable filament for his incandescent light bulbs, for example, lead him and his co-workers to scan 6 000 candidate materials over a period of three years (1878-1880) [91]. They ultimately selected carbonized cotton threads, although tungsten filaments have nowadays been found to be even more suitable. Edison's high-throughput investigation is one of the first and best-known examples of large-scale materials searches in an experimental context. Such methods have today become a useful tool to screen large search spaces, saving precious time and money and allowing for a shorter time to market [91]. The same advantages can be listed for computational approaches, which are even faster than experimental ones. Presently, large computational databases exist for domains varying from biophysics [92] over astronomy [93] to chemistry [94]. In materials science, high-throughput studies have already been applied to many classes of compounds and properties (see further). The success of high-throughput methods has even given rise to the Materials Genome Initiative for Global Competitiveness, a US governmental program that specifically aims to accelerate the discovery of advanced materials by computational means [95].

Although computational high-throughput studies have much in common with their experimental counterparts, the computational aspect raises new questions and causes the field to develop into an independent research domain [96, 97]. Essential is the use of an automated workflow, which poses challenges at each of the four stages of the investigation: input, calculations, storage and analysis (see Figure 1.8).

Input Even the vast computational resources that are nowadays available are not unlimited. Practical high-throughput studies are therefore restricted to finite numbers of candidates. These candidates can be chosen beforehand (intrinsic searches) or interactively over the course of the investigation (extrinsic searches) [97]. It depends on the problem at hand which method is most appropriate.

Intrinsic searches examine a predefined set of compounds, so any optimization translates into a screening of previously calculated results. It is therefore essential to start from a suitable database of input structures. This

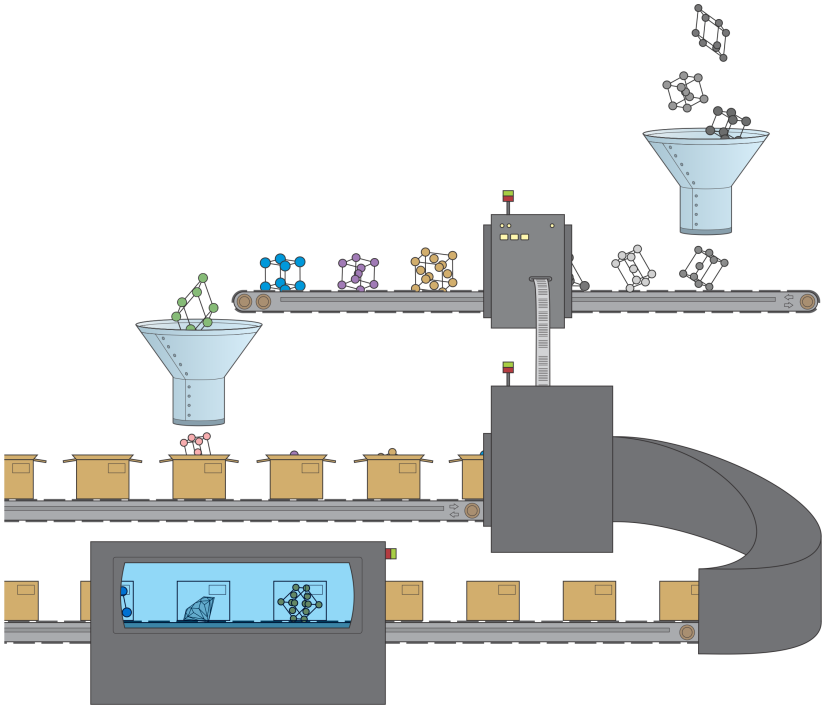


Figure 1.8: The computational conveyer belt [98]. High-throughput simulations are characterized by four (preferably automated) stages: input (upper right funnel), calculations (processing unit), data storage (boxes) and analysis (X-ray scanner). The purpose of these simulations is to identify valuable solutions (diamond) in vast search spaces.

information can be retrieved from experimental repositories (such as ICSD [99] or COD [100]), but that is not required. By selecting experimentally unknown materials, much physical insight can be obtained as well and even new discoveries may be made. Intrinsic searches are applicable to a large range of materials properties. On the other hand, they only allow studying the manually selected input structures. Unexpected solutions are hence a priori excluded.

When a particular materials property needs to be optimized and the crystal structure is not known beforehand, extrinsic searches are more appropriate. They allow screening the entire combinatorial space independently, updating the set of input materials as the investigation proceeds. Unfortunately, unless the energy is minimized, there is no guarantee that the resulting structures are also stable. Most methods are therefore still limited to structure predictions, looking for the crystals with the lowest energy for a given composition. Nevertheless, other materials properties have already

been optimized irrespective of stability, such as the band gap, density and hardness [101–103].

A random search [104] is the most straightforward extrinsic search method. It selects one arbitrary compound at a time and relaxes it until a local energy minimum is obtained. This method is often inefficient, however, as it can take quite some time until the entire search space is sampled adequately. Indeed, the many possible compositions and crystal structures generate numerous local extrema in the energy landscape, hindering the detection of the global optimum. Several methodologies have therefore been developed to direct the screening effort towards more promising areas, sampling the most interesting parts of the search space in more detail. The simulated-annealing method [105, 106], for example, performs a random walk through search space and accepts the move from one local energy minimum to another based on how well the old and the new minimum fulfil the objective f. The technique strongly resembles a Metropolis algorithm [107], and the probability of moving follows a Boltzmann-like distribution $\exp\left(-\frac{\Delta f}{T}\right)$. By decreasing the temperature-like parameter T in an appropriate way, the global optimum is eventually found. Franceschetti and Zunger applied the method to maximize the optical band gap of $Al_{0.25}Ga_{0.75}As$ [101], for example. Evolutionary algorithms [108] are another way to guide the sampling of the search space. These methods have become increasingly popular in computational materials design, especially genetic algorithms[4] [109], and several independent implementations are already available [110–114]. Evolutionary methods aim to mimic natural selection by considering an initial generation, which survives, cross-breeds and mutates based on its fitness with respect to the objectives. Such approaches have been shown to solve highly multidimensional problems like materials design very efficiently. We also mention principal-component analyses [115], which establish heuristic rules for crystal structure prediction based on previously obtained results (using partial least-squares calculations, for example [116]). The method extracts inherent dependencies between the different structures, such that the simulation of only a few structures already allows to predict the properties of a much larger structure set. Finally, other extrinsic approaches exist as well, which draw inspiration from computer sciences or physics-based considerations (e.g. [117, 118]).

Calculations Current *ab initio* techniques have become sufficiently accurate and robust to apply them to a wide range of systems without much human intervention. Several computational environments have therefore been

[4] Genetic algorithms are one particular implementation of vector-based evolutionary algorithms (as opposed to the optimization of computer programs, for example).

developed to steer calculations in a fully automated way [119–123]. Nevertheless, unsuspected errors may occur and the high-throughput framework needs to be able to deal with them. It should moreover enable complicated workflows in parallellized environments, starting both serial and parallel calculations and extracting the necessary intermediate results to direct the computational effort towards the ultimate goal. Finally, in addition to these machine-independent requirements, high-throughput environments are also confronted with more prosaic issues, such as infrastructure failures (calling for good back-up mechanisms), queues, available wall time and optimal use of allotted CPU power.

Storage Large-scale calculations produce huge amounts of data. These data obviously need to be stored somewhere, but fortunately, disk space is one of the few computational resources that comes cheap. A more pressing issue is the need to store the data in an intelligent way. While small databases can be perused quite easily, the 'big data' associated with many high-throughput studies require a more efficient way to query the database. This consideration has been addressed by several script- [123, 124] and web-based repositories [125–128]. Such initiatives are not only motivated by fast database access, however. There are many advantages to a complete documentation of the computational parameters, storing both input and output files. This information can serve as a reference for beginners in a certain field or code, and allows pinpointing methodological assumptions or even mistakes. ESTEST [129] and Quixote [130] are two tools that are designed for this very purpose. The collection of high-quality results may also prove useful for external stakeholders or serve for benchmarking purposes. The US National Institute for Standards and Technology has a long tradition in collecting such information, for example (e.g. [131]). Benchmark data not only allow assessing the computational accuracy, but can also help identifying experimental outliers [132, 133].

Analysis The creation of a large database of first-principles results does not automatically lead to useful results. It is necessary to search the generated data for valuable knowledge, be it by optimizations or by pattern recognition. To facilitate this extraction of information, several powerful tools have been developed, gathered under the common denominator 'data mining techniques'. One of their main goals is to extract hidden trends and patterns. This, for example, allows establishing so-called quantitative structure-property relationships (QSPR) that express macroscopic quantities of interest in terms of computationally accessible predictors. This is particularly useful when the macroscopic quantity is determined by a complex in-

terplay of different multiscale phenomena and is therefore hard to obtain computationally. The melting temperature can be predicted much more easily using a QSPR in terms of the cohesive energy, for example [134]. It is even not always necessary to know the QSPR explicitly. We already described simulated-annealing techniques, evolutionary algorithms and principal-component analyses as means to heuristically exploit the intrinsic correlations between different structures (see earlier). These methods can also be applied to property optimizations. More examples of QSPR-based data mining involve neural networks [135] or other machine learning techniques [136]. In the identification of patterns, visual aids can moreover provide an added value, as is proven by Pettifor structure maps [137] and Ashby charts [138].

Unfortunately, most of these methods are directed towards one-dimensional optimization problems. When only a single criterion needs to be met, it is straightforward how the quality of two different solutions should be compared. Practical design situations involve multiple requirements, however, and before any of the data mining techniques described above can be applied, it needs to be agreed on how a particular solution can be the best candidate with respect to several objectives. This question lies at the basis of an entire research area, called multicriterion decision making (MCDM) [139, 140]. Depending on whether the space of possible alternatives is discrete and finite or continuous and infinite, two branches are distinguished: multi-attribute decision analysis and multi-objective optimization problems, respectively. As this work focuses on the selection of crystal structures from a database, the techniques we will highlight here will mostly be applicable to the former research domain.

A useful concept when faced with multiple objectives is that of Pareto optimality (or Pareto efficiency). Although originally an economic theory [141], it is now also used by engineers and computer scientists, since design problems or database searches require many goals to be met at the same time as well. Whether a solution is Pareto optimal or not, depends on how it compares to the other candidates. For non-Pareto solutions, it is always possible to find an alternative that performs better in all respects, but switching between Pareto solutions requires some trade-off. For a candidate solution $\tilde{\boldsymbol{x}}$, this condition can be mathematically expressed as

$$\tilde{\boldsymbol{x}} \in \mathcal{P} \;\Leftrightarrow\; \nexists\, \boldsymbol{x} : \begin{cases} f^{(k)} \geq \tilde{f}^{(k)}\ \forall k = 1, ..., N \\ \exists\, k_0 : f^{(k_0)} > \tilde{f}^{(k_0)} \end{cases} \tag{1.23}$$

where $\mathcal{P}$ represents the set of Pareto solutions (or Pareto set in short) and $f^{(k)}$ the N maximizable objective functions. Figure 1.9 illustrates the concept for a hypothetical data set in two dimensions, with maximizable objective functions X and Y. It shows how the Pareto set creates a boundary between

feasible (yet inefficient) solutions and unattainable ones. This boundary is therefore also referred to as the Pareto frontier or, in the context of databases, the skyline [142, 143]. Indeed, in discrete cases, the Pareto frontier takes the appearance of a city skyline. More generally, the Pareto frontier forms an N-dimensional hypersurface that delimits the region of all feasible solutions.

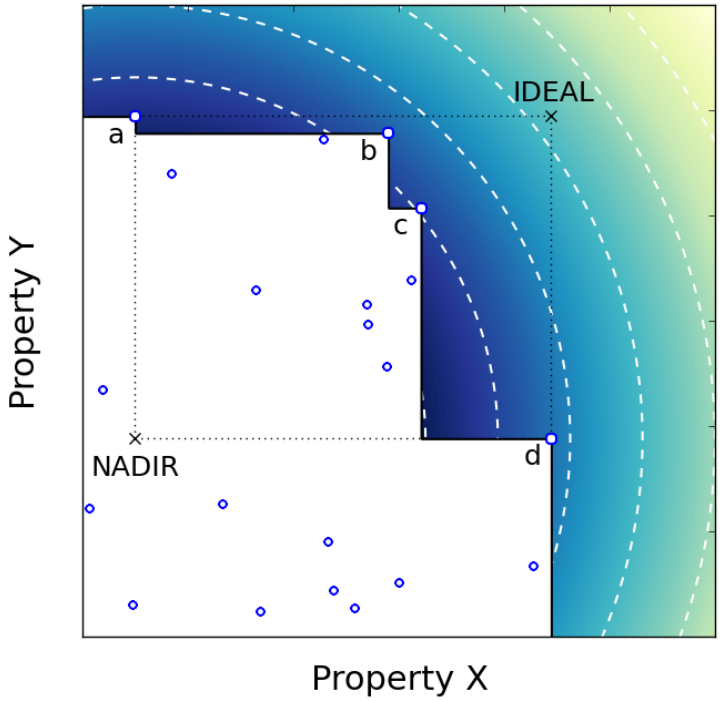


Figure 1.9: Hypothetical 2D data set with maximizable objectives X and Y. The Pareto frontier (or skyline) is indicated by a full black line and is determined by the Pareto set (a,b,c,d). Data points to the left of it (white area) are feasible (yet suboptimal), while data points to the right (coloured area) are unattainable. The Pareto set is delimited by the ideal and the nadir point, which may serve to define a value function. A set of indifference curves is shown for when the value function corresponds to the distance from the nadir point (dashed white lines).

Pareto analyses have already proven useful for many applications. They can be combined with evolutionary algorithms [144], for example, and have even been applied to computational materials design [145, 146], although the latter studies are still uncommon. Pareto optimizations only offer a limited solution, however. For large databases, the Pareto set itself becomes large as well, and additional means of refinement are needed. The decision maker is essential in these so-called 'post-Pareto methods', since any selection beyond

the Pareto set is inherently arbitrary. In such cases, it is best to formalize the used preference structure to exclude human bias. This can be done prior to the decision-making process or in an interactive way (e.g. [147]).

A non-interactive post-Pareto approach is most straightforwardly achieved by defining a value function (also called a utility function or preference function). It maps the $\mathbb{R}^N$ objective space onto a range in $\mathbb{R}$ and hence allows tackling the multidimensional optimization as a regular, one-dimensional problem. In full N-dimensional space, the contour surfaces of the value function are referred to as 'indifference surfaces', since they represent all points that are equally preferable to the decision maker. The most optimal solution then corresponds to the point where the Pareto frontier and the outer indifference surface are tangent to each other. An appropriate value function can be created in several ways, but many approaches are based on the distance to a fixed point, such as the nadir point or the ideal point [139] (see Figure 1.9). It is even possible to express the value function in terms of percentiles [148] or to iterate over different value functions to gain some independence over its exact form [149]. Other techniques are not based on value functions, and involve cluster analyses [149], extending the Pareto concept to a more stringent formalism (such as sorted-Pareto efficiency [150] or Pareto efficiency of order k [151]) or even game theory [152].

1.3.4 What about fusion?

In nuclear fusion research, high-throughput computational studies are still mainly uncharted territory. Regular simulations, on the other hand, are already widely applied to fusion materials, as they offer insights inaccessible to experiments and do not require the fabrication of the investigated compound [10, 35, 36, 153]. Indeed, the EFDA road map explicitly lists theory and modelling as an important tool to extrapolate available physics and guide the experiments of IFMIF [12]. Computational activities in fusion materials research cover two distinct areas. On the one hand, the characteristics of defect-free compounds are of interest. For iron-chromium or tungsten, for example, many materials properties are still poorly understood, such as phase stability and ductility. A second point of focus is the study of radiation damage. When high-energy neutrons impinge on a plasma-facing material, they impart most of their energy onto a so-called primary knock-on atom. This atom can be sputtered away from the surface, but it may also start an entire collision cascade. In the latter case, a multitude of defects and defect clusters are left behind, which migrate, interact and evolve over time. Modelling radiation damage is hence also important to understand the behaviour of fusion materials.

The properties of both pure and irradiated materials are determined over

several length scales. Modelling efforts may therefore provide valuable information at each of these scales. Multiscale approaches are also available in literature, although less common. We refer to an article of Fu and co-workers, for example, who used *ab initio* methods and kinetic Monte Carlo to explain the resistivity changes in electron-irradiated α-iron [32].

Much computational work has been performed with density-functional theory. It predicts several atomic-scale quantities at a much higher accuracy than interatomic potentials do, without increasing the computational load too dramatically. Typical quantities of interest are the properties of point defects and defect clusters. Their formation, migration and interaction with other defects and impurities are essential input data for mesoscale models, both in ordered [154] and disordered solids [155]. The structures of interest may range from simple vacancies or self-interstitial atoms to the interaction of helium and hydrogen with microstructural features [156]. Even an atom's displacement threshold energy (the minimum energy to knock an atom off its lattice site) provides valuable input for simulations of collision cascades [157]. This does not mean that only calculations of defected crystals possess added value, however. Phase stabilities [158] and tensile strengths [10] are instrumental in the search for more ductile tungsten alloys, for example. With a powerful computational infrastructure available, dislocations can moreover be studied as well [159]. Their formation energies, core configurations and Peierls barriers are particularly useful to provide insight into dislocation stability and mobility, two key aspects in plastic deformation. For more examples on the use of density-functional theory in modelling radiation damage, we refer to the comprehensive review of Dudarev [153].

2

Objective and research overview

For the first wall of the demonstration fusion reactor DEMO, tungsten is the main candidate (see Sections 1.1 and 1.2). This is thanks to its favourable high-temperature properties, such as a high melting temperature and a good thermal conductivity, and its resilience under neutron irradiation. On the other hand, tungsten's room-temperature brittleness complicates its conversion into components and compromises its long-term stability. Alloying with other elements might yield a more suitable compound, but it is not *a priori* clear which dopants would have such a favourable effect.

This doctoral work aims to computationally screen a large set of tungsten alloys and evaluate their applicability for first-wall purposes. Such a search reverses the traditional sense of materials design, by choosing a material as a function of some desired properties, rather than optimizing an existing one. By following a computational approach, we are moreover able to assess the effects of all possible dopants and identify the most promising candidates, which is experimentally impracticable. The resulting selection of optimal compounds may then be examined further in lab tests.

Studying experimentally unknown alloys imposes some restrictions on the computational search. It is necessary to start from atomic-scale quantities, for example, since not enough information is available to immediately proceed to larger length scales. We use DFT (Section 1.3.2) because of its fast and accurate treatment of large numbers of compounds. Afterwards, one option is to extract macroscopic results from a multiscale approach (see Section 1.3.1), but we instead opt to use QSPRs (which we will also refer to as semi-empirical relations). This enables us to determine macroscopic ma-

terials properties at a computationally low cost. Unfortunately, the lack of experimental data prevents a direct validation of our results. We therefore carefully calibrate our predictions by means of well known benchmark systems, providing us with generally applicable error estimates and proofs of principle.

The project consists of three steps, for each of which a methodology needs to be established. First, the computational framework is to be verified and validated, ensuring the agreement between theoretical and experimental results. In a second phase, several properties are calculated for a small subset of materials (the tungsten binaries), which allow developing and testing an analysis tool for the corresponding multicriterion problem (Section 1.3.3). Finally, these methods can be extended to all ternary tungsten alloys of interest, using a high-throughput workflow. The subsequent sections briefly discuss each of these three steps and explain the role of the corresponding publications, either published (Chapters 3-4) or in preparation (Chapter 5).

2.1 Validation and verification by means of the elemental crystals

Much information on the behaviour of tungsten alloys is already present at the atomic scale. Quantities like the lattice constants, energetic stability or elastic parameters also determine phenomena at larger length scales. We therefore investigated several materials properties at the *ab initio* level, using density-functional theory (DFT) (Section 1.3.2) with the GGA functional of Perdew, Burke and Ernzerhof (PBE) [52]. This approach allows investigating a wide range of alloying elements for tungsten, without requiring experimental information. Moreover, many macroscopic properties can reliably be expressed in terms of first-principles predictors. These quantitative structure-property relationships (QSPRs) skip the entire chain of multiscale modelling and directly relate macroscopic effects to atomic-scale parameters. We particularly focused on two high-temperature quantities that are important for fusion applications: the melting temperature T_m and the thermal expansion coefficient α. Indeed, even the high heat fluxes in fusion reactors should not cause a suitable first-wall material to melt, as liquid droplets might enter the plasma and endanger its stability. A low thermal expansion is also recommended, as the high temperatures could lead to significant volume changes and subsequently large thermal stresses.

Unfortunately, the predictions for T_m and α are not exact. DFT-based QSPRs involve several computational levels, and each of them adds to the deviation between theory and experiment. A first source of errors is already contained in the code that is used to generate the DFT values. As each im-

plementation solves the governing equations differently – by using different basis set types, atomic potentials or mathematical algorithms – different amounts of noise are introduced. Even when the computational settings are all but converged (using massive numbers of k points and huge cut-off energies, for example), these *numerical errors* cannot be neglected. Fortunately, most differences are small, but there are some exceptions, especially when inappropriate or inadequate (pseudo)potentials are employed.

A second level of computational uncertainties arises when extracting *ab initio* predictors from DFT energy calculations. Typical quantities are cohesive energies, equilibrium volumes or elastic constants. Their predicted values deviate from experimental results, because practical exchange-correlation functionals are only approximate guesses of the exact (yet unknown) form. The PBE functional, for example, is known to underbind crystals, yielding too large volumes and too small bulk moduli. The observed deviations from experiment are therefore inherent to the used flavour of DFT, and we will refer to them as *intrinsic errors*. They possess both a systematic component (such as the underbinding of PBE) and a part that varies unpredictably for different compounds.

Finally, the use of a QSPR constitutes a major approximation, and therefore gives rise to additional deviations from experiment. These *semi-empirical errors* are caused by the simplified relation between atomic-scale and macroscopic properties. Although a one-to-one relationship is expected to exist, it cannot be expressed analytically and only a multiscale approach is able to take into account all governing phenomena. The difference between QSPR-based values and experimental observations is due to this reduced dimensionality.

Error bars are needed to quantify each of these deviations, but contrary to experimental research, computational error bars are often absent or ill-defined. In **papers 1 and 2** (Chapter 3), we applied generally applicable and statistically sound procedures to quantify the three sources of error in DFT-based QSPRs. Using a benchmark set with 71 elemental ground-state crystals, we analysed the difference between computational results, or between PBE-based predictions and experimental values, considering several materials properties. We found the errors to increase for each computational level: they are smallest for numerical errors, moderate for intrinsic errors and largest for semi-empirical errors. Reliably predicting materials properties from first principles hence requires keeping each of these errors under control, and in particular the semi-empirical and intrinsic ones.

Numerical errors were expressed in terms of a newly proposed Δ gauge, which defines the difference between two implementations by their equations of state (Table 2.1). As one of the first tools to compare numerical errors, it has already received considerable attention, especially by developers

of codes and (pseudo)potentials. A community-wide effort is presently being set up to characterize as many solid-state DFT codes and potentials as possible, and to eliminate unacceptable differences. The current state of this initiative can be followed on our website [160]. Intrinsic and semi-empirical errors, on the other hand, were defined in terms of systematic deviations, residual error bars and outliers. This yielded quantifiable measures of the errors from PBE and several QSPRs (Table 2.1), and also allowed to correct for these errors. When predicting macroscopic properties, we moreover compared the results of QSPRs with entirely first-principles results from literature, which used *ab initio* molecular dynamics for the melting temperature or quasiharmonic phonon theory for the thermal expansion. In the latter case, we found some QSPR approaches to even outperform the more theoretical (and computationally more intensive) phonon calculations, which are gener-

Table 2.1: Overview of the errors quantified in papers 1 and 2 (Chapter 3). We distinguish between systematic errors ($Y = \beta_1 X + \beta_0$, with X representing PBE values and Y experimental ones) and residual error bars. Specific classes of materials that cannot be predicted with PBE or a QSPR are mentioned in the final column (N/A). Symbols refer to compounds with dominant correlation (C), London dispersion (D) or a low coordination number (L).

	β_1	β_0	error bar	N/A
Numerical errors (paper 1)				
$\Delta(\text{PAW})_{\text{VASP}}$ [kJ/mol]	–	–	0.19	–
$\Delta(\text{PAW})_{\text{GPAW}}$ [kJ/mol]	–	–	0.32	–
Intrinsic errors (paper 1)				
ΔE_{coh} [kJ/mol]	1.000	–	30	CD
V_0 [Å^3/atom]	0.964	–	1.1	CD
B_0 [GPa]	1.049	–	15	D
B_1 [–]	0.952	–	0.7	L
C_{ij} [GPa]	1.020	–	23	–
Semi-empirical errors (paper 2)				
α_V^S [10^{-5}/K]	0.852	-0.6	1.7	DL
α_V^{DM} [10^{-5}/K]	1.056	-0.7	1.8	DL
α_V^{VZ} [10^{-5}/K]	1.354	-0.8	2.1	DL
α_V^G [10^{-5}/K]	0.894	-0.8	1.9	D
α_V^{Ts} [10^{-5}/K]	1.324	-1.4	2.4	CD
$T_m(B_0)$ [K]	1.000	0	466	–
T_m^G [K]	1.042	8	283	–

ally considered to yield the best obtainable results within DFT. Specialized GGA functionals may offer more accurate phonon results, but substantial improvements are only possible by going beyond GGA.

2.2 Multicriterion optimization by means of the tungsten binaries

The quantification of electronic-structure errors is only the first prerequisite for the computational design of tungsten alloys. Once we have a good idea about the attainable degree of accuracy, we also need some means of comparing the large numbers of results. Suitable fusion materials should fulfil several objectives, such as an acceptable ductility and a high thermal resistance, so the computational study needs to be succeeded by some kind of multicriterion decision making (MCDM) (see Section 1.3.3, 'Analysis'). Pareto optimizations are particularly useful, but they are not able to reduce the entire database to a single (or a few) compound(s). Most available post-Pareto methods suffer from the same issue, and the few approaches that do not, are equally applicable to non-Pareto compounds, thus failing to adequately represent the special status of Pareto optimality (e.g., value functions).

In **paper 3** (Chapter 4), we selected a small test set of tungsten alloys to investigate more suitable MCDM approaches. This test set contained all tungsten binaries with four different dopant levels (50 %, 25 %, 12.5 % and 6.25 %), for which several materials properties were computed (such as volume, bulk modulus and shear modulus). Regular Pareto optimizations indeed yielded a considerable number of solutions, preventing an easy selection of the most optimal compounds. We therefore defined our own post-Pareto criterion, the minimum win fraction or mwf. This novel criterion ranks the Pareto solutions based on the trade-off between their properties and allows to select those few compounds which outperform the rest (see Figure 2.1). To demonstrate its applicability and intuitive character, we first applied the mwf to a few classic materials problems, such as the simultaneous optimization of mass density and price (for large-weight applications) and that of thermal resistance, hardness and price (for high-impact purposes). In the former case, iron, lead and tungsten were not unexpectedly predicted as the most suitable candidates, while for the latter mwf study, the performance of tungsten carbide was highlighted, a commonly used material in drill bits. After this initial validation, we applied our post-Pareto analysis to a fusion-based design problem, in which we optimized the Cauchy pressure (ductility), cohesive energy (thermal stability) and price. Here, W-Ta and W-V were ranked highest among our database of binary tungsten alloys. However, experimental data suggest these compounds to be even more brittle than pure tungsten.

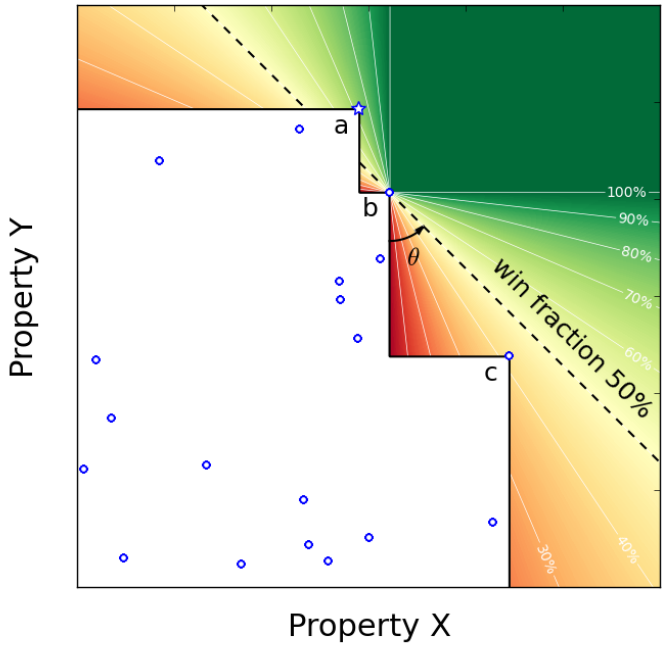


Figure 2.1: Win fraction with respect to Pareto solution *b*, ranging from 0 % (red, worst tradeoff) over 50 % ($\theta = 45°$, all solutions equally preferred) to 100 % (green, best tradeoff). The minimum win fraction compares a Pareto solution to all other ones and takes the minimum of all win fractions. For the given Pareto front, *a* outranks *b* and *b* outranks *c*. *a* hence possesses the largest mwf and *c* the lowest.

We conclude that despite the frequent use of the Cauchy pressure as a predictor for ductility, it is not able to describe the plastic behaviour of tungsten and its alloys. In all other respects, the developed methodology appears to work perfectly. It is even more generally applicable, since the mwf may prove useful for any design problem where multiple competing criteria have to be optimized.

2.3 High-throughput computations by means of the tungsten ternaries

The tungsten binaries only serve as a proof of principle for our post-Pareto ranking, as most binaries have already been investigated for fusion purposes. For plasma-facing components, the added value of computational materials

design lies in the treatment of more complex systems. The combinatorial space of tungsten alloys is infinite, however, so we require a relevant yet manageable subset to screen for fusion-based objectives. A high-throughput methodology also needs to be constructed to perform calculations in an automated way. For these large numbers of compounds, any (significant) human intervention is unpractical.

Chapter 5 establishes such a high-throughput approach for the analysis of the tungsten ternaries. Even for this subset of materials, an exhaustive characterization exceeds realistic computational resources, so we limited ourselves to one composition ($W_{30}XY$) and 20 possible dopants (the 4d and 5d transition metals). These 210 candidate compounds were subjected to a predefined workflow, using Python-based packages for the manipulation and parallellization on high-performance computing clusters [119, 161]. First, we assessed the relative positions of the two dopant atoms, corresponding to six possible 32-atom unit cells. Then, the most stable configuration was further investigated, extracting information on the equation of state, elastic constants, band structure and density of states.

These results led to a number of interesting trends, in which the d electrons play an essential role. Indeed, transition metals are to a large extent determined by the occupation of the d band, where the first five electrons occupy bonding states and the last five are assigned to antibonding ones. Compounds with half-filled d bands are therefore optimal, displaying a maximum bond order. It causes the cohesive energy to reach a maximum when the average number of d electrons is 5, preferably when the energy difference between the d bands of the alloying elements, and hence the associated charge transfer, is small as well. It also gives rise to small volumes and large bulk moduli, although slightly shifted towards the end of the d block due to atomic-size effects. Even formation energies are more favourable, with the most stable compounds when the dopants have a tungsten-like electron configuration and a bcc ground state. The total of these effects causes tungsten ternaries with half-filled d bands to take the most prominent places in our mwf-based Pareto analyses.

When simultaneously optimizing the thermal expansion, ductility and price of $W_{30}XY$ compounds, the combination of dopants from the outer ends of d block yielded the highest-ranked solutions, no matter what ductility criterion was used. The preference over mid-d-block elements, which also possess 5 d electrons on average, is due to their lower price. A particularly promising material is $W_{30}ZrAg$, which was predicted to significantly reduce the brittleness of pure tungsten and which ranked among the three best Pareto solutions for three out of four considered ductility predictors. Further investigations will be necessary to assess whether the accuracy of our ductility criteria suffices.

3

Validation and verification of first-principles predictions

PAPER 1

"Error estimates for solid-state density-functional theory predictions: an overview by means of the ground-state elemental crystals"

K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier

Crit. Rev. Solid State **39**, 1-24 (2014)

Critical Reviews in Solid State and Materials Sciences, 39:1–24, 2014
Published with license by Taylor & Francis
ISSN: 1040-8436 print / 1547-6561 online
DOI: 10.1080/10408436.2013.772503

Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals

K. Lejaeghere,[1] V. Van Speybroeck,[1] G. Van Oost,[2] and S. Cottenier[1,3,*]

[1]*Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*
[2]*Department of Applied Physics, Ghent University, Sint-Pietersnieuwstraat 41, BE-9000 Ghent, Belgium*
[3]*Department of Materials Science and Engineering, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*

Predictions of observable properties by density-functional theory calculations (DFT) are used increasingly often by experimental condensed-matter physicists and materials engineers as data. These predictions are used to analyze recent measurements, or to plan future experiments in a rational way. Increasingly more experimental scientists in these fields therefore face the natural question: what is the expected error for such a first-principles prediction? Information and experience about this question is implicitly available in the computational community, scattered over two decades of literature. The present review aims to summarize and quantify this implicit knowledge. This eventually leads to a practical protocol that allows any scientist—experimental or theoretical—to determine justifiable error estimates for many basic property predictions, without having to perform additional DFT calculations.

A central role is played by a large and diverse test set of crystalline solids, containing all ground-state elemental crystals (except most lanthanides). For several properties of each crystal, the difference between DFT results and experimental values is assessed. We discuss trends in these deviations and review explanations suggested in the literature.

A prerequisite for such an error analysis is that different implementations of the same first-principles formalism provide the same predictions. Therefore, the reproducibility of predictions across several mainstream methods and codes is discussed too. A quality factor Δ expresses the spread in predictions from two distinct DFT implementations by a single number. To compare the PAW method to the highly accurate APW+lo approach, a code assessment of VASP and GPAW (PAW) with respect to WIEN2k (APW+lo) yields Δ-values of 1.9 and 3.3 meV/atom, respectively. In both cases the PAW potentials recommended by the respective codes have been used. These differences are an order of magnitude smaller than the typical difference with experiment, and therefore predictions by APW+lo and PAW are for practical purposes identical.

Keywords error estimate, density-functional theory, benchmarking, code comparison

Table of Contents

*E-mail: Stefaan.Cottenier@UGent.be

1. INTRODUCTION

Density-functional theory (DFT)[1,2] remains one of the most popular methods to treat both model systems and realistic materials in a quantum mechanical way.[3–8] In condensed-matter physics, DFT is not only used to understand the observed behavior of solids, but increasingly more to predict characteristics of compounds that have not yet been determined experimentally (for example in Ma et al.,[9] Hautier et al.,[10] and Jain et al.[11]). In both cases, the first-principles results provide a point of reference, either to analyze data from measurements or to plan future experiments. It is therefore essential to have a quantitative idea of the expected deviation between a DFT prediction of a certain property and the corresponding experimental value. Error estimates are routinely provided in experimental physics, but in DFT applications this is a much less common practice. When confronted with a disagreement between theory and experiment, one usually resorts to higher-order levels of theory instead.[12–14]

As such, DFT is an exact reformulation of quantum physics, and does not involve any approximation. From a purely theoretical point of view, it should lead to exact predictions, with no need for an error estimate. In practice, however, one requires an educated guess for an essential ingredient of DFT: the exchange-correlation functional (hereafter referred to as "functional"). Apart from this main approximation, there are some other features that go beyond DFT in the way it is usually applied, such as the failure of the Born-Oppenheimer approximation[15,16] or high-Z radiative and other corrections from quantum electrodynamics.[17–19] They affect the results to some extent as well, but they are generally much less important than the particular choice of the exchange-correlation functional. For any of these choices, DFT predictions will not agree perfectly with experimental observations. Deviations of this kind will be referred to as the *intrinsic error* for a particular functional.

Regardless of the difference with experiment, however, all predictions should be independent of how the DFT (Kohn-Sham) equations are solved numerically: each of the many available DFT implementations should give identical results for the same functional. In reality, there will be some scatter in the predictions of different codes,[20,21] as each of them introduces a distinct amount of numerical noise. This second source of fluctuations leads to a *numerical error*. It is therefore also important to assess to what extent predicted properties vary among DFT approaches.

The goal of this article is to quantify the knowledge about these two kinds of DFT errors—intrinsic and numerical—and, where relevant, to review the physics behind them.

A way to obtain insight into computational errors is by means of benchmark studies, examining the performance of different implementations and functionals for a large set of materials and properties. For atomic and molecular benchmark sets the intrinsic errors have already been assessed in great detail, often with the aim of selecting the best functional for a particular property.[20,22–26] Similar studies exist for solid-state DFT as well,[27–42] but they are mostly limited to a small number of properties and/or compounds. In addition, their focus is often on understanding the differences between functionals, so they do not lead to quantitative and universally applicable error estimates with respect to experiment. A benchmark that is really comprehensive

should meet two criteria: the number of elements that is included in the test set should be sufficiently large and the crystal structures in the set should be sufficiently diverse. This guarantees the transferability of the benchmark conclusions with respect to both the intrinsic and numerical errors. A natural choice to construct such an extensive test set emerges from the periodic table of elements. By taking the ground-state crystal structures of all elements,[43] the two criteria for a comprehensive solid-state benchmark set are simultaneously fulfilled. All elements are included, thus trivially fulfilling the first requirement. In addition, the corresponding crystal structures range from simple hexagonal and cubic configurations to low-symmetry geometries, like orthorhombic and monoclinic cells. Of course, extrapolating the obtained insights to more complex materials, such as multicomponent compounds, requires some care. It is not impossible, however, as will be shown, for example, in Section 3.2.

An additional advantage of using all elemental crystals follows from the periodic table's inherent ability to display trends and correlations. The systematic behavior of observable quantities along periods or groups is well known, but the deviations between DFT predictions and experimental values also appear to follow such trends (Section 3). In particular, the largest errors are restricted to distinct regions. So apart from providing a complete test set, the classification of elements in the periodic table allows for an easy visualization and interpretation of the data. Furthermore, the elemental materials are among the best known and most studied materials on Earth. Experimental data collections are, hence, rather easy to find, and one can assume with sufficient confidence that the reported data are accurate.

By means of the ground-state elemental crystals, the present review offers an overview of the power and limitations of solid-state DFT calculations. Although this knowledge is often implicitly available in the computational community, most of it is scattered over two decades of literature. The current work therefore summarizes and quantifies this implicit knowledge into a practical protocol, that will allow any scientist—experimental or theoretical—to provide justifiable error estimates for many basic property predictions. Intrinsic errors follow from a statistical analysis of the deviations between DFT predictions and experimental values for a given functional. Subsets of materials for which the deviations are particularly large are identified, and the reasons for this behavior are discussed. Numerical errors, on the other hand, express to what extent two independent DFT approaches produce identical predictions. We will focus on the correspondence between two particular methods, APW+lo and PAW, by means of representative mainstream codes. For the PAW method, the use of different atomic potentials can have large effects, but by considering only the sets of potentials recommended by each code, we hope to establish a general idea of the PAW error.

Instead of performing an extensive level-of-theory study for various functionals, we will focus on one typical example within the generalized gradient approximation of DFT (GGA). For this, the PBE functional is chosen, because it is known to yield good results for solids of a wide range of elements and properties.[39] Moreover, its popularity[44] guarantees the comparibility and applicability of our results. Other GGA functionals are expected to display approximately the same behavior, except maybe for very specific material classes. Kurth et al.,[27] for instance, showed how four GGA functionals provide similar trends for both the equilibrium volume and the bulk modulus. Of course, the determination of the error estimates is not limited to GGAs. The presented methodology is also applicable to other functionals (such as LDA or hybrid functionals) or first-principles approaches (such as Hartree-Fock, GW or RPA[45]).

This article is organized as follows. Section 2 describes the computational procedure for all properties under consideration and discusses the prerequisites for a sound comparison between theory and experiment. Within Section 3 the differences between DFT-GGA predictions and experimental values are assessed (intrinsic errors), whereas Section 4 focuses on the method and code dependence of the theoretically determined properties (numerical errors).

2. PREDICTING EXPERIMENTAL PROPERTIES BY MEANS OF DFT

2.1 Computational Recipes

DFT computations for five distinct sets of materials properties will be discussed. They can be divided into energetic (ΔE_{coh}) and elastic quantities (V_0, B_0, B_1, C_{ij}).[46] Of course many more properties may be determined by means of DFT, but the quantities introduced here are directly available from straightforward total energy calculations.

The **cohesive energy** or atomization energy ΔE_{coh} is a popular benchmark quantity.[20,22,27–29,32,33,39] Expressed as an energy difference per atom, it is defined as

$$\Delta E_{coh} = -(E_0 - E_{at}). \quad [1]$$

Here, E_0 represents the energy per atom of the compound under investigation in its ground state, i.e., at 0 K and without external stress. One can determine it by means of a standard pressure optimization procedure, or by fitting a few $E(V)$ data points to an empirical equation of state (EOS) and extracting the equilibrium energy analytically. In this article, the latter option was chosen, using a common third-order Birch-Murnaghan relation:[47]

$$E(V) = E_0 + \frac{9V_0B_0}{16}\left\{\left[\left(\frac{V_0}{V}\right)^{2/3} - 1\right]^3 B_1 + \left[\left(\frac{V_0}{V}\right)^{2/3} - 1\right]^2\left[6 - 4\left(\frac{V_0}{V}\right)^{2/3}\right]\right\}. \quad [2]$$

V_0 represents the equilibrium volume, B_0 the bulk modulus, and B_1 its pressure derivative. Other equations of state exist as well, but no significant difference with respect to ground-state properties is to be expected.

E_{at}, on the other hand, is the energy of one isolated atom in its electronic ground state. Since many solid-state DFT codes only allow for the use of periodic boundary conditions, the isolated atom needs to be calculated in a periodic unit cell as well. In the present computations, a free atom is placed in a big orthorhombic unit cell, such that every atom is surrounded by at least 15 Å of vacuum. In this way, one can sufficiently suppress spurious interactions between periodic images (<1 meV). The orthorhombic symmetry is chosen over, e.g., a simpler cubic one, to avoid physically incorrect spherical states. After all, the use of a unit cell forces the electron density to assume the same symmetry as the lattice.[48] This is, in most cases, only possible by means of partial occupation of the different electron orbitals, which is not physical. Lowering the crystal symmetry counteracts this phenomenon and should lead to strictly integer occupation numbers. However, some atoms end up with partially filled states, even when this approach is applied. In such cases, the occupation numbers have to be fixed manually before looking for the usual, self-consistent solution. In this article, only experimental ground-state electron configurations are used: even when DFT predicts a different configuration to be more stable[49] (e.g., for W), the experimental occupation numbers are taken in order to guarantee a meaningful comparison to measurements. Only for spin-orbit coupled calculations for the Pb atom is it not possible to impose the experimental electronic state. The PBE ground state $^1S_0$ is therefore used, and not the experimental $^3P_0$ state.

The negative sign in Equation (1) causes positive cohesive energies to correspond to stable phases (with respect to atomic decohesion). The other sign convention, however, is commonly used as well.

One of the key physical properties of a given compound is its **volume**. In first-principles calculations the equilibrium volume per atom V_0 can be obtained easily. One either employs an optimization routine or fits some $E(V)$ points to an empirical equation of state. This is similar to the procedure used to determine E_0, and again the latter option is chosen in this work.

The **bulk modulus** is closely related to the $E(V)$ behavior as well. It is proportional to the curvature of the equation of state at the equilibrium volume:

$$B_0 = -V\left.\frac{\partial P}{\partial V}\right|_{V=V_0} = V\left.\frac{\partial^2 E}{\partial V^2}\right|_{V=V_0}. \quad [3]$$

It represents the resistance of the unloaded material to volume change, and hence to uniform pressure. Because it is linked to the curvature of the $E(V)$ relation, B_0 is a numerically sensitive quantity. A small deviation at a few data points is already able to change its value noticeably, especially when the bulk modulus is small (shallow EOS). This is increasingly so when only a narrow volume range is inspected.

B_1 stands for the **derivative of the bulk modulus** with respect to pressure, evaluated at the equilibrium volume:

$$B_1 = \left.\frac{\partial B}{\partial P}\right|_{V=V_0} = \frac{\partial}{\partial P}\left(V\frac{\partial^2 E}{\partial V^2}\right)\Bigg|_{V=V_0}. \quad [4]$$

It is a third-order derivative of the energy and hence describes effects that are one order higher even than the bulk modulus. It is related to the volume-dependence of the $E(V)$ curvature. B_1 is therefore the most sensitive elastic quantity discussed in this study. Again, both the bulk modulus and its pressure derivative are obtained from fitting an EOS to calculated $E(V)$ data points.

The mechanical behavior of a crystal cannot be described solely by means of the bulk modulus. When anisotropic deformations are applied, other **elastic constants** come into play as well. The full set of these constants makes up the stiffness matrix $\boldsymbol{C}$. It represents a tensor of rank 2 and relates (small) cell strains to the corresponding stresses via Hooke's law $\boldsymbol{\sigma} = \boldsymbol{C} \cdot \boldsymbol{\epsilon}$. C is a symmetric 6×6 matrix, containing 21 independent constants at the most. In the case of hexagonal crystals, five distinct values remain (C_{11}, C_{12}, C_{33}, C_{13}, and C_{44}), while for cubic compounds there are only three (C_{11}, C_{12}, and C_{44}). The C_{ij} parameters can also be translated into more general elastic moduli, such as Young's modulus E, the shear modulus G, and Poisson's ratio ν. Even the bulk modulus can be obtained from a simple combination of the C_{ij}. In addition, the elastic constants are known to relate to structural stability and various other important physical properties.[50, 51]

Several methods are available to obtain the elastic constants from first principles, either by relating energy and strain[52] or stress and strain.[53, 54] In most cases, a stress-based procedure is preferred, because it is inherently faster. However, it requires an *ab initio* code that can determine the stress tensor. In a first step the cell pressure components are then extracted for a minimal set of deformed geometries. Together with the corresponding strains, this results in a system of linear equations. Solving that system yields the required elastic constants. When it is important to obtain an accurate value of C_{ij}, one should construct an overdetermined system, by applying the same strain sets at different magnitudes. The elastic constants can then be retrieved by using a least-squares method.

2.2. Comparing Theory and Experiment

When a DFT prediction is compared to a number from experiment, the corresponding ambient conditions should be as identical as possible. This means that, in the first place, the experimental result should refer to 0 K. Moreover, the measurement should be corrected for zero-point vibrational effects, which are not present in standard DFT calculations. The following paragraphs discuss how to extrapolate the experimental values to absolute zero and correct them for zero-point vibrations.

For the **cohesive energy** it takes little effort to match up theory and experiment consistently. Experimental data at low temperatures are, in most cases, available. Only the zero-point

energy ζ hinders a direct comparison between 0 K and experiment. From quantum mechanics this quantity is known to be $\frac{3}{2}\hbar\langle\omega\rangle$, with $\langle\omega\rangle$ the average phonon frequency. The latter can be estimated from Debye theory, where it is proportional to the maximum vibrational frequency, and hence to the Debye temperature Θ_D. The zero-point energy correction becomes[55]

$$\zeta = \frac{9}{8}k_B\Theta_D. \qquad [5]$$

Theoretical cohesive energies can only be compared to experiment if this contribution is added to the experimental values (added, due to the chosen sign convention in Equation (1)).

When no experimental value is available, Θ_D can be estimated. Here, the Debye-Grüneisen approximation[56]

$$\Theta_D = 0.617\frac{\hbar}{k_B}(6\pi^2)^{1/3}V_0^{1/6}\left(\frac{B_0}{M}\right)^{1/2} \qquad [6]$$

will be used. Both V_0 and the mass M are expressed per particle, corresponding to a single atom for most materials. For dimeric crystals, however, the diatomic molecule is chosen as a unit of repetition. The regular, room temperature experimental values for B_0 and V_0 are filled in, except when the difference with low temperature results (see further) is significant. This is the case for H, Cl, Br, and I.

Thermal **volume** corrections consist of two parts. Assuming to have a room temperature measurement at one's disposal, the first step consists in accounting for thermal expansion from absolute zero to ambient temperature:

$$\frac{\Delta V^{(1)}}{V} = \int_0^{T_{rt}} \alpha_V(T)\mathrm{d}T; \qquad [7]$$

$\alpha_V(T)$ represents the temperature-dependent volume expansion coefficient. It is zero at 0 K and $\alpha_{V,rt}$ at room temperature (T_{rt}). Since $\Delta V^{(1)}$ constitutes only a small correction with respect to the total volume V, Equation (7) will be approximated here as

$$\frac{\Delta V^{(1)}}{V} \approx \int_0^{T_{rt}} \alpha_{V,rt}\frac{T}{T_{rt}}\mathrm{d}T = \frac{\alpha_{V,rt}T_{rt}}{2}. \qquad [8]$$

In a limited number of cases the experimental expansion coefficient is not known. It can then be estimated from an empirical correlation to the "moleculization" energy:[57]

$$\alpha_{V,rt} = 3 \times \frac{48.14\cdot 10^{-6}\mathrm{eV/K/atom}}{\Delta E_{mol}}. \qquad [9]$$

ΔE_{mol} is defined as the energy difference per atom between the crystalline material and its gas-like molecules. For elements with an atomic gas phase, it reduces to the atomization energy (cohesive energy). In the absence of experimental data on both $\alpha_{V,rt}$ and ΔE_{mol}, Equation (9) is completed with DFT values.

A second modification is again due to zero-point effects. Because of the volume-dependence of the zero-point energy ζ, the equilibrium volume is shifted slightly. According to Alchagirov et al.,[37,55] this small difference per atom amounts to

$$\Delta V^{(2)} = \frac{(B_1-1)\zeta}{2B_0} = \frac{9}{16}(B_1-1)\frac{k_B\Theta_D}{B_0}. \qquad [10]$$

Dacorogna and Cohen[58] proposed an alternative definition of the zero-point volume shift. They obtain a similar formula, but with B_1 instead of B_1-1 in Equation (10). However, the mathematical expression is preceded by some significant simplifications. When calculating zero-point effects it is therefore advisable to use Equation (10) instead, especially when B_1 is small.

For the **bulk modulus** thermal effects should be taken into account as well. A first contribution originates in the thermal expansion of the material. Similar to $\Delta V^{(1)}$, a correction $\Delta B^{(1)}$ can be determined too. Roughly approximating the relevant behavior, one can write:[59]

$$\Delta B^{(1)} = B_1\cdot P(\Delta V^{(1)}) = -B_0B_1\frac{\Delta V^{(1)}}{V_0}. \qquad [11]$$

On the other hand, the effect of zero-point vibrations on the bulk modulus boils down to[55]

$$\frac{\Delta B^{(2)}}{B_0} = -\frac{\Delta V^{(2)}}{V_0}\left[\frac{1}{2}(B_1-1) + \frac{2}{B_1-1}\left(\frac{2}{9}-\frac{1}{3}B_1-\frac{1}{2}B_0B_2\right)\right], \qquad [12]$$

where B_2 stands for the second-order derivative of the bulk modulus with respect to pressure. It is a highly sensitive parameter and very difficult to extract from a few $E(V)$ data points. In addition, B_2 is not included in Equation (2) and a higher-order Birch-Murnaghan fit should be applied. Instead, the present work will use the intrinsic Birch-Murnaghan value:

$$(B_0B_2)^{BM} = B_0\frac{\partial^2}{\partial P^2}\left(V\frac{\partial^2 E^{BM}}{\partial V^2}\right)\Big|_{V=V_0} = -\frac{143}{9}+7B_1-B_1^2. \qquad [13]$$

There are other possibilities as well,[55] ranging from a different equation of state to an accurate numerical determination of B_2. In order to establish the small correction $\Delta B^{(2)}$, however, this more consistent approach suffices.

Since it is already hard to accurately measure a high-order parameter like B_1 or to determine it from first principles, zero temperature modifications will often be negligible compared to experimental or computational errors. B_1 is therefore not adjusted to incorporate thermal expansion or zero-point effects.

No thermal corrections are applied to the elastic constants C_{ij} either. One can, however, imagine a modification similar to that of Dacorogna and Cohen[58] for the bulk modulus:

$$\Delta C_{ij}^{(m)} = \frac{\partial C_{ij}}{\partial P}\cdot P(\Delta V^{(m)}) = -B_0\frac{\partial C_{ij}}{\partial P}\frac{\Delta V^{(m)}}{V_0} \qquad [14]$$

with $m = 1$ to account for thermal expansion and $m = 2$ for zero-point effects. Unfortunately, experimental data about the pressure derivative of the elastic constants are scarce.

3. INTRINSIC ERRORS

3.1. Test Set Preparation

In order to establish statistically justified intrinsic error estimates, the ground-state elemental crystals at 0 K will be used as a benchmark set. Pettifor[60] lists these crystal structures, based on an overview by Villars and Daams.[43] However, in some cases literature suggests another phase to be even more stable at low temperatures. In order to ensure the use of 0 K cell geometries as much as possible, such an alternate structure is taken for boron,[61] nitrogen,[62] oxygen,[63] and sulfur.[64] Table 1 presents an overview of all structures used in the current test set.

Using a 0 K benchmark set entails two distinct advantages. On the one hand, some elements only crystallize just above absolute zero. Collecting both 300 K and 0 K compounds in one set might then seem a bit inconsistent. On the other hand, this approach facilitates the extrapolation from the experimental temperature to 0 K, as there are no phase transformations along the way.

All structures are considered in their stress-free ground state. This means that, when the space group allows some freedom in the internal positions, an optimization with respect to the total energy is necessary. This optimization procedure calls for a fast and well-accepted DFT algorithm. The projector augmented-wave method[65,66] (PAW) as implemented in VASP[8,67] (version 5.2.2) fulfills both criteria.

The elemental crystal structures have therefore been relaxed by means of this code, using the recommended PAW atomic potentials listed in the manual.[68] A force convergence criterion of 0.01 eV/Å was set. All calculations were performed using the tetrahedron method with Blöchl corrections,[69] while the reciprocal space was sampled by means of a Monkhorst-Pack grid.[70] Further computational details for the calculations are given in the online Supplementary Material, posted on the publisher's website.

The equilibrium structure has been obtained in two stages. For the determination of the equilibrium volume a uniformly spaced 13-point EOS (up to $V_0 \pm 6\%$) was calculated and fitted to a least-squares third-order Birch-Murnaghan relation (see Section 2.1). Only for a number of shallow $E(V)$ curves—in particular for H, N, S, the halogens, and the noble gases—an increased volume range turned out to be necessary. For each of the 13 crystal volumes, the atomic positions and the cell shape have been individually optimized. In a second step, the crystal has been reinitialized at the fitted V_0 and has then been optimized again.

These optimized crystal structures form the definitive test set (submitted to the COD[71] and ICSD[72] crystallographic databases). For each of them, most of the properties discussed in Section 2.1 have been determined in order to quantify the difference between PBE and experimental values (Section 3.2.3). The DFT part of the comparison has been performed by means of

TABLE 1
Ground-state crystal structures for all elements up to radon. Both the space group number and the Pearson notation are given (with hR*x* standing for *x* atoms in the *hexagonal* setting of the rhombohedral unit cell)

H																	He
194																	194
hP4																	hP2
Li	Be											B	C	N	O	F	Ne
166	194											166	194	205	12	15	225
hR9	hP2											hR36	hP4	cP8	mS4	mS8	cF4
Na	Mg											Al	Si	P	S	Cl	Ar
166	194											225	227	64	70	64	225
hR9	hP2											cF4	cF8	oS8	oF128	oS8	cF4
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
229	225	194	194	229	229	217	229	194	225	225	194	64	227	166	152	64	225
cI2	cF4	hP2	hP2	cI2	cI2	cI58	cI2	hP2	cF4	cF4	hP2	oS8	cF8	hR6	hP3	oS8	cF4
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
229	225	194	194	229	229	194	194	225	225	225	194	139	227	166	152	64	225
cI2	cF4	hP2	hP2	cI2	cI2	hP2	hP2	cF4	cF4	cF4	hP2	tI2	cF8	hR6	hP3	oS8	cF4
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
229	229	194	194	229	229	194	194	225	225	225	139	194	225	166	221		225
cI2	cI2	hP2	hP2	cI2	cI2	hP2	hP2	cF4	cF4	cF4	tI2	hP2	cF4	hR6	cP1		cF4

VASP, using the settings mentioned earlier. They allow for the convergence of all energy differences up to a few meV per atom at the most. For O and Cr (antiferromagnetic), Mn (ferrimagnetic), Fe, Co, and Ni (ferromagnetic), spin polarization has been taken into account, while for the heaviest elements (as from Lu) spin-orbit contributions have been incorporated. At that point relativistic effects beyond the scalar-relativistic approach become important, as will be shown later (Table 7).

The analysis in Section 3.2.3 will not show the raw calculated data, but will rather elaborate on the deviation between theory and experiment. The first-principles results and the thermally corrected experimental numbers[40,43,46,62,73–105] have been included in the Supplementary Material. A tabulation of calculated and experimental values for the elastic constants C_{ij} was published before by Shang et al.[40] for most of the present benchmark set. In Section 3.2.3 their data are used. Only for the experimental numbers of Ba did we find more realistic results elsewhere.[98,99] Since the authors only considered bcc, fcc, and hcp structures, this implies that for Li and Na a different geometry was applied than in the rest of this work (bcc instead of hR9). Moreover their results are based on a PW91 functional,[106,107] rather than the PBE approximation employed in the rest of this work. Although these GGA approaches yield different results in a few situations,[108] they are in most cases very similar and for the elastic constants no significant deviation is expected.

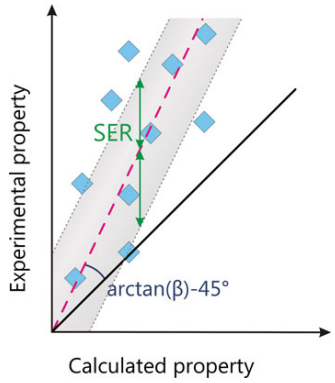


FIG. 1. The intrinsic error between experiment and DFT can be decomposed in a systematic deviation and a residual error bar. Systematic differences are quantified by comparing the linear regression line $X = \beta T$ (red, dashed) to the bisector $X = T$ (black, full). The residual error bar (green arrows) is defined as the standard deviation of the regression errors (SER). (Color figure available online.)

3.2. Statistical Analysis

3.2.1. Linear Regression

Benchmark studies usually analyze the difference between DFT and experiment statistically. The most common characteristics investigated are the mean error (signed) and the mean absolute error (unsigned). However, this approach implicitly assumes that the offset between DFT predictions and experimental results is the same for large and small values. For strictly positive quantities a relative shift seems more reasonable. The present analysis, therefore, explicitly treats relative deviations, in addition to the remaining scatter on that trend. This is done by means of a linear regression between DFT data and experimental results.

The linear regression is performed by means of a least-squares fit, from which we obtain the slope as well as the scatter with respect to the regression line.[109] The model hence presumes that a perfect correlation between experiment X and theory T exists, distorted by a random error ϵ centered around a zero mean: $X = \beta T + \epsilon$. If the exact exchange-correlation functional were known, it would approximately lead to $\epsilon = 0$ and $\beta = 1$. In practice, comparing the least-squares estimate of β to 1 offers a good measure of any systematic deviations, while the standard deviation of ϵ (also denoted as standard error of the regression or SER) expresses the residual error bar (see Figure 1).

Although the experimental community commonly employs the nomenclature "systematic errors" and "nonsystematic errors," we will avoid using the second term. Nonsystematic errors suggest a certain degree of randomness, which is not present in DFT. When one performs the same experiment several times, the results are spread around a mean value. In DFT, repeating the same calculation always yields identical results. In contrast to experimental error analysis, the spread of deviations in DFT results becomes apparent only when predictions for many different compounds are compared to experiment, i.e., when a benchmark set is used. The DFT scatter is then caused by some (subsets of) crystals being described better by the functional than others. Intrinsic error bars are hence fundamentally different from experimental error bars as well. The systematicness of DFT also appears when studying results within a certain family of materials: similar systems behave almost identically, proving the DFT spread not to behave randomly at all. An excellent example is found in literature, where the correspondence with experiment displays much less scatter when chemically similar compounds are involved.[42]

As an additional side note, one should be aware that experimental error bars have not been taken into account in any way. In fact, the statistical model should be $X + \eta = \beta T + \epsilon$, where η represents an additional (but uncorrelated) zero-mean perturbation. This does not only affect the comparison between individual DFT and experimental results, but also influences the values of the intrinsic systematic deviations and residual error bars that are presented in this work. By considering a test set that is sufficiently large, such as in the present study, one may hope that these effects level out. In addition, the elemental crystals belong to the most intensively studied materials, such that the

experimental errors are much smaller than for regular compounds. Without full knowledge of the experimental errors, however, we can only state that the SER provides an upper limit for the real PBE spread σ_ϵ. A possible solution consists in comparing DFT values to results from highly accurate many-body techniques instead. Such high-precision data are not available for many of the materials considered here, however, and to calculate them ourselves exceeds the scope of the current article.

3.2.2. Eliminating Outliers

A full statistical analysis of all elemental data cannot be performed straightforwardly. Some subsets of elements strongly distort the agreement between DFT and experiment. More meaningful error estimates are obtained when the most striking outliers are removed from the data set. Since the deviating behavior is often caused by a bad description of some underlying physical mechanism, most of them will be grouped in subsets of similar compounds. Instead of removing one outlier at a time from the data set, we choose only to remove entire structure types at once. This way, individual materials that behave well for the wrong reasons, are excluded as well, avoiding bias towards smaller errors.

A decomposition of the test set into eight subsets is proposed, based on some common physical properties of the corresponding elemental crystals (Figure 2). They are: (1) alkali and alkaline earth metals; (2) nonmagnetic transition metals; (3) magnetic materials; (4) correlation-dominated materials; (5) high-coordination p block compounds; (6) low-coordination p block compounds; (7) molecular crystals; and (8) noble gases. Subsets 7 and 8 correspond to materials where dispersion interactions are essential. (Color figure available online.)

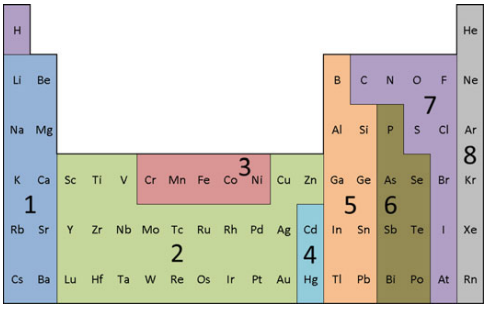


FIG. 2. Decomposition of the periodic table into smaller subsets of elements, based on common physical properties of the corresponding ground-state crystals: (1) alkali and alkaline earth metals; (2) nonmagnetic transition metals; (3) magnetic materials; (4) correlation-dominated materials; (5) high-coordination p block compounds; (6) low-coordination p block compounds; (7) molecular crystals; and (8) noble gases. Subsets 7 and 8 correspond to materials where dispersion interactions are essential. (Color figure available online.)

p block compounds; (7) molecular crystals; and (8) noble gases. Obviously, for some boundary elements, the most appropriate subset can be matter for discussion, but the classification in Figure 2 explains most trends for the intrinsic errors in a satisfactory manner (see further).

These eight subsets of elemental materials are representative for more complex (multicomponent) crystals as well. They provide prototype systems for particular bond types and physical phenomena, such as London dispersion (subsets 7 and 8), magnetism (subset 3), and electronic correlation (subset 4). Observations of DFT performance for these eight subsets will therefore carry over to multicomponent compounds.

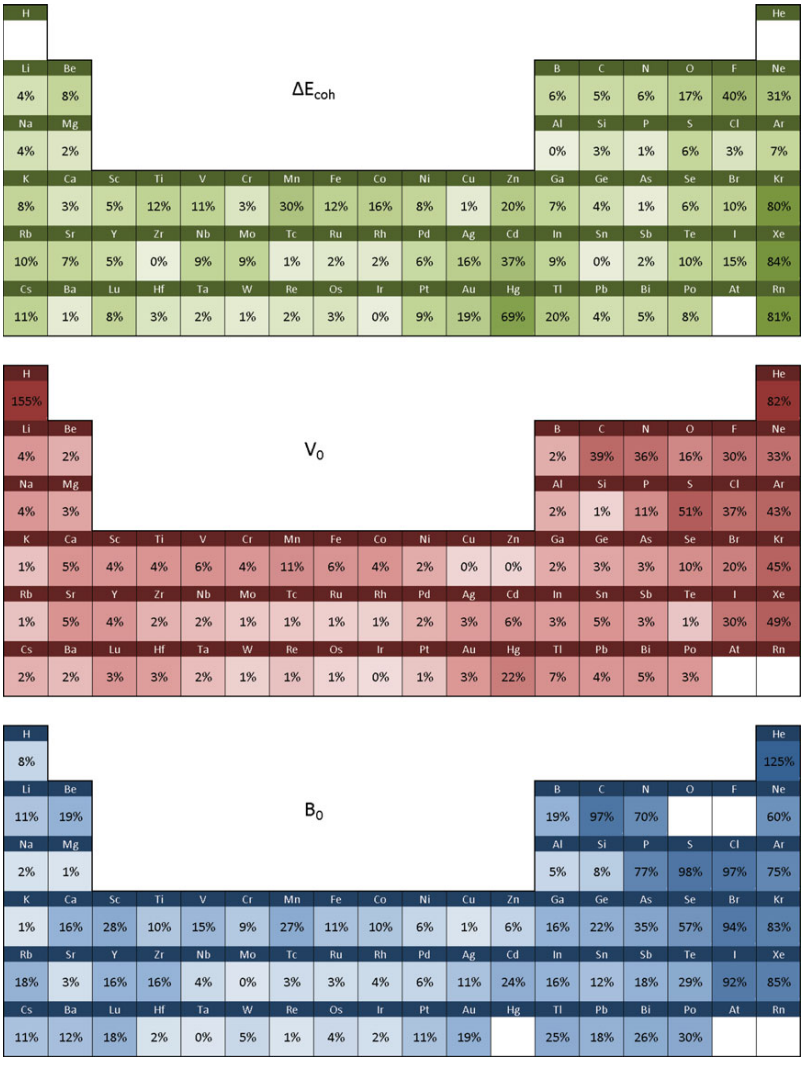
In order to eliminate deviating subsets in an objective way, the following procedure has been used. All subsets from Fig. 2 that have half or more of their elements differing significantly from the dominating trend, have been excluded. A two-sided p-value of 10% is maintained. In other words, a data point is considered to deviate substantially when the (signed) relative residual error (expressing the difference between the DFT regression value and the experimental one) belongs to the outer 10% of a normal distribution. This approach was repeated in an iterative way: after the elimination of each subset the significance criterion has been reestablished, until no deviating subsets remained. For solids belonging to an excluded category, PBE is not expected to provide reliable property predictions.

This selection criterion has been visualized in Tables 2 and 3. For each elemental crystal the relative residual error $|x_{exp} - x_{reg}|/x_{exp}$ is shown. Large numbers suggest a significant deviation from the regression line and hence allow for the recognition of outliers. Because these differences are displayed in the shape of the periodic table, they allow for easy identification of deviating subsets*. A color code has been added to improve intuition, with the darkest shades corresponding to the largest deviations. The deviations with respect to the elastic constants represent the mean absolute errors over C_{11}, C_{12}, C_{33}, C_{13}, and C_{44}.

Another way to visualize the assessment procedure is presented in Figure 3. In order to get a better view on systematic deviations, both the linear regression line (dashed red) and the first quadrant bisector (full black, representing a perfect match between theory and experiment) have been added for all accepted elements (see Table 4). As an additional quality indicator, the Pearson product-moment correlation coefficient r has been included for each property,[113] with $r = 1$ indicating a perfect positive correlation. Data points corresponding to omitted

*The graphical representations in this work employ the conventional, medium-form periodic table. Hydrogen is kept in group IA (contrary to the vivid discussion in, e.g., Scerri[110]) and lutetium in group IIIB.[111] This increases the intuitive character of the results. Insights should be conveyed at a glance and many researchers are most familiar with the standard format of the periodic table. For the same reason, the two-dimensional representation is preferred over some 1D alternatives.[112]

TABLE 2
Relative residual errors (of the VASP-PBE regression results with respect to the thermally corrected experimental values) for ΔE_{coh} [46,73,74] (green), V_0 [43,75–82] (red), and B_0 [46,62,83–90] (blue) of the elemental crystals. The darkest shades correspond to the largest errors (color figure available online)



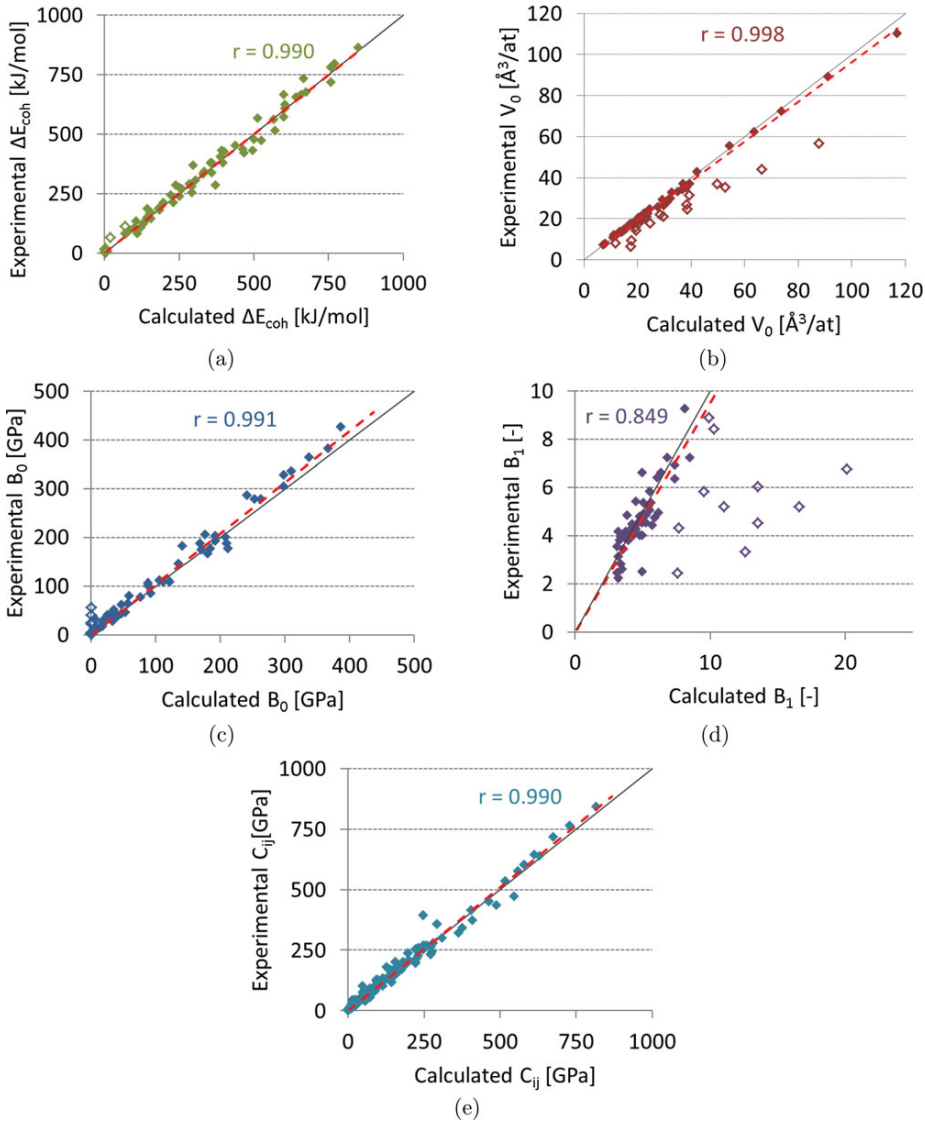
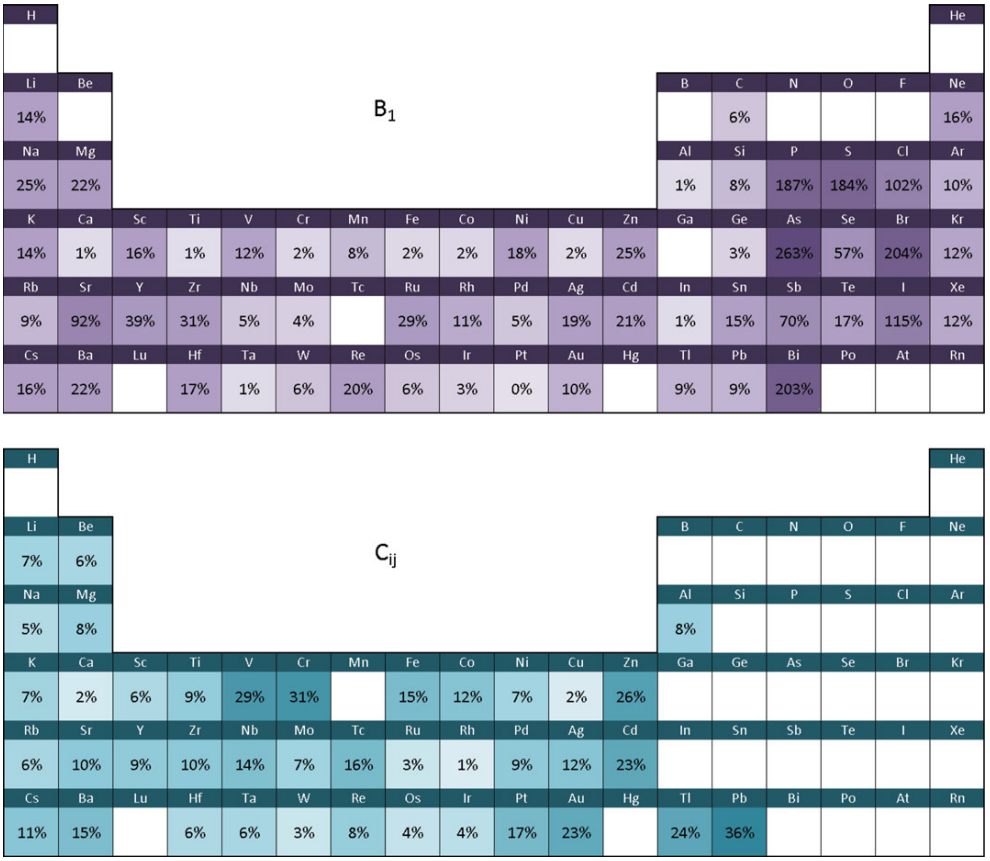


FIG. 3. Linear regression (dashed red) between the (thermally corrected) experimental and theoretical results (VASP-PBE / -PW91, see text) for the cohesive energy,[46,74] equilibrium volume,[43,75–82] bulk modulus,[46,62,83–90] pressure derivative of the bulk modulus,[91–97] and elastic constants.[40,98,99] The full, black line stands for $x_{exp} = x_{th}$. r represents the Pearson product-moment correlation coefficient for all elements included in the regression (filled symbols). The criterion for excluding certain elements from the fit (open symbols) is discussed in the text. (Color figure available online.)

TABLE 3
Relative residual errors (of the VASP-PBE/-PW91 regression results with respect to the thermally corrected experimental values) for B_1[91–97] (purple) and C_{ij}[40,98,99] (PW91, cyan) of the elemental crystals. The darkest shades correspond to the largest errors (color figure available online)



subsets, on the other hand, have been represented by an open symbol.

3.2.3. Predicting Experiment

Using all crystals that survived this selection procedure (filled symbols in Figure 3), a least-squares linear regression can now be computed for all properties from Section 2.1. Table 4 summarizes the resulting intrinsic errors in terms of relative systematic deviations ($1 - \beta$) and residual error bars (SER). Systematic errors are mentioned for PBE with respect to experiment, so a positive number implies PBE to overestimate that property. Each percentage is expressed relative to the PBE result, allowing for a straightforward calculation of the regression value ($x_{reg} = x_{th} - (1 - \beta)x_{th}$). Between brackets the significance level of $\beta = 1$ is mentioned. It represents the two-sided p-value when a null hypothesis of $\beta = 1$ is assumed: if there

really were no deviation at all, a small p-value would indicate that finding an even more extreme result would be highly unlikely. For the residual error bars a 95% confidence interval is given. The last column lists the subsets of elements that were excluded from the regression analysis by the selection procedure described before. For this the naming convention from Figure 2 has been used.

This statistical treatment makes several implicit assumptions. One of the most important premises is the use of a relative error over a constant offset. After all, for strictly positive quantities, such as V_0, an invariable shift seems counterintuitive. The impact of such an error is much larger when the investigated property is small. In addition, when using relative systematic deviations, the difference from $\beta = 1$ indeed matters. For the equilibrium volume ($p = 6 \cdot 10^{-11}$) and the bulk modulus ($p = 5 \cdot 10^{-4}$) the deviation from the bisector $x_{exp} = x_{th}$ is clearly significant. For other properties this is not always that obvious, but due to physical connections with V_0 and B_0, it is relevant to consider systematic deviations there as well.

A second remark concerns the nature of the intrinsic residual error bars. The numbers in Table 4 were computed assuming a normal distribution for the random error ϵ. This results in an absolute residual error bar. Indeed, a Pearson's χ^2-test does not contradict the applicability of a normal distribution to the intrinsic random errors. A null hypothesis assuming a Gaussian distribution around zero yields a 1-sided p-value of 0.21 for the volume. Within the DFT community, however, relative error bars are often implicity assumed. As a consequence, small volumes are expected to be predicted much more accurately, for example. According to our analysis, the matter appears to be not that simple, as is shown in Figure 4. In Figure 4a, the relative residual errors are plotted against the volume obtained by the least-squares fit. The overall decreasing trend suggests that the

TABLE 4
Systematic deviations $1-\beta$ and intrinsic error bars (SER) for the VASP-PBE/-PW91 (see text) properties presented in Tables 2 and 3, compared with experiment. The significance of the systematic deviation from $x_{exp} = x_{th}$ is indicated between brackets, by means of the two-sided p-value (low p-values for high significance). For the standard error of the regression a 95% confidence interval is given in terms of the upper and lower limit (superscript and subscript). Subsets containing a lot of outliers have been excluded from the data set by means of the procedure mentioned in the text (notation from Figure 2)

	$1-\beta$	SER	Excluded
ΔE_{coh} [kJ/mol]	−0.0% (0.99)	30^{+7}_{-4}	4, 8
V_0 [Å^3 /atom]	+3.6% (10^{-10})	$1.1^{+0.2}_{-0.2}$	4, 7, 8
B_0[GPa]	−4.9% (10^{-3})	15^{+4}_{-2}	7, 8
B_1[–]	+4.8% (0.03)	$0.7^{+0.2}_{-0.1}$	6, 7
C_{ij} [GPa][40]	−2.0% (0.01)	23^{+3}_{-2}	—

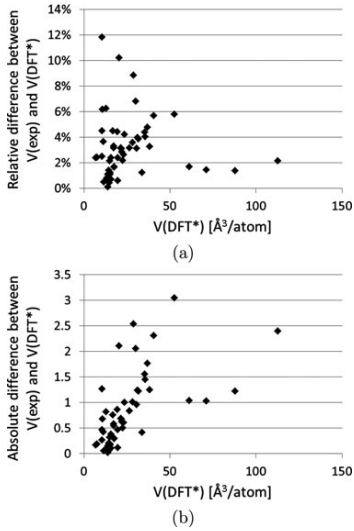


FIG. 4. Relative (a) and absolute (b) differences of the DFT regression data (DFT*) with respect to thermally corrected experimental values[43,75–82] for the equilibrium volume.

DFT errors are best described in terms of absolute error bars. Figure 4b displays the absolute residual errors instead. A rough linear correlation emerges, that implies that on the contrary a relative error bar is more appropriate. Both conclusions are compatible, by assuming a relative residual error bar of about 3% for small to median volumes (<50 Å^3/atom) and an absolute residual error bar of approximately 1.5 Å^3/atom for larger volumes (with the exception of a few outliers[†]). The number of data points is not sufficiently large for this finding to be really convincing, however. We therefore prefer to adopt the overall absolute error bar of 1.1 Å^3/atom (Table 4) as being valid for all volumes. Admittedly, for most small volumes the intrinsic residual error bars are then overestimated.

For any given material, the information in Table 4 can now be used to determine a meaningful estimate of a certain property. As an example, we consider the bulk modulus of diamond. This material is not included in the test set (the ground-state crystal

[†]The fitting error is primarily correlated to the bond type. Although badly performing subsets of materials have already been excluded from the test set, some crystal types are described significantly better by DFT than others, especially those with strong bonds (covalent p-bonds, half-filled d-shell elements). Their strong bonds also lead to more compact structures, which explains the smaller errors for smaller volumes (Figure 4b). Other subsets of crystals with similar volumes perform worse, however, and give rise to the larger relative errors in Figure 4a.

structure of carbon is graphite), but it can be assigned to subset 5, the high-coordination *p*-block compounds (similar to Si, Ge, and Sn). Subset 5 does not belong to one of the excluded subsets for B_0, and the intrinsic error for the bulk modulus prediction of diamond should therefore be representative. A bare VASP-PBE computation yields 434.8 GPa. When taking into account that PBE bulk moduli are systematically too small by 4.9% (Table 4), this value increases to 456.0 GPa. By means of Equations (12), (10), and (6), zero-point corrections (−17.6 GPa) are added back in, yielding 438.4 GPa. Using the appropriate intrinsic residual error bar of 15 GPa (Table 4), the final result becomes 438 ± 15 GPa. This is the most accurate DFT-PBE prediction of the experimental bulk modulus at 0 K, including an error bar on the computed value. In comparison, the experimental value extrapolated to absolute zero (Equations (11) and (8)) amounts to 443 GPa.[46] This number remains neatly within the error bar and is indeed closer to the regression-corrected bulk modulus than to the bare DFT value. A similar procedure can be used for all properties in Table 4. Table 5 offers a few more examples.

Table 4 is based on elemental solids only. One has to verify that the results from this statistical analysis are transferable to multicomponent materials. A good test case in that respect is the collection of thirty-one binary compounds for which Haas et al.[37,38] calculated lattice parameters by means of PBE. When we take their DFT results, the experimental volume falls outside the confidence interval for seven crystals. In all of these cases, the PBE volume is too large. By taking into account the systematic overestimation of 3.6%, however, only in two cases the experimental value exceeds the predicted range. Both of these compounds are ionically bound (NaF and NaCl), a bond type that has not been considered in any of the proposed subsets (Figure 2). Although the multicomponent test set may therefore not contain all materials types and although it only relates to the atomic volumes, these results already strongly support the transferability of our error estimates.

For one example from Haas et al., GaAs, the intrinsic error contributions are listed in Table 5. It illustrates that the systematic deviation really matters if it is the goal to get as close as possible to the experimental value. This also appears from the mean absolute difference between experiment and the 31 theoretical predictions by Haas et al. When one does not apply the relative deviation from Table 4, this number amounts to 0.72 σ_V (0.78 Å^3/atom), while for the regression values it is only 0.37 σ_V (0.40 Å^3/atom).

3.3. Agreement with Experiment

3.3.1. Errors per Materials Type

Because Tables 2 and 3 are shaped like the periodic table, the color code immediately allows singling out the areas where PBE breaks down. It leads to a number of subsets (Figure 2) which can be eliminated from the test set. They are listed in Table 4. For these elements, PBE performs significantly worse than usual, mostly because some key physical phenomenon is not described (well) by the functional. In this subsection, these localized error zones are discussed in more detail, as well as the mechanisms on which they are based.

A first, well-known example of the failure of PBE is the class of **dispersion-governed compounds**. Although some more advanced DFT approaches address this issue specifically,[114,115] regular GGA functionals do not describe London forces. This translates into a decreased cohesion, and hence an inflated volume and underestimated bulk modulus (see also Section 3.3.2).

Although London forces have been demonstrated to play a role in other structures as well,[116–118] the most important crystals that suffer from this shortcoming, belong to the nonmetals (subsets 7 and 8). They include the noble gases, the dimeric crystals, graphite, and sulfur. In these materials the London dispersion interaction governs the bonding between atoms, diatomic molecules, graphene sheets, and 8-membered rings, respectively. Nevertheless, it is essential to realize that both the element type and the crystal structure contribute to the importance of dispersion. It is perfectly plausible that a certain element behaves badly in structure A, while there are no problems when it assumes structure B. This can be illustrated nicely by means of carbon. The dispersion forces between the graphene layers in graphite give rise to a large discrepancy between DFT and experiment. Diamond on the other hand follows the same behavior as neighboring (semi)metallic elements or even outperforms them (Table 6).

For the molecular crystals (subset 7) the PBE cohesive energy is larger than the experimental value, contrary to the expectation. This is due to the overestimation of the intramolecular

TABLE 5
Demonstration of the use of intrinsic errors (Table 4) to improve DFT-PBE predictions and assess their reliability

	V_0(W) [Å^3/atom]	B_0(diamond) [GPa]	V_0(GaAs) [Å^3/atom]
PBE (bare)	16.28	434.8	23.73[37]
systematic deviation	15.69 (−3.6%)	456.0 (+4.9%)	22.87 (−3.6%)
zero-point correction	15.71 (+0.02)	438.4 (−17.6)	23.00 (+0.13)[37]
residual error bar	**15.7 ± 1.1**	**438 ± 15**	**23.0 ± 1.1**
experiment (0 K)	**15.8**[37]	**443**[46]	**22.5**[37]

TABLE 6
Residual errors (of the VASP-PBE regression results with respect to the zero-kelvin extrapolated experimental values[43,46,73,74,83]) for two allotropes of carbon

	ΔE_{coh} [kJ/mol]	V_0 [Å^3/at]	B_0[GPa]
graphite	39 (5%)	3.2 (39%)	54 (97%)
diamond	13 (2%)	0.1 (2%)	5 (1%)

bond strength (see, e.g., Lany[36] and Table II of Paier et al.[20]), which covers up any influence of the lack of dispersion. Elastic properties, on the other hand, are in most cases not affected by intramolecular effects and show a similar behavior as for the remaining nonmetals.

The **magnetic materials** (subset 3) stand out as well, predominantly with respect to ΔE_{coh}. Although the use of the generalized gradient approximation and a correct atomic reference state have already reduced the gap between theory and experiment substantially,[48] the remaining difference cannot be neglected. Current GGA funtionals are not able to describe magnetic compounds very well. Manganese illustrates this nicely. Its intricate magnetic state[119] has been approximated by assuming only collinear magnetism, but this does not explain the observed differences. The cohesive energy, for example, would be higher in its correct ground state, leading to an even more pronounced deviation from experiment. An explanation is found with Singh and Ashkenazi,[120] who noticed that GGAs overestimate the magnetic energy. This is caused by the increased number of degrees of freedom in spin-polarized systems (two spins), while the number of physical relations the GGA must fulfill stays the same.

The discrepancies between theory and experiment are not caused by the DFT functional alone, however. For some magnetic elements the applied thermal extrapolations are no longer valid, because of phase transition effects in the vicinity of the Curie or Néel temperature. Experimental chromium is a good example, displaying large magnetic distortions of the thermal expansion coefficient near 311 K.[121] A relation as simple as that of Equation (8) cannot capture these complex underlying phenomena.

The transition metals with (nearly) full *d* shells sometimes deviate from experiment as well (subset 4). The effects are smaller than for the previous two classes of materials, but they are unmistakably present, especially in terms of the cohesive energy and the elastic constants. One can attribute this phenomenon to **electronic correlation**. For Zn, Cd, and Hg a full-fledged many-body treatment has indeed convincingly shown the influence of *d* electron correlation on ΔE_{coh} and the potential energy landscape.[13,122,123] Data in Tables 2 and 3 even imply that similar (but smaller) effects show up in other elements at the end of the *d* block, such as in Pd, Ag, Pt, and Au. In noble metals, dispersion phenomena play an important role too,[117] however, and it is not immediately clear how much of the remaining discrepancy can be attributed to electronic correlation. Since the influence of correlation in these elements appears limited, only Cd an Hg have been assigned to subset 4.

It seems that at the end of the *d* block, the high number of localized *d* electron pairs in combination with a small interatomic distance and a close-packed environment enhances correlation effects. Moreover, Philipsen and Baerends[48] suggested that at the very beginning and the very end of the 3*d* transition metals the GGA exchange energy drops, causing the electron correlation to gain importance. Any fortuitous cancellation of errors between exchange and correlation has therefore disappeared for the transition metals at the border of the *d* block.

These correlation effects appear to be of a mainly anisotropic nature, since all elastic constants C_{ij} are affected, but the deviations of V_0 and B_0 are less pronounced. This is also suggested by Wedig et al.[123] for Zn and Cd, where a different interlayer and intralayer behavior is observed.

Relativistic effects are expected to strongly influence heavy elements. VASP therefore makes use of the scalar-relativistic Kohn-Sham equations by default.[124] The major remaining contribution is due to spin-orbit coupling. However, it is shown by Philipsen and Baerends[125] that this does not change physical properties substantially. Only for gold and bismuth a distinct change is reported, but without closing the gap between theory and experiment entirely. The remaining difference for Au is primarily due to correlation and dispersion effects, as was already suggested above. For the 6*p* elements, on the other hand, spin-orbit coupling really plays an important role. Some key properties for the 5*d* and 6*p* compounds have been calculated, both with and without spin-orbit coupling (Table 7). It is immediately clear that, starting from the end of the 5*d* block, a spin-orbit treatment becomes indispensable. Hence, for all 5*d* and 6*p* elements this contribution has been included, except for C_{ij}.[40]

3.3.2. Errors per Property

The previous section shows that PBE is not "complete": some features just cannot be described by a simple GGA functional. One can exclude the affected materials (outliers) beforehand, however, and limit the analysis to those cases where PBE should perform well. The intrinsic errors from Table 4 are applicable to these crystals. Table 4 then shows that the PBE error estimates largely depend on what property is considered. The behavior of the residual error bar and the systematic deviation from experiment can be traced back to both the functional and the numerical determination of that particular property. Nevertheless, it is important to note that, although the overall error estimate can be linked to theoretical aspects, the correspondence to experiment for a single compound depends on the experimental accuracy as well. This is especially true for higher-order properties (such as the elastic constants and their derivatives), which are generally measured at a lower precision than those from (quasi)direct measurements (the lattice constants or

TABLE 7
Relative residual errors (of the VASP-PBE regression results with respect to the zero-kelvin extrapolated experimental values[43,46,82]) for Ag and the 5*d* and 6*p* materials, both with (SO) and without spin-orbit coupling (n-SO)

	ΔE_{coh}		V_0		B_0	
	n-SO	SO	n-SO	SO	n-SO	SO
Ag	16%	16%	3%	3%	11%	11%
Lu	7%	8%	3%	3%	18%	18%
Hf	1%	3%	3%	3%	2%	2%
Ta	1%	2%	2%	2%	1%	0%
W	0%	1%	1%	1%	3%	5%
Re	3%	2%	2%	1%	3%	1%
Os	1%	3%	0%	1%	1%	4%
Ir	4%	0%	0%	0%	0%	2%
Pt	7%	9%	1%	1%	9%	11%
Au	22%	19%	4%	3%	21%	19%
Hg	75%	69%	29%	22%		
Tl	6%	20%	9%	7%	27%	25%
Pb	44%	4%	4%	4%	10%	18%
Bi	15%	5%	2%	5%	17%	26%
Po	59%	8%	2%	3%	73%	30%
Rn	83%	81%				

the cohesive energy, for example). This is illustrated by the sometimes large spread on the data in Knittle's overview of B_1 values.[91]

From a computational viewpoint, however, the equilibrium **volumes** offer the best results among all considered quantities. After eliminating the outliers (listed in Table 4 and represented in Figure 3b by open symbols), an almost perfect correlation is obtained. Even so, the regression line does not coincide with the first quadrant bisector. Table 4 shows that the cell volumes are consistently too large by approximately 4%. This deviation is a well-known property of any GGA,[126] including PBE. It originates in a systematic underestimation of the bond strength (underbinding), resulting in slightly larger volumes. More particularly, GGAs favor inhomogeneous systems, with large (reduced) density gradients. Small unit cells, which have a more evenly distributed electron density, are therefore energetically less preferable. This phenomenon especially affects open structures, where the high-gradient tails of the valence electron orbitals become nonnegligible.[127]

Immediately linked to this observation is the underestimation of the **bulk modulus**. More weakly bound structures will be more easily compressible, leading to smaller B_0 values. Just like the too large predicted volumes, it is common behavior for GGA functionals.[126] On the other hand, PBE bulk moduli are predicted with a larger uncertainty than the volumes. The intrinsic residual error remains in most cases below 10–15%, however (Table 2). The magnitude of this difference is mostly due to the sensitivity of the $E(V)$ curvature.

Since B_0 and the other **elastic constants** are closely related, the intrinsic errors with respect to the C_{ij} parameters are of a comparable scale. The bulk moduli are larger on average, which leads to slightly smaller relative errors (Tables 2 and 3). However, a good correlation is found in both cases, with a similar value of r for the elastic constants (Figure 3e) and the bulk moduli (Figure 3c).

For the **cohesive energy** PBE yields very good results as well. The intrinsic error bar of 30 kJ/mol is of the same order of magnitude as the rms error found by Lany[36] for PBE heats of formation of semiconductors and insulators (0.24 eV/atom). It is therefore representative for PBE energy differences between chemically different compounds. As mentioned before, for similar systems the intrinsic residual error bar is much smaller.[42] This also explains the success of evolutionary algorithms. They are based on energy differences of the order of a few meV per atom,[128–131] but some results have already been confirmed experimentally nevertheless.[9]

Contrary to V_0 and B_0, there is now no systematic under- or overestimation compared to experiment. The typical underbinding of GGA does not show as conclusively in ΔE_{coh}. This is due to the magnetic materials and the molecular crystals. As mentioned before, GGA functionals bias solutions towards magnetism for the former and overestimate the intramolecular contribution in the latter. In both cases this causes the cohesive energy to oppose the dominating trend. Without the crystals from subsets 3 and 7 in the test set, the cohesive energies would have been underestimated by 2% instead (p-value of 0.008). This behavior is in accordance with the expected underbinding of GGA.

Since the **bulk modulus derivative** B_1 is a higher-order parameter than B_0, the errors are expected to be one order worse as well. Although this is certainly the case, eliminating the outliers substantially improves the results (Figure 3d). However, even when they are removed, the resulting correlation coefficient ($r = 0.849$) remains significantly lower than for any other property already discussed.

B_1 appears to be overestimated with respect to experiment. This systematic deviation is significant, although the p-value may not show it conclusively (Table 4). It is again caused by GGA underbinding. As mentioned before, large volumes are favored due to their substantial density gradients. GGA hence lowers the energies of bigger cells most and straightens out the equation of state. This causes the $E(V)$ line to alter its decreasing curvature even more rapidly, increasing the rate of change of the bulk modulus with pressure (and volume), B_1. It also explains the deviating behavior of crystals with a low coordination, such as the molecular crystals. In these compounds the tails of the electron wave functions dominate the interstitial space, leading to considerable density gradients. The increase of the sensitive parameter B_1 is then enhanced even further.

4. NUMERICAL ERRORS

The previous section describes the intrinsic PBE errors for five different properties, based on a statistical treatment. They are used in a protocol which allows experimentalists and theoreticians to correct the bare DFT-PBE values for the observed systematic deviation from experiment and which quantifies the uncertainty on the obtained predictions (Table 5). A prerequisite for such a protocol is that different DFT implementations provide the same predictions: using different algorithms to solve the same (Kohn-Sham) equations should ideally lead to identical solutions. In practice, different amounts of noise are inevitably introduced in the predictions, even when numerical convergence has been achieved for each individual code. This scatter is due to several aspects of the solution algorithm. It can be due to its nature (e.g., the kind of basis set or the frozen-core approximation), its specific ingredients (e.g., the chosen pseudopotential) or its use of particular routines for standard tasks (e.g., Fourier transform routines). In order to guarantee the reproducibility of the intrinsic errors in Table 4, one needs to examine to what extent these issues affect the DFT computations. Once again, a reliable benchmark can be established using the ground-state elemental crystals.

The current section describes a procedure to express the difference between predictions from independent solid-state DFT approaches in a quantitative way, yielding a numerical error estimate. It will be used to examine differences between the PAW and APW+lo method, representing both methods by suitable mainstream codes (see further). However, the difference between codes can be attributed to other aspects as well, as was mentioned earlier. The influence of standard task routines is most likely small, but we will also use two PAW codes with different PAW atomic potentials, which can have quite drastic effects on the DFT results. All codes are therefore considered here with their recommended potentials. These can be thought of as representative for the quality of the investigated code. Although it is not our primary goal here, the procedure that will be described can also be used to select better performing PAW atomic potentials. The same holds for pseudopotentials in the case of plane-wave codes.

4.1. Test Set Preparation

The present implementation assessment starts from one reference code, the all-electron program WIEN2k[132] (version 11.1). It uses the APW+lo basis set,[133, 134] which is considered to be a standard for the numerical accuracy of solid-state DFT. WIEN2k predictions can therefore be considered to yield the exact results for a given functional, as long as numerical accuracy is achieved[37] (large basis set and dense k-mesh, see Supplementary Material for more details). Two codes are compared to this reference code: VASP[8,67](version 5.2.2) and GPAW[135–137] (version 0.8.0), both using the PAW method.[65] GPAW calculates all wave functions, densities, and potentials as grid-based quantities, while VASP uses a plane-wave basis set. All calculations with these programs are performed by means of the potentials recommended by the respective developers: the 2010 recommended PAW potentials for VASP and the 0.6 atomic set-ups for GPAW. Detailed computational parameters are summarized in the Supplementary Material.

For reasons of uniformity and comparability, the same PBE functional has been selected for all three codes. It is used in a protocol that seeks to evaluate a particular DFT approach in an easily reproducible manner. The VASP-optimized ground-state crystal (Section 3.1) serves as a starting point for each computation and from a 7-point equation of state (0.94–1.06 V_0) the properties of interest (E_0, V_0, B_0, and B_1) are extracted. All geometries are kept frozen (the cell shape and relative atomic positions are kept fixed at their initialvalues), instead of allowing for relaxation changes. This not only lowers the computational load, it also restricts the code evaluation to the implementation of DFT-PBE itself. Indeed, the task of optimizing the cell shape or internal positions belongs to another computational layer, on top of the task of solving the DFT equations for a given rigid geometry. This section aims to examine how different implementations compare with respect to the DFT-PBE procedure only. It does not intend to study how close every individual approach comes to experiment.

In the same spirit, some other modifications of the hitherto employed test set have been made. All calculations have been limited to the scalar-relativistic part (using the Koelling-Harmon approach[124]). By neglecting the spin-orbit contribution, an additional secondary algorithm implementation is avoided. The computational procedure also becomes more uniform this way, since all elements are now treated on equal terms. Because no spin-orbit coupling is added to the system's Hamiltonian, it suffices to use non spin-orbit geometries as a starting point for the seven-point equation of state.

A simplified unit cell has been selected for Mn and S as an additional means of lowering the computational effort. Manganese is treated in an antiferromagnetic fcc phase (space group 225, cF4), while for sulfur the β Po phase is imposed (space group 166 or hR3). These geometries are physically relevant, as they can be found in the Mn and S phase diagram, respectively.[138,139]

All other elements have been kept at the structure previously optimized by VASP (Section 3.1), in order to conserve the large diversity of the input set. The CIF files for all crystals in this code benchmark set are available in the Supplementary Material.

4.2. Agreement between Implementations

The procedure mentioned above results in a large collection of numbers for each code (71 elements $\times$ 4 properties). It is not convenient to compare them directly, however. Because of the different units involved, a coherent approach would require the use of relative deviations. Tables 2 and 3, on the other hand, show that each property corresponds to a different magnitude of relative error. This scale is mainly determined by the computational procedure and therefore does not alter substantially

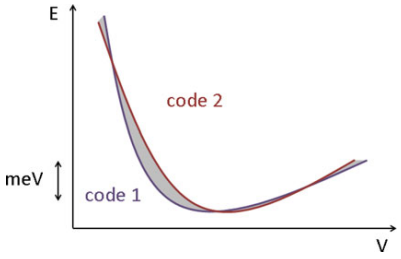


FIG. 5. The EOS parameters can differ significantly, while the $E(V)$ curves themselves are very similar. In that case the area between the two functions is a better indicator of the overall deviation. (Color figure available online.)

when shifting from a code-experiment comparison to an intercode assessment. A single numerical error value, expressing the difference between two particular DFT methods by means of one number, can be obtained by applying a weighted average. As all properties of interest depend on the equation of state, it is most straightforward to compare the $E(V)$ curves produced by different approaches directly. The dispersion-governed compounds illustrate this strategy well. Since their $E(V)$ curves are very shallow, small deviations in the bulk modulus will inflate the relative error considerably. However, the equations of state as such can be very similar, the two curves at no point differing by more than a few meV per atom (Figure 5). For that reason a numerical error estimate Δ is defined as follows:

$$\Delta = \left\langle \sqrt{\frac{\int \Delta E^2(V) \mathrm{d}V}{\Delta V}} \right\rangle. \qquad [15]$$

In other words, the rms energy difference between the $E(V)$ curves of these particular programs is averaged over all elemental crystals. Δ hence provides an intuitive measure of the energy distance between equations of state.

Because different codes sometimes employ different reference energies E_0, depending on the concept, all equations of state are set to zero at their equilibrium volume. An alternative solution would entail the calculation of cohesive energies, in order to provide a common reference for the equilibrium energy. However, not all programs allow for an easy manipulation of the electronic configuration of atoms. Moreover, the computational load would increase considerably.

The computation of Δ can be automated quite easily. The fitted Birch-Murnaghan equation allows Equation (15) to be written in an analytical form. Only V_0, B_0, and B_1 are then needed for both codes under investigation. The resulting expressions have been added in the Appendix for convenience. The WIEN2k data necessary for a code comparison have been provided in the Supplementary Material.

The interval of integration is linked to the reference data. In view of how the $E(V)$ parameters are determined, the intercode difference is to be integrated between $V_{0,\mathrm{WIEN\,2k}} \pm 6\%$. ΔV hence corresponds to 0.12 $V_{0,\mathrm{WIEN\,2k}}$. By definition, $\Delta(\mathrm{APW{+}lo})_{(\mathrm{WIEN2k})}$ becomes zero.

The rms energy differences between the equations of state predicted by $\mathrm{APW{+}lo}_{(\mathrm{WIEN2k})}$ and $\mathrm{PAW}_{(\mathrm{VASP})}$, or $\mathrm{APW{+}lo}_{(\mathrm{WIEN2k})}$ and $\mathrm{PAW}_{(\mathrm{GPAW})}$, are represented in Table 9. They show that most critical elements are characterized by approximately half-filled d levels. Such numerical errors can amount to up to 8.3 meV/atom for $\mathrm{PAW}_{(\mathrm{VASP})}$ (Tc) and 20.9 meV/atom for $\mathrm{PAW}_{(\mathrm{GPAW})}$ (Ru). This agrees with physical intuition, because these crystals are among the least compressible. Their equations of state are very steep, and relatively small modifications of the parameters can strongly change the energy. The least sensitive elements are for the same reason located near the alkali metals and the noble gases (0–0.7 meV/atom numerical error) (see Table 8). Only in comparison to experiment the latter group of materials stands out, but this is because PBE grossly overestimates the rare-gas volumes.

When averaging the numbers in Table 9 over all elements, the numerical error of each DFT approach can be determined for the given set of recommended PAW potentials. Δ is 1.9 meV/atom for $\mathrm{PAW}_{(\mathrm{VASP})}$, while for $\mathrm{PAW}_{(\mathrm{GPAW})}$ it is 3.3 meV/atom. This agreement between implementations is an order of magnitude better than the difference with experimental results. To show this, a similar energy difference between DFT-PBE and experiment is computed. It uses experimental values as the reference situation, while the method under test is the full-fledged version of $\mathrm{PAW}_{(\mathrm{VASP})}$. This means that the $E(V)$ parameters have been taken from Tables 2 and 3. The deviations per element are presented in Tab. 9, leading up to a Δ-factor of 23.5 meV/atom. This difference in magnitude can also be observed with the $E(V)$ characteristics themselves. Figure 6 shows the distribution of volume errors between two codes and with respect to experiment. Again, the spread is much larger in the latter case.

$\Delta(\mathrm{PAW})_{(\mathrm{VASP})}$ does not change noticeably when the number of elements is reduced to that of $\mathrm{PAW}_{(\mathrm{GPAW})}$. This shows that GPAW and VASP, while both using the same PAW method, do not produce entirely identical results. This variation most likely originates in the different quality of the atomic potentials and the different type of basis functions used. However, in comparison to

TABLE 8
Comparison between codes for two extreme situations: large (Tc an Ru) and small (Ar) numerical errors Δ_i. V_0 is given in Å^3/atom, B_0 in GPa, and Δ_i in meV/atom. B_1 is dimensionless

		V_0	B_0	B_1	Δ_i
Tc	$\mathrm{APW{+}lo}_{(\mathrm{WIEN2k})}$	14.47	301.4	4.56	0
	$\mathrm{PAW}_{(\mathrm{VASP})}$	14.60	298.5	4.55	8.3
Ru	$\mathrm{APW{+}lo}_{(\mathrm{WIEN2k})}$	13.81	315.4	4.96	0
	$\mathrm{PAW}_{(\mathrm{GPAW})}$	14.09	310.9	4.87	20.9
Ar	$\mathrm{APW{+}lo}_{(\mathrm{WIEN2k})}$	52.21	0.7	7.84	0
	$\mathrm{PAW}_{(\mathrm{VASP})}$	52.65	0.8	7.35	0.1
	$\mathrm{PAW}_{(\mathrm{GPAW})}$	52.66	0.8	3.27	0.1

TABLE 9

Rms energy differences Δ_i between the equations of state predicted by APW+lo$_{(WIEN2k)}$ and PAW$_{(VASP)}$ (green), APW+lo$_{(WIEN2k)}$ and PAW$_{(GPAW)}$ (red), and experiment and PAW$_{(VASP)}$ (blue) for the ground-state elemental crystals. All values are expressed in meV per atom. The darkest shades correspond to the largest errors. The average numerical error Δ is shown for each code at the header of the table (color figure available online)

$\Delta(PAW)_{(VASP)}$ = 1.9 meV/atom

H																	He
0.1																	0.0
Li	Be											B	C	N	O	F	Ne
0.2	0.1											0.3	0.3	10.6	8.3	1.5	0.1
Na	Mg											Al	Si	P	S	Cl	Ar
0.0	0.7											0.3	2.0	3.8	3.3	4.0	0.1
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
0.1	0.2	0.4	0.9	1.3	3.1	1.4	3.4	3.4	2.0	0.4	0.3	0.2	2.4	1.7	1.5	1.5	0.1
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.1	0.1	0.5	2.7	7.3	5.5	8.3	2.3	5.4	4.4	4.1	1.4	0.4	0.2	0.1	0.5	0.9	0.1
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
0.3	0.7	4.3	1.2	1.0	3.5	4.3	3.8	1.9	2.5	5.9	0.5	0.4	0.6	0.4	0.4		0.0

$\Delta(PAW)_{(GPAW)}$ = 3.3 meV/atom

H																	He
0.2																	0.0
Li	Be											B	C	N	O	F	Ne
0.2	2.6											0.2	0.3	0.2	10.4	1.0	0.2
Na	Mg											Al	Si	P	S	Cl	Ar
0.1	0.3											0.2	0.6	1.2	2.2	0.9	0.1
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
0.0	1.0	0.4	3.7	9.3	2.3	3.7	1.0	1.3	3.1	2.5	0.4	1.7	1.5	1.0	2.3	2.8	0.0
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.1	1.5		1.2	3.2	3.8		20.9	14.5	3.3	4.4	0.1	0.3	0.1		8.4	3.3	
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
0.3	0.6			10.2	19.3		18.5	13.2	6.3	6.8			0.8	0.4			

$\Delta(exp)$ = 23.5 meV/atom

H																	He
Li	Be											B	C	N	O	F	Ne
0.4													17.6				17.1
Na	Mg											Al	Si	P	S	Cl	Ar
0.2	1.5											3.9	13.7	19.0	44.0	46.1	38.6
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
1.4	3.0	1.7	3.0	13.1	1.6	47.7	10.6	3.3	6.2	13.5	7.6		28.2	17.1	11.1	22.3	54.6
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
1.3	2.4	0.7	8.2	14.8	26.5		30.4	44.6	49.1	39.3	34.0	23.6	40.7	25.1	9.9	68.1	59.8
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
4.0	2.5		2.8	16.8	35.6	29.9	65.2	50.7	54.9	64.0		42.4	32.9	32.0			

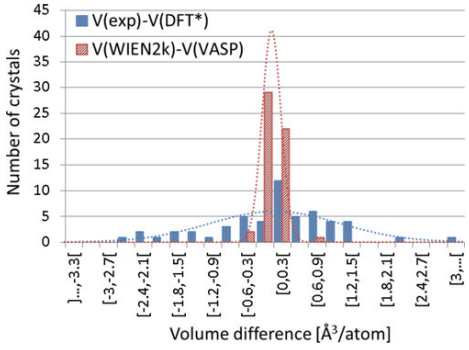


FIG. 6. Intrinsic (PBE regression vs. experiment) and numerical errors ($PAW_{(VASP)}$ vs. $APW{+}lo_{(WIEN2k)}$) for the equilibrium volume of the ground-state elemental crystals, using the subsets of elements that have been shown to perform well for PBE (see Table 4). A normal distribution has been fitted to both data sets (dotted line). (Color figure available online.)

experiment, the differences are negligible. The intrinsic residual error bars and regression slopes provided in Table 4 can therefore be applied to DFT-PBE results, irrespective of which approach was used to calculate them.

This comparison of three DFT implementations can easily be extended. Ideally every solid-state DFT approach should be tested in the same way, and have its Δ-value computed. As such tests are preferably performed by specialists in the individual codes, all input CIF files have been made available in the Supplementary Material and at arXiv,[140] as well as some ready-made post-processing scripts. In addition, the ASE developers have implemented a framework for performing the necessary calculations.[135,141] On the CMM website[142] an updated overview will be maintained of all Δ-factors reported to us. Such information not only provides insight into the reproducibility of the intrinsic errors of Table 4, but can also guide users to select a method for a specific task, at least as far as accuracy of energy vs. volume relations is concerned.

5. CONCLUSIONS

Using the ground-state elemental crystals as a test set, DFT-PBE computational errors have been reviewed. Errors intrinsic to the functional were quantified for five materials properties, describing energetic (ΔE_{coh}) and elastic (V_0, B_0, B_1, C_{ij}) quantities. They explain the deviation of DFT predictions from experiment. Numerical errors, due to the implementation of the DFT scheme into a computer code, were studied for the PAW method (VASP and GPAW), and were expressed with respect to the reference APW+lo method (WIEN2k). Both types of errors have been discussed for PBE, one of the most widely applied functionals in solid-state DFT. The results are expected to be representative of GGA in general.

Each of the five properties has been assessed with respect to the ground-state elemental crystals. The correspondence to experiment has been analyzed statistically, leading to a decomposition of the intrinsic error into systematic deviations and residual error bars. These intrinsic errors have been shown to agree with some generally known GGA traits. The typical underbinding of GGA has been reproduced and quantified, for example. Table 10 presents a summary of our results, as well as a similar analysis for other properties that are available from data sets in the literature. Contrary to Table 4, however, Table 10 presents systematic deviations in terms of the experimental value: $(x_{th} - x_{exp})/x_{exp} = 1/\beta - 1$. This expresses more intuitively to what extent DFT varies from experiment: a $1/\beta - 1$ of $+1\%$ for example means that PBE overestimates the experimental result by 1%.

Based on the quantification of intrinsic errors, a computational recipe has been presented which allows to correct bare DFT-PBE results for the systematic deviation from experiment, and which attaches meaningful error estimates to the obtained predictions (Table 5). An examination of 31 binary compounds not included in the benchmark set[37,38] indicates that our analysis carries over to multicomponent crystals. Errors can hence be estimated straightforwardly for PBE predictions already available from literature.

The overall agreement between VASP-PBE and experiment is quite good, but some subsets of elements perform better than others. DFT predictions for magnetic materials and correlation-dominated compounds deviate significantly from experimental values, for example, especially with respect to the cohesive energy. Long-range interaction is another issue. Although some solutions exist to incorporate London dispersion into DFT, such as the DFT-D[114] or vdW-DF2 method,[115] regular GGAs do not describe dispersion-governed crystal types well. Bulk moduli are found to be consistently underestimated, while predictions for both the volume and the pressure derivative of the bulk modulus are systematically too large. Results for heavy elements are acceptable as long as spin-orbit coupling is added, starting from the 6*p* block. Based on these observations, some general guidelines have been summarized in Table 10 as to what categories of materials will not be described well. Some classes were not or only marginally represented in the elemental benchmark set, such as ionic or strongly correlated compounds. For these, a similar study using an extended benchmark set, by including some binary ionic compounds and transition metal oxides, would be useful.

All conclusions with respect to the intrinsic PBE errors can only be universally applicable when it does not matter how the DFT formalism is implemented. Such numerical errors should be much smaller than intrinsic ones. By means of a quality factor Δ, which conveys exactly this information, $APW{+}lo_{(WIEN2k)}$ has been compared to $PAW_{(VASP)}$ ($\Delta = 1.9$ meV/atom) and $PAW_{(GPAW)}$ ($\Delta = 3.3$ meV/atom), both for their recommended sets of atomic potentials. The rms energy distance between equations of state from different methods indeed appears to be an order of magnitude smaller than the gap between theory and experiment (see also Table 10). The intrinsic systematic

TABLE 10
Systematic deviations $1/\beta-1$ and residual error bars for DFT-GGA predictions compared with experiment (intrinsic errors) and between codes (numerical errors). Δ represents the average rms energy difference between the equations of states of two codes. All data were (re)analyzed in the present study, except for the error bar for ΔE_{evol}, which is based on a proof of principle (see Section 3.3.2). Subsets of materials to which the error estimates do not apply are mentioned explicitly

	Intrinsic errors (this work, VASP-PBE)		
	systematic deviation	residual error bar	not applicable to:
ΔE_{coh}[kJ/mol]	-0.0%	30	dominant correlation; pure dispersion
V_0 [Å^3 /atom]	$+3.8\%$	1.1	dominant correlation; dispersion
B_0[GPa]	-4.7%	15	dispersion
B_1[$-$]	$+5.0\%$	0.7	low coordination number
	Intrinsic errors (other works, VASP-PW91/-PBE/-PBE+U)		
C_{ij} [GPa][40]	-2.0%	23	
ΔE_{form} [kJ/mol][36]	-13.1%	15	
ΔE_{react} [kJ/mol][42]	-4.8%	3.2	chemically dissimilar
ΔE_{evol} [kJ/mol][9,128–131]		$\lesssim$1	chemically or structurally dissimilar
	Numerical errors (this work, PBE)		
Δ(PAW)$_{\text{(VASP)}}$ [kJ/mol]		0.19	—
Δ(PAW)$_{\text{(GPAW)}}$ [kJ/mol]		0.32	—
⋮		⋮	

deviations and residual error bars presented in Table 10 can hence be applied to PBE predictions regardless of the computational approach. This is useful when discussing the implications of DFT results in an experimental context. This accuracy review is to be considered as a starting point only.

The presented statistical procedure is applicable to other functionals or methods as well. It would be useful to determine the intrinsic systematic deviations and residual error bars for, e.g., LDA or hybrid functionals. A comparison to results from high-level many-body techniques would even allow for the elimination of the influence of experimental errors. Another extension would be to take into account experimental error bars in the statistical analysis. For the relevant properties of the elemental materials, this requires an extensive literature search for the most accurately known values and their error bars. These error bars are not commonly available in tabulations in the literature and are therefore beyond the scope of this work. With respect to the assessment of numerical errors, we invite both code developers and users to determine the quality factor Δ for their code as well. It not only guarantees the transferability of the intrinsic errors to all codes, but also provides a criterion to evaluate the accuracy of a particular DFT approach. A comprehensive list of Δ-factors[142] can then serve as a guideline through the maze of available DFT methods.

SUPPLEMENTARY MATERIAL

See Supplementary Material posted on the publisher's website, for computational details, the full calculated and experimental data and the files necessary for a code comparison.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research – Flanders (FWO) and by the Research Board of Ghent University. Stefaan Cottenier acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. Calculations were carried out using the Stevin Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Hercules Foundation, and the Flemish Government (EWI Department). The authors thank Martijn Marsman for his useful insights on the spin-orbit coupled Pb atom, Fabien Tran for a helpful discussion on zero-point corrections, and Manuel Richter for some interesting comments with respect to scalar-relativistic techniques. They acknowledge the implementation by Marcin Dułak of the code comparison database into ASE,[135] and thank him for valuable remarks with respect to the Δ-factor computation script. This work has benefited a lot from remarks and input given by many colleagues at conference presentations, as well as from the extensive and constructive comments provided by three anonymous referees.

REFERENCES

1. P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.*, **136**, B864–B871 (1964).

2. W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.,* **140**, A1133–A1138 (1965).
3. M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne, First-principles simulation: ideas, illustrations and the CASTEP code, *J. Phys. Condens. Matt.*, **14**, 2717–2744 (2002).
4. W. K. Liu, E. G. Karpov, S. Zhang, and H. S. Park, An introduction to computational nanomechanics and materials, *Comput. Meth. Appl. Mech. Eng.*, **193**, 1529–1578 (2004).
5. R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, 2004.
6. K. Capelle, A bird's-eye view of density-functional theory, *Braz. J. Phys.*, **36**, 1318–1343 (2006).
7. J. Hafner, C. Wolverton, and G. Ceder (Guest Editors), Topical issue of the MRS Bulletin: 'Density Functional Theory in Materials Research', *MRS Bull.*, **31** (2006).
8. J. Hafner, *Ab-Initio* Simulations of Materials Using VASP: Density-Functional Theory and Beyond, *J. Comput. Chem.*, **29**, 2044–2078 (2008).
9. Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, Transparent dense sodium, *Nature*, **458**, 182–185 (2009).
10. G. Hautier, A. Jain, S. P. Ong, B. Kang, C. Moore, R. Doe, and G. Ceder, Phosphates as lithium-ion battery cathodes: An evaluation based on high-throughput *ab initio* calculations, *Chem. Mater.*, **23**, 3495–3508 (2011).
11. A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, A high-throughput infrastructure for density functional theory calculations, *Comp. Mater. Sci.*, **50**, 2295–2310 (2011).
12. L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, Accurate surface and adsorption energies from many-body perturbation theory, *Nature Mater.*, **9**, 741–744 (2010).
13. N. Gaston, D. Andrae, B. Paulus, U. Wedig, and M. Jansen, Understanding the hcp anisotropy in Cd and Zn: the role of electron correlation in determining the potential energy surface, *Phys. Chem. Chem. Phys.*, **12**, 681–687 (2010).
14. A. Bil, B. Kolb, R. Atkinson, D. G. Pettifor, T. Thonhauser, and A. N. Kolmogorov, van der Waals interactions in the ground state of $Mg(BH_4)_2$ from density functional theory, *Phys. Rev. B*, **83**, 224103 (2011).
15. D. N. Beratan and J. J. Hopfield, Failure of the Born-Oppenheimer and Franck-Condon approximations for long distance electron transfer rate calculations, *J. Chem. Phys.*, **81**, 5753–5759 (1984).
16. I. B. Bersuker, Limitations of density functional theory in application to degenerate states, *J. Comput. Chem.*, **18**, 260–267 (1997).
17. P. Pyykkö, The physics behind chemistry and the periodic table, *Chem. Rev.*, **112**, 371–384 (2012).
18. E. Engel, A. Facco Bonetti, S. Keller, I. Andrejkovics, and R. M. Dreizler, Relativistic optimized-potential method: Exact transverse exchange and Møller-Plesset-based correlation potential, *Phys. Rev. A*, **58**, 964–992 (1998).
19. E. Engel and U. Lechner, Exchange-correlation and QED effects from a density functional based level shift approach, *Chem. Phys.*, **311**, 209–226 (2005).
20. J. Paier, R. Hirschl, M. Marsman, and G. Kresse, The Perdew-Burke-Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set, *J. Chem. Phys.*, **122**, 234102 (2005).
21. A. Kiejna, G. Kresse, J. Rogal, A. De Sarkar, K. Reuter, and M. Scheffler, Comparison of the full-potential and frozen-core approximation approaches to density-functional calculations of surfaces, *Phys. Rev. B*, **73**, 035404 (2006).
22. J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, **77**, 3865–3868 (1996).
23. S. Kotochigova, Z. H. Levine, E. L. Shirley, M. D. Stiles, and C. W. Clark, Local-density-functional calculations of the energy of atoms, *Phys. Rev. A*, **55**, 191–199 (1997).
24. S. Kotochigova, Z. H. Levine, E. L. Shirley, M. D. Stiles, and C. W. Clark, Erratum: Local-density-functional calculations of the energy of atoms, *Phys. Rev. A*, **56**, 5191–5192 (1997).
25. Y. Zhao and D. G. Truhlar, Density functionals with broad applicability in chemistry, *Acc. Chem. Res.*, **41**, 157–167 (2008).
26. L. Goerigk and S. Grimme, A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions, *Phys. Chem. Chem. Phys.*, **13**, 6670–6688 (2011).
27. S. Kurth, J. P. Perdew, and P. Blaha, Molecular and solid-state tests of density functional approximations: LSD, GGAs and meta-GGAs, *Int. J. Quantum Chem.*,**75**, 889–909 (1999).
28. L. Vitos, B. Johansson, J. Kollár, and H. L. Skriver, Exchange energy in the local Airy gas approximation, *Phys. Rev. B*, **62**, 10046–10050 (2000).
29. V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Tests of a ladder of density functionals for bulk solids and surfaces, *Phys. Rev. B*, **69**, 075102 (2004).
30. V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Erratum: tests of a ladder of density functionals for bulk solids and surfaces, *Phys. Rev. B*, **78**, 239907(E) (2008).
31. J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional, *J. Chem. Phys.*, **123**, 174101 (2005).
32. J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán, Screened hybrid density functionals applied to solids, *J. Chem. Phys.*, **124**, 154709 (2006).
33. F. Tran, R. Laskowski, P. Blaha, and K. Schwarz, Performance on molecules, surfaces, and solids of the Wu-Cohen GGA exchange-correlation energy functional, *Phys. Rev. B*, **75**, 115131 (2007).
34. B. Grabowski, T. Hickel, and J. Neugebauer, *Ab initio* study of the thermodynamic properties of nonmagnetic elementary fcc metals: exchange-correlation-related error bars and chemical trends, *Phys. Rev. B*, **76**, 024309 (2007).
35. M. Ropo, K. Kokko, and L. Vitos, Assessing the Perdew-Burke-Ernzerhof exchange-correlation density functional revised for metallic bulk and surface systems, *Phys. Rev. B*, **77**, 195445 (2008).
36. S. Lany, Semiconductor thermochemistry in density functional calculations, *Phys. Rev. B*, **78**, 245207 (2008).
37. P. Haas, F. Tran, and P. Blaha, Calculation of the lattice constant of solids with semilocal functionals, *Phys. Rev. B*, **79**, 085104 (2009).

38. P. Haas, F. Tran, and P. Blaha, Erratum: Calculation of the lattice constant of solids with semilocal functionals, *Phys. Rev. B*, **79**, 209902(E) (2009).
39. G. I. Csonka, J. P. Perdew, A. Ruzsinszky, P. H.T. Philipsen, S. Lebègue, J. Paier, O.A. Vydrov, and J. G. Ángyán, Assessing the performance of recent density functionals for bulk solids, *Phys. Rev. B*, **79**, 155107 (2009).
40. S. L. Shang, A. Saengdeejing, Z. G. Mei, D. E. Kim, H. Zhang, S. Ganeshan, Y. Wang, and Z.K. Liu, First-principles calculations of pure elements: Equations of state and elastic stiffness constants, *Comput. Mater. Sci*., **48**, 813–826 (2010).
41. R. Nazarov, T. Hickel, and J. Neugebauer, Vacancy formation energies in fcc metals: Influence of exchange-correlation functionals and correction schemes, *Phys. Rev. B*, **85**, 144118 (2012).
42. G. Hautier, S. P. Ong, A. Jain, C. J. Moore, and G. Ceder, Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability, *Phys. Rev. B*, **85**, 155208 (2012).
43. P. Villars and J. L.C. Daams, Atomic-environment classification of the chemical elements, *J. Alloys Comp*., **197**, 177–196 (1993).
44. http://www.marcelswart.eu/dft-poll.
45. W. H. Dickhoff and D. Van Neck, *Many-Body Theory Exposed!,* 2nd ed., World Scientific, Singapore (2008).
46. C. Kittel, *Introduction to Solid State Physics,* 8th ed., John Wiley & Sons, Inc., New York (2005).
47. F. Birch, Finite elastic strain of cubic crystals, *Phys. Rev*., **71**, 809–824 (1947).
48. P. H. T. Philipsen and E. J. Baerends, Cohesive energy of 3*d* transition metals: Density functional theory atomic and bulk calculations, *Phys. Rev. B*, **54**, 5326–5333 (1996).
49. E. J. Baerends, V. Branchadell, and M. Sodupe, Atomic reference energies for density functional calculations, *Chem. Phys. Lett*., **265**, 481–489 (1997).
50. D. E. Kim, S. L. Shang, and Z. K. Liu, Effects of alloying elements on elastic properties of Ni by first-principles calculations, *Comput. Mater. Sci*., **47**, 254–260 (2009).
51. N. Miao, B. Sa, J. Zhou, and Z. Sun, Theoretical investigation on the transition-metal borides with Ta_3B_4-type structure: A class of hard and refractory materials, *Comput. Mater. Sci*., **50**, 1559–1566 (2011).
52. Y. Le Page and P. Saxe, Symmetry-general least-squares extraction of elastic coefficients from *ab initio* total energy calculations, *Phys. Rev. B*, **63**, 174103 (2001).
53. Y. Le Page and P. Saxe, Symmetry-general least-squares extraction of elastic data for strained materials from *ab initio* calculations of stress, *Phys. Rev. B*, **65**, 104104 (2002).
54. S. L. Shang, Y. Wang, and Z. K. Liu, First-principles elastic constants of α- and θ-Al_2O_3, *Appl. Phys. Lett*., **90**, 101909 (2007).
55. A. B. Alchagirov, J. P. Perdew, J. C. Boettger, R. C. Albers, and C. Fiolhais, Energy and pressure versus volume: Equations of state motivated by the stabilized jellium model, *Phys. Rev. B*, **63**, 224115 (2001).
56. V. L. Moruzzi, J. F. Janak, and K. Schwarz, Calculated thermal properties of metals, *Phys. Rev. B*, **37**, 790–799 (1988).
57. Y. Tsuru, Y. Shinzato, Y. Saito, M. Shimazu, M. Shiono, and M. Morinaga, Estimation of linear thermal expansion coefficient from cohesive energy obtained by ab-initio calculation of metals and ceramics, *J. Ceram. Soc. Jpn*., **118**, 241–245 (2010).
58. M. M. Dacorogna and M. L. Cohen, First-principles study of the structural properties of alkali metals, *Phys. Rev. B*, **34**, 4996–5002 (1986).
59. R. Gaudoin and W. M. C. Foulkes, *Ab initio* calculations of bulk moduli and comparison with experiment, *Phys. Rev. B*, **66**, 052104 (2002).
60. D. G. Pettifor, *Bonding and Structure of Molecules and Solids*, Oxford University Press, Oxford (1995).
61. A. Masago, K. Shirai, and H. Katayama-Yoshida, Crystal stability of α- and β- boron, *Phys. Rev. B*, **73**, 104102 (2006).
62. T. A. Scott, Solid and liquid nitrogen, *Phys. Rep*., **27C**, 89–157 (1976).
63. Y. A. Freiman and H. J. Jodl, Solid oxygen, *Phys. Rep*., **401**, 1–228 (2004).
64. W. I. F. David, R. M. Ibberson, S. F. J. Cox, and P. T. Wood, Order-disorder transition in monoclinic sulfur: a precise structural study by high-resolution neutron powder diffraction, *Acta Crystallogr. Sect. B*, **62**, 953–959 (2006).
65. P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B*, **50**, 17953–17979 (1994).
66. G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B*, **59**, 1758–1775 (1999).
67. G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci*., **6**, 15–50 (1996).
68. http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html (April 23, 2009 version).
69. P. E. Blöchl, O. Jepsen, and O.K. Andersen, Improved tetrahedron method for Brillouin-zone integrations, *Phys. Rev. B*, **49**, 16223–16233 (1994).
70. H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, *Phys. Rev. B*, **13**, 5188–5192 (1976).
71. S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, Crystallography Open Database – an open-access collection of crystal structures, *J. Appl. Crystallogr.*, **42**, 726–729 (2009).
72. A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design, *Acta Crystallogr. Sect. B*, **58**, 364–369 (2002).
73. N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements,* 2nd ed., Butterworth-Heinemann, Oxford (1997).
74. T. Tohei, A. Kuwabara, F. Oba, and I. Tanaka, Debye temperature and stiffness of carbon and boron nitride polymorphs from first principles calculations, *Phys. Rev. B*, **73**, 064304 (2006).
75. L. H. Bolz, M. E. Boyd, F. A. Mauer, and H. S. Peiser, A reexamination of the crystal structures of α and β nitrogen, *Acta Crystallogr.*, **12**, 247–248 (1959).
76. L. Meyer, C. S. Barrett, and S. C. Greer, Crystal structure of α-fluorine, *J. Chem. Phys*., **49**, 1902–1907 (1968).
77. D. N. Batchelder, D. L. Losee, and R. O. Simmons, Measurements of lattice constant, thermal expansion, and isothermal compressibility of neon single crystals, *Phys. Rev*., **162**, 767–775 (1967).
78. B. M. Powell, K. M. Heal, and B. H. Torrie, The temperature dependence of the crystal structures of the solid halogens, bromine and chlorine, *Mol. Phys*., **53**, 929–939 (1984).

79. O. G. Peterson, D. N. Batchelder, and R. O. Simmons, Measurements of X-ray lattice constant, thermal expansivity, and isothermal compressibility of argon crystals, *Phys. Rev.*, **150**, 703–711 (1966).
80. D. L. Losee and R. O. Simmons, Thermal-expansion measurements and thermodynamics of solid krypton, *Phys. Rev.*, **172**, 944–957 (1968).
81. D. R. Sears and H. P. Klug, Density and expansivity of solid xenon, *J. Chem. Phys.*, **37**, 3002–3006 (1962).
82. C. A. Swenson, Phase transition in solid mercury, *Phys. Rev.*, **111**, 82–91 (1958).
83. E. Kim and C. Chen, Calculation of bulk modulus for highly anisotropic materials, *Phys. Lett. A*, **326**, 442–448 (2004).
84. Y. Endoh, G. Shirane, and J. Skalyo, Jr., Lattice dynamics of solid neon at 6.5 and 23.7 K, *Phys. Rev. B*, **11**, 1681–1688 (1975).
85. E.-F. Düsing, W. A. Grosshans, and W. B. Holzapfel, Equation of state of solid chlorine and bromine, *J. Phys. Colloques (Paris)*, **45**, 203–206 (1984).
86. G. J. Keeler and D. N. Batchelder, Measurement of the elastic constants of argon from 3 to 77°K, *J. Phys. C*, **3**, 510–522 (1970).
87. H. Fujihisa and K. Takemura, Stability and the equation of state of α-manganese under ultrahigh pressure, *Phys. Rev. B*, **52**, 13257–13260 (1995).
88. J. Skalyo, Jr., Y. Endoh, and G. Shirane, Inelastic neutron scattering from solid krypton at 10°K, *Phys. Rev. B*, **9**, 1797–1803 (1974).
89. D. L. Price, J. M. Rowe, and R. M. Nicklow, Lattice dynamics of grey tin and indium antimonide, *Phys. Rev. B*, **3**, 1268–1279 (1971).
90. N. A. Lurie, G. Shirane, and J. Skalyo, Jr., Temperature dependence of the zero-sound elastic constants of crystalline xenon, *Phys. Rev. B*, **9**, 2661–2669 (1974).
91. E. Knittle, Static compression measurements of equations of state, in *Mineral Physics and Crystallography: A Handbook of Physical Constants*, Ahrens, T. J., Ed., American Geophysical Union Press, 98–142 (1995).
92. S. S. Batsanov, *Effects of Explosions on Materials: Modification and Synthesis under High Pressure Shock Compression*, Springer, New York (1994).
93. M. W. Guinan and D. J. Steinberg, Pressure and temperature derivatives of the isotropic polycrystalline shear modulus for 65 elements, *J. Phys. Chem. Solids*, **35**, 1501–1512 (1974).
94. C. S. Menoni, J. Z. Hu, and I. L. Spain, Germanium at high pressures, *Phys. Rev. B*, **34**, 362–368 (1986).
95. H. J. Beister, K. Strössner, and K. Syassen, Rhombohedral to simple-cubic phase transition in arsenic under pressure, *Phys. Rev. B*, **41**, 5535–5543 (1990).
96. T. Kenichi, Bulk modulus of osmium: High-pressure powder x-ray diffraction experiments under quasihydrostatic conditions, *Phys. Rev. B*, **70**, 012101 (2004).
97. J. H. Li, S. H. Liang, H. B. Guo, and B. X. Liu, Four-parameter equation of state and determination of the thermal and mechanical properties of metals, *J. Alloys Compd.*, **431**, 23–31 (2007).
98. L. G. Hector, Jr., J. F. Herbst, W. Wolf, P. Saxe, and G. Kresse, *Ab initio* thermodynamic and elastic properties of alkaline-earth metals and their hydrides, *Phys. Rev. B*, **76**, 014121 (2007).
99. U. Buchenau, M. Heiroth, H. R. Schober, J. Evers, and G. Oehlinger, Lattice dynamics of strontium and barium, *Phys. Rev. B*, **30**, 3502–3505 (1984).
100. D. R. Lide (ed.) *CRC Handbook of Chemistry and Physics,* 89th ed., CRC Press, Boca Raton, Florida (2008–2009).
101. J. B. Nelson and D. P. Riley, The thermal expansion of graphite from 15°C. to 800°C.: Part I. Experimental, *Proc. Phys. Soc. London*, **57**, 477–486 (1945).
102. R. W. Keyes, The electrical properties of black phosphorus, *Phys. Rev.*, **92**, 580–584 (1953).
103. J. Wallis, I. Sigalas, and S. Hart, Determination of the thermal expansion of orthorhombic sulfur, *J. Appl. Crystallogr.*, **19**, 273–274 (1986).
104. R. R. Pawar and V. T. Deshpande, Anisotropic thermal expansion of arsenic, *J. Mater. Sci.*, **5**, 1061–1062 (1970).
105. M. S. Anderson and C. A. Swenson, Experimental compressions for sodium, potassium, and rubidium metals to 20 kbar from 4.2 to 300 K, *Phys. Rev. B*, **28**, 5395–5418 (1983).
106. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B*, **46**, 6671–6687 (1992).
107. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B*, **48**, 4978 (1993).
108. A. E. Mattsson, R. Armiento, P. A. Schultz, and T. R. Mattsson, Nonequivalence of the generalized gradient approximations PBE and PW91, *Phys. Rev. B*, **73**, 195123 (2006).
109. M. H. DeGroot and M. J. Schervish, *Probability and Statistics,* 3rd ed., Addison-Wesley, Boston (2002).
110. E. Scerri, Explaining the periodic table, and the role of chemical triads, *Found. Chem.*, **12**, 69–83 (2010).
111. E. Scerri, Which elements belong in group 3? *J. Chem. Educ.*, **86**, 1188 (2009).
112. P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.-H. Pao, B. Igelnik, M. Oxley, B. Bakshi, P. Chen, and S. Iwata, Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number, *J. Alloys Compd.*, **317-318**, 26–38 (2001).
113. J. L. Rodgers and W. A. Nicewander, Thirteen ways to look at the correlation coefficient, *Amer. Statist.*, **42**, 59–66 (1988).
114. S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, *J. Comput. Chem.*, **27**, 1787–1799 (2006).
115. K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Higher-accuracy van der Waals density functional, *Phys. Rev. B*, **82**, 081101(R) (2010).
116. N. G. Pace, G. A. Saunders, and Z. Sümengen, The elastic constants and interatomic binding forces in arsenic, *J. Phys. Chem. Solids*, **31**, 1467–1476 (1970).
117. D. D. Richardson and J. Mahanty, Van der Waals contribution to the binding energy of noble metals, *J. Phys. C Solid State Phys.* **10**, 3971–3976 (1977).
118. S. Appalakondaiah, G. Vaitheeswaran, S. Lebègue, N. E. Christensen, and A. Svane, Effect of van der Waals interactions on the structural and elastic properties of black phosphorus, *Phys. Rev. B*, **86**, 035105 (2012).
119. D. Hobbs, J. Hafner, and D. Spišák, Understanding the complex metallic element Mn. I. Crystalline and noncollinear magnetic structure of α-Mn, *Phys. Rev. B*, **68**, 014407 (2003).

120. D. J. Singh and J. Ashkenazi, Magnetism with generalized-gradient-approximation density functionals, *Phys. Rev. B*, **46**, 11570–11577 (1992).
121. R. B. Roberts, G. K. White, and E. Fawcett, Thermal expansion of Cr and Cr-V alloys, *Physica B (Utrecht)*, **119**, 63–67 (1983).
122. N. Gaston, B. Paulus, K. Rosciszewski, P. Schwerdtfeger, and H. Stoll, Lattice structure of mercury: Influence of electronic correlation, *Phys. Rev. B*, **74**, 094102 (2006).
123. U. Wedig, M. Jansen, B. Paulus, K. Rosciszewski, and P. Sony, Structural and electronic properties of Mg, Zn, and Cd from Hartree-Fock and density functional calculations including hybrid functionals, *Phys. Rev. B*, **75**, 205123 (2007).
124. D. D. Koelling and B. N. Harmon, A technique for relativistic spin-polarized calculations, *J. Phys. C Solid State*, **10**, 3107–3114 (1977).
125. P. H.T. Philipsen and E. J. Baerends, Relativistic calculations to assess the ability of the generalized gradient approximation to reproduce trends in cohesive properties of solids, *Phys. Rev. B*, **61**, 1773–1778 (2000).
126. V. Ozoliņš and M. Körling, Full-potential calculations using the generalized gradient approximation: Structural properties of transition metals, *Phys. Rev. B*, **48**, 18304–18307 (1993).
127. P. Haas, F. Tran, P. Blaha, K. Schwarz, and R. Laskowski, Insight into the performance of GGA functionals for solid-state calculations, *Phys. Rev. B*, **80**, 195109 (2009).
128. A. R. Oganov and C. W. Glass, Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications, *J. Chem. Phys.*, **124**, 244704 (2006).
129. N. L. Abraham and M. I. J. Probert, A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction, *Phys. Rev. B*, **73**, 224104 (2006).
130. N. L. Abraham and M. I. J. Probert, Improved real-space genetic algorithm for crystal structure and polymorph prediction, *Phys. Rev. B*, **77**, 134117 (2008).
131. A. N. Kolmogorov, S. Shah, E. R. Margine, A. F. Bialon, T. Hammerschmidt, and R. Drautz, New superconducting and semiconducting Fe-B compounds predicted with an *ab initio* evolutionary search, *Phys. Rev. Lett.*, **105**, 217003 (2010).
132. P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka, and J. Luitz, WIEN2k, An augmented plane wave + local orbitals program for calculating crystal properties, Karlheinz Schwarz, Techn. Universität Wien, Austria (1999).
133. E. Sjöstedt, L. Nordström, and D. J. Singh, An alternative way of linearizing the augmented plane-wave method, *Solid State Commun.*, **114**, 15–20 (2000).
134. G. K. H. Madsen, P. Blaha, K. Schwarz, E. Sjöstedt, and L. Nordström, Efficient linearization of the augmented plane-wave method, *Phys. Rev. B*, **64**, 195134 (2001).
135. S. R. Bahn and K. W. Jacobsen, An object-oriented scripting interface to a legacy electronic structure code, *Comput. Sci. Eng.*, **4**, 56–66 (2002).
136. J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Real-space grid implementation of the projector augmented wave method, *Phys. Rev. B*, **71**, 035109 (2005).
137. J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K.H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method, *J. Phys. Condens. Matt.*, **22**, 253202 (2010).
138. E. Rapoport and G. C. Kennedy, Phase diagram of manganese to 40 kbars, *J. Phys. Chem. Solids*, **27**, 93–98 (1966).
139. H. Luo, R. G. Greene, and A. L. Ruoff, β-Po phase of sulfur at 162 GPa: X-ray diffraction study to 212 GPa, *Phys. Rev. Lett.*, **71**, 2943–2946 (1993).
140. K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, The ground state elemental crystals as a benchmark set for solid state DFT: intrinsic accuracy and code comparison, (2012), arXiv:1204.2733 [cond-mat.mtrl-sci].
141. M. Dułak, Private Communication.
142. http://molmod.ugent.be/DeltaCodesDFT.

APPENDIX: CALCULATING THE Δ-FACTOR

For a particular compound, the energy difference between the equations of state of WIEN2k (w) and a code under investigation (c) can be evaluated analytically, using the Birch-Murnaghan relation of Equation (2). Some mathematical manipulations yield

$$\int_{V_i}^{V_f} (E^c(V) - E^w(V))^2 \mathrm{d}V = F(V_f) - F(V_i), \qquad [A.16]$$

where the primitive function $F(V)$ can be written as a power series in $V^{-1/3}$:

$$F(V) = \sum_{n=-2}^{4} x_n V^{-(2n+1)/3}. \qquad [A.17]$$

The coefficients x_n are given by

$$x_n = -\frac{3}{2n+1} \sum_{i+j=n+2} \left(a_i^c - a_i^w\right)\left(a_j^c - a_j^w\right) \qquad [A.18]$$

with $i, j \in \{0, 1, 2, 3\}$. Then x_{-1}, for example, becomes

$$x_{-1} = 6\left(a_1^c - a_1^w\right)\left(a_0^c - a_0^w\right). \qquad [A.19]$$

The constants a_i are the coefficients of the Birch-Murnaghan equation in its polynomial form:

$$a_3 = \frac{9V_0^3 B_0}{16}(B_1 - 4) \qquad [A.20]$$

$$a_2 = \frac{9V_0^{7/3} B_0}{16}(14 - 3B_1) \qquad [A.21]$$

$$a_0 = \frac{9V_0^{5/3} B_0}{16}(3B_1 - 16) \qquad [A.22]$$

$$a_0 = \frac{9V_0 B_0}{16}(6 - B_1). \qquad [A.23]$$

When evaluating Equation (A.18), a_i^c stands for the coefficient of the code under test, while a_i^w means it corresponds to the WIEN2k equation of state.

PAPER 2

"*Ab initio* based thermal property predictions at a low cost: An error analysis"

K. Lejaeghere, J. Jaeken, V. Van Speybroeck, and S. Cottenier

Phys. Rev. B **89**, 014304 (2014)

PHYSICAL REVIEW B **89**, 014304 (2014)

Ab initio based thermal property predictions at a low cost: An error analysis

Kurt Lejaeghere,[1] Jan Jaeken,[1] Veronique Van Speybroeck,[1] and Stefaan Cottenier[1,2]
[1]*Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*
[2]*Department of Materials Science and Engineering, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*
(Received 8 November 2013; published 28 January 2014)

Ab initio calculations often do not straightforwardly yield the thermal properties of a material yet. It requires considerable computational efforts, for example, to predict the volumetric thermal expansion coefficient α_V or the melting temperature T_m from first principles. An alternative is to use semiempirical approaches. They relate the experimental values to first-principles predictors via fits or approximative models. Before applying such methods, however, it is of paramount importance to be aware of the expected errors. We therefore quantify these errors at the density-functional theory level using the Perdew-Burke-Ernzerhof functional for several semiempirical approximations of α_V and T_m, and compare them to the errors from fully *ab initio* methods, which are computationally more intensive. We base our conclusions on a benchmark set of 71 ground-state elemental crystals. For the thermal expansion coefficient, it appears that simple quasiharmonic theory, in combination with different approximations to the Grüneisen parameter, provides a similar overall accuracy as exhaustive first-principles phonon calculations. For the melting temperature, expensive *ab initio* molecular-dynamics simulations still outperform semiempirical methods.

DOI: 10.1103/PhysRevB.89.014304 PACS number(s): 65.40.De, 64.70.dj, 06.20.Dk, 71.15.Mb

I. INTRODUCTION

Density-functional theory (DFT) [1–3] is used increasingly often to support experimental research. It moreover allows predicting a compound's characteristics without raw materials or expensive apparatus. Even conditions inaccessible to experiment, such as those in the Earth's core, are not particularly more troublesome than ambient pressure and temperature [4]. Unfortunately, not all properties directly follow from simple DFT simulations. Static DFT calculations are practically limited to 0 K, complicating the determination of thermal properties.

Two such properties are the volumetric thermal expansion coefficient α_V and the melting temperature T_m. They can be predicted from an *ab initio* approach, but only by going beyond single-point DFT computations, using either high-temperature molecular dynamics [5,6] or phonon-based calculations [7]. These quickly become computationally intensive. Alternatively, semiempirical approximations provide the same information at a much lower cost [8–12]. They significantly reduce the calculation time and effort by relating α_V and T_m to much simpler DFT predictors, such as the cohesive energy or the pressure derivative of the bulk modulus. These methods are based on approximative models or fits to experiment.

A semiempirical approach is often the only pragmatic way to ensure a fast interaction between experimental and computational research. This is especially true when the design of new materials is concerned. Only when the resulting accuracy does not suffice or when the physical mechanisms behind a particular compound are unclear, may more expensive calculations be justified. Semiempirical approximations do lead to additional errors, however. These errors originate from the less-than-perfect correlation between the investigated property and its predictor. In particular, one simple expression often does not fully represent the physics behind a material's behavior. Describing it with a predictor hence oversimplifies reality, causing semiempirical predictions to deviate from the experimental values.

The semiempirical relations for α_V and T_m should therefore be treated with caution. Although they expedite calculations substantially, this benefit must be weighed against the introduction of additional (non-DFT) errors. To decide how useful a particular semiempirical approach is, its errors need to be quantified. For that reason, we performed an error analysis similar to that in Ref. [13], which was dedicated to properties directly from DFT, such as crystal volume or cohesive energy. The procedure investigates the agreement between experimental values and (semi)theoretical predictions for 71 elemental crystals and characterizes the remaining discrepancies in terms of systematic deviations and residual error bars. Applied to α_V and T_m, such a procedure allows comparing the different semiempirical approximations and evaluating the difference with strictly first-principles methods.

The remainder of this paper is therefore structured as follows. Section II outlines which purely theoretical approaches and semiempirical relations can be employed to predict α_V and T_m, while Sec. III elaborates on the computational aspects of our predictor calculations and error analyses. These methods are then applied in Sec. IV to provide reliable error estimates for the semiempirical approaches. The results are also compared to more advanced predictions. Section V summarizes the most important conclusions.

II. HIGH-LEVEL THEORY AND SEMIEMPIRICAL ALTERNATIVES

A. Thermal expansion coefficient

The thermal expansion coefficient nicely illustrates that a purely theoretical framework leads to computationally expensive simulations. High-precision results are obtained in the quasiharmonic approximation, by calculating the volume-dependent phonon spectrum $\omega_i(V)$ as a function of wave vector and mode [7]. Using thermodynamical formulas, this allows establishing the temperature-dependent free energy, which reaches a minimum at its equilibrium volume V_0. The shift of

V_0 as a function of T yields the volumetric thermal expansion coefficient $\alpha_V(T)$. However, the phonon spectrum consists of the eigenvalues of the dynamical matrix, and this matrix includes the derivatives of the forces with respect to all atomic displacements. Unit cells with M inequivalent atoms therefore require the assessment of three independent displacements for each of these atoms. For small unit cells, periodic boundary conditions moreover cause the displaced atoms to interact with their periodic images. To prevent such undesired interactions, the unit cell needs to be expanded to several times its original size. Finally, only stringent computational settings yield numerically converged forces. The overall result is a set of (at least) $3M$ supercell calculations, each of which is computationally intensive due to its considerable cell size and required accuracy.

The method proposed by Tsuru *et al.* [8] offers a much simpler alternative. Inspired by the link between lattice expansion and crystalline cohesion, the authors fitted inverse relations between the experimental linear thermal expansion coefficient $\alpha_l(T)$ and the DFT-PW91 [14,15] cohesive energy ΔE_{coh}. At room temperature (T_{rt}), they obtained

$$\alpha_V^{Ts} = 3\,\alpha_l^{Ts} = 3\,\frac{48.14\times10^{-6}\ \text{eV/K/atom}}{\Delta E_{\text{coh}}} \tag{1}$$

by fitting to a number of pure metals, as well as binary oxides, nitrides, borides, and carbides.

Other semiempirical approaches are based on thermodynamical considerations. In the quasiharmonic approximation, it can be shown that [7]

$$\alpha_V = \frac{\gamma C_V}{B_0 V_0} \tag{2}$$

with

$$\gamma = \frac{\sum \gamma_i C_{V,i}}{C_V} \tag{3}$$

and

$$\gamma_i = -\frac{\partial \ln \omega_i}{\partial \ln V}. \tag{4}$$

Here, B_0 stands for the equilibrium bulk modulus, C_V for the isochoric heat capacity, and γ for the overall Grüneisen parameter. The index i again labels the different phonon modes, such that within the 1D Einstein model:

$$C_{V,i} = k_B\left(\frac{\hbar\omega_i}{k_B T}\right)^2 \frac{\exp\left(\frac{\hbar\omega_i}{k_B T}\right)}{\left[\exp\left(\frac{\hbar\omega_i}{k_B T}\right)-1\right]^2} \tag{5}$$

(with k_B the Boltmann constant). The key quantities in Eq. (2), however, are the Grüneisen parameters γ_i, which express the volume dependence of the phonon frequencies and can hence be considered as a measure of the anharmonicity of the crystal. Indeed, purely harmonic crystals only contain volume-independent phonons, so thermal expansion does not occur there [$\alpha_V(T) = 0, \forall\, T$].

Equation (2) can be directly completed with information from phonon calculations, using Eq. (4). More approximatively, however, it is also possible to determine γ from the pressure derivative of the bulk modulus B_1. Such a relation effectively replaces thermal effects by an overall (pressurelike) bulk effect and therefore does not require computing phonons.

Several formulations are available for $\gamma(B_1)$, the most popular ones being those of Slater [9], Dugdale and MacDonald [10], and Vashchenko and Zubarev [11]:

$$\gamma^S = -\tfrac{1}{6} + \tfrac{1}{2}B_1, \tag{6}$$

$$\gamma^{\text{DM}} = -\tfrac{1}{2} + \tfrac{1}{2}B_1, \tag{7}$$

$$\gamma^{\text{VZ}} = -\tfrac{5}{6} + \tfrac{1}{2}B_1. \tag{8}$$

They yield

$$\alpha_V^S = \left(-\frac{1}{6} + \frac{1}{2}B_1\right)\frac{C_V}{B_0 V_0}, \tag{9}$$

$$\alpha_V^{\text{DM}} = \left(-\frac{1}{2} + \frac{1}{2}B_1\right)\frac{C_V}{B_0 V_0}, \tag{10}$$

$$\alpha_V^{\text{VZ}} = \left(-\frac{5}{6} + \frac{1}{2}B_1\right)\frac{C_V}{B_0 V_0}. \tag{11}$$

γ can also be derived from the relation between a material's binding energy and atomic separation. By employing a universal binding-energy relation for metals [16], Guinea and co-workers [12] found an expression that does not explicitly depend on B_1:

$$\gamma^G = 0.38\,\frac{r_{\text{WS},0}}{l_{\text{TF}}} \tag{12}$$

with $r_{\text{WS},0}$ the equilibrium Wigner-Seitz radius and l_{TF} the Thomas-Fermi screening length:

$$l_{\text{TF}} = \sqrt{\frac{\Delta E_{\text{coh}}}{12\pi r_{\text{WS},0} B_0}}. \tag{13}$$

Completing Eq. (2) with Eq. (12) then yields a relation similar to that of Tsuru *et al.* [Eq. (1)]:

$$\alpha_V^G = 10.35\left(\frac{l_{\text{TF}}}{r_{\text{WS},0}}\right)\left(\frac{k_B}{\Delta E_{\text{coh}}}\right). \tag{14}$$

Contrary to Eq. (1), however, the prefactor is now compound-dependent.

B. Melting temperature

Just like α_V, the melting temperature T_m cannot be obtained from a simple DFT calculation either. Moreover, the quasiharmonic approximation fails at temperatures that high, so even most phonon-based results become inadequate. *Ab initio* melting temperatures therefore require molecular-dynamics calculations of the solid and/or liquid sample near the melting point. Several versions of this strategy are available. It is possible to investigate the evolution of a one- or two-phase sample into a solid or liquid monophase [5], for example, and map the phase diagram by changing the ambient conditions. Alternatively, the coexistence between both phases can be studied as well [6], maintaining a biphase sample at all times. In addition to these two techniques, other approaches allow establishing the melting temperature too [17], but they have

until now not been possible without using fitted potentials, and can therefore not be designated as purely "first principles" yet.

In the monophase methods, where only one phase remains at the end of the simulation, T_m follows from tuning the temperature and pressure, and monitoring the resulting state (solid or liquid). This can be done using NVE, NVT, or NPT ensembles, where either the particle number N, the volume V, the energy E, the temperature T or the pressure P are held constant. The melting temperature is then found as the transition point between the two states of matter. It has been shown that the most accurate monophase results are obtained when the initial configuration contains a solid-liquid interface [18], the so-called two-phase approach [19–21]. When this is not the case, nucleation of the competing phase is hindered, leading to superheated or undercooled states and hence to substantial errors on the predicted melting temperature. One-phase calculations do offer the smallest computational effort, however, and two strategies exist to remedy the superheating issue in solid-to-liquid simulations. The Z method was proposed by Belonoshko and co-workers [22] and retrieves the *ab initio* melting temperature by letting the system relax from its superheated melting point to the NVE ensemble equilibrium [23]. The method is named after the characteristic shape of the corresponding isochores. The void method, on the other hand, introduces voids into the solid sample [24], facilitating the melting process in a similar way as the solid-liquid interface does in the two-phase approach. However, as far as we know, void-induced melting calculations have not yet been attempted at the DFT level.

Contrary to the monophase methods, the necessary melting information can also be extracted from the first-principles solid-liquid equilibrium. This is the so-called coexistence method [25]. Here, an NVE [6] or NPH [26] ensemble is simulated for a biphase medium (containing both the solid and the liquid). With the NVE ensemble, each starting energy characterizes another point (P,T) of the phase equilibrium. This allows establishing the melting curve point by point. The NPH ensemble, on the other hand, allows to straightforwardly select a predefined pressure, for which the corresponding melting temperature is obtained.

In stark contrast to these rather intricate procedures, the melting temperature is known to correlate well with the cohesive energy [27]. Both properties are closely linked to the material's bond strength, and ΔE_{coh} readily follows from routine DFT calculations. A theoretical proportionality has been proposed by Guinea and co-workers [12], based on their universal binding-energy relation for metals [16]. They assumed the rms displacement of the atoms at T_m to equal the Thomas-Fermi screening length and obtained

$$T_m = 0.032\frac{\Delta E_{\text{coh}}}{k_B}. \tag{15}$$

In addition to this theoretical equation, more empirical forms exist as well. Li *et al.*, for example, found a proportionality similar to Eq. (15) not only to be valid for pure metals, but even to approximate the melting behavior of intermetallic structures such as CsCl-type compounds [28] and Laves phases [29]. Tateno mentions a proportionality between the cohesive energy and the melting temperature too, but he includes additional materials properties in the prefactor [30]. In this study, we restricted the ΔE_{coh} proportionality to a relation similar to Eq. (15), i.e., a single linear function of the cohesive energy with a constant prefactor.

The melting temperature also correlates well with elastic moduli. Indeed, just like the cohesive energy, they express the resistance of the material against deformation, which is needed to initiate melting. Fine *et al.*, for example, used this connection to predict $\langle C_{ii}\rangle$ from T_m [31]. Kittel moreover observes that "melting temperatures and bulk moduli vary roughly as the cohesive energies" [27], which implies that bulk moduli and melting temperatures are related. In this work, we therefore examined B_0 as a predictor for T_m as well [see further, Eq. (17)].

III. METHODOLOGY

We evaluated the performance of all semiempirical relations for α_V and T_m (Sec. II) using the error analysis method by Lejaeghere *et al.* [13]. This procedure is based on a large test set, containing 71 ground-state elemental crystals up to radon (not including the lanthanides), for which the correspondence with experiment is checked. It results in quantitative estimates for the systematic and residual errors.

A. Computational details

For the elemental test set, all calculations were performed using the projector augmented-wave method (PAW) [32,33] in the Vienna *ab initio* simulation package (VASP) [34,35] (version 5.2.2). Spin polarization was taken into account for Fe, Co, and Ni (ferromagnetic), O and Cr (antiferromagnetic), and Mn (ferrimagnetic), while for the heaviest elements (as from the 5*d* transition metals), spin-orbit coupling was included. We used the atomic potentials recommended by the VASP manual [36], which we also listed in Ref. [37]. For the description of exchange and correlation, the functional by Perdew, Burke, and Ernzerhof (PBE) [38] was selected. As it is one of the most popular functionals to date [39], an error assessment for PBE is of more use to the community than for most other functionals. Admittedly, Eq. (1) was originally fitted to PW91 data, but results from this functional are very similar to those from PBE [38], so we directly applied Eq. (1) to PBE as well.

To ensure that all error estimates are independent of the code settings, it is essential to numerically converge each DFT-based property. The cutoff energy was therefore set at 400 eV, except for He, B, C, N, O, F, and Ne, where it was 600 eV. We moreover took $6750/N$ k points in the first Brillouin zone for each N-atom cell, using a Monkhorst-Pack grid [40]. The shape and atomic positions of each unit cell were relaxed until all forces were smaller than 10^{-2} eV/Å, and the self-consistent electronic cycles were converged up to 10^{-4} eV. Combined, such settings led to a numerical convergence of 1 meV/atom in the total energy. These results were largely independent of the used code as well, since different implementations seldom change the equations of state more than a few meV/atom [13,41].

Using these computational parameters, α_V and T_m were determined semiempirically for each of the test set crystals. This required only a limited number of DFT predictors (Sec. II): the cohesive energy ΔE_{coh}, the bulk modulus B_0, the equilibrium volume per atom V_0, and the pressure derivative of the bulk modulus B_1. When necessary, the

isochoric heat capacity C_V was set to its high-temperature limit $3k_B$. For some elements the cohesive energy was slightly modified. In particular, all semiempirical relations use ΔE_{coh} as a predictor for the crystalline cohesion [Eqs. (1), (14), and (15)], but for dimeric crystals, it is mainly the cohesion between diatomic molecules that matters. These entities are maintained throughout expansion, until even after melting. A good predictor for α_V or T_m should therefore incorporate that information. To emphasize the difference from ΔE_{coh}, we named this quantity the "moleculization" energy ΔE_{mol}:

$$\Delta E_{\text{mol}} = -(E_{\text{solid}} - E_{\text{gaslike}}). \tag{16}$$

For most compounds, it does correspond to the atomization (or cohesive) energy, but for dimeric crystals, it yields a more suitable predictor.

The moleculization energy was determined for all elemental compounds by subtracting the energy of isolated atoms or molecules from that of the equilibrium crystal structure. Since VASP employs periodic boundary conditions, that isolated particle can only be simulated in a periodic environment as well. All atoms and molecules were therefore calculated in an orthorhombic cell of approximately $15 \times 15 \times 15$ Å^3. These dimensions allowed to sufficiently suppress the unphysical interaction between periodic images (<1 meV/atom). In most cases this yielded reliable single-particle energies, but when fractional energy-level occupancies were found or when the predicted ground-state electron configuration differed from experiment, the self-consistent cycle was manually constrained. Only for the spin-orbit-coupled calculation of the Pb atom, this was not possible, so there the PBE ground state $^1S_0$ was used instead of the experimental $^3P_0$ state.

The mechanical parameters V_0, B_0, and B_1 were computed from a 13-point equation of state, homogeneously spaced between $V = 0.94V_0$ and $V = 1.06V_0$. Each of these crystals was optimized with respect to the cell shape and ionic positions, followed by a single-point energy calculation. The data points were then fitted to a four-parameter Birch-Murnaghan equation [42], from which the relevant properties were extracted. Although some semiempirical equations were meant to be completed with room-temperature values, such as Eq. (2), we assumed thermal corrections on the input data to play only a minor role.

B. Error analysis

For both properties, α_V and T_m, reliable error estimates were determined by applying a least-squares linear regression $Y = \beta_1 X + \beta_0 + \epsilon$ to the experimental values Y as a function of the (DFT-based) semiempirical predictions X [43] (see Fig. 1). The regression slope β_1 and the intercept β_0 express the systematic deviation between experiment and theory, while the standard error of the regression (SER), equal to the standard deviation of the zero-centered, normally distributed error ϵ, provides an error bar for the fit.

If numerical convergence is achieved in all DFT calculations, then β_1, β_0, and the SER are intrinsic to the used model, with an ideal model yielding $\beta_1 = 1$, $\beta_0 = 0$, and SER $= 0$. In practice, however, there is both a systematic deviation and a residual scatter, which can be attributed to the semiempirical relation and to the DFT predictions themselves. Some data

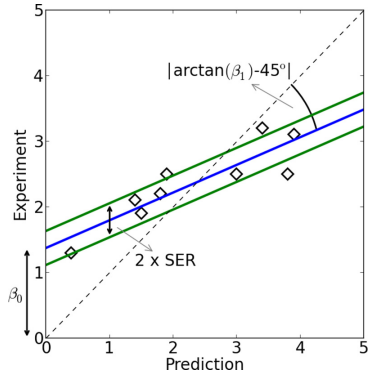


FIG. 1. (Color online) The error on semiempirical predictions can be quantified by the systematic deviation (β_0 and β_1) and the residual error bar (SER), which follow from a least-squares linear regression between the experimental and theoretical values. For this hypothetical data set, the middle full line (in blue) represents the regression curve, and the dashed line represents the first-quadrant bisector.

points even deviate so strongly that they become outliers with respect to the general trend. The resulting distortion of the linear regression substantially complicates the search for a realistic fit. To deal with such deviating data points, an exclusion criterion was developed to identify materials which the semiempirical predictions fail to appropriately describe. A flowchart of the method is presented in Fig. 2, consisting of two parts: as it is hard to automatically identify outliers when they distort the regression curve, a RANSAC procedure [44] was first performed to minimize the influence of outliers on the fit (upper part of Fig. 2); RANSACs are hard to combine with statistical considerations, however, so the resulting regression line was then used as a starting point for a more refined removal of outliers (bottom row of Fig. 2).

RANSAC algorithms identify the most probable regression curve by comparing many inlier-outlier combinations. To avoid bias, that process is randomized. For our 50- to 70-point data sets, we randomly took 100 two-point samples (Fig. 2: Nos. 1 to 100, first row) and collected inliers (filled symbols) in an iterative fashion (reversed arrows). This was done by enforcing a threshold of seven times the SER (using the overall SER in the first step) and adding all points that lay closer to the fit (second row). Then the regression line and the SER were re-established (third row), and the procedure was repeated until no more points could be added. Using such a flexible threshold allows detecting the most important outliers, while losing none of the inliers. Afterwards, the quality of each of the 100 fits was defined as the median error of the inliers, and only the fit with the smallest median of squared residuals was withheld.

The resulting set of preliminary inliers was used to remove additional outliers in a statistically rigorous way, again working in an iterative and elementwise fashion (bottom row of Fig. 2): we excluded the compound with the most extreme externally studentized residual and redid the fit, for as long as that residual belonged to the outer 0.1% of a

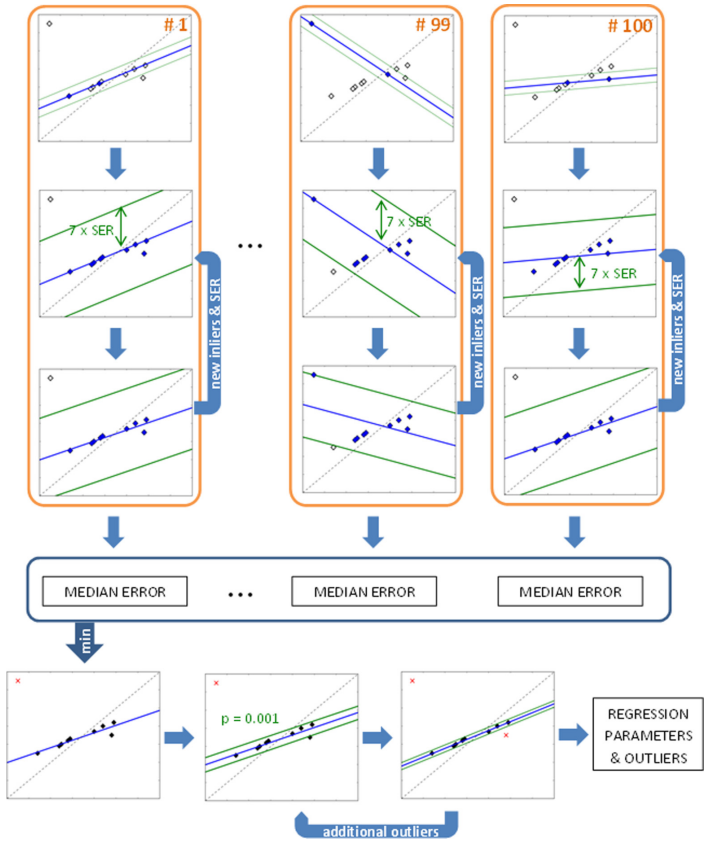


FIG. 2. (Color online) Flowchart of the applied error analysis procedure for a hypothetical data set. First a RANSAC approach [44] is used (top four rows), comparing 100 randomly initiated regression lines (Nos. 1 to 100) and retaining the one with the smallest median error. Afterwards, that resulting fit is employed for a statistically more rigourous outlier exclusion (bottom row). A more elaborate explanation is provided in the text.

Student's t distribution (two-sided p value) [43]. Although such a stringent significance threshold might not get rid of all outliers, it prevents inliers from being excluded and the remaining data set already yields reliable estimates for the regression parameters.

IV. RESULTS AND DISCUSSION

A. Thermal expansion coefficient

By fitting experimental results [45–51] to predicted ones, we estimated the errors introduced by the different semiempirical α_V approximations. This yielded a measure of the systematic and residual errors (Sec. III B). The results are summarized in Table I and are essentially independent of the used code (see Ref. [37] for a comparison of error estimates for the Slater approximation using VASP and GPAW [52–54]). Individual residual errors are shown in Tables II and III and Fig. 3, also indicating the outliers, which were excluded from the linear regression (hatched elements or open symbols). A complete overview of all theoretical and experimental numbers is presented in Ref. [37].

1. Elementwise deviations

Table I and Fig. 3 confirm that the correspondence between the different semiempirical predictions and experiment is not perfect. First, there is some systematic deviation, as most regression slopes and intercepts deviate from the ideal values ($\beta_1 = 1, \beta_0 = 0$). The two-sided p values with respect to a null hypothesis of $\beta_1 = 1$, for example, are smaller than 0.1% for

TABLE I. Error estimates for the semiempirically predicted volumetric thermal expansion coefficients α_V (in 10^{-5} K^{-1}) and melting temperatures T_m (in degrees Kelvin), based on a least-squares regression $Y = \beta_1 X + \beta_0$ between experimental data Y and semiempirical data X for the ground-state elemental crystals. 95% confidence intervals have been provided for the standard error of the regression (SER) and the regression parameters β_1 and β_0. The columns headed $P(\beta_1 = 1)$ and $P(\beta_0 = 0)$ represent the two-sided probabilities that, if in fact $\beta_1 = 1$ or $\beta_0 = 0$, respectively, the current regression parameters are obtained. The final column displays the number of outliers (see Sec. III B) with respect to the total number of test set crystals.

Approximation		SER	β_1	β_0	$P(\beta_1 = 1)$	$P(\beta_0 = 0)$	Number of outliers
α_V (Slater)	[Eq. (9)]	$1.7^{+0.4}_{-0.3}$	0.852 ± 0.041	-0.6 ± 0.8	4×10^{-9}	0.12	10/56
α_V (Dugdale-MacDonald)	[Eq. (10)]	$1.8^{+0.5}_{-0.3}$	1.056 ± 0.051	-0.7 ± 0.8	0.03	0.07	10/56
α_V (Vashchenko-Zubarev)	[Eq. (11)]	$2.1^{+0.5}_{-0.4}$	1.354 ± 0.072	-0.8 ± 1.0	6×10^{-13}	0.10	10/56
α_V (Guinea *et al.*)	[Eq. (14)]	$1.9^{+0.5}_{-0.3}$	0.894 ± 0.037	-0.8 ± 0.8	4×10^{-7}	0.05	2/56
α_V (Tsuru *et al.*)	[Eq. (1)]	$2.4^{+0.6}_{-0.4}$	1.324 ± 0.061	-1.4 ± 1.0	1×10^{-14}	0.01	3/56
T_m (via B_0)	[Eq. (17)]	466^{+96}_{-68}	1.000 ± 0.003	0 ± 198	1.00	1.00	0/68
T_m (Guinea *et al.*)	[Eq. (15)]	283^{+58}_{-41}	1.042 ± 0.003	8 ± 113	0.00	0.89	0/68

all but one of the α_V approximations: if the correct regression slope were 1 nevertheless, then the probability of obtaining test set data as extreme as ours would be less than 0.1%. Second, the remaining scatter cannot be neglected either, as appears from the residual errors and outliers (Tables II and III). Indeed, all SERs lie between 1.7 and 2.5×10^{-5} K^{-1}, which is in some cases of the same order of magnitude as the expansion coefficients themselves (see Fig. 3). To explain the largest deviations, we look at a few elements in more detail.

Several approximations of the thermal expansion coefficient rely on accurate DFT values for the volume and the bulk modulus. However, for some subsets of elements, PBE is not able to predict those properties correctly [13]. This is particularly the case for the correlation-dominated materials, the molecular crystals and the noble gases. Indeed, C, S, and Cd were indicated as outliers or have large residual errors in all α_V approximations based on Eq. (2). For the Slater, Dugdale-MacDonald, and Vashchenko-Zubarev approaches, reliable B_1 values are needed as well. Given PBE's inaccurate B_1 predictions for the low-coordination p block compounds and Sr, these α_V methods do not adequately describe P, As, Se, Sb, Te, Bi, and Sr either. Finally, the approximations of Guinea *et al.* and Tsuru *et al.* depend on the cohesive energy. For ΔE_{coh}, the correlation-dominated compounds and the noble gases fall short. It explains why the α_V predictions for Cd, and to a lesser extent Zn, deviate considerably.

However, not all differences between the theoretical and the experimental results are due to the accuracy with which the DFT-based predictors are determined. The semiempirical approximations themselves can also introduce errors, as they do for Cr, for example. We cannot expect any of the investigated approximations to yield meaningful results for this material, as it displays a magnetic phase transformation around room temperature [55]. Such phenomena are not included in any of the equations, which purely relate thermal expansion to vibrational or bonding types of interactions. Strikingly, the diamond-type structures (Si, Ge and Sn) yield unreliable results as well, while from a DFT point of view, the necessary predictors are determined up to a high degree of accuracy. Here too, the experimental expansion behavior contains some strange features, with negative α_V values over a considerable temperature range, but explicit phonon calculations are able to reproduce it (see later, Table V). This indicates that the semiempirical approaches are to blame. They fail to describe the negative Grüneisen parameters of the low-frequency phonon modes, which give rise to the initial low-temperature compression [56]. While the experimental results for γ incorporate these negative contributions, the semiempirical approximations do not. An estimate for γ that is based on bulk properties, only represents the behavior of average phonons, rather than the average phonon behavior: at low temperatures, the low-energy phonons are occupied the most, and their behavior can deviate significantly from the behavior of an average phonon. Bulk γ values hence fail when the low-frequency phonon modes differ too much from the rest. Negative Grüneisen parameters are not only found with the diamond-type compounds, but occur whenever bonds are strengthened during expansion. This happens in certain bending modes [56,57], leading to phonon frequencies that increase with volume. However, overall bond strengthening only occurs when the vibrational mode does not hinder neighboring atoms. The diamond structure, with its tetragonal environment, is ideally suited for such a behavior, but so are compounds with a low coordination number in general. The bad thermal expansion predictions of Se, Sb, and such are therefore as much due to the semiempirical relations themselves as to the unreliable PBE values for B_1. Using experimental values for B_1 does indeed not improve the quality of the predictions (see Tables 2.5 and 2.6 in Ref. [37]).

2. *Comparison between methods*

Despite similar shortcomings in each of the α_V approximations, there is a clear difference in overall quality. Systematic deviations can be corrected for—this is equivalent to proposing a new semiempirical equation—but the remaining scatter is intrinsic to the particular set of predictors and the form of the semiempirical function. The SERs (Table I) suggest that the relations by Slater, Dugdale and MacDonald, and Guinea *et al.* yield the most accurate results. These SERs are not always based on the same number of data points, however. In particular, less outliers are excluded from the regression when the methods of Guinea *et al.* and Tsuru *et al.* are evaluated, even though they have some classic outliers only slightly below the exclusion threshold (see Sec. III B). Indeed, Table III

TABLE II. (Color online) (Absolute) elementwise deviations between the experimental [45–51] room-temperature volumetric thermal expansion coefficients $\alpha_V(T_{rt})$ and least-squares linear regression results based on semiempirical predictions in the Slater [Eq. (9)], Dugdale-MacDonald [Eq. (10)], and Vashchenko-Zubarev [Eq. (11)] approximations (in 10^{-5} K^{-1}). The darkest shades correspond to the largest errors (common color code for Tables II and III) and hatched areas indicate outliers.

$\alpha_{V,rt}$(Slater)

H																	He
Li	Be											B	C	N	O	F	Ne
3.4	1.3											1.1	928				
Na	Mg											Al	Si	P	S	Cl	Ar
3.7	0.4											1.6	2.5	140	2530		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
3.2	0.4	0.4	0.2	0.6	3.9	2.6	0.3	1.2	0.8	0.5	1.1	1.8	3.3	32	106		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.6	6.3	0.1	0.3	0.7	0.3	1.0	0.6	0.7	0.6	0.2	2.4	0.6	3.8	9	23		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
6.9	2.0	0.5	0.1	0.7	0.5	1.1	0.6	0.8	0.9	0.9		1.5	2.0	10	1.1		

$\alpha_{V,rt}$(DM)

H																	He
Li	Be											B	C	N	O	F	Ne
2.3	0.9											1.0	1070				
Na	Mg											Al	Si	P	S	Cl	Ar
4.6	0.5											1.5	2.4	166	3034		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
3.5	0.9	0.1	0.4	0.8	4.5	2.2	0.2	1.2	0.7	0.3	1.9	2.0	3.5	38	125		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.8	7.0	0.4	0.1	0.9	0.4	1.1	0.7	0.6	0.5	0.2	3.5	0.1	4.0	11	27		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
7.9	1.3	0.3	0.3	0.8	0.6	1.2	0.7	0.8	0.8	0.7		2.2	1.7	12	0.7		

$\alpha_{V,rt}$(VZ)

H																	He
Li	Be											B	C	N	O	F	Ne
0.3	0.3											0.9	1270				
Na	Mg											Al	Si	P	S	Cl	Ar
6.3	0.7											1.3	2.4	204	3760		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
4.3	1.6	0.2	0.6	0.9	5.4	1.7	0.1	1.1	0.5	0.0	3.0	2.3	3.8	47	150		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
1.6	7.9	0.8	0.1	1.0	0.5	1.1	0.6	0.5	0.2	0.8	5.0	1.0	4.3	13	33		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
10	0.2	0.0	0.5	0.9	0.7	1.2	0.7	0.7	0.7	0.3		3.3	1.2	15	0.2		

shows that the low-coordination compounds and molecular crystals also perform badly for these methods. To compare all approximations on an equal footing, we looked at the SERs of the common inliers [58] (Table IV, set II). These reduced to 1.7, 1.7, 1.9, 1.6, and 1.9×10^{-5} K^{-1} for the Slater, Dugdale-MacDonald, Vashchenko-Zubarev, Guinea and Tsuru relations respectively. Among the different approximations to α_V, the one by Guinea *et al.* therefore performs best. It incorporates

TABLE III. (Color online) (Absolute) elementwise deviations between the experimental [45–51] room-temperature volumetric thermal expansion coefficients $\alpha_V(T_n)$ and least-squares linear regression results based on semiempirical predictions in the approximations of Guinea *et al*. [Eq. (14)] and Tsuru *et al*. [Eq. (1)] (in 10^{-5} K^{-1}). The darkest shades correspond to the largest errors (common color code for Tables II and III) and hatched areas indicate outliers.

$\alpha_{V,rt}$(Guinea)

H																	He
Li	Be											B	C	N	O	F	Ne
1.2	1.4											0.1	25				
Na	Mg											Al	Si	P	S	Cl	Ar
2.9	1.2											2.7	2.1	3.7	26		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
3.7	1.4	0.6	0.1	0.5	1.4	3.8	1.0	1.6	1.5	1.3	0.9	0.4	2.1	2.3	0.7		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.4	3.2	0.5	0.3	0.7	0.2	1.0	0.7	1.0	0.9	1.2	3.4	3.4	2.8	2.4	2.8		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
3.8	3.3	0.7	0.2	0.7	0.5	1.1	0.8	0.9	1.1	1.3		0.3	2.4	3.1	0.7		

$\alpha_{V,rt}$(Tsuru)

H																	He
Li	Be											B	C	N	O	F	Ne
3.3	0.4											0.8	1.3				
Na	Mg											Al	Si	P	S	Cl	Ar
5.0	3.9											2.7	2.1	5.8	20		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
4.4	1.9	0.2	0.4	0.6	1.9	2.9	1.0	1.5	1.4	0.8	6.9	0.5	2.0	0.0	11		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.2	3.7	0.0	0.0	0.8	0.3	0.7	0.4	0.6	0.3	0.7	15	2.7	3.1	2.4	1.5		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
3.7	2.6	0.3	0.1	0.8	0.5	0.8	0.5	0.5	0.4	0.5		2.3	0.2	3.8	3.4		

the largest number of bulk properties and is not based on B_1, which can not always be reliably obtained from PBE. The other methods do rely on B_1 or contain too few input parameters to fully describe thermal expansion.

3. *Source of the errors*

As mentioned earlier, the SER reflects the influence of two distinct effects. On the one hand, the use of DFT-PBE values for the predictors introduces some errors, since the agreement with experiment is not perfect. On the other hand, no semiempirical method can capture all aspects of reality perfectly. Both issues give rise to an appreciable deviation from the regression line. Their respective contributions to the total SER can be investigated by replacing the PBE predictors by experimental numbers [e.g., by completing Eq. (1) with experimental cohesive energies]. The resulting error is then entirely due to the (semi)empirical relation. By comparing to the original SERs, the influence of using DFT values instead of experimental ones can be evaluated. Such an analysis has been performed in Table IV. For a fair comparison between the PBE- and experiment-based predictions, equal test sets were considered by only assessing elements that are inliers for all of the methods (Table IV, set III). These common inliers are listed in Ref. [37].

The resulting SERs show that taking the predictors from PBE or from experiment does not change the overall scatter on the regression line much. In all cases, the values are very similar, which suggests that the main contribution to the SER is due to the semiempirical relation. The table also shows that the specific error value strongly depends on the size of the used test set. This is because ideal inliers or outliers do not exist, and smaller test sets will inevitably decrease the corresponding error. For the smallest considered set (set III), all semiempirical relations perform equally well. The method of Guinea *et al*. still has the widest applicability, however, since its SERs for sets II and III lie very closely together.

4. *Comparison to high-level theory*

Although some semiempirical approximations are better than others, fully first-principles methods are still expected to yield the most accurate results. The applicability of semiempirical methods hence largely depends on the difference with

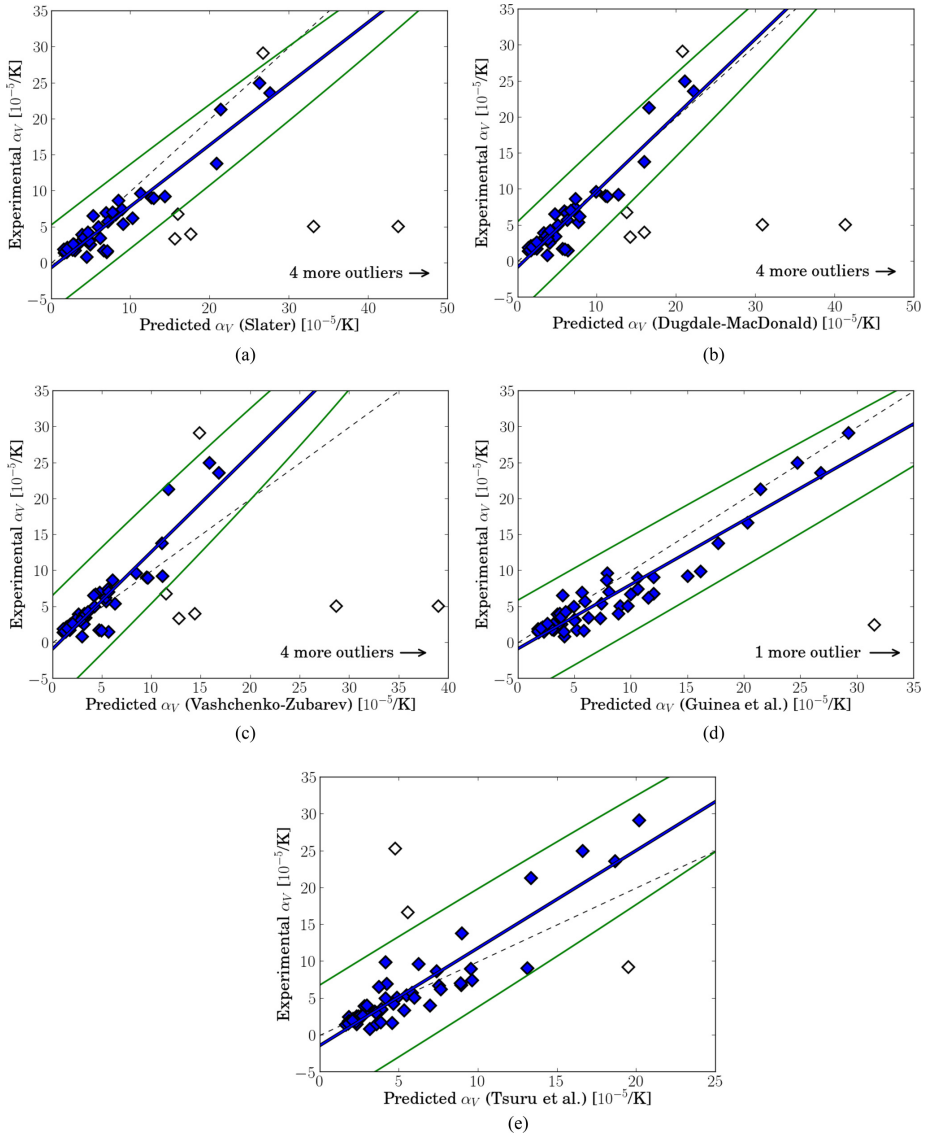


FIG. 3. (Color online) Least-squares linear regression curves (middle full line, in blue) between experimental [45–51] and semiempirically predicted values for the room-temperature volumetric thermal expansion coefficients of the ground-state elemental crystals. Outliers are depicted as open symbols and correspond with two-sided p values smaller than 0.1% for the externally studentized residuals (upper and lower full lines, in green). The dashed line represents the first-quadrant bisector, with which the regression line would coincide if all predictions were perfect.

TABLE IV. Standard errors of the regression (SER) for the semiempirically predicted volumetric thermal expansion coefficients α_V (in 10^{-5} K^{-1}) and melting temperatures T_m (in degrees Kelvin), based on a least-squares regression $Y = \beta_1 X + \beta_0$ between experimental data Y and semiempirical data X for the ground-state elemental crystals. Set I represents the full data set for each of the PBE-based methods (excluding their respective outliers) and is identical to Table I. Sets II and III, on the other hand, only contain the common inliers between all PBE-based predictions (II) or between all PBE- and experiment-based predictions (III) of either α_V or T_m (these sets are listed in Ref. [37]).

Approximation		PBE - set I	PBE - set II	PBE - set III	Exp - set III
α_V (Slater)	[Eq. (9)]	1.7	1.7	1.4	1.3
α_V (Dugdale-MacDonald)	[Eq. (10)]	1.8	1.7	1.4	1.3
α_V (Vashchenko-Zubarev)	[Eq. (11)]	2.1	1.9	1.5	1.5
α_V (Guinea *et al.*)	[Eq. (14)]	1.9	1.6	1.5	1.6
α_V (Tsuru *et al.*)	[Eq. (1)]	2.4	1.9	1.4	1.4
T_m (via B_0)	[Eq. (17)]	466	466	473	496
T_m (Guinea *et al.*)	[Eq. (15)]	283	283	295	282

those high-accuracy predictions. As mentioned in Sec. II A, the thermal expansion coefficient can be determined more rigorously from quasiharmonic phonon theory. Grabowski *et al.*, for example, applied it to the nonmagnetic fcc elemental metals using the PBE functional [59]. They still noticed considerable errors, however, especially for silver and gold. These relatively large errors for the phonon approach were also observed by Souvatzis and Eriksson [60], using PW91 calculations for the 4*d* transition metals. In both studies, the errors on the phonon predictions were of the same magnitude or even larger than the errors on most semiempirical predictions discussed here. Souvatzis and Eriksson suggested the LDA functional to outperform GGAs for transition metals, but only by fortuitous cancellation of errors. In fact, only more advanced functionals substantially improve phonon-based results for the transition metals [61]. PBE predictions are more reliable for non-transition-metal compounds, as is shown by Al and Pb in the study of Grabowski *et al.*, and by our own results for Si and K (Table V, using phonopy [62]; see Ref. [37] for further computational details).

Table V compares some PBE-based phonon results to semiempirical predictions. Systematic deviations were determined and corrected for by establishing a linear regression between these 12 experimental and theoretical data points. Hence, instead of the phonon-based or semiempirical values X themselves, the regression-corrected numbers $\beta_1 X + \beta_0$ are listed. Even for the phonon approach, a small systematic deviation was found, since GGA functionals are known to underbind crystals [63], yielding too large volumes and expansion coefficients.

At first sight, the SER for quasiharmonic phonon theory seems to be best, but this impression is entirely due to the results for silicon. Indeed, as discussed before, all semiempirical approaches fail to reliably represent its thermal expansion. When excluding the influence of Si on the SER, the semiempirical methods offer a similar (or even better) accuracy as the PBE-based phonon predictions, even for nontransition metals (Table V). It is therefore important to realize that explicit phonon calculations only have added value when the semiempirical relations break down and when the PBE functional itself is still valid (e.g., when no dispersion interactions are present). In all other cases, the thermal expansion coefficient can safely be calculated from a semiempirical approach, as this often provides the same degree of accuracy, but at a much smaller computational cost. When that accuracy is insufficient, phonon calculations may prove more suitable, but only by exploring better functionals [61]. The difference in computational effort will increase even further, however.

B. Melting temperature

The melting temperature can be estimated from both the cohesive energy [Eq. (15)] [12] and the bulk modulus [27]. To our knowledge, no explicit relation has been published for the

TABLE V. Volumetric thermal expansion coefficient (in 10^{-5} K^{-1}) according to experiment [45], the (regression-corrected) quasiharmonic phonon approach and the (regression-corrected) approximations by Slater [Eq. (9)], Dugdale-MacDonald [Eq. (10)], Vashchenko-Zubarev [Eq. (11)], Guinea *et al.* [Eq. (14)], and Tsuru *et al.* [Eq. (1)]. The applied linear regressions $Y = \beta_1 X + \beta_0$ are based on the experimental numbers Y and the raw theoretical values X for these 12 compounds, and correct for systematical deviations. Transition metals are distinguished from other element types.

Compound	Exp [45]	QHA	Slater	DM	VZ	Guinea	Tsuru
Ni	4.02	3.23 [64]	3.79	3.74	3.67	3.52	2.70
Cu	4.95	4.76 [59]	5.17	5.24	5.37	4.81	4.55
Rh	2.46	2.44 [59]	2.12	1.97	1.73	2.35	1.85
Pd	3.54	3.60 [59]	3.41	3.41	3.43	3.63	4.14
Ag	5.67	6.57 [59]	6.34	6.70	7.30	5.81	7.17
Ir	1.92	1.90 [59]	1.38	1.16	0.81	1.77	1.22
Pt	2.64	2.94 [59]	2.08	1.95	1.73	2.40	2.26
Au	4.26	5.82 [59]	3.91	4.03	4.24	4.04	5.32
Al	6.93	6.10 [59]	6.18	6.21	6.27	5.55	4.63
Pb	8.67	7.84 [59]	7.65	7.94	8.43	7.80	9.69
Si	0.78	0.67	3.78	3.59	3.28	3.95	2.99
K	24.99	24.95	25.03	24.88	24.55	25.20	24.31
Applied β_1		0.869	0.978	1.230	1.652	1.031	1.588
Applied β_0		−0.1	−0.6	−1.0	−1.7	−0.3	−2.0
SER with Si		0.73	1.10	1.07	1.12	1.15	1.35
SER[a] without Si		0.77	0.58	0.64	0.83	0.59	1.22

[a]Strictly speaking, the systematic deviation should be redetermined when the SER without Si is required. However, this one crystal does not influence the regression line much, and we are only interested in a measure of the effect of Si on the SER.

TABLE VI. (Color online) (Absolute) elementwise deviations between the experimental [27,65,66] melting temperatures T_m and least-squares linear regression results based on semiempirical predictions in the approximations of Eq. (17) and of Guinea *et al*. [Eq. (15)] (in degrees Kelvin). The darkest shades correspond to the largest errors (common color code).

$T_m(B_0)$

H																	He
472																	
Li	Be											B	C	N	O	F	Ne
142	76											162		426	440	442	467
Na	Mg											Al	Si	P	S	Cl	Ar
175	141											181	479	341	97	322	405
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
176	487	887	544	237	204	698	191	416	338	242	263	574	248	398	25	230	372
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
193	468	980	881	864	264	444	497	325	32	23	189	345	270	170	95	112	326
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
197	447	1068	1145	1233	777	18	337	516	413	300	354	124	178	122	233		

$T_m(\Delta E_{coh})$

H																	He
3																	
Li	Be											B	C	N	O	F	Ne
176	112											49		50	38	37	9
Na	Mg											Al	Si	P	S	Cl	Ar
56	332											402	83	474	798	146	39
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
8	362	203	178	94	560	22	74	228	149	5	258	724	243	68	522	209	98
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
5	409	179	302	59	471	171	62	47	386	264	299	473	731	151	218	308	142
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
17	263	344	65	214	270	401	210	1	23	127	144	17	163	267	109		

latter, however. Because the data suggest a linear correlation (Fig. 2.3 in Ref. [37]), we fitted a linear regression between the experimental values of T_m [27,65,66] and the PBE values of B_0, yielding

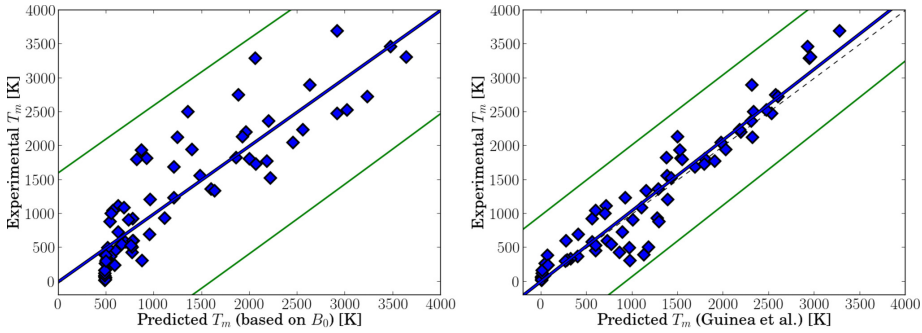
$$T_m = 482.8 + 8.172\, B_0. \tag{17}$$

Of course, there will be no systematic deviation for this equation by definition, but the effects of using PBE are expected to be similar as with the bulk modulus [13]. All other error characteristics remain useful, and are listed in Table I, together with the errors for $T_m(\Delta E_{coh})$. These values again do not depend much on the used code [37]. The individual residual errors are shown in Table VI and are visualized in Fig. 4.

1. Comparison between methods

Predicting a material's melting temperature from B_0 clearly yields less accurate results than using the cohesive energy. The SER is almost double in the first case (Table I). One possible reason is that Eq. (17) employs the relation between T_m and B_0 to capture the link between elasticity and melting behavior, while this link could also be expressed more generally. To check this, we fitted the melting temperature to a linear combination of the elastic constants C_{ij}, using PBE data for the cubic and hexagonal elemental crystals (see Table 2.8 in Ref. [37]). The resulting SER was indeed smaller than for a fit to B_0 (489 K instead of 512 K), but only slightly. We can therefore assume the relation between the melting temperature and the bulk modulus to be sufficiently representative of the physical connection between melting and elasticity.

The different quality of the two semiempirical approaches to T_m must hence be explained differently. Since the effect of using DFT-based predictors is again small (Table IV), that difference is mainly due to the semiempirical approaches themselves. In that respect, it is important to realize that melting goes beyond a purely elastic behavior. When a compound starts melting, the supplied heat needs to break all bonds. This is a process which is well described by the cohesive energy, since it relates to the initial and final stages of the bond breaking. Elastic properties only describe the initial stage of bond stretching. To account for non-negligible deformations, anharmonic effects need to be included [12,67], and a function of B_0 does not contain that information.



(a) (b)

FIG. 4. (Color online) Least-squares linear regression curves (middle full line, in blue) between experimental [27,65,66] and semiempirically predicted values for the melting temperatures of the ground-state elemental crystals. Outliers are depicted as open symbols and correspond with two-sided p values smaller than 0.1% for the externally studentized residuals (upper and lower full lines, in green). The dashed line represents the first-quadrant bisector, with which the regression line would coincide if all predictions were perfect.

2. *Comparison to high-level theory*

Not only the comparison between approximations is of interest, but so is their overall quality. This again requires looking at fully first-principles methods. Unfortunately, the number of high-quality results for the melting temperature is quite small due to the huge computational effort associated with *ab initio* molecular dynamics (AIMD). Table VII lists relevant data for five elemental compounds (i.e., based on a PBE or PW91 functional and at ambient pressure). We only provided values from the coexistence method (see Sec. II B), as these are the most frequently encountered AIMD results, and because it might be imprudent to treat all high-level approaches on the same footing. To correct for systematic deviations, we fitted separate regression lines to these five materials. Even for AIMD results, some deviation was found (and corrected for), since PBE underbinds crystals and therefore affects the simulated melting behavior. Note that the AIMD value for Li is based on the bcc structure (which is the structure at melting), while the semiempirical predictions relate to an hR9 crystal. The cohesive energy and the bulk modulus are very similar in both structures, however (a 0.1 kJ/mol and a 0.2 GPa difference, respectively), leading to almost identical melting temperature predictions.

Table VII shows that most T_m calculations are quite accurate, especially those from molecular dynamics or the relation by Guinea *et al.* Indeed, not only do the (regression-corrected) AIMD results correspond almost perfectly to the experimental values, the quality of Eq. (15) is also extremely good, particularly when the simplicity of the underlying computation is considered. Contrary to the predictions of α_V, however, the more fundamental method (in this case AIMD) remains the most accurate for T_m.

TABLE VII. Melting temperature (in degrees Kelvin) from experiment, (regression-corrected) fully *ab initio* molecular dynamics calculations (AIMD), and (regression-corrected) semiempirical relations in terms of ΔE_{coh} [Guinea *et al.*, Eq. (15)] and B_0 [Eq. (17)]. The applied linear regressions $Y = \beta_1 X + \beta_0$ are based on the experimental numbers Y and the raw theoretical values X for these five compounds, and correct for systematical deviations.

Compound	Exp [27]	AIMD	Guinea	via B_0
Li	454	522 [68]	300	535
Al	934	870[a] [6]	1153	1314
Si	1687	1664 [69]	1676	1455
Ta	3293	3316 [70]	3257	2735
W	3695	3690 [71]	3676	4023
Applied β_1		1.052	1.258	1.502
Applied β_0		43	−450	−360
SER		57	156	456

[a]PW91 instead of PBE functional.

V. CONCLUSIONS

Using a test set with all ground-state elemental crystals (up to Rn, not including the lanthanides), we quantified the accuracy of several semiempirical approaches to the thermal expansion coefficient α_V and the melting temperature T_m. The discrepancy between theoretical predictions and experimental values was expressed in terms of a systematic and a residual error. The systematic deviation was determined from the slope of a linear regression curve between experiment and theory, while the remaining error bar was defined as the standard error of the regression. Outliers were flagged based on their externally studentized residuals, allowing us to identify those compounds for which the α_V or T_m predictions were certainly not reliable anymore. This unreliability could be due to the failure of the semiempirical approach itself or to the limited accuracy of the underlying DFT predictor (with the largest contribution from the former effect, see Table IV). The results from the error analyses are presented in Table I and are summarized below.

For the thermal expansion coefficient, we assessed multiple predictions based on the Grüneisen parameter γ, as well as an inverse proportionality to the cohesive energy. They all yielded very good results, particularly the former approximations, with the best behavior when γ was derived from a universal binding-energy relation for metals [12]. There were a few exceptions, however, as some elemental crystals did not perform well for any of the semiempirical approaches. Several deviations from experiment were caused by the failure of the used PBE functional (when dispersion interactions were present, for example, such as for graphite). On the other hand, it was also shown that neither an expression in γ nor in ΔE_{coh} could reproduce effects which were caused by anomalous low-frequency phonon modes, as most notably happens in Si. In such materials, more advanced calculations are necessary, e.g., using quasiharmonic phonon theory. Nevertheless, except for these few special cases, the accuracy of most semiempirical approximations is as good or even better than for explicit phonon calculations (Table V).

The melting temperature was approximated by both a relation to the cohesive energy and to the bulk modulus, with the latter method yielding the largest residual errors. B_0 only relates to elastic phenomena, while melting involves strongly anharmonic effects. By contrast, the quality of the correlation to ΔE_{coh} was quite good. Requiring only a limited computational effort, its results lay about 300 K from the experimental values. *Ab initio* molecular dynamics does offer even better results (Table VII), but at a much higher cost.

Fully first-principles calculations are often regarded as the only means to obtain high-accuracy predictions for materials properties. Nevertheless, some semiempirical approximations offer a cheap alternative with the same quality of results (such as for α_V) or with only marginally larger errors (such as for T_m). These methods are particularly interesting for materials design, where time can be an issue.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research–Flanders (FWO) by means of a fellowship and a project (G.0402.11N) and by the Research Board of Ghent University. Stefaan Cottenier acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. Calculations were carried out using the Stevin Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Hercules Foundation, and the Flemish Government (EWI Department).

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).
[4] D. Alfè, M. J. Gillan, and G. D. Price, Nature (London) **401**, 462 (1999).
[5] T. Ogitsu, E. Schwegler, F. Gygi, and G. Galli, Phys. Rev. Lett. **91**, 175502 (2003).
[6] D. Alfè, Phys. Rev. B **68**, 064423 (2003).
[7] M. T. Dove, *Introduction to lattice dynamics* (Cambridge University Press, Cambridge, 1993).
[8] Y. Tsuru, Y. Shinzato, Y. Saito, M. Shimazu, M. Shiono, and M. Morinaga, J. Ceram. Soc. Jpn. **118**, 241 (2010).
[9] J. C. Slater, Phys. Rev. **57**, 744 (1940).
[10] J. S. Dugdale and D. K. C. MacDonald, Phys. Rev. **89**, 832 (1953).
[11] V. Ya. Vashchenko and V. N. Zubarev, Sov. Phys. Solid State **5**, 653 (1963).
[12] F. Guinea, J. H. Rose, J. R. Smith, and J. Ferrante, Appl. Phys. Lett. **44**, 53 (1984).
[13] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, Crit. Rev. Solid State **39**, 1 (2014).
[14] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
[15] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **48**, 4978 (1993).
[16] J. H. Rose, J. Ferrante, and J. R. Smith, Phys. Rev. Lett. **47**, 675 (1981).
[17] D. Alfè, M. J. Gillan, and G. D. Price, J. Chem. Phys. **116**, 6170 (2002).
[18] J. Bouchet, F. Bottin, G. Jomard, and G. Zérah, Phys. Rev. B **80**, 094102 (2009).
[19] J. D. Kubicki and A. C. Lasaga, Am. J. Sci. **292**, 153 (1992).
[20] J. Mei and J. W. Davenport, Phys. Rev. B **46**, 21 (1992).
[21] A. B. Belonoshko, Geochim. Cosmochim. Acta **58**, 4039 (1994).
[22] A. B. Belonoshko, N. V. Skorodumova, A. Rosengren, and B. Johansson, Phys. Rev. B **73**, 012201 (2006).
[23] A. B. Belonoshko, L. Burakovsky, S. P. Chen, B. Johansson, A. S. Mikhaylushkin, D. L. Preston, S. I. Simak, and D. C. Swift, Phys. Rev. Lett. **100**, 135701 (2008).
[24] J. F. Lutsko, D. Wolf, S. R. Phillpot, and S. Yip, Phys. Rev. B **40**, 2841 (1989).
[25] A. J. C. Ladd and L. V. Woodcock, Mol. Phys. **36**, 611 (1978).
[26] S. Yoo, X. C. Zeng, and S. S. Xantheas, J. Chem. Phys. **130**, 221102 (2009).
[27] C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Wiley, Hoboken, 2005).
[28] C. Li and P. Wu, Chem. Mater. **14**, 4833 (2002).
[29] C. Li, J. L. Hoe, and P. Wu, J. Phys. Chem. Solids **64**, 201 (2003).
[30] J. Tateno, Solid State Commun. **10**, 61 (1972).
[31] M. E. Fine, L. D. Brown, and H. L. Marcus, Scr. Metall. **18**, 951 (1984).
[32] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[33] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[34] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[35] J. Hafner, J. Comput. Chem. **29**, 2044 (2008).
[36] VASP *manual*, http://www.vasp.at (April 23, 2009 version).
[37] See Supplementary Material at http://link.aps.org/supplemental/10.1103/PhysRevB.89.014304 for more computational details, as well as the full calculated and experimental data.
[38] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[39] DFT poll, http://www.marcelswart.eu/dft-poll

[40] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[41] Comparing Solid State DFT Codes, Basis Sets and Potentials, http://molmod.ugent.be/DeltaCodesDFT

[42] F. Birch, Phys. Rev. **71**, 809 (1947).

[43] N. R. Draper and H. Smith, *Applied regression analysis*, 3rd ed., Wiley Series in Probability and Statistics (Wiley-Interscience, New York, 1998).

[44] M. A. Fischler and R. C. Bolles, Comm. ACM **24**, 381 (1981).

[45] *CRC Handbook of Chemistry and Physics*, edited by D. Lide, 89th ed. (CRC Press, Boca Raton, Florida, 2008-2009).

[46] J. B. Nelson and D. P. Riley, Proc. Phys. Soc. London **57**, 477 (1945).

[47] R. W. Keyes, Phys. Rev. **92**, 580 (1953).

[48] J. Wallis, I. Sigalas, and S. Hart, J. Appl. Crystallogr. **19**, 273 (1986).

[49] R. R. Pawar and V. T. Deshpande, J. Mater. Sci. **5**, 1061 (1970).

[50] M. S. Anderson and C. A. Swenson, Phys. Rev. B **28**, 5395 (1983).

[51] J. A. C. Marples and C. C. Koch, Phys. Lett. A **41**, 307 (1972).

[52] S. R. Bahn and K.W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).

[53] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).

[54] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, J. Phys.: Condens. Matter **22**, 253202 (2010).

[55] G. K. White, Aust. J. Phys. **14**, 359 (1961).

[56] C. H. Xu, C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. B **43**, 5024 (1991).

[57] C. Lind, Materials **5**, 1125 (2012).

[58] In total, 12 possible outliers were excluded from all data sets, 11 of which are indicated in Tables II and III. The last one is Zn, which only appears after the first 11 outliers are removed from the data set of α_V(Tsuru). It is therefore omitted from the other data sets as well.

[59] B. Grabowski, T. Hickel, and J. Neugebauer, Phys. Rev. B **76**, 024309 (2007).

[60] P. Souvatzis and O. Eriksson, Phys. Rev. B **77**, 024110 (2008).

[61] A. Dal Corso, J. Phys.: Condens. Matter **25**, 145401 (2013).

[62] A. Togo, F. Oba, and I. Tanaka, Phys. Rev. B **78**, 134106 (2008).

[63] V. Ozoliņš and M. Körling, Phys. Rev. B **48**, 18304 (1993).

[64] M. Pozzo and D. Alfè, Phys. Rev. B **88**, 024111 (2013).

[65] N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements*, 2nd ed. (Butterworth-Heinemann, Oxford, 1997).

[66] R. L. Mills and E. R. Grilly, Phys. Rev. **101**, 1246 (1956).

[67] J. H. Rose, J. R. Smith, F. Guinea, and J. Ferrante, Phys. Rev. B **29**, 2963 (1984).

[68] E. R. Hernández, A. Rodriguez-Prieto, A. Bergara, and D. Alfè, Phys. Rev. Lett. **104**, 185701 (2010).

[69] S. Yoo, S. S. Xantheas, and X. C. Zeng, Chem. Phys. Lett. **481**, 88 (2009).

[70] L. G. Wang and A. van de Walle, Phys. Chem. Chem. Phys. **14**, 1529 (2012).

[71] L. G. Wang, A. van de Walle, and D. Alfè, Phys. Rev. B **84**, 092102 (2011).

4

A post-Pareto multicriterion optimization tool

PAPER 3

"Ranking the stars: A refined Pareto approach to computational materials design"

K. Lejaeghere, S. Cottenier, and V. Van Speybroeck

Phys. Rev. Lett. **111**, 075501 (2013)

PRL 111, 075501 (2013) PHYSICAL REVIEW LETTERS week ending 16 AUGUST 2013

Ranking the Stars: A Refined Pareto Approach to Computational Materials Design

Kurt Lejaeghere,[1] Stefaan Cottenier,[1,2] and Veronique Van Speybroeck[1,*]

[1]*Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*

[2]*Department of Materials Science and Engineering, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*

(Received 2 April 2013; published 14 August 2013)

We propose a procedure to rank the most interesting solutions from high-throughput materials design studies. Such a tool is becoming indispensable due to the growing size of computational screening studies and the large number of criteria involved in realistic materials design. As a proof of principle, the binary tungsten alloys are screened for both large-weight and high-impact materials, as well as for fusion reactor applications. Moreover, the concept is generally applicable to any design problem where multiple competing criteria have to be optimized.

DOI: 10.1103/PhysRevLett.111.075501 PACS numbers: 81.05.Zx, 07.05.Kf, 28.52.Fa, 62.20.−x

In recent years, materials design has benefitted considerably from computational searches [1]. The increase in computing power has allowed the screening of large numbers of hypothetical compounds for several well-defined design criteria. However, interpreting the output from such investigations is not a trivial task, especially when multiple objectives are optimized simultaneously. The field of multicriteria decision making [2,3] proposes the use of the so-called Pareto set to drastically reduce the number of materials under consideration. Multicriteria decision making notes that in the case of competing requirements, a single "best" solution does not exist. Instead, a number of solutions can be shown to outperform the rest, forming the Pareto set $\mathcal{P}$. Compared to such a Pareto solution $\tilde{\boldsymbol{x}}$, none of the alternative $\boldsymbol{x}$ improve all of the decision criteria $f^{(k)}$ simultaneously:

$$\tilde{\boldsymbol{x}} \in \mathcal{P} \Leftrightarrow \nexists\, \boldsymbol{x}\text{:} \begin{cases} f^{(k)} \geq \tilde{f}^{(k)} & \forall\ k = 1, \ldots, N \\ \exists\ k_0\text{:}\ f^{(k_0)} > \tilde{f}^{(k_0)}. \end{cases} \quad (1)$$

Here, we assume the design problem to be described by N normalized, maximizable objective functions $f^{(k)}$ with $k = 1, \ldots, N$, relating to each Pareto solution a set of coordinates. The assembled Pareto points are therefore also called the Pareto front or skyline, as they outline a hypersurface in N-dimensional space (full black line in Fig. 1).

Pareto optimality has already been successfully applied to computational materials design [4,5]. Unfortunately, as the dimensionality of the problem increases, so does the size of the Pareto set $\mathcal{P}$. It then becomes prohibitively time consuming to study every single Pareto compound in more detail. Several post-Pareto analysis methods therefore try to reduce the number of candidates even further [6–8], but none of them offer a quantitative ordering. This Letter, on the other hand, proposes a mathematically founded procedure that allows the *ranking* of the Pareto compounds, thus identifying the most optimal compromises with respect to the design requirements. It is, however, also more generally applicable to any multicriterion design problem, ranging from space technology [9] over urban studies [10] to linguistics [11].

Figure 1 illustrates how a ranking can be based on the tradeoff between two Pareto solutions, using a hypothetical 2D data set. If we consider an arbitrary Pareto point, such as b, its bottom left quadrant (hatched area) contains only suboptimal (dominated) points, while solutions outside this region offer a tradeoff with the properties of b, either because they improve X (e.g., c) or because they improve Y (e.g., a). The beneficial part of this tradeoff increases as we move away from the dominated quadrant, corresponding to a larger angle θ. When $\theta = 45°$, none of the points is particularly preferred: every loss ΔY is now compensated by an equivalent gain in ΔX, or vice versa. For angles

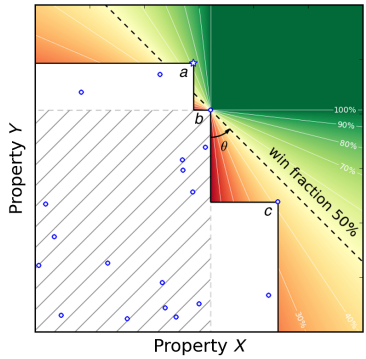


FIG. 1 (color online). Win fraction [Eq. (2)] with respect to Pareto solution b (wf_b), ranging from 0% (red, worst tradeoff) to 100% (green, best tradeoff). The quadrant dominated by point b has been hatched. At a deviation $\theta = 45°$ from the skyline, all solutions are equally valuable (50% win fraction). For the given two-dimensional Pareto front, however, a outranks b and b outranks c.

PRL 111, 075501 (2013) PHYSICAL REVIEW LETTERS week ending 16 AUGUST 2013

larger than 45°, on the other hand, the tradeoff favors the new data point. When applied to Fig. 1, this means that a outranks b, while b outranks c.

To distinguish between two Pareto solutions, we therefore define the win fraction (wf) of $\tilde{\boldsymbol{x}}_j$ with respect to $\tilde{\boldsymbol{x}}_i$:

$$\mathrm{wf}_i(j) = \frac{\sum_k |\tilde{f}_j^{(k)} - \tilde{f}_i^{(k)}| \times \mathcal{H}(\tilde{f}_j^{(k)} - \tilde{f}_i^{(k)})}{\sum_k |\tilde{f}_j^{(k)} - \tilde{f}_i^{(k)}|}, \qquad (2)$$

where $\mathcal{H}(t)$ stands for the Heaviside step function (1 for $t > 0$ and 0 for $t < 0$). The denominator of Eq. (2) represents the total tradeoff between the two Pareto solutions, while in the nominator, only the positive terms remain. The win fraction $\mathrm{wf}_i(j)$ hence expresses what part of the total tradeoff between $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}_j$ favors $\tilde{\boldsymbol{x}}_j$. Since each part of the tradeoff favors one of both solutions, $\mathrm{wf}_i(j)$ and $\mathrm{wf}_j(i)$ always sum to 100%. In 2D, the win fraction is directly related to the skyline angle θ, as can be seen from the win-fraction contour lines in Fig. 1.

A ranking factor for a given solution $\tilde{\boldsymbol{x}}_j$ now follows from the minimum of the win fraction (mwf) with respect to all other Pareto compounds $\tilde{\boldsymbol{x}}_i$: $\mathrm{mwf}(j) = \min_i[\mathrm{wf}_i(j)]$. A large mwf value indicates that $\tilde{\boldsymbol{x}}_j$ provides a considerably better tradeoff than any other Pareto solution. In this way, an mwf-based ordering identifies the most striking features of the skyline.

We first apply this ranking criterion to two classic engineering problems, as a proof of principle: the design of large-weight and of high-impact materials. In both examples, the same search space is investigated, containing 285 ordered body-centered cubic (bcc) binary tungsten alloys with doping levels of 50%, 25%, 12.5%, and 6.25%. This includes all possible alloying elements up to Rn (without the lanthanides) and pure W. For each of these compounds, a case-specific set of properties was calculated by means of density-functional theory [12], using the projector augmented-wave formalism [13,14] in VASP [15,16]. All calculations were performed with the Perdew-Burke-Ernzerhof exchange-correlation functional [17]. Before the Pareto analysis, the value ranges of all properties were normalized to the corresponding range of the elemental crystals [18] and scaled with a weight factor. This provides the mwf the flexibility to express the relative importance of each objective, if user-specific preferences require so. In the examples discussed here, all weights were set to unity (equally important objectives). A summary of further computational settings and the calculated properties is given in the Supplemental Material [19].

As a first example, consider the search for an inexpensive weight, useful as a counterbalance, a press, or a wrecking ball. Acceptable candidate compounds should then combine a large mass density with a good price. The mass density (as well as the bulk modulus in a later example) was extracted from a Birch-Murnaghan fit [20] to a 13-point equation of state (between $V_0 \pm 6\%$), while price data were obtained from the Chemicool Web site, based on prices from Goodfellow [21]. The corresponding skyline is displayed in Fig. 2 (top). In addition, a buffer zone of one error bar is indicated by a dashed line to show the influence of computational uncertainties [18] (see the Supplemental Material [19]). By considering this buffer zone, we avoid overlooking compounds that belong to the *experimental* Pareto set, yet—due to computational inaccuracies—not to the *computational* Pareto set.

In Fig. 2, pure W intuitively appears as the most prominent Pareto solution: increasing the mass density comes at a tremendous cost (a 35-fold price increase), while relaxing the constraints on the mass density only marginally improves the price. Indeed, this also appears from its mwf (73%), which overshadows all other Pareto compounds. When the price of tungsten is considered to be too high, however, alloying offers a solution. The minimum win fractions of WPb (27%), W_3Pb (25%), W_7Pb (24%), and $W_{15}Pb$ (21%) are ranked just below pure W. When a mass density increase is desired, on the other hand, WIr is the highest-ranked alloy in the list, with an mwf of 10%.

Even when the search space is extended with 71 elemental crystals [18], tungsten remains the most interesting Pareto solution, with an mwf of 54%. It is now directly

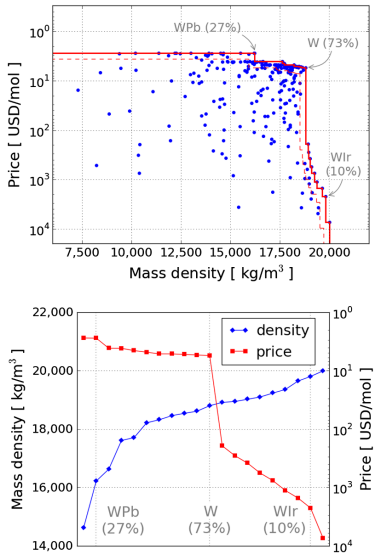


FIG. 2 (color online). Pareto analysis of the bcc binary tungsten alloys with respect to the mass density and the molar price. Top: skyline plot with a one-error-bar uncertainty zone (dashed line). Bottom: evolution of the objectives within the Pareto set. Three high-ranked solutions are indicated with their minimum win fractions.

PRL 111, 075501 (2013) PHYSICAL REVIEW LETTERS week ending 16 AUGUST 2013

followed by pure lead (46%) and pure iron (43%). Both are commonly used materials for large-weight applications, favored over tungsten because of their considerably lower price. They illustrate that an mwf-based ranking indeed identifies the most promising options.

In a second test case, we focus on the hardness and temperature resistance of tungsten. Although W itself already performs very well for these properties, using alloying elements might nevertheless yield an overall improvement. We therefore screen our search space for a good bulk modulus (hardness), cohesive energy (thermal resistance), and price. Materials that meet all of these objectives can be used in drill bits or armor-piercing projectiles, for example.

A 3D representation of the skyline (left-hand panel of Fig. 3) allows us to discern the two most prominent Pareto solutions, WC (in green) and pure W (in red). The minimum win fractions correctly suggest them as highest-ranked materials, with mwf values of 56% and 44%, respectively. The third Pareto compound WRe already performs much worse, with an mwf of only 13%. WC and W are indeed used commercially. Tungsten carbide is the harder material, both in its metastable cubic structure and in the ground-state hexagonal phase [22], while tungsten itself is more temperature resistant. Another commonly employed material, iron, enters the Pareto set at an mwf of 32% when we extend the search space with 71 elemental crystals.

These two test cases demonstrate that our procedure yields meaningful results for well-known examples. A last case study, in the domain of nuclear fusion research, shows its strength in tackling cutting-edge design problems. Because of the harsh operating conditions inside fusion reactors, materials development is a critical aspect there [23,24]. This particularly holds for the in-vessel components, which directly come into contact with the plasma. The divertor, for example, filters ions from the plasma but only by withstanding huge heat and particle fluxes [25]. Most designs for it are based on tungsten because of the associated high-temperature resistance and low activation [26]. Unfortunately, the ductile-to-brittle transition temperature of tungsten is too high for structural applications: in the low-temperature range of operating conditions, tungsten is in the brittle regime. This has initiated an ongoing search for ductilizing alloying elements [25].

Our test set of binary W alloys can contribute to this screening of potential divertor compounds. A suitable divertor material needs to improve the ductility of tungsten while maintaining its favorable high-temperature resistance and acceptable price. Evaluating these three objectives for each of the tungsten alloys may therefore provide the fusion community with some attractive candidate reactor materials.

Since at a defect-free atomic scale, ductility is an intrinsic property of the material, we use the Cauchy pressure P_C to estimate it. Large values of P_C indicate large nonpairwise contributions to the interatomic forces, suggesting an isotropic material [27]. Indeed, directional bonds hinder shear and result in brittle fracture more easily, while isotropic bonds allow for a certain plasticity. For cubic crystals, P_C equals the elastic constant difference $C_{12} - C_{44}$. In our W alloy test set, these elastic constants were extracted from the relation between a set of 2% strains and the corresponding stresses.

Contrary to the two previous test cases, we only consider the mechanically stable structures here. Mechanical stability, as prescribed by the Born criteria [28], is governed by the elastic parameters, and crystals that do not meet these requirements produce highly unrealistic Cauchy pressures.

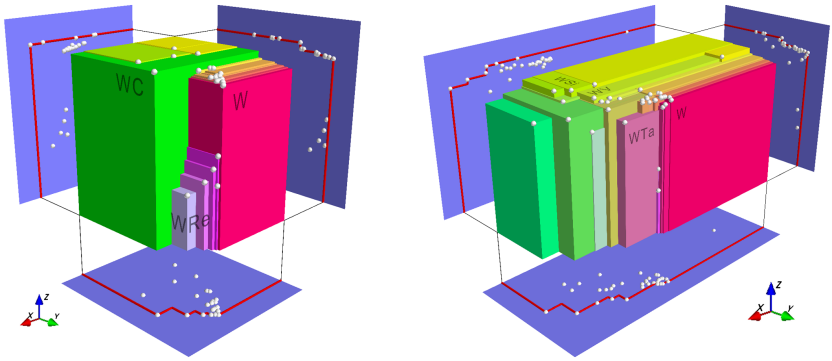


FIG. 3 (color online). Skyline plot and its projection onto the three coordinate planes [33] for the bcc binary tungsten alloys. Left: with respect to the bulk modulus (x), cohesive energy (y), and logarithmic price ($-z$). Right: with respect to the Cauchy pressure (x), cohesive energy (y), and logarithmic price ($-z$).

As this affects the Pareto front substantially, it is best to exclude them. An overview of the stability of the tungsten alloys is given in the Supplemental Material [19].

A Pareto analysis of the three objectives—Cauchy pressure, cohesive energy, and price—retains 36 possible solutions (Table I). They are depicted in Fig. 3 (right). If we moreover require the Pareto solutions to be energetically stable (with respect to phase segregation), this number can be reduced even further. Energetic stability mainly refers to the formation energy. When the formation energy exceeds several times the estimated error on the computed value (0.15 eV/at, Ref. [18]), it can safely be concluded that this compound is unlikely to be (meta)stable. This is the case for W_7C, W_3P, and $W_{15}C$, for example, which can therefore be neglected in the Pareto set of Table I. *A posteriori* analyses of the Pareto set are not necessarily limited to density-functional theory quantities. Other design requirements, such as avoiding radioactive activation, could be taken into account in a similar fashion.

The mwf procedure does not rank pure tungsten highest. At 22%, WTa provides a better tradeoff, yielding an improved Cauchy pressure (128 GPa versus 64 GPa), while the cohesive energy is almost the same (differing by only 30 kJ/mol). In addition, other doping levels have similar effects. For three of the four studied concentrations, the W-Ta system is Pareto efficient, and $W_{15}Ta$ lies only 6% of an error bar from the Pareto front. All W-Ta alloys are mechanically and energetically stable, even at 0 K [19]. Tantalum is three times more expensive than tungsten, however.

Table I also suggests vanadium as an interesting possibility. Although the cohesive energy deteriorates by up to 140 kJ/mol, combining tungsten with high concentrations of vanadium improves its intrinsic ductility even further than with tantalum ($P_C = 138$ GPa for WV). In addition, using vanadium lowers the price significantly, a 75% cut compared to W. All four W-V crystals belong to the Pareto set and are energetically and mechanically stable at 0 K [19].

Contrary to Ta and V, other high-mwf alloying elements yield ambiguous results. Silicon-doped alloys, for example, all belong to the Pareto set and alter the behavior of pure tungsten in the same way as vanadium. Unfortunately, they have quite high formation energies. The magnetic 3*d* transition metals rank well, too. Except for chromium, they all give rise to intrinsically less brittle metals. Mechanical and energetic stability is sometimes an issue for these materials, however. Rhenium, which is experimentally found to lower the ductile-to-brittle transition temperature of tungsten [29], does not appear from this Pareto analysis. Our calculations indeed find an improved Cauchy pressure but suggest that the same improvement can be achieved at better tradeoff conditions.

W-Ta and W-V compounds were investigated theoretically before, but based on the intrinsic ductility alone, their promising performance was not observable [30]. Other studies found Ta and V to embrittle W [25,31]. This discrepancy between the intrinsic and macroscopic ductility of tungsten alloys remains to be elucidated, especially in contrast to observations for other materials systems (e.g., for stainless steel [32]).

TABLE I. Pareto set for the bcc binary tungsten alloys with respect to the Cauchy pressure, cohesive energy, and price. The Pareto compounds are ranked according to the mwf.

mwf		mwf		mwf	
58.1%	W_7C	10.2%	$W_{15}V$	2.5%	W_7Ni
41.9%	W_3P	9.0%	W_7V	2.4%	W_7Ge
22.1%	WTa	8.6%	WV	2.4%	W_3Fe
19.0%	W	8.0%	WFe	2.1%	W_3V
18.9%	W_7Ta	7.1%	$W_{15}Si$	2.0%	W_7Os
16.4%	W_3Ta	3.6%	W_3Si	1.8%	$W_{15}Cr$
14.8%	$W_{15}Co$	3.5%	W_3Sb	0.8%	W_7Cr
14.6%	$W_{15}Ni$	3.4%	$W_{15}Fe$	0.7%	$W_{15}O$
14.2%	$W_{15}C$	3.4%	W_7Fe	0.4%	$W_{15}Rh$
13.9%	W_7Si	3.1%	WC	0.4%	W_3Co
10.7%	WSi	2.8%	$W_{15}P$	0.3%	$W_{15}N$
10.6%	WMn	2.5%	W_3Ni	0.0%	W_7O

All three test cases show that the minimum win fraction orders the Pareto set in an intuitive way. Contrary to conventional Pareto approaches, optimal designs can be compared to each other and ranked, drastically reducing the number of candidates that require closer inspection. Using materials design as an example, the procedure has been shown to yield meaningful results for both classic engineering problems and state-of-the-art applications. It can hence serve as an enabler for discoveries and developments in many research areas.

This work is supported by the Fund for Scientific Research—Flanders (FWO) and by the Research Board of Ghent University. S. C. acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. Calculations were performed using the Stevin Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Hercules Foundation, and the Flemish Government (EWI Department).

*Veronique.VanSpeybroeck@UGent.be

[1] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, Nat. Mater. **12**, 191 (2013).
[2] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs* (Cambridge University Press, Cambridge, England, 1993).
[3] K. M. Miettinen, *Nonlinear Multiobjective Optimization*, International Series in Operations Research and Management Science (Kluwer Academic, Boston, 1999).
[4] T. Bligaard, G. H. Jóhannesson, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, Appl. Phys. Lett. **83**, 4527 (2003).

[5] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, and J. K. Nørskov, Nat. Mater. **5**, 909 (2006).
[6] I. Das, Struct. Optim. **18**, 30 (1999).
[7] G. K. Kao and S. H. Jacobson, Comput. Optim. Applic. **40**, 73 (2008).
[8] N. Lopez, O. Aguirre, J. F. Espiritu, and H. A. Taboada, in *Proceedings of the 41st International Conference on Computers & Industrial Engineering, Los Angeles, California, 2011*, edited by M. I. Dessouky, http://www.usc.edu/dept/ise/caie/34th/conference/website/documents/cover.html.
[9] N. Assadian and S. H. Pourtakdoust, Adv. Space Res. **45**, 398 (2010).
[10] S. Liu, R. Tao, and C. M. Tam, Habitat Int. **37**, 155 (2013).
[11] M. Choudhury, V. Jalan, S. Sarkar, and A. Basu, *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, edited by J. Nerbonne, M. T. Ellison, and G. Kondrak (Association for Computational Linguistics, Prague, Czech Republic, 2007), pp. 65–74.
[12] R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, England, 2004).
[13] P. E. Blöchl, Phys. Rev. B **50**, 17 953 (1994).
[14] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[15] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[16] J. Hafner, J. Comput. Chem. **29**, 2044 (2008).
[17] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[18] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, arXiv:1204.2733 [Crit. Rev. Solid State Mater. Sci. (to be published)].
[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.111.075501 for additional computational details and a full overview of the calculated values.
[20] F. Birch, Phys. Rev. **71**, 809 (1947).
[21] Chemicool Periodic Table (February 2012), http://www.chemicool.com; Goodfellow Group, http://www.goodfellow.com.
[22] D. V. Suetin, I. R. Shein, and A. L. Ivanovskii, Phys. Status Solidi B **245**, 1590 (2008).
[23] A.-A. F. Tavassoli, J. Nucl. Mater. **302**, 73 (2002).
[24] I. Cook, Nat. Mater. **5**, 77 (2006).
[25] M. Rieth *et al.*, J. Nucl. Mater. **432**, 482 (2013).
[26] G. A. Cottrell, R. Pampin, and N. P. Taylor, Fusion Sci. Technol. **50**, 89 (2006).
[27] D. G. Pettifor, Mater. Sci. Technol. **8**, 345 (1992).
[28] M. Born, J. Chem. Phys. **7**, 591 (1939).
[29] G. A. Geach and J. E. Hughes, in *Plansee Seminar Proceedings*, edited by F. Benesovsky (Pergamon, London, 1955).
[30] D. Nguyen-Manh, M. Muzyk, K. J. Kurzydłowski, N. L. Baluc, M. Rieth, and S. L. Dudarev, Key Eng. Mater. **465**, 15 (2011).
[31] H. Li, S. Wurster, C. Motz, L. Romaner, C. Ambrosch-Draxl, and R. Pippan, Acta Mater. **60**, 748 (2012).
[32] L. Vitos, P. A. Korzhavyi, and B. Johansson, Nat. Mater. **2**, 25 (2003).
[33] P. Ramachandran and G. Varoquaux, IEEE Comput. Sci. Eng. **13**, 40 (2011).

5

Computational screening of tungsten alloys for nuclear fusion

A high-throughput computational approach for the design of first-wall materials in nuclear fusion reactors

The previous papers provide the necessary methodology to screen the tungsten alloys for nuclear fusion purposes. A preliminary investigation was even performed in paper 3 (Chapter 4) for the binary tungsten alloys, although mainly to validate our multicriterion optimization protocol. In practice, many binary compounds have already been considered for fusion materials, while the properties of more complex alloys are still unknown. The number of possibilities then becomes quite large, however. Even when only considering the ternaries, a high-throughput approach is needed, leading to several complications before our multicriterion procedure can be applied. The high degree of parallelism (in the form of the different ternary compounds) necessitates specialized scripts that efficiently distribute the calculations over all available computing nodes. On the other hand, the sheer number of calculations requires resorting to the most powerful computers available. Even on the Tier-1 cluster muk with its 528 16-core nodes (see Appendix A), typically allotted computing times are insufficient to evaluate all possible ternaries within a reasonable time.

We therefore limited ourselves to 4d and 5d transition metal dopants in $W_{30}XY$ alloys. For nuclear fusion applications, transition metal alloys are the

most probable candidates, and the concentration of 3 % is in the range of interest[1]. We moreover focused on the 4d and 5d elements, because they allow us to study physical phenomena in d-type compounds more systematically. 3d dopants may distort the observed materials behaviour significantly, especially because of the potential occurrence of magnetism. In addition, different magnetic states are notoriously difficult to stabilize – while substantially influencing several materials properties (e.g. [162, 163]) – and also increase the computational effort considerably. In first instance, they have therefore been excluded from this search.

The studied $W_{30}XY$ cells were constructed from $4 \times 4 \times 4$ conventional bcc supercells, on which fcc sublattices of the two alloying elements were superimposed. The remaining freedom in the relative positions of these two sublattices allowed for six possible configurations. In order of increasing X-Y distance, we refer to them as pos1 up to pos6 (see Figure 5.1). Pos1 corresponds to X and Y in nearest-neighbour positions, while pos5 and pos6 are highly symmetric crystals where the alloying elements surround each other in tetrahedral and octahedral environments, respectively. All of these numerous alloy structures were extensively investigated and the resulting data were screened with the multicriterion analysis from paper 3. Such information allows a study of how fusion-relevant materials properties depend on the alloy composition and may direct the experimental effort on fusion alloys towards the most interesting candidate systems.

Workflow

We performed all DFT computations with the Vienna *Ab initio* Simulation Package (VASP) (version 5.3.3) [164, 165]. This was done in the PAW formalism [83, 166] (see Section 1.3.2) using the April 2012 potentials as recommended by the September 2013 manual [74]. We distinguished between low-, moderate- and high-accuracy settings, which are listed in Table 5.1. In all calculations, the reciprocal space was sampled with a Monkhorst-Pack grid [75]. Both magnetism and spin-orbit coupling were neglected, as no 3d (magnetism) or 6p elements (spin-orbit coupling) are included, for which such a treatment is indispensable (see, e.g., Chapter 3, paper 1, Table 7).

To characterize the different tungsten ternaries, we followed an elaborate high-throughput scheme. This procedure is schematically illustrated in Figure 5.2. First, we assessed the energetic stability of the different configurations (pos1 to pos6, middle part of Figure 5.2). This was done by an initial volume optimization at low accuracy, followed by a 5-point $E(V)$ fit between

[1] Plansee sells a W-Re alloy that contains 5 % of Re, for example [27]. Moreover, the high neutron fluxes in nuclear fusion reactors would generate up to 6 % of Re and 3 % of Os in the first wall due to transmutations [18].

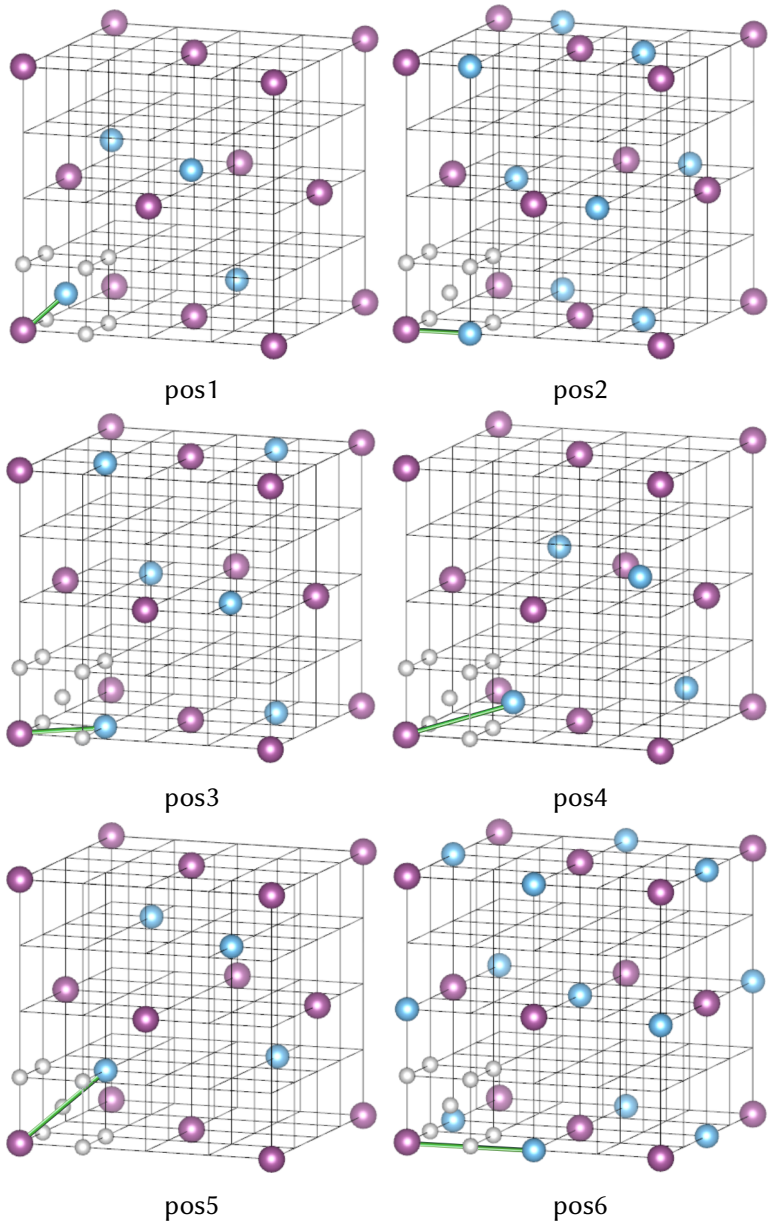


Figure 5.1: Six possible configuration for the $W_{30}XY$ supercell (displayed in their conventional unit cell). The blue and purple atoms represent the alloying elements, while the white atoms represent tungsten (depicted for one bcc unit cell only). For each structure, one X-Y nearest-neighbour distance is indicated in green.

Table 5.1: Computational settings for the three levels of accuracy in our high-throughput investigation of the 4d and 5d tungsten ternaries: cutoff energy, k point sampling, convergence threshold in the self-consistent solution of the Kohn-Sham equations and convergence threshold in structural relaxations.

	low	moderate	high
E_{cut}	300 eV	400 eV	400 eV
k sampling	$3 \times 3 \times 3$	$7 \times 7 \times 7$	$9 \times 9 \times 9$
ΔE(SCF)	10^{-3} eV	10^{-4} eV	10^{-6} eV
Δ(geom)	0.1 eV/Å	0.001 eV	0.01 eV/Å

± 6 % of that preliminary volume (calculated at moderate settings). We used a 4-parameter Birch-Murnaghan equation of state [167]:

$$E(V) = E_0 + \frac{9V_0B_0}{16}\left\{\left[\left(\frac{V_0}{V}\right)^{2/3} - 1\right]^3 B_1 + \left[\left(\frac{V_0}{V}\right)^{2/3} - 1\right]^2 \left[6 - 4\left(\frac{V_0}{V}\right)^{2/3}\right]\right\} \quad (5.1)$$

The fit hence yielded estimates for the equilibrium volume V_0, the bulk modulus B_0 and the pressure derivative of the bulk modulus B_1, as well as the equilibrium energy E_0.

In a second step, the structure with the lowest energy (highlighted in green in Figure 5.2) was selected for a more exhaustive characterization. Not only was the equation of state redetermined with 7 volume points instead of 5 (not depicted in Figure 5.2), the resulting equilibrium volume was also used as a starting point for the calculation of elastic constants, densities of states (DOS) and band structures (right part of Figure 5.2, from top to bottom). All of these simulations were performed with the high-accuracy settings. The density of states and band structure readily followed from standard DFT calculations, where the band structure corresponds to the energy spectrum $\epsilon_{\boldsymbol{k},n}$ along the Brillouin zone edges (Figure 5.3) and the DOS $n(\epsilon)$ represents the Brillouin-zone-averaged number of Kohn-Sham states at an energy ϵ:

$$n(\epsilon) = \frac{1}{V_{BZ}} \sum_n \int_{BZ} \delta(\epsilon - \epsilon_{\boldsymbol{k},n})\, \mathrm{d}\boldsymbol{k} \quad (5.2)$$

The elastic constants, on the other hand, were extracted from Hooke's law for cubic materials [169]:

$$\boldsymbol{\sigma} = \boldsymbol{C} \cdot \boldsymbol{\epsilon} \quad (5.3)$$

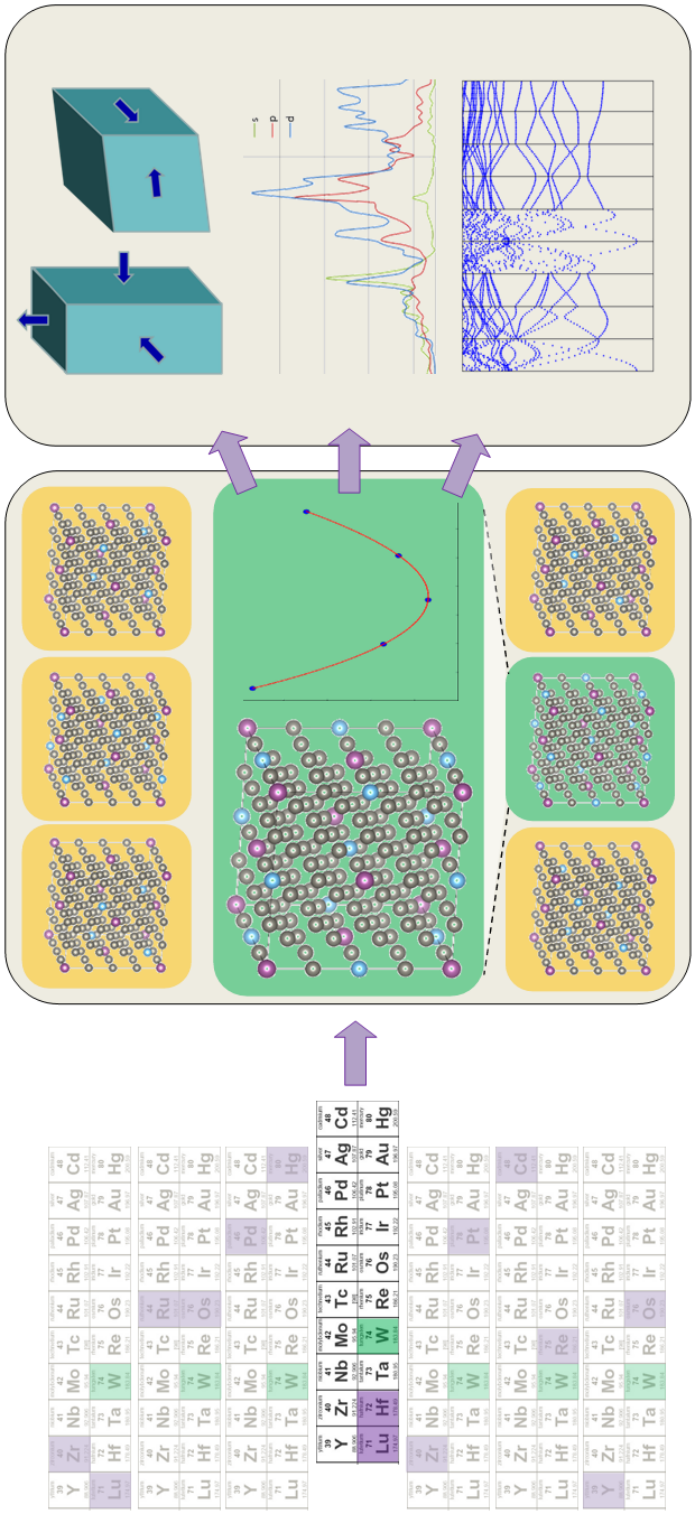


Figure 5.2: Schematic representation of the high-throughput workflow for the simulation of ternary tungsten alloys (see text)

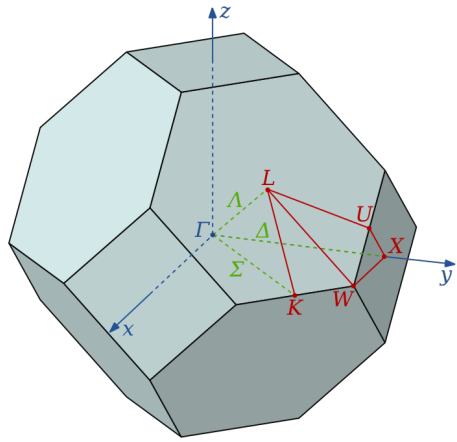


Figure 5.3: High-symmetry points and lines along the edges of the first Brillouin zone of the impurity fcc sublattice [168].

or

$$\begin{pmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \sigma_{yz}\\ \sigma_{zx}\\ \sigma_{xy}\end{pmatrix} = \begin{pmatrix}C_{11} & C_{12} & C_{12} & 0 & 0 & 0\\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0\\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0\\ 0 & 0 & 0 & C_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & C_{44} & 0\\ 0 & 0 & 0 & 0 & 0 & C_{44}\end{pmatrix} \begin{pmatrix}\epsilon_{xx}\\ \epsilon_{yy}\\ \epsilon_{zz}\\ 2\epsilon_{yz}\\ 2\epsilon_{zx}\\ 2\epsilon_{xy}\end{pmatrix} \tag{5.4}$$

with $\boldsymbol{\epsilon}$ the strain, $\boldsymbol{\sigma}$ the stress and $\boldsymbol{C}$ the stiffness tensor in Voigt notation. In particular, we applied a volume-conserving tetragonal and monoclinic deformation [170, 171]:

$$\boldsymbol{\epsilon}_{tetra} = \left(\frac{1}{\sqrt[3]{1+\delta}} - 1, \frac{1}{\sqrt[3]{1+\delta}} - 1, (\sqrt[3]{1+\delta})^2 - 1, 0, 0, 0\right)^T \tag{5.5}$$

$$\boldsymbol{\epsilon}_{mono} = \left(0, 0, \frac{\delta^2}{4-\delta^2}, 0, 0, \delta\right)^T \tag{5.6}$$

in which the δ parameter allowed tuning the strain. It was not necessary to include a third deformation that isotropically expands the material: this operation is characterized by the bulk modulus $B_0 = (C_{11} + 2C_{12})/3$, which was already determined from the equation of state. The stiffness constants C_{ij} could then be determined from the elastic energy [172], but instead, we preferred a stress-based scheme [173, 174], in which the C_{ij} parameters were directly derived from (5.4). Although the most accurate results are obtained by applying different magnitudes of strain and finding the best matching elastic constants, our high-throughput philosophy required a minimum computational effort. We therefore chose to investigate only one set of strains,

corresponding with a deformation δ of 2 % in both cases. This choice is expected to offer a good compromise between precision (smaller strains are more prone to noise) and linear elasticity (large strains do not fulfil Hooke's law anymore).

These extensive calculations were performed for all 4d and 5d $W_{30}XY$ transition-metal compounds. Although this translated into only 20 possible alloying elements and although only one concentration was investigated, contrary to the example in Section 1.3.1, the number of calculations should not be underestimated. Indeed, there were 190 compounds for which 6 geometries needed to be tested by means of a 5-point equation of state, followed by a 7-point equation of state and the calculation of a DOS, band structure and two deformations. For 20 more materials ($W_{31}X$ and W), the same procedure was required, except for the structure comparison. In total, this boils down to 9 170 DFT simulations at low, moderate or high computational settings, some of which were relaxation runs and some of which were spin-polarized. It is therefore clear that an efficient high-throughput framework was required to automate these calculations. We encoded the entire workflow in an extensive home-made script, using the Atomic Simulation Environment (ASE) [119] to manipulate the different structures and manage the in- and output. The calculations themselves were handled with Scalable Concurrent Operations in Python (SCOOP) [161]. This Python package allows distributing different high-throughput subroutines over the available nodes and coordinates the parallellization (see Figure 5.4). It enabled us to investigate several compounds at once, for example, or to treat the six $W_{30}XY$ configurations independently. On the other hand, the different volumes of an equation of states were still calculated sequentially to benefit from previously established charge density files.

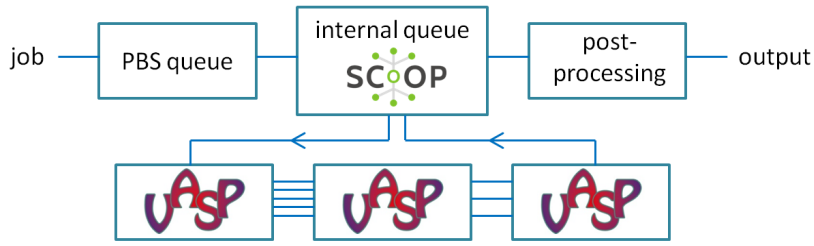


Figure 5.4: Typical usage of SCOOP [161] for high-throughput calculations. In addition to the queuing system of the supercomputer itself, SCOOP provides an internal queue, which distributes job subroutines over previously appointed computing nodes. This allows a flexible reallocation of resources (in our case for VASP calculations [164, 165]).

Results and discussion

The results presented here were mainly calculated in the framework of a VSC Tier-1 pilot project, for which 4 053 node days were granted on muk (1.6 million core hours or almost 8 days of full cluster use). The following paragraphs present an analysis of these results. We particularly focus on trends and patterns and demonstrate the use of data mining techniques, illustrated by our minimum win fraction.

As mentioned in Section 1.3.3, the identification of patterns is often facilitated by using appropriate means of visualization. Many results in the upcoming paragraphs are therefore depicted with charts similar to Pettifor for structure maps [137]. These charts resemble two-dimensional matrices, where every matrix element stands for the combination of two alloying elements, X (vertical axis) and Y (horizontal axis), and is shaded according to a property-dependent colour scale. Because of the symmetry in X and Y, only half of the chart is completed. Typical for Pettifor structure maps, and for our charts, is that the elements are ordered according to the Mendeleev number, which is based on a chemical similarity scale [175]. For the 4d and 5d transition metals, the Mendeleev number reduces to an ordering of the elements according to their group rather than their period in the periodic table (Figure 5.5).

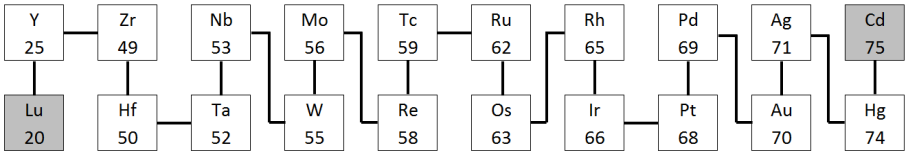


Figure 5.5: Mendeleev number M [137] for 4d and 5d transition metals.

Configurational stability The first step in characterizing any tungsten ternary consists in finding its most stable configuration. Figure 5.1 showed the six possible structures for the two alloying elements in $W_{30}XY$. The actual ground state, as well as the relative stability of the different geometries, depend on the interaction energy of the two dopant atoms in a tungsten environment. We can therefore extract information about this interaction by examining how favourable pos1 to pos6 are compared to each other. Figure 5.6 displays a structure map for all 4d and 5d ternaries.

We notice some distinct trends. When one of the two dopants contains more d electrons than tungsten, for example, it prefers to assume a nearest-neighbour position with respect to the other dopant. Only when the former element has a full d band ($N_d = 10$, which corresponds to s^0d^{10}, s^1d^{10} or s^2d^{10}) and the latter a less than half filled d band, pos4 is the most stable

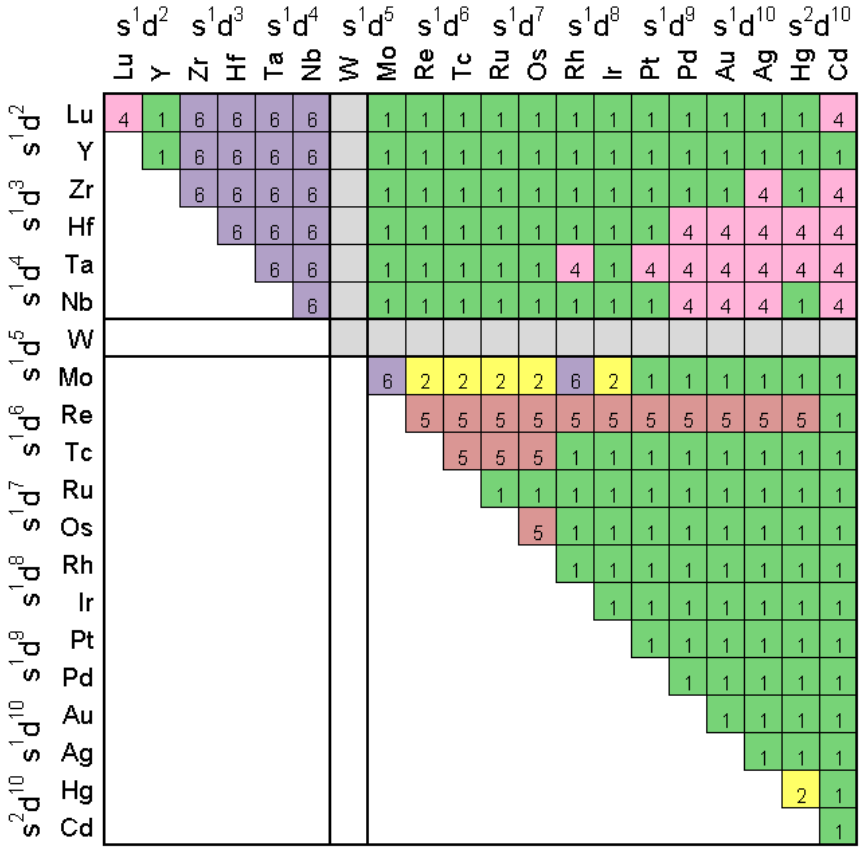


Figure 5.6: Most stable configuration of $W_{30}XY$, where 1 stands for X and Y in nearest-neighbour positions and 6 for X and Y at the most distant lattice sites (see Figure 5.1). Colours have moreover been added to guide the eye. The alloying elements X (vertical) and Y (horizontal) are ordered by their Mendeleev number, and tungsten has been highlighted.

configuration. When both impurities are d deficient, pos6 is the most stable geometry, with each impurity octahedrally surrounding the other at a large distance. Finally, when Re or Tc is included in the ternary cell, pos5 often seems to be the most favourable configuration. We will see that these varying preferential positions are also reflected on an absolute energy scale (see further). Note that in Figure 5.6, we have shaded the $W_{31}X$ cells grey. Although numerical effects will always select one geometry as the most stable one, all six possibilities for $W_{31}X$ are actually equivalent, and none of the configurations (or all of them) is the most stable one.

To demonstrate the stability of the different configurations more clearly, we have depicted the energies for 4d dopants as surfaces in Figure 5.7. These

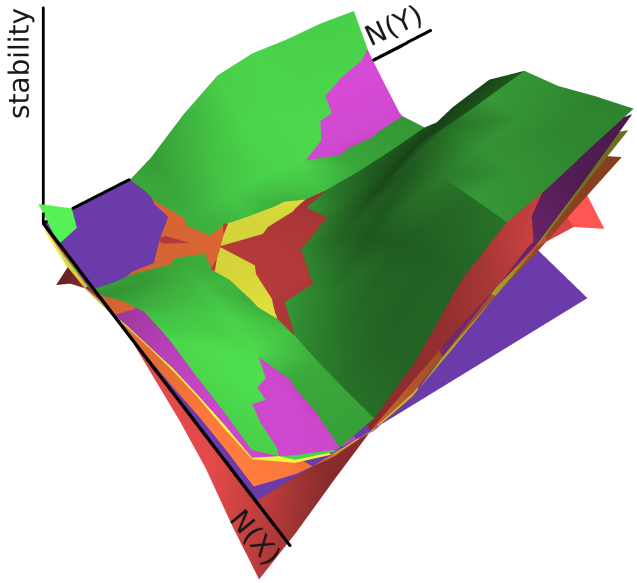


Figure 5.7: Structural stability of the $W_{30}XY$ configurations relative to pos6 (purple plane), where X and Y are 4d elements and pos1 to pos5 are depicted by green, yellow, orange, pink and red surfaces, respectively. Values are plotted as a function of the numbers of d electrons, N(X) and N(Y), and high z values correspond to a better stability.

energies have been represented relative to the configuration where the alloying elements are furthest apart (pos6). We indeed see the same behaviour as in Figure 5.6, with pos1 and pos4 dominating the ternaries with large numbers of d electrons, and pos6 occurring predominantly for d deficient dopants. When investigating the relative stability of metastable geometries, we notice that the energetics of many configurations do not differ much. This is especially the case for the tungsten group (W, Mo), but that is no surprise: all geometries are equally stable when one of the dopants is tungsten itself. The effect of Mo is almost identical to that of pure tungsten because of the equal number of d electrons. Another intuitive phenomenon is the opposite reaction of pos1 and pos6. Indeed, the impurities either want to be located close to each other, or as far away as possible. Chemically compatible dopants will therefore prefer pos1 and have a very unfavourable pos6, while for chemically incompatible species, it is the other way round. The behaviour of pos5 is less easy to explain. In all corners of Figure 5.7 (when both impurities have full or nearly empty d bands) it is significantly less stable than the other geometries. Only when both dopants have many d electrons, is pos6 even more unfavourable. In the other cases, the instability of pos5 with respect to pos6

must be explained from symmetry effects, since both dopants are already at a substantial distance from each other in both pos5 and pos6. Rather, it seems that the octahedral environment is more advantageous than the tetrahedral one. Indeed, the d orbitals themselves are also octahedrally symmetric and this symmetry is more pronounced when only one orbital is (un)occupied. Another surprising observation is the relative stability of pos4 with respect to pos3, even in areas where the impurities clearly want to be near each other. An explanation from symmetry is plausible here too (the pos4 cell is of a lower symmetry than the pos3 one), but less easily identified.

Equation of state Fitting the $E(V)$ data points to a Birch-Murnaghan equation of state offers much information. Not only does it estimate several relevant materials properties (V_0, B_0, B_1, E_0), the quality of the fit also relates to the correctness of the calculations. In high-throughput workflows, it is sometimes hard to discern whether a simulation has gone according to plan. A breakdown of the DFT code is a clear indicator, but even when the calculation is completed successfully, the results are sometimes unreliable. The most likely cause is convergence to a metastable electronic state for one or more volume points. Fortunately, such events can be signalized by inspecting the coefficient of determination R^2 of the EOS fit, which should be close to 1. An example is provided in Figure 5.8 (top) and displays the logarithm of $1 - R^2$ for the pos5 configuration. One compound immediately stands out: $W_{30}YRh$. Indeed, its coefficient of determination deviates by more than 10^{-2} from 1. This appears to be due to the convergence of one volume point to a metastable electronic configuration. When such problems are identified, they can easily be corrected (Figure 5.8, bottom).

Figure 5.9 shows the variation of the equilibrium volume V_0, the bulk modulus B_0, the pressure derivative of the bulk modulus B_1 and the cohesive energy ΔE_{coh} as a function of the dopants. The smallest volumes appear to occur when both alloying elements are located in the group of osmium (s^1d^7), as do the largest bulk moduli. This is common behaviour, as B_0 is approximately inversely proportional to V_0. The pressure derivative of the bulk modulus, on the other hand, reaches a minimum when at least one element from the group of tungsten (s^1d^5), tantalum (s^1d^4) or hafnium (s^1d^3) is added. Finally, the highest cohesive energies are found when both dopants have an electron configuration similar to tungsten.

Many of these patterns can be explained from the unsaturated, covalent-like d bonds of the transition metals. Gelatt *et al.* showed that the principal contribution to the cohesion of transition metals is due to the filling of the d band with N_d electrons [176]. The first 5 electrons then occupy bonding states, while the last 5 electrons are added to antibonding ones. This causes the material to reach its optimal stability for $N_d = 5$. Indeed, when taking

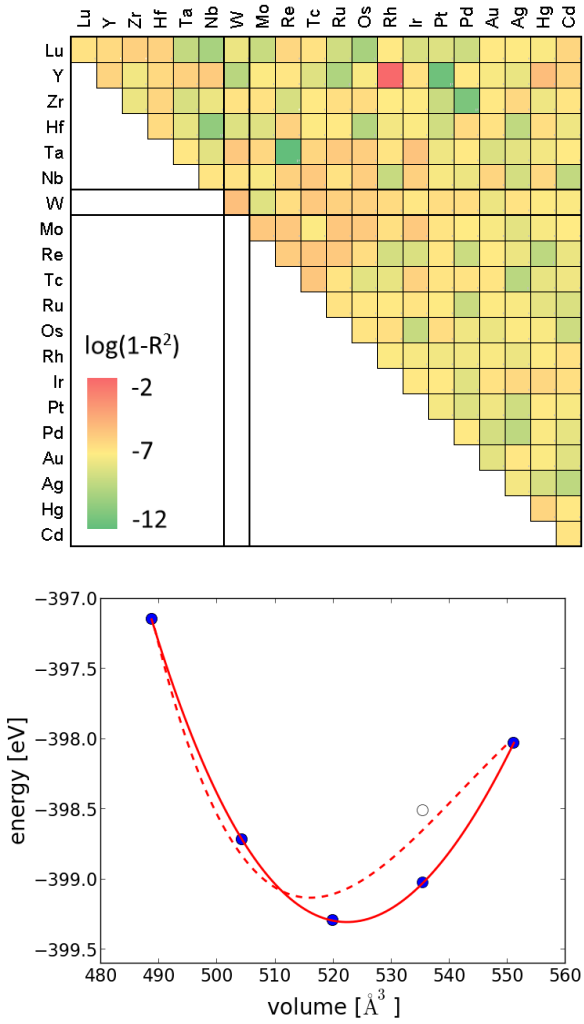


Figure 5.8: Top: $\log(1 - R^2)$ for the equation of state of pos5 $W_{30}XY$, with R^2 the coefficient of determination. The alloying elements X (vertical) and Y (horizontal) are ordered by their Mendeleev number, and tungsten has been highlighted. Bottom: Equation of state of $W_{30}YRh$ in the pos5 configuration. Both the DFT data points and the fitted Birch-Murnaghan curve are displayed, before (open symbols, dashed line) and after (filled symbols, full line) fixing issues with a metastable state.

into account that in bulk, one s electron is promoted to the d band [177], this explains the high cohesive energy for the tungsten group. The small volumes (and large bulk moduli), on the other hand, are shifted towards larger d electron numbers. It is caused by the inward contraction of the d orbitals as N_d

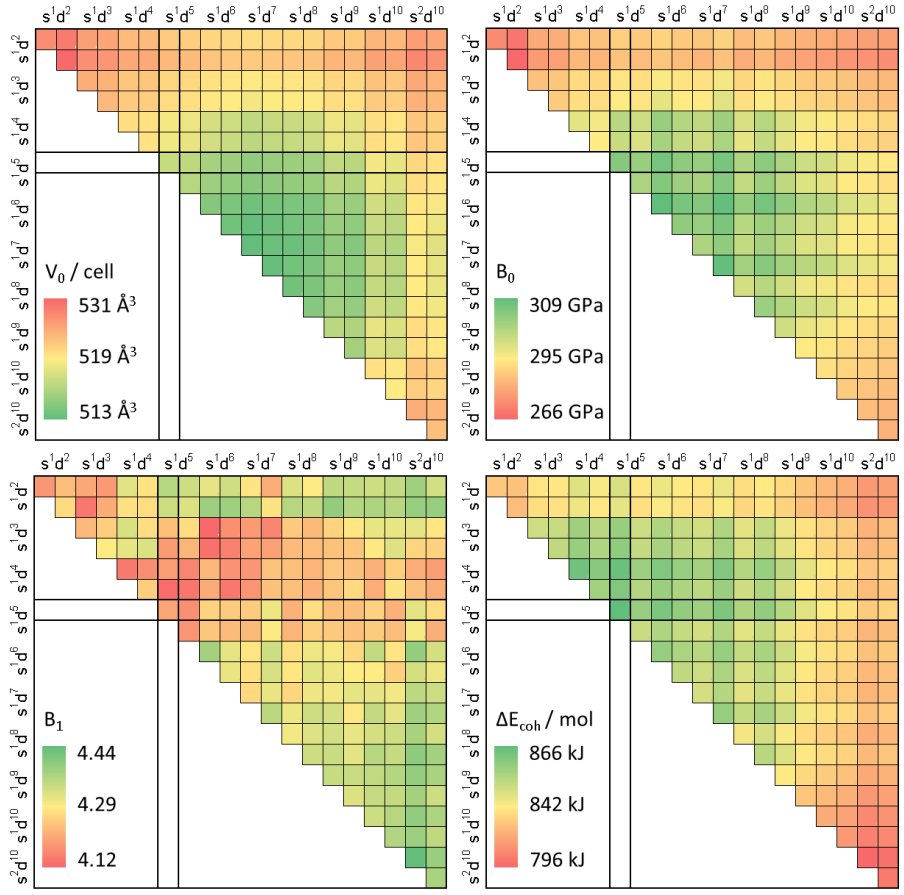


Figure 5.9: Equilibrium volume V_0, bulk modulus B_0, pressure derivative of the bulk modulus B_1 and cohesive energy ΔE_{coh} of the most stable configuration of $W_{30}XY$. The alloying elements X (vertical) and Y (horizontal) are ordered by their Mendeleev number, and tungsten has been highlighted.

(and the nuclear charge) increases, which in turn causes the interaction range of the atoms to drop. This enables atoms with high numbers of d electrons to come closer to each other than the stability of the d band would justify. These arguments are further supported by tight-binding theory for equiatomic, disordered XY alloys (see Appendix B). When using tight-binding theory in the second-moment approximation with a rectangular d-band model and exponential bond and repulsion integrals, the major trends are already quite well reproduced (Figure 5.10). Especially for the cohesive energy, a closed expression (B.18) exists, which allows looking at the different contributions for al-

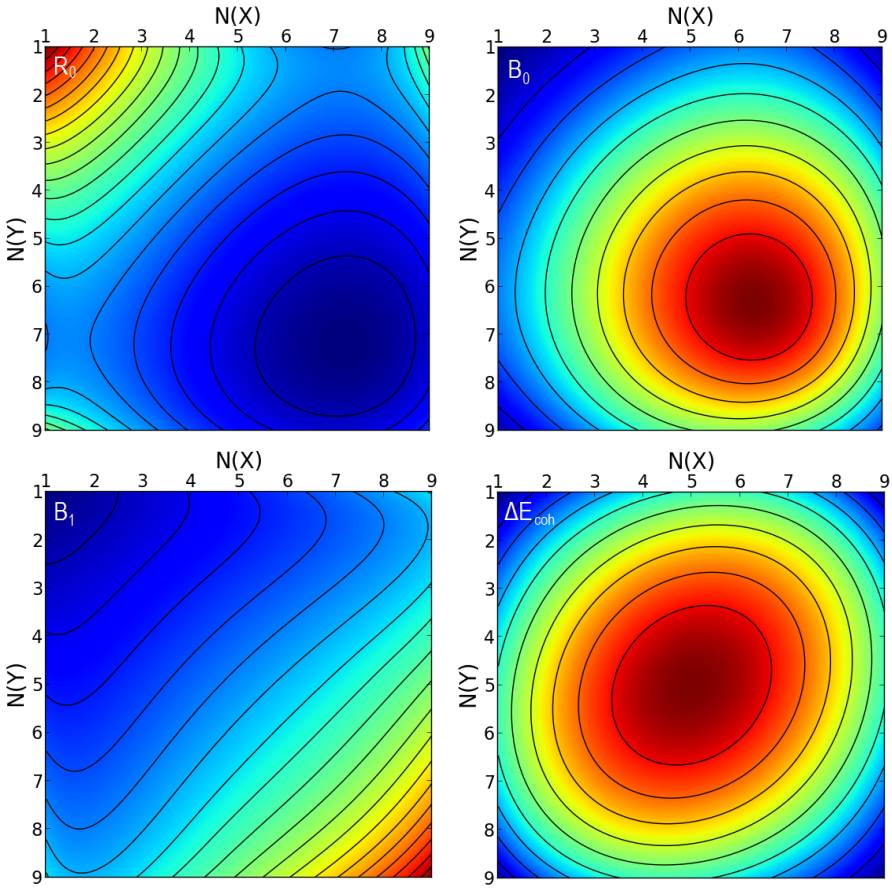


Figure 5.10: Interatomic distance (B.17) (top left), bulk modulus (B.19) (top right), bulk modulus derivative (B.20) (bottom left) and cohesive energy (B.18) (bottom right) from tight-binding theory in the second-moment approximation, applied to equiatomic, disordered XY alloys. A rectangular d band and exponential bond and repulsion integrals were assumed. The plots display the variation of these four properties as a function of the number of d electrons for X and Y, with a colour scale ranging from blue (small) to red (large). They are therefore directly comparable to the DFT results from Figure 5.9.

loys in more detail:

$$\Delta E_{coh}^{TB}(\mathrm{XY}) \propto \bar{N}^2(10-\bar{N})^2 - \frac{25}{3}(\Delta N)^2 \tag{5.7}$$

with $\bar{N}$ the average number of d electrons and ΔN the difference. The formula shows that the stability of alloys is driven by two competing effects:

on the one hand, half-filled bands are indeed very stable (first term), but on the other hand, large electron transfers ΔN from one alloying element to another are unfavourable due to the mismatch in atomic energies (second term). The same phenomena are expected to play a role for the other EOS parameters.

Although tight-binding theory indeed reproduces the most important trends for V_0, B_0 and ΔE_{coh}, the behaviour of B_1 seems to deviate somewhat. This can be due to any of the model assumptions. We expect d electrons in XY and W_{30}XY to behave approximately the same – tungsten has a half-filled d band, so the effect of ΔN_{WA} and ΔN_{WB} will not change the privileged position of $\bar{N} = 5$ – but there is also some crucial difference: W_{30}XY is equiatomic nor disordered. The local geometry around the alloying elements can therefore play a substantial role. Moreover, B_1 represents a third-order derivative of the energy, and is hence very sensitive to the exact shape of the interatomic potential. An exponential shape is then probably not accurate enough[2].

Energetic considerations Figure 5.9 discussed the cohesive energy as a measure of the crystal stability. It expresses the energy required to bring isolated atoms together in a lattice. When alloy formation is concerned, other energetic quantities are relevant as well (Figure 5.11). The formation energy from the ground-state elemental phases ΔE_f, for example,

$$\Delta E_f = E(W_{30}XY) - 30E_0(W) - E_0(X) - E_0(Y) \tag{5.8}$$

indicates whether the alloy is stable with respect to phase segregation. Tight-binding formulas [178] suggest this quantity to show a similar behaviour to ΔE_{coh}, with the highest stability (lowest energies) when the d band is half filled and when the electron numbers of the constituents do not differ too much. Instead, Figure 5.11 locates the most stable alloys at $N_d < 5$. The reason for this apparent deviation is the elemental crystal structure. When calculating the formation energy from the bcc phases of the pure elements ΔE_{bcc} (Figure 5.11, right),

$$\Delta E_{bcc} = E(W_{30}XY) - 30E_{bcc}(W) - E_{bcc}(X) - E_{bcc}(Y) \tag{5.9}$$

the negative energy differences are scattered quite symmetrically around $N_d = 5$. However, only for the elements with $N_d \leq 5$, the bcc symmetry

[2]Nevertheless, a tight-binding expression for B_1 works quite well for the pure elements. Under the same assumptions as before, it yields (B.13): $B_1 = 1 + \kappa R_0$, with $\kappa \propto N_d$ the inverse interaction length of the bond integral and R_0 the equilibrium interatomic distance. As a result, B_1 is predicted to increase monotonically with the number of d electrons like $B_{1,0} - \ln(10 - N_d)$, a behaviour that is indeed observed from DFT for the 4d and 5d elements (see paper 1, Supplementary Material Table 2.4)

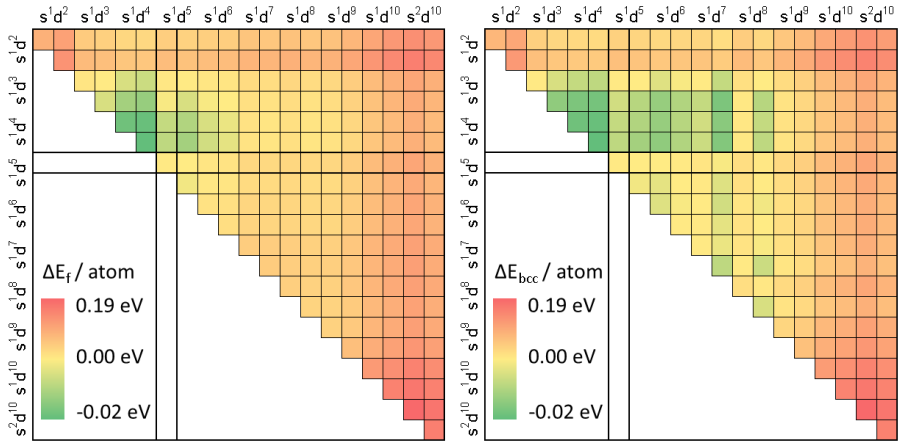


Figure 5.11: Formation energy from the ground state ΔE_f and formation energy from the bcc phase ΔE_{bcc} for the most stable configuration of $W_{30}XY$. The alloying elements X (vertical) and Y (horizontal) are ordered by their Mendeleev number, and tungsten has been highlighted.

is also (near) the ground state. For the other elements, another structure is more favourable, raising the formation energy ΔE_f by the structural energy difference between the ground state and bcc.

Density of states As described in (5.2), the density of states expresses the number of energy levels at a given Kohn-Sham energy, averaged over the first Brillouin zone. Higher DOS values suggest that more electrons can be accommodated. The practical occupation of the energy levels follows a Fermi-Dirac distribution, which takes the shape of a Heaviside step function $\mathcal{H}(E_F - E)$ at zero kelvin, but is smeared out at higher temperatures. The most important changes in the electron occupation therefore take place near the Fermi energy E_F.

Because of its position at the boundary between valence and conduction states, the density of states at the Fermi energy substantially influences electronic effects. Band gaps at the Fermi energy give rise to insulators or semiconductors, for example, while nonzero DOS values suggest a metallic character. The density of states at the Fermi level $n(E_F)$ is also an essential part of the electronic heat capacity C_V:

$$C_V = \frac{1}{3}\pi^2 k_B^2 n(E_F) T \tag{5.10}$$

Since $W_{30}XY$ alloys possess very similar crystal lattices, and hence nearly equal ionic contributions to the heat capacity, changes in the heat capacity

mainly come from changes in this $\mathcal{C}_V$. For fusion materials, a large heat capacity corresponds to a better thermal resistance of the plasma-facing components. Equation (5.10) therefore shows that for fusion applications, a high density of states at the Fermi level is recommended.

$n(E_F)$ is displayed in Figure 5.12 for our tungsten ternaries. It is small at both high and low numbers of d electrons, and large in between ($5 < N_d < 10$). This behaviour can be explained when looking at a few illustrative DOS curves (Figure 5.13). For low numbers of d electrons, the d band is broad and lies high in energy, as it does not feel the attractive potential of the nucleus much due to the screening by filled lower-lying shells ($W_{30}Y_2$). As we progress through the d series, however, not only the number of d electrons increases, but also the nuclear charge. The d band is then increasingly drawn towards lower energies while the s and p bands remain in place. This causes the increasingly narrow and high d DOS to first cross the Fermi energy – leading to a maximum $n(E_F)$ ($W_{30}Ru_2$) – and then move to even lower energies, making the density of states at the Fermi level small again ($W_{30}Ag_2$). The (electronic) heat capacity follows the same behaviour [see (5.10)].

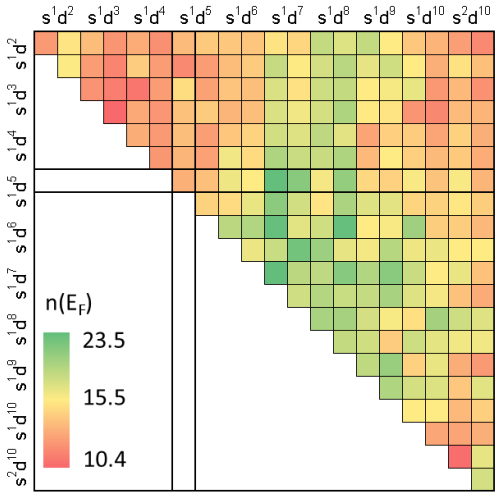


Figure 5.12: Density of states at the Fermi level for the most stable configuration of $W_{30}XY$. The alloying elements X (vertical) and Y (horizontal) are ordered by their Mendeleev number, and tungsten has been highlighted.

Pareto analysis While the previous paragraphs focused on the different trends among the tungsten ternaries, we are mostly interested in the extrema when designing new materials. Outliers from the trends are also very valuable, as they arise from unexpected physical phenomena, which may render

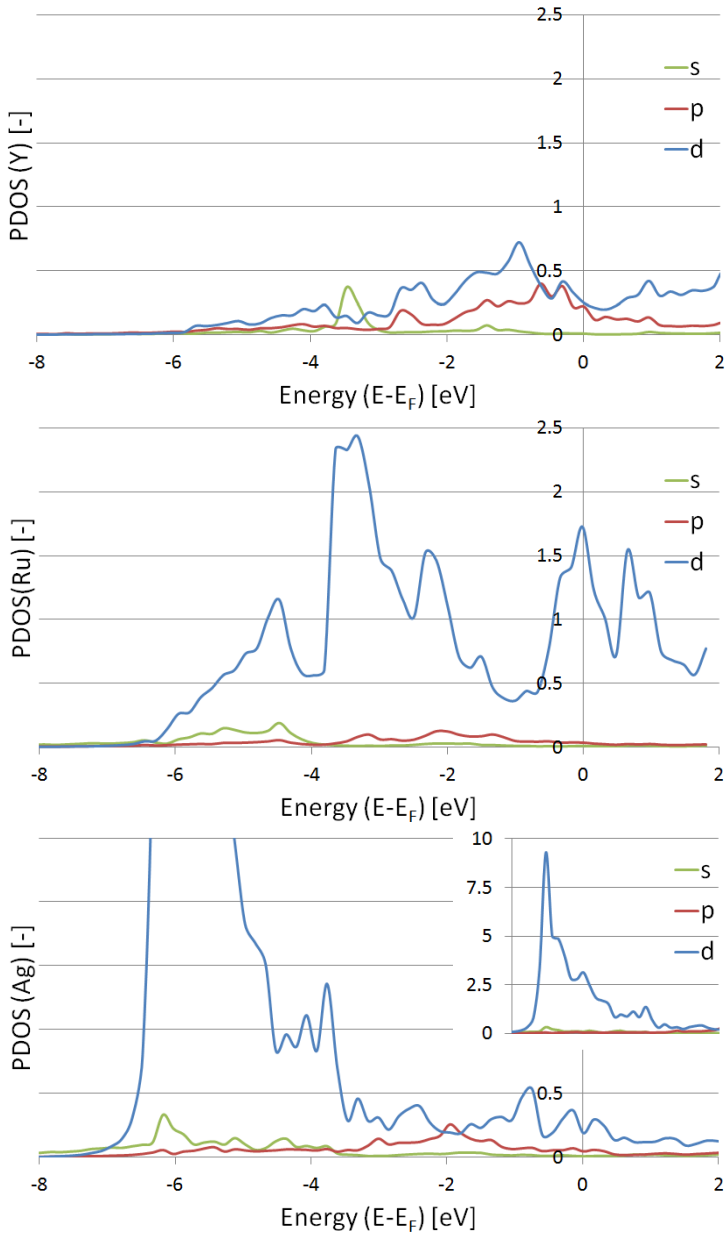


Figure 5.13: Partial densities of states of the dopant atoms in $W_{30}Y_2$, $W_{30}Ru_2$ and $W_{30}Ag_2$ (with inset).

a material more suitable for the application at hand. For the ductilization of tungsten, most intuitive options have already been exhausted during previous experimental screenings. Outliers are therefore our best hope for new dis-

coveries, although the trend extrema may provide some hitherto overlooked solutions as well.

To identify the most promising tungsten alloys, we performed a Pareto analysis and applied our novel minimum win fraction (mwf) criterion in a similar way to paper 3 (Chapter 4). While the latter study concerned the simultaneous optimization of the cohesive energy ΔE_{coh} (thermal stability), Cauchy pressure P_C (ductility) and molar price p, we replaced the maximization of ΔE_{coh} by the minimization of the thermal expansion coefficient α in the approximation of Guinea *et al.* [134]. Indeed, a low thermal expansion is essential in nuclear fusion reactors, because it prevents the generation of large thermal stresses. We moreover found the approximation of Guinea *et al.* to provide the same quality as quasiharmonic phonon theory (see Chapter 3, paper 2). On the other hand, we previously discovered the Cauchy pressure to predict ductility incorrectly. We therefore considered four different descriptors for the ductility and assessed the differences between their Pareto sets. In particular, ductile compounds are linked to a high Cauchy pressure $P_C = C_{12} - C_{44}$ [179] or B/G ratio [180], as well as a low C_{44} or prelogarithmic energy $\mathcal{E}$ [181]. The former two quantities express the contribution of non-pairwise forces in the crystalline cohesion (and hence the covalent character) [182], while the latter two are closely correlated to the Peierls barrier height, which determines the dislocation mobility [181]. For the prelogarithmic energy, a slip-plane-dependent formula should actually be used [44], but we applied the common polycrystalline approximation $\frac{1}{2}Gb^2$ with b the Burgers vector [183].

Figure 5.14 shows the skyline when ductility is represented by the Cauchy pressure. We notice that there is not much variation with respect to the expansion coefficient, so the Pareto optimization will be mostly determined by P_C and the price. An analogous effect occurs for the other ductility criteria, so all four Pareto frontiers are similar. Nevertheless, the particular compounds in the Pareto set sometimes differ substantially. Table 5.2 lists all solutions with an average mwf value of 5 % or more. As in paper 3 (Chapter 4), the values of the different objectives were normalized to the corresponding ranges for the elemental crystals, or to those of the binary tungsten alloys (from paper 3) when this information was not available. We now decreased the importance of the price to only one tenth, however, because the price of the ternaries still varies over several orders of magnitude, while α and all ductility predictors do not differ as much. Indeed, when we do not introduce this reduced preference for the price, the somewhat lower cost of $W_{30}Hg_2$ ranks it highest for the set (α, P_C, price), while it has a significantly worse α and P_C than $W_{30}ZrAg$. Decreasing the importance of the price resolves this issue.

Several interesting observations can be made from Table 5.2. In general, we see some similarity between the different ductility criteria, as most Pareto

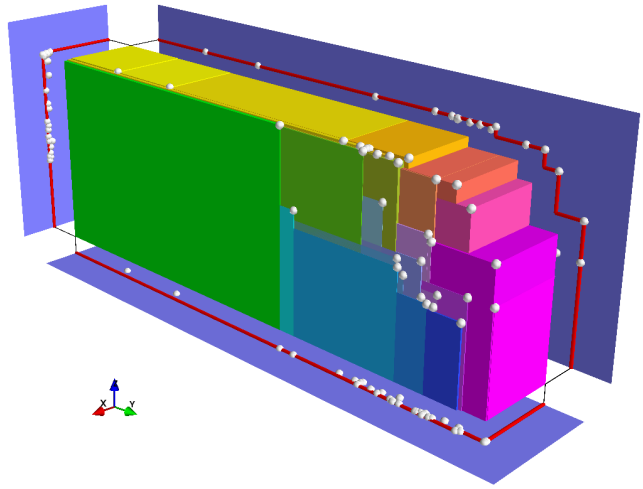


Figure 5.14: Pareto set for the simultaneous optimization of the ternary tungsten alloys with respect to the volumetric thermal expansion coefficient (x), the Cauchy pressure (y) and minus the logarithm of the molar price (z).

compounds occur in more than one set. Only the most prominent solutions for C_{44} and $\mathcal{E}$ do not appear to overlap nor coincide with any of the other ductility predictors. Nevertheless, Weinberger *et al.* demonstrated both predictors to display an excellent correlation with the Peierls barrier height [181]. Our apparently contradicting results are probably because of the large numerical inaccuracy when determining C_{44}. In our high-throughput analysis, an efficient compromise was sought between accuracy and computational effort, but for C_{44}, higher levels of precision may be required. A second point of interest is the apparent optimality of combining alloying elements from the outer ends of the d block (such as Zr-Ag, Ag-Hf and Cd-Lu). In terms of stability (ΔE_{coh}) and hardness (B_0), an average number of d electrons close to 5 or 6 is favourable, but the large differences between the two dopants should have an adverse effect (see Figures 5.9 and 5.10). In our Pareto search, that disadvantage may have been compensated by a lower price. We also notice the (re)occurrence of the W-Ta alloying system, which was already identified in paper 3. Although its importance varies from one ductility criterion to the other, their is clearly a theoretical basis to expect an increased ductility in these compounds. The experimental refutation of this result indicates the influence of some more elaborate effect, going beyond dislocation movements in defect-free monocrystalline materials. Some other striking compounds in all four sets are $W_{30}ReOs$, which will form naturally over time in fusion reactors because of neutron-induced transmutation, and $W_{30}Re_2$,

Table 5.2: Pareto solutions with the highest mwf ranking when simultaneously optimizing the ternary tungsten alloys with respect to the volumetric thermal expansion coefficient, one of four ductility criteria (P_C, B/G, C_{44} and $\mathcal{E}$) and the logarithm of the molar price (with a weight factor of only one tenth). The compounds were ordered according to their average mwf.

compound	mwf(P_C)	mwf(B/G)	mwf(C_{44})	mwf($\mathcal{E}$)
$W_{30}ZrAg$	50.6 %	53.6 %		49.1 %
$W_{30}Hg_2$	14.2 %	46.4 %		18.5 %
$W_{30}AgHf$	49.4 %	12.0 %		10.5 %
$W_{30}Ta_2$	15.9 %	42.4 %	1.85 %	4.07 %
$W_{30}Lu_2$			60.9 %	0.60 %
$W_{30}PdTa$	44.9 %	6.49 %		2.26 %
$W_{30}CdLu$				50.9 %
$W_{31}Zr$	5.04 %	30.5 %	9.76 %	4.01 %
$W_{30}TaHg$	9.09 %	34.7 %	0.16 %	2.50 %
$W_{30}YZr$			39.1 %	
$W_{30}AgTa$	8.83 %	21.7 %		1.85 %
$W_{30}ZrCd$	1.49 %	19.9 %	2.53 %	7.32 %
$W_{31}Hg$	5.96 %	22.5 %	0.19 %	2.36 %
$W_{31}Ta$	5.28 %	22.4 %	1.01 %	1.82 %

which is experimentally used to improve the ductility of tungsten[3]. Finally, one compound in particular stands out with respect to three out of four ductility predictors: $W_{30}ZrAg$. Some of its properties are compared to those of pure tungsten in Table 5.3, taking into account the systematic and residual error estimates from papers 1 and 2 (Chapter 3). We indeed notice $W_{30}ZrAg$ to outperform tungsten for all ductility predictors. Its formation energy is positive, on the other hand, but within the corresponding error bar. Further studies are required to investigate whether our DFT analysis is indeed on to something promising.

A note of caution One note of caution is to be added to the previous observations. Although several trends clearly arise from the computation of the 4d and 5d ternary tungsten cells, numerical errors have not been taken into account yet. The Birch-Murnaghan fit, for example, can add some noise to the predicted properties. Shifting the volume range, changing the number

[3] $W_{30}ReOs$ and $W_{30}Re_2$ are not shown in Table 5.2 because of their low mwf values: 3.48 % and 2.52 % (P_C), 3.51 % and 1.40 % (B/G), 0.19 % and 0.24 % (C_{44}) and 0.68 % and 0.41 % ($\mathcal{E}$).

Table 5.3: Comparison between $W_{30}ZrAg$ and W, taking into account systematic and residual errors where available (see papers 1 and 2 in Chapter 3). An asterisk (*) indicates that systematic errors have corrected for.

	$W_{30}ZrAg$	W
ΔE_{coh} [kJ/mol]	831* $\pm$ 30	866* $\pm$ 30
V_0 [Å^3/atom]	15.7* $\pm$ 1.1	15.6* $\pm$ 1.1
B_0 [GPa]	300* $\pm$ 15	321* $\pm$ 15
B_1 [–]	4.1* $\pm$ 0.7	4.0* $\pm$ 0.7
ΔE_f [eV/atom]	0.10* $\pm$ 0.16	0.00* $\pm$ 0.00
α_V [10^{-5}/K]	0.9* $\pm$ 1.9	0.8* $\pm$ 1.9
price [USD/mol]	7.48	5.42
P_C [GPa]	69	56
B/G [–]	2.2	2.0
C_{44} [GPa]	132* $\pm$ 23	150* $\pm$ 23
$\mathcal{E}$ [eV/Å]	3.11	3.57

of volume points or adjusting the volume interval affects the fitted equation of state parameters. The effect of the number of volume points in the Birch-Murnaghan fit can already be estimated by comparing the EOS of the most stable configuration to that of the same structure in the original assessment of pos1 to pos6 (see Figure 5.2). In the former case, 7 volume points were used, while the latter only counted 5. The central volume was moreover slightly shifted, because the more accurate EOS fit was centred around the fitted equilibrium volume of the less accurate one. These combined effects did not lead to that large differences, but deviations still amounted to up to 0.17 A^3/cell for V_0, 0.022 eV/cell for E_0, 2 GPa for B_0 and 0.4 for B_1. Especially for B_1, this raises the question whether the observed trends are significant, or a simple matter of numerical noise. However, the smooth behaviour of B_1 in Figure 5.9 indicates that that the errors are smaller when comparing similar equations of state. Note that in comparing the 5- and 7-point EOS fits, the number of k points ($7 \times 7 \times 7$ to $9 \times 9 \times 9$) and the numerical accuracy were changed as well (10^{-4} eV to 10^{-6} eV), but preliminary convergence tests suggest the effects to be small.

6
Conclusions and outlook

Conclusions

We presented a high-throughput computational screening of the tungsten alloys, in which we searched for promising candidates for nuclear fusion applications. Tungsten is one of the best candidates for application in plasma-facing components because of its low activation and favourable thermal properties (such as a high melting temperature, a high thermal conductivity and a low thermal expansion). The room-temperature brittleness of pure tungsten prevents using it on a large scale, however, and compromises its long-term reliability. Alloying tungsten with other elements may mitigate or resolve this issue, while retaining the beneficial properties of the host material. Since a thorough experimental screening study of all possible alloys would not be affordable, both in time and in cost, a computational approach offers an interesting alternative. It allows investigating much larger numbers of tungsten compounds at a much faster rate, significantly expediting the materials design process.

Several levels of theory are available to perform this search. We chose to apply so-called quantitative structure-property relationships, which directly relate atomic-scale features to the macroscopic materials behaviour instead of requiring a multiscale approach. We worked within the framework of density-functional theory, which yields accurate electronic-structure predictions without demanding too large computational efforts. In particular, we applied the generalized-gradient approximation in the implementation of Perdew, Burke and Ernzerhof, as it provides a good accuracy and speed

for the description of solids. Because of the immense combinatorial space associated with all possible tungsten alloys – with respect to their crystal structure, constituents and compositions – we limited ourselves to bcc-based supercells with one or two alloying elements and in only a few concentrations. Their characterization involved three major steps: (i) a quantification of the expected deviations between computational results and experimental values, (ii) the development of a multicriterion decision tool to deal with the conflicting requirements of nuclear fusion, and (iii) the construction of a high-throughput environment to automate the large number of simulations.

Computational materials design is only possible when the underlying uncertainties are well understood. In a first stage, we therefore benchmarked the computational errors, which arise at three distinct levels. Numerical errors are due to the implementation of the theoretical formalism in practical algorithms. We assessed them by defining a Δ gauge to express the deviation between equations of states of different methods or codes. Intrinsic errors result from the use of an approximate functional within density-functional theory, in our case that of Perdew, Burke and Ernzerhof. We applied a statistical analysis to distinguish systematic deviations from the remaining differences between theory and experiment. It yielded reproducible and generally valid error bars for these types of calculations. Finally, semi-empirical errors are present when a macroscopic property is related to a microscopic predictor, based on a fit between experimental values and theoretical materials properties. This reduction of dimensionality, from the full description of all physics to a simple formula with only a few variables, constitutes a major simplification. Its effects were estimated with a similar statistical approach as for the intrinsic errors, focusing on the melting temperature and the thermal expansion coefficient.

An appropriate fusion material needs to fulfil many competing criteria. Computationally screening the tungsten alloys hence not only requires accurate property predictions, but also a procedure to optimize the multiple objectives involved. In the second stage of our investigation, we developed a post-Pareto ranking to select a limited number of promising solutions from a large data set. This 'minimum win fraction' takes into account the trade-off between the different materials properties and thus identifies the most favourable candidate compounds. We included preference-dependent weight factors, which provide the minimum win fraction with a considerable flexibility to deal with case-specific circumstances.

These two studies paved the way for a computational screening of the tungsten alloys with respect to several fusion-related objectives. Whereas the first stage of this work focused on the ground-state elemental crystals, for which much experimental information is available, and the second stage investigated the binary tungsten alloys, as a test set for the multicriterion

methodology development, this final step encompassed a large number of tungsten ternaries. Before the previously proposed procedures could be applied, a Python-based high-throughput framework was first constructed to automate the calculations. It manages the preparation and parallellization of density-functional calculations, making use of the available computing power as efficiently as possible. We have applied this high-throughput approach to 210 ternary tungsten compounds, from which we were able to derive some intriguing results. In particular, the most suitable alloys for fusion purposes appear to posses one dopant with little d electrons and one dopant with an (almost) full d shell. This is the case in $W_{30}ZrAg$, for example.

Outlook

In the near future, we plan to investigate the influence of 3d transition metal dopants with our DFT high-throughput procedure. Although computationally not trivial, this may give rise to interesting effects, not only because 3d elements are significantly smaller than their 4d or 5d counterparts, but also because Cr, Mn, Fe, Co and Ni are often spin-polarized. Indeed, magnetism can lead to large and unexpected changes in the materials properties. For fusion purposes this is less relevant, as most compounds will be above their Curie or Néel temperature, but it may provide insight into the effects of magnetic inclusions in transition metal alloys. In a preliminary test calculation, for example, we noticed that magnetism is able to change the relative stability of the different crystal structures (see Figure 5.1): in nonmagnetic $W_{30}TiMn$, the pos1 configuration is most stable, while including spin polarization changes the preference to a pos4 configuration (followed by pos6). The treatment of 3d dopants is also of importance for our Pareto analyses. Most 3d elements are inexpensive. Incorporating them into our database may therefore modify our currently established Pareto front and mwf ranking.

A second extension of this work relates to the used ductility criteria. In paper 3, we showed the Cauchy pressure to be too simplistic to predict the ductile behaviour of tungsten alloys. We addressed this issue in our study of the ternaries by considering four different predictors. Unfortunately, the C_{44} parameter proved numerically too sensitive for reliable conclusions. In addition, a polycrystalline approximation was applied to the prelogarithmic energy. Although this approximation is common practice in experimental work, it may influence our Pareto results and may even change the character of some compounds from ductile to brittle or the other way round. An improvement of our model is therefore possible by calculating the exact prelogarithmic energies, using anisotropic continuum theory [44].

Another methodological assumption is that of ordered tungsten alloys.

In reality, most alloys are characterized by a lattice over which the host and dopant atoms are randomly distributed. This has a large impact on the materials properties. The formation energy of a disordered compound can differ significantly from that of an ordered phase, for example (see, e.g., [184]). The influence on the ductility predictors therefore also needs to be further investigated. Although the most straightforward approach is to consider ensembles of large disordered supercells, simulations of disorder are also possible in other density-functional-theory formalisms. They include the virtual crystal approximation [185], special quasirandom structures [186], the coherent-potential approximation [187] or a cluster expansion [188]. Especially the special quasirandom structure method is promising, as it is most easily combined with high-throughput schemes. Of course, the lack of experimental information for many multicomponent tungsten alloys again requires thorough benchmark calculations for better-studied materials systems. This ensures that these advanced simulations yield realistic predictions for fusion purposes.

We also foresee applications beyond the screening of tungsten alloys for nuclear fusion. Indeed, some of the developed methods are even useful beyond computational materials design. The minimum win fraction, for example, can be employed for the optimization of any data-driven problem subject to competing requirements, providing it with a broad multidisciplinary character. It shows that a construction project with gypsum board for the ceiling, matt emulsion paint for the walls and laminated flooring combines the best cost efficiency and lowest CO_2 emission [189]. It demonstrates how over time, Bengali dialects have evolved to an efficient compromise between ease of pronunciation and acoustic diversity [190]. It confirms that it is possible to devise a Hohmann-type transfer from the Earth to the Moon that outperforms all genuine two-step trajectories with respect to flight time and the number of fuel-consuming maneuvers [191]. The range of possible applications for the minimum win fraction is therefore almost unlimited. Whenever it is essential to select the best solution under multiple requirements, it may offer an unbiased alternative for human decision making and assist in formalizing our preferences.

Last, but certainly not least, we aim to continue our efforts with respect to the validation and verification of DFT codes and methods. As our work was one of the first on this topic, we have already received much response from developers of (pseudo)potentials and density-functional theory codes. It inspired us to start the DeltaCodesDFT initiative [160], in which already 16 codes participate and which aims to provide a broad overview of the performance of solid-state density-functional-theory methods. Table 6.1 provides a snapshot of this collaboration (March 2014). We have also become involved in a community-wide discussion between developers of codes and potentials.

Table 6.1: Symmetrized Δ values (in meV/atom) comparing different codes and methods (DeltaCodesDFT version 3.0 [160]). (L)APW (yellow), PAW (red) and pseudopotential implementations (blue) are grouped.

	WIEN2k	VASP-GW	VASP-5.3	VASP-5.2	Abinit-0.9	GPAW-0.9	GPAW-0.6	OpenMX	DACAPO	Abinit-FHI
WIEN2k	-	0.8	0.7	2.1	1.3	1.5	3.8	2.0	6.2	14.5
VASP-GW	0.8	-	0.9	2.1	1.6	1.9	4.0	2.2	6.1	15.1
VASP-5.3	0.7	0.9	-	1.8	1.4	1.7	3.7	2.1	6.3	14.5
VASP-5.2	2.1	2.1	1.8	-	2.4	2.5	2.8	1.7	6.1	13.3
Abinit-0.9	1.3	1.6	1.4	2.4	-	0.7	3.6	2.3	6.0	14.2
GPAW-0.9	1.5	1.9	1.7	2.5	0.7	-	3.6	2.5	6.4	14.3
GPAW-0.6	3.8	4.0	3.7	2.8	3.6	3.6	-	3.0	7.6	13.0
OpenMX	2.0	2.2	2.1	1.7	2.3	2.5	3.0	-	6.4	13.6
DACAPO	6.2	6.1	6.3	6.1	6.0	6.4	7.6	6.4	-	18.4
Abinit-FHI	14.5	15.1	14.5	13.3	14.2	14.3	13.0	13.6	18.4	-

† *WIEN2k by S. Cottenier (APW+lo with small muffin-tin radii and re-determined core-valence separation); VASP-GW by K. Lejaeghere (plane-wave PAW with GW-ready potentials); VASP-5.3 by K. Lejaeghere (plane-wave PAW with April 2012 potentials); VASP-5.2 by K. Lejaeghere (plane-wave PAW with pre-2012 potentials); Abinit-0.9 by F. Jollet and M. Torrent (plane-wave PAW with GPAW-0.9 potentials); GPAW-0.9 by M. Dułak (plane-wave PAW with 0.9 potentials); GPAW-0.6 by M. Dułak (grid-based PAW with 0.6 potentials); OpenMX by T. Ozaki (pseudo-atomic orbitals with Morrison-Bylander-Kleinman norm-conserving pseudopotentials); DACAPO by M. Dułak (plane waves with Vanderbilt ultrasoft pseudopotentials (v2)); Abinit-FHI by M. Dułak (plane waves with FHI Troullier-Martins norm-conserving pseudopotentials)*

This coordinated effort intends to re-assess and refine the available methodologies. For pseudopotential methods or PAW approaches, for example, it is essential that each code yields the same results for the same potentials, or that the remaining differences are understood. The availability of our APW+lo data also allows investigating the quality of the potentials them-

selves, and has rekindled the search for high-accuracy (pseudo)potentials. Finally, our WIEN2k data set has exposed unexpected differences between several (L)APW codes. Developers are now examining how an ultimate accuracy can be reached, using the Δ gauge as a guideline. Because of this apparent need for numerical error assessments, and at the request of several developers, a refined Δ criterion is now underway. It will eliminate the use of an absolute reference for code comparisons, treating all methods on equal terms. We are also working on a new and improved data set for WIEN2k, in which the computational settings border the extreme (Table 6.1 uses these results). For future developments, we aim to extend the currently used benchmark set of elemental crystals to multicomponent compounds, which will improve the overall transferability of Δ-based conclusions. Other functionals may also prove useful for further investigations, as to date, all comparisons are limited to the functional of Perdew, Burke and Ernzerhof. As a last point of interest, more advanced code comparison tools may be worth looking into as well. A Δ_1 criterion was recently defined to address some shortcomings of the original Δ [192], but not all issues seem to have been resolved yet.

A

Computational infrastructure

On a global scale, four levels of computational infrastructure are available. Although some calculations can be performed on personal computers (Tier-3), more powerful machines are provided at an institute-wide (Tier-2), regional (Tier-1) or international scale (Tier-0). For researchers at Ghent University, these clusters are provided by HPC UGent (Stevin supercomputer infrastructure [193]), the Flemish Supercomputer Centre (Tier-1 Flemish supercomputer infrastructure [194]) and PRACE (pan-European supercomputer infrastructure [195]) (Figure A.1).

Figure A.1: Top three tiers of computational infrastructure in Europe [193].

In this dissertation, computational resources were used at the Tier-3, Tier-2 and Tier-1 level, for which the available infrastructure has evolved extensively over the last few years. Table A.1 therefore summarizes the most important clusters and their specifications at the time of this work (spring 2014) [193, 196]. The machines with InfiniBand connections (IB) are mostly meant for highly parallellized or memory-intensive calculations, while gastly, haunter and raichu are directed towards large numbers of jobs, using one node or less. The former clusters

hence focus on *capability computing* (trying to solve large problems that regular computers cannot), and the latter on *capacity computing* (using computational resources as efficient as possible). The Tier-1 supercomputer muk, on the other hand, combines the best of both worlds. It primarily aims for massively parallel use, but it is equally suitable for large numbers of medium-scale calculations. We successfully applied for a Tier-1 pilot project in which 4 053 node days were granted for the high-throughput characterization of the tungsten ternaries (see Chapter 5).

Table A.1: Overview of the computational infrastructure used for this work (status in spring 2014) [196].

	year	cores	network	clock	RAM/node
molmod66	2010	4	–	3.00 GHz	3.9 GiB
gengar	2008	1536†	IB DDR	2.5 GHz	16 GiB
gastly	2009	448	Ethernet	2.26 GHz	12 GiB
haunter	2010	1336	Ethernet	2.26 GHz	12 GiB
gulpin	2011	1088	IB 2x QDR	2.4 GHz	64 GiB
dugtrio	2011	192	IB 2x QDR	3.06 GHz	96 GiB
raichu	2012	1024	Ethernet	2.6 GHz	32 GiB
delcatty	2013	2560	IB FDR	2.6 GHz	64 GiB
muk	2013	8448	IB FDR	2.6 GHz	64 GiB

† Original number of cores.

The performance of VASP on the different clusters is illustrated in Figures A.2 and A.3. These plots show that the Tier-1 infrastructure muk indeed outperforms all Tier-2 machines. It possesses the single-node efficiency of raichu, gastly and haunter, but its InfiniBand connections also allow for an almost ideal scaling behaviour up to 100 cores, similar to gengar. The same performance is expected for the smaller cluster delcatty. The main limitation when using VASP on muk is the available memory. VASP is very memory-intensive, and we have so far not succeeded in simulating cells with more than 1728 atoms (using a single-point calculation on 128 16-core nodes).

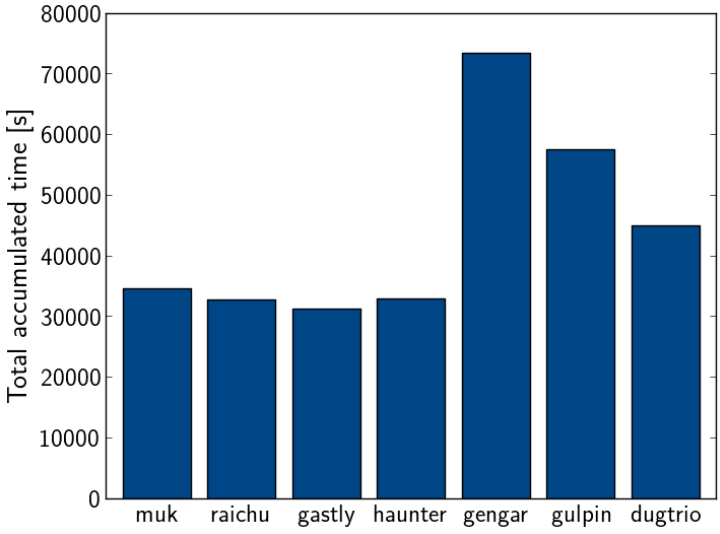


Figure A.2: Total computation time on one full node (real time $\times$ number of cores) for a benchmark VASP calculation of a Ge [100] nanowire, presented for all available Tier-2 and Tier-1 machines. Only delcatty has not been included, but its performance should be identical to that of muk, as their architectures are the same.

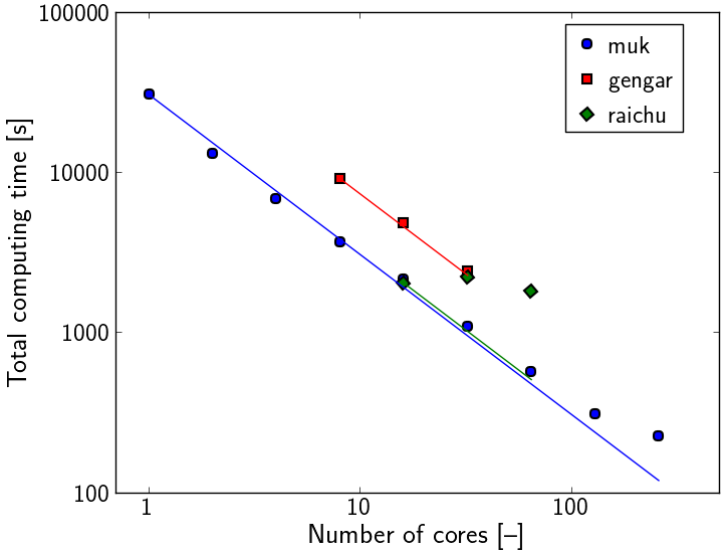


Figure A.3: Scaling behaviour of the wall time of muk, gengar and raichu for a benchmark VASP calculation of a Ge [100] nanowire. The corresponding ideal scaling laws $\propto N^{-1}$ are represented by full lines.

B

Tight-binding formalism for transition-metal compounds

As mentioned in Section 1.3.2, the full many-body Schrödinger equation is almost impossible to solve without some approximations. In the framework of this dissertation, we focused on the use of density-functional theory, but there are other possibilities. A more intuitive (yet more approximate) formalism for solids is that of tight binding (TB). TB models already originated in the 1950s with the seminal work of Slater and Koster [197] and proved very successful in explaining the structures and properties of a wide range of solids. The treatment of transition metals was facilitated by additional developments, such as the second-moment approximation [198, 199] and the rectangular d band model [200], which even allowed treating elastic constants [201] and disordered alloys [202]. Below, we present some of the most important results, using the notation by Pettifor in his 1995 book [178].

In essence, tight-binding models assume the electron wave functions to be tightly bound to their corresponding atoms, displaying only little interactions with the rest of the crystal. The ionic potential can therefore be decomposed in localized contributions, $v(\boldsymbol{r}) = \sum_{\boldsymbol{R}} v_{at}(\boldsymbol{r} - \boldsymbol{R})$ and the wave functions written as linear combinations of atomic orbitals (LCAO):

$$\psi_{\boldsymbol{k},n}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}} \psi_{at,n}(\boldsymbol{r} - \boldsymbol{R}) \tag{B.1}$$

with $\boldsymbol{R}$ the ionic positions and N the number of atoms in the unit cell. Note that (B.1) also fulfils Bloch's theorem. Tight-binding theory goes beyond the LCAO approximation, however. When completing the Schrödinger equation

with equation (B.1), this gives rise to matrix elements

$$\int \psi^*_{at,n}(\boldsymbol{r}-\boldsymbol{R})\, v_{at}(\boldsymbol{r}-\boldsymbol{S})\, \psi^*_{at,n}(\boldsymbol{r}-\boldsymbol{T})\, \mathrm{d}\boldsymbol{r} \tag{B.2}$$

TB also assumes the localized nature of the electron wave functions to allow neglecting all three-centre terms and crystal field contributions (wherein both local orbitals feel a neighbouring potential). Only integrals of the form

$$\int \psi^*_{at,n}(\boldsymbol{r}-\boldsymbol{R})\, v_{at}(\boldsymbol{r})\, \psi^*_{at,n}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \tag{B.3}$$

then remain, the so-called atomic bond integrals h, which are a function of the bond strength and the local coordination.

When applied to transition metals, five contributions to the bond energy can be identified: (i) the atomic preparation energy ($\mathrm{s}^2\mathrm{d}^n \rightarrow \mathrm{s}^1\mathrm{d}^{n+1}$), (ii) the renormalization energy (atom $\rightarrow$ solid), (iii) the sp band energy, (iv) the d bond energy and (v) the sp-d hybridization energy. Gelatt *et al.* showed only the d bond to contribute significantly, however [176]:

$$U_{bond} = \int^{E_F} (E - E_d) n_d(E)\, \mathrm{d}E \tag{B.4}$$

with E_d the centre of the d band and n_d the d band density of states. Indeed, the first half of the d band consists of bonding states, while the last half has an antibonding character. Practical TB calculations are possible by assuming a rectangular d band:

$$n_d(E) = \begin{cases} 10/W & |E - E_d| < W/2 \\ 0 & |E - E_d| > W/2 \end{cases} \tag{B.5}$$

in which the band width W can be determined from the second-moment approximation. This theorem states that the second moment of the DOS equals the trace of the squared Hamiltonian, which can be written in terms of tight-binding bond integrals. For d orbitals in elemental crystals, W becomes

$$W = \sqrt{\frac{12z}{5}\left(h^2_{dd\sigma} + 2h^2_{dd\pi} + 2h^2_{dd\delta}\right)} = \sqrt{12z}|h| \tag{B.6}$$

with z the coordination number. The d bond energy then yields

$$U_{bond} = -\sqrt{\frac{3z}{100}}|h| N_d(10 - N_d) \tag{B.7}$$

where N_d stands for the number of d electrons. This formula shows the unsaturated character of the d bond: the energy per bond U_{bond}/z decreases

like $1/\sqrt{z}$, indicating that the electrons are smeared out over the different bonds, and that increasing their number weakens the strength of each individual bond.

The stability of transition-metal compounds can be calculated from (B.7) using advanced calculations of the bond integral h as a function of the interatomic distance R. More intuitive results can be obtained when using pair potentials. Since the bond integrals are approximately related to the Pauli repulsion like h^2 [203], a Buckingham potential [204] allows writing the dependence on the bond distance in an exponential form:

$$U = U_{rep} + U_{bond} = \frac{1}{2} z a N_d^2 \, \mathrm{e}^{-2\kappa R} - \sqrt{\frac{3z}{100}} b N_d^2 (10 - N_d) \, \mathrm{e}^{-\kappa R} \tag{B.8}$$

where h was assumed proportional to the charge density, and hence to N_d. The prefactors a and b are independent of N_d or R, while the decay rate κ linearly increases with the number of d electrons. From (B.8), the structural properties of elemental crystals can be assessed. Minimizing U with respect to the interatomic distance yields:

$$R_0 = \frac{1}{\kappa} \ln \left[\frac{10 a \sqrt{z}}{\sqrt{3} b (10 - N_d)} \right] \tag{B.9}$$

$$W(R_0) = \frac{3b^2}{5a} N_d (10 - N_d) \tag{B.10}$$

$$\Delta E_{coh} = \frac{3b^2}{200a} \left[N_d (10 - N_d) \right]^2 \tag{B.11}$$

$$B_0 = \frac{2\sqrt{2}\kappa^2}{9 R_0} \Delta E_{coh} \tag{B.12}$$

$$B_1 = 1 + \kappa R_0 \tag{B.13}$$

A similar approach can be applied to binary compounds. When assuming the A-B bond integral in disordered, equiatomic AB alloys h_{AB} to equal the geometric mean of the elemental values $\sqrt{h_{AA} h_{BB}}$, the band width becomes

$$W_{AB}^2 = \bar{W}_{AB}^2 + 3 (\Delta E)^2 \tag{B.14}$$

with $\bar{W}_{AB} = [W_A(R_{AB}) + W_B(R_{AB})]/2$ and ΔE the energy mismatch between the two d bands. However, the creation of a single d band causes a charge transfer from one element to the other, while in a metal, local charge neutrality must be ensured. The d bands hence shift their position compared to elemental compounds – and their band width through (B.14) as well – until no net charge difference is present anymore. This corresponds with

$$\Delta E = - \frac{\bar{W}_{AB} \, \Delta N}{\sqrt{\frac{9}{25} [\bar{N}(10 - \bar{N})]^2 - 3 (\Delta N)^2}} \tag{B.15}$$

where $\bar{N}$ represents the average number of d electrons and ΔN the difference. When incorporated in (B.4), this yields

$$U_{bond}^{AB} = -\sqrt{\frac{3z}{100}}\bar{h}_{AB}\bar{N}(10-\bar{N})\cdot\sqrt{1-\frac{25}{3}\frac{(\Delta N)^2}{[\bar{N}(10-\bar{N})]^2}} \tag{B.16}$$

with $\bar{h}_{AB} = (h_{AA}+h_{BB})/2$. We indeed notice that the average bond energy is reduced by a factor proportional to ΔN due to the compensated charge transfer. The structural properties of the AB alloys can again be determined using pair potentials. R_0 is now the solution of a transcendent equation:

$$N_A\,\mathrm{e}^{-\kappa_A R_0} + N_B\,\mathrm{e}^{-\kappa_B R_0} = \frac{\sqrt{3}b}{5a\sqrt{z}}\sqrt{\bar{N}^2(10-\bar{N})^2 - \frac{25}{3}(\Delta N)^2} \tag{B.17}$$

yielding

$$\Delta E_{coh} = \frac{3b^2}{200a}\left[\bar{N}^2(10-\bar{N})^2 - \frac{25}{3}(\Delta N)^2\right] \tag{B.18}$$

$$B_0 = \frac{\sqrt{2}az}{36R_0}\left(\kappa_A N_A\,\mathrm{e}^{-\kappa_A R_0} + \kappa_B N_B\,\mathrm{e}^{-\kappa_B R_0}\right)^2 \tag{B.19}$$

$$B_1 = 1 + R_0\frac{\kappa_A^2 N_A\,\mathrm{e}^{-\kappa_A R_0} + \kappa_B^2 N_B\,\mathrm{e}^{-\kappa_B R_0}}{\kappa_A N_A\,\mathrm{e}^{-\kappa_A R_0} + \kappa_B N_B\,\mathrm{e}^{-\kappa_B R_0}} \tag{B.20}$$

C

Supplementary Material

Supplementary Material of Paper 1

Error estimates for solid-state density-functional theory predictions: an overview by means of the ground-state elemental crystals

K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier
Critical Reviews in Solid State and Materials Sciences (2014)
Supplementary Material

All citations in this document refer to the list that is given in the end, and not to the bibliography of the main paper.

1. Computational parameters

In the main article three DFT codes are evaluated. The current section lists the most important settings, applied to all test set compounds.

VASP

cut-off energy
400 eV for most elements
600 eV for He, B, C, N, O, F, and Ne

self-consistent cycle
10^{-4} eV energy convergence criterion

reciprocal space representation
6750 / N k-points for N-atom cells

FFT grid
default (2 G_{cut} x 2 G_{cut} x 2 G_{cut} ; G_{cut} being the cut-off electronic wave vector)

augmentation charge grid
default (4 G_{cut} x 4 G_{cut} x 4 G_{cut})

potential files
see Tab. 1.1

WIEN2k

muffin tin radius
2.30 a.u. when this does not cause the atomic spheres to overlap
else: all but touching spheres in a cell with volume V_0 - 6%

spherical harmonic truncation
lmax = 14 inside the muffin-tin spheres
lmax = 6 for matrix elements with contributions outside of the muffin-tin sphere

core-valence separation
see Tab. 1.2

product of the muffin tin radius and the maximum wave vector
see Tab. 1.2

Fourier expansion truncation
G_{max} equal to approximately 3 G_{min}

FFT mesh parameters
automatically calculated, by means of a multiplicative factor of 4

reciprocal space representation
30 000 k points per unit cell volume of 120 a.u. (full Brillouin zone)

energy level broadening
Fermi-Dirac distribution corresponding to $k_B T = 0.001$ Ry

self-consistent cycle
$5 \cdot 10^{-7}$ *e* charge convergence criterion

GPAW

grid spacing
0.075 Å

reciprocal space representation
6750 / N k-points for N-atom cells

self-consistent cycle
0.5 meV per valence electron energy convergence criterion
$1 \cdot 10^{-4}$ *e* per valence electron charge convergence criterion

energy level broadening
Fermi-Dirac distribution corresponding to $k_B T = 0.1$ eV

Tab. 1.1 VASP potentials used for the elemental crystals

	potential file
H	PAW_PBE H 15Jun2001
He	PAW_PBE He 05Jan2001
Li	PAW_PBE Li_sv 23Jan2001
Be	PAW_PBE Be 06Sep2000
B	PAW_PBE B 06Sep2000
C	PAW_PBE C 08Apr2002
N	PAW_PBE N 08Apr2002
O	PAW_PBE O 08Apr2002
F	PAW_PBE F 08Apr2002
Ne	PAW_PBE Ne 05Jan2001
Na	PAW_PBE Na_pv 05Jan2001
Mg	PAW_PBE Mg 05Jan2001
Al	PAW_PBE Al 04Jan2001
Si	PAW_PBE Si 05Jan2001
P	PAW_PBE P 17Jan2003
S	PAW_PBE S 17Jan2003
Cl	PAW_PBE Cl 17Jan2003
Ar	PAW_PBE Ar 07Sep2000
K	PAW_PBE K_sv 06Sep2000
Ca	PAW_PBE Ca_pv 06Sep2000
Sc	PAW_PBE Sc_sv 07Sep2000
Ti	PAW_PBE Ti_pv 07Sep2000
V	PAW_PBE V_pv 07Sep2000
Cr	PAW_PBE Cr_pv 07Sep2000
Mn	PAW_PBE Mn_pv 07Sep2000
Fe	PAW_PBE Fe 06Sep2000
Co	PAW_PBE Co 06Sep2000
Ni	PAW_PBE Ni 06Sep2000
Cu	PAW_PBE Cu 05Jan2001
Zn	PAW_PBE Zn 06Sep2000
Ga	PAW_PBE Ga_d 06Sep2000
Ge	PAW_PBE Ge_d 06Sep2000
As	PAW_PBE As 06Sep2000
Se	PAW_PBE Se 06Sep2000
Br	PAW_PBE Br 06Sep2000
Kr	PAW_PBE Kr 07Sep2000
Rb	PAW_PBE Rb_sv 06Sep2000
Sr	PAW_PBE Sr_sv 07Sep2000
Y	PAW_PBE Y_sv 06Sep2000
Zr	PAW_PBE Zr_sv 07Sep2000
Nb	PAW_PBE Nb_pv 08Apr2002
Mo	PAW_PBE Mo_pv 08Apr2002
Tc	PAW_PBE Tc_pv 06Sep2000
Ru	PAW_PBE Ru 06Sep2000
Rh	PAW_PBE Rh 06Sep2000
Pd	PAW_PBE Pd 05Jan2001
Ag	PAW_PBE Ag 06Sep2000

Cd	PAW_PBE Cd 06Sep2000
In	PAW_PBE In_d 06Sep2000
Sn	PAW_PBE Sn_d 06Sep2000
Sb	PAW_PBE Sb 06Sep2000
Te	PAW_PBE Te 08Apr2002
I	PAW_PBE I 08Apr2002
Xe	PAW_PBE Xe 07Sep2000
Cs	PAW_PBE Cs_sv 08Apr2002
Ba	PAW_PBE Ba_sv 06Sep2000
Lu	PAW_PBE Lu_3 06Sep2000
Hf	PAW_PBE Hf_pv 06Sep2000
Ta	PAW_PBE Ta_pv 07Sep2000
W	PAW_PBE W_pv 06Sep2000
Re	PAW_PBE Re 17Jan2003
Os	PAW_PBE Os_pv 20Jan2003
Ir	PAW_PBE Ir 06Sep2000
Pt	PAW_PBE Pt 05Jan2001
Au	PAW_PBE Au 06Sep2000
Hg	PAW_PBE Hg 06Sep2000
Tl	PAW_PBE Tl_d 06Sep2000
Pb	PAW_PBE Pb_d 06Sep2000
Bi	PAW_PBE Bi_d 06Sep2000
Po	PAW_PBE Po_d 25May2007
Rn	PAW_PBE Rn 28Aug2006

Tab. 1.2 WIEN2k core-valence separation and RK_{max} used for the elemental crystals

	highest core	lowest valence	RK_{max}
H		1s	5.0
He		1s	10.0
Li		1s	10.0
Be		1s	10.0
B	1s	2s	7.5
C	1s	2s	7.5
N	1s	2s	7.5
O	1s	2s	7.5
F	1s	2s	8.0
Ne	1s	2s	10.0
Na	1s	2s	10.0
Mg	1s	2s	10.0
Al	2s	2p	10.0
Si	2s	2p	10.0
P	2s	2p	9.0
S	2p	3s	10.0
Cl	2p	3s	8.0
Ar	2p	3s	10.0
K	2p	3s	10.0
Ca	2p	3s	10.0
Sc	2p	3s	10.0
Ti	2p	3s	10.0
V	2p	3s	10.0
Cr	2p	3s	10.0
Mn	2p	3s	10.0
Fe	2p	3s	10.0
Co	2p	3s	10.0
Ni	2p	3s	10.0
Cu	2p	3s	10.0
Zn	2p	3s	10.0
Ga	3p	3d	10.0
Ge	3p	3d	10.0
As	3p	3d	10.0
Se	3p	3d	10.0
Br	3p	3d	10.0
Kr	3p	3d	10.0
Rb	3d	4s	10.0
Sr	3d	4s	10.0
Y	3d	4s	10.0
Zr	3d	4s	10.0
Nb	3d	4s	10.0
Mo	3d	4s	10.0
Tc	3d	4s	10.0
Ru	3d	4s	10.0
Rh	3d	4s	10.0
Pd	3d	4s	10.0
Ag	3d	4s	10.0

Cd	3d	4s	10.0
In	4s	4p	10.0
Sn	4s	4p	10.0
Sb	4s	4p	10.0
Te	4s	4p	10.0
I	4s	4p	10.0
Xe	4s	4p	10.0
Cs	4p	4d	10.0
Ba	4p	4d	10.0
Lu	4d	5s	12.0
Hf	4d	5s	12.0
Ta	4d	5s	12.0
W	4d	5s	12.0
Re	4d	5s	12.0
Os	4d	5s	12.0
Ir	4d	5s	12.0
Pt	4d	5s	12.0
Au	4d	5s	12.0
Hg	4d	5s	12.0
Tl	4f	5p	12.0
Pb	4f	5p	12.0
Bi	4f	5p	12.0
Po	4f	5p	12.0
Rn	4f	5p	12.0

2. Difference with experiment

Tab. 2.1 Cohesive energy (in kJ/mol) by means of VASP-PBE and from experiment, with ($\Delta E_{coh,exp}$) and without ($\Delta E_{coh,exp}^{0}$) correcting for the zero-point energy ζ , and the Debye temperature Θ_D (in K) necessary for the calculation of ζ

	$\Delta E_{coh,VASP}$	$\Delta E_{coh,exp}$	$\Delta E_{coh,exp}^{0}$	Θ_D	ζ
H	219				
He	1.10				
Li	155	161	158 [1]	344 [1]	3.2
Be	360	333	320 [1]	1440 [1]	13.5
B	600	569	561 [1]	805 [4]	7.5
C	758	719	715 [2]	402 [3]	3.8
N	503	475	474 [1]	70 [4]	0.7
O	294	252	251 [1]	93 [4]	0.5
F	114	81	81 [1]	65 [4]	0.5
Ne	1.80	2.62	1.92 [1]	75 [1]	0.7
Na	104	108	107 [1]	158 [1]	1.5
Mg	145	149	145 [1]	400 [1]	3.7
Al	331	331	327 [1]	428 [1]	4.0
Si	439	452	446 [1]	645 [1]	6.0
P	336	333	331 [1]	231 [4]	2.2
S	294	277	275 [1]	182 [4]	1.7
Cl	141	137	135 [1]	142 [4]	2.0
Ar	9.16	8.60	7.74 [1]	92 [1]	0.9
K	83.9	91.0	90.1 [1]	91 [1]	0.9
Ca	185	180	178 [1]	230 [1]	2.2
Sc	400	379	376 [1]	360 [1]	3.4
Ti	528	472	468 [1]	420 [1]	3.9
V	571	516	512 [1]	380 [1]	3.6
Cr	390	401	395 [1]	630 [1]	5.9
Mn	372	286	282 [1]	410 [1]	3.8
Fe	468	417	413 [1]	470 [1]	4.4
Co	496	428	424 [1]	445 [1]	4.2
Ni	466	432	428 [1]	450 [1]	4.2
Cu	335	339	336 [1]	343 [1]	3.2
Zn	106	133	130 [1]	327 [1]	3.1
Ga	254	274	271 [1]	320 [1]	3.0
Ge	361	375	372 [1]	374 [1]	3.5
As	286	287.9	285.3 [1]	282 [1]	2.6
Se	251	238	237 [1]	90 [1]	0.8
Br	131	120	118 [1]	104 [4]	1.6
Kr	2.34	11.9	11.2 [1]	72 [1]	0.7
Rb	74.7	82.7	82.2 [1]	56 [1]	0.5
Sr	156	167	166 [1]	147 [1]	1.4
Y	402	425	422 [1]	280 [1]	2.6
Zr	604	606	603 [1]	291 [1]	2.7
Nb	669	733	730 [1]	275 [1]	2.6
Mo	603	662	658 [1]	450 [1]	4.2

Tc	658	665	661 [1]	387 [4]	3.6
Ru	644	656	650 [1]	600 [1]	5.6
Rh	567	558	554 [1]	480 [1]	4.5
Pd	357	379	376 [1]	274 [1]	2.6
Ag	240	286	284 [1]	225 [1]	2.1
Cd	71.5	114	112 [1]	209 [1]	2.0
In	223	244	243 [1]	108 [1]	1.0
Sn	306	305	303 [1]	200 [1]	1.9
Sb	261	267	265 [1]	211 [1]	2.0
Te	233	212	211 [1]	153 [1]	1.4
I	125	109	107 [1]	87 [4]	1.6
Xe	2.69	16.5	15.9 [1]	64 [1]	0.6
Cs	69.0	78.0	77.6 [1]	38 [1]	0.4
Ba	182	184	183 [1]	110 [1]	1.0
Lu	395	430	428 [1]	210 [1]	2.0
Hf	606	623	621 [1]	252 [1]	2.4
Ta	766	784	782 [1]	240 [1]	2.2
W	852	863	859 [1]	400 [1]	3.7
Re	760	779	775 [1]	430 [1]	4.0
Os	770	793	788 [1]	500 [1]	4.7
Ir	676	674	670 [1]	420 [1]	3.9
Pt	514	566	564 [1]	240 [1]	2.2
Au	300	370	368 [1]	165 [1]	1.5
Hg	20	66	65 [1]	71.9 [1]	0.7
Tl	146	183	182 [1]	78.5 [1]	0.7
Pb	189	197	196 [1]	105 [1]	1.0
Bi	200	211	210 [1]	119 [1]	1.1
Po	157	145	144 [1]	92 [4]	0.9
Rn	3.78	19.6	19.5 [1]	15 [4]	0.1

Tab. 2.2 Equilibrium volume (in $Å^3$/atom) by means of VASP-PBE and from experiment, with ($V_{0,exp}$) and without ($V_{0,exp}{}^0$) thermally correcting for $\Delta V^{(1)}$ and $\Delta V^{(2)}$ (see article), and the thermal volume expansion coefficient at room temperature $\alpha_{V,rt}$ (in 10^{-5}/K) necessary for the calculation of $\Delta V^{(1)}$

	$V_{0,VASP}$	$V_{0,exp}$	$V_{0,exp}{}^0$	$\Delta V^{(1)}$	$\Delta V^{(2)}$	$\alpha_{V,rt}$
H	17.41	6.58	19.03 [5]		12.45	
He	17.73	9.40	31.7 [5]		22.33	
Li	20.28	20.41	20.99 [5]		0.58	13.8 [14]
Be	7.92	7.82	8.11 [5]	0.04	0.25	3.39 [14]
B	7.25	7.16	7.29 [5]	0.03	0.11	2.49 [14]
C	11.66	8.06	8.82 [5]	0.03	0.73	2.4 [15]
N	29.65	21.09	22.47 [6]		1.39	
O	19.18	15.98	17.36 [5]		1.38	
F	19.52	14.47	16.05 [7]		1.58	
Ne	24.69	17.85	22.2 [8]		4.35	
Na	37.07	37.18	37.74 [5]		0.56	21.3 [14]
Mg	22.85	22.65	23.24 [5]	0.26	0.33	7.44 [14]
Al	16.49	16.27	16.6 [5]	0.17	0.16	6.93 [14]
Si	20.45	19.82	20.02 [5]	0.02	0.17	0.78 [14]
P	21.38	18.50	18.99 [5]	0.28	0.21	9.9 [16]
S	38.48	24.63	25.52 [5]	0.43	0.46	11.19 [17]
Cl	38.18	26.95	27.53 [9]		0.58	
Ar	52.67	35.55	37.2 [10]		1.65	
K	73.85	72.20	75.72 [5]	2.84	0.68	24.99 [14]
Ca	42.17	42.95	43.63 [5]	0.44	0.25	6.69 [14]
Sc	24.66	24.77	25 [5]	0.11	0.12	3.06 [14]
Ti	17.37	17.50	17.64 [5]	0.07	0.07	2.58 [14]
V	13.49	13.81	13.92 [5]	0.05	0.06	2.52 [14]
Cr	11.83	11.82	12 [5]	0.03	0.15	1.47 [14]
Mn	11.10	11.97	12.2 [5]	0.12	0.11	6.51 [14]
Fe	11.37	11.64	11.78 [5]	0.06	0.08	3.54 [14]
Co	10.88	10.96	11.08 [5]	0.06	0.06	3.9 [14]
Ni	10.94	10.81	10.93 [5]	0.07	0.06	4.02 [14]
Cu	12.03	11.65	11.81 [5]	0.09	0.08	4.95 [14]
Zn	15.35	14.86	15.21 [5]	0.21	0.14	9.06 [14]
Ga	20.36	19.15	19.47 [5]	0.16	0.16	5.4 [14]
Ge	23.91	22.44	22.64 [5]	0.06	0.14	1.74 [14]
As	22.73	21.23	21.52 [5]	0.16	0.13	5.04 [18]
Se	29.84	26.22	27.27 [5]	0.68	0.37	16.61 [14]
Br	39.29	31.47	31.89 [9]		0.42	
Kr	66.44	44.04	45 [11]		0.96	
Rb	91.24	89.16	92.84 [5]	3.28	0.40	23.55 [19]
Sr	54.53	55.60	56.32 [5]	0.57	0.15	6.75 [14]
Y	32.92	32.95	33.18 [5]	0.16	0.07	3.18 [14]
Zr	23.53	23.18	23.28 [5]	0.06	0.04	1.71 [14]
Nb	18.34	17.97	18.07 [5]	0.06	0.04	2.19 [14]
Mo	15.92	15.51	15.58 [5]	0.03	0.04	1.44 [14]
Tc	14.60	14.22	14.3 [5]	0.05	0.04	2.1 [1,20]

Ru	13.84	13.45	13.57 [5]	0.04	0.08	1.92 [14]
Rh	14.18	13.57	13.67 [5]	0.05	0.05	2.46 [14]
Pd	15.44	14.56	14.69 [5]	0.08	0.05	3.54 [14]
Ag	18.05	16.85	17.06 [5]	0.15	0.06	5.67 [14]
Cd	23.05	21.01	21.44 [5]	0.30	0.14	9.24 [14]
In	27.54	25.70	26.17 [5]	0.38	0.09	9.63 [14]
Sn	36.86	33.97	34.16 [5]	0.08	0.11	1.62 [14]
Sb	31.79	29.68	29.97 [5]	0.15	0.14	3.3 [14]
Te	34.98	33.30	33.94 [5]	0.26	0.38	5.04 [14]
I	49.78	37.04	42.63 [5]	5.10	0.49	79.7 [20]
Xe	87.78	56.87	57.6 [12]		0.73	
Cs	116.96	110.32	115.79 [5]	5.05	0.41	29.1 [14]
Ba	63.55	62.29	62.99 [5]	0.58	0.12	6.18 [14]
Lu	29.32	29.27	29.5 [5]	0.13	0.10	2.97 [14]
Hf	22.42	22.30	22.41 [5]	0.06	0.05	1.77 [14]
Ta	18.30	17.93	18.01 [5]	0.05	0.03	1.89 [14]
W	16.28	15.80	15.86 [5]	0.03	0.03	1.35 [14]
Re	14.96	14.62	14.7 [5]	0.04	0.04	1.86 [14]
Os	14.44	13.85	13.91 [5]	0.03	0.03	1.53 [14]
Ir	14.61	14.06	14.14 [5]	0.04	0.04	1.92 [14]
Pt	15.78	15.02	15.11 [5]	0.06	0.03	2.64 [14]
Au	18.05	16.82	16.97 [5]	0.11	0.04	4.26 [14]
Hg	28.27	22.25	22.55 [13]		0.30	
Tl	31.17	27.99	28.41 [5]	0.38	0.03	8.97 [14]
Pb	32.27	29.86	30.34 [5]	0.39	0.08	8.67 [14]
Bi	38.27	35.13	35.38 [5]	0.21	0.04	4.02 [14]
Po	39.62	36.93	37.43 [5]	0.40	0.10	7.05 [14]
Rn	89.60					

Tab. 2.3 Bulk modulus (in GPa) by means of VASP-PBE and from experiment, with ($B_{0,exp}$) and without ($B_{0,exp}{}^{0}$) thermally correcting for $\Delta B^{(1)}$ and $\Delta B^{(2)}$ (see article)

	$B_{0,VASP}$	$B_{0,exp}$	$B_{0,exp}{}^{0}$	$\Delta B^{(1)}$	$\Delta B^{(2)}$
H	0.4	0.5	0.2 [1]		-0.27
He	0.9	0.42	0.08 [1]		-0.34
Li	13.8	13.1	11.6 [1]	-0.84	-0.62
Be	122.8	108.1	100.3 [1]	-1.65	-6.20
B	210.5	185.9	178 [1]	-2.65	-5.27
C	1.6	55.6	33.8 [21]	-1.08	-20.71
N	0.8	2.69	2.16 [22]		-0.53
O	1.4				
F	1.6				
Ne	1.1	2.8	1.1 [23]		-1.70
Na	7.7	7.9	6.8 [1]	-0.90	-0.22
Mg	36.5	38.6	35.4 [1]	-1.90	-1.31
Al	77.3	77.1	72.2 [1]	-3.34	-1.60
Si	88.8	101.3	98.8 [1]	-0.51	-1.97
P	7.3	33.2	30.4 [1]	-2.03	-0.77
S	0.4	21.3	17.8 [1]	-2.02	-1.48
Cl	1.3	40.8	11.7 [24]	-28.37	-0.73
Ar	0.8	3.30	2.68 [25]		-0.62
K	3.6	3.8	3.2 [1]	-0.49	-0.06
Ca	17.5	15.9	15.2 [1]	-0.47	-0.18
Sc	54.3	44.5	43.5 [1]	-0.56	-0.48
Ti	112.5	107.4	105.1 [1]	-1.38	-0.87
V	181.4	165.8	161.9 [1]	-2.53	-1.41
Cr	177.0	204.6	190.1 [1]	-2.89	-11.57
Mn	212.3	174.7	158 [26]	-10.18	-6.52
Fe	185.9	175.1	168.3 [1]	-4.11	-2.70
Co	208.4	198.4	191.4 [1]	-4.77	-2.23
Ni	193.7	192.5	186 [1]	-4.49	-1.97
Cu	136.7	144.3	137 [1]	-4.96	-2.31
Zn	57.9	64.7	59.8 [1]	-3.58	-1.29
Ga	48.3	60.2	56.9 [1]	-2.15	-1.15
Ge	58.8	79.4	77.2 [1]	-0.96	-1.23
As	25.5	40.8	39.4 [1]	-0.98	-0.46
Se	4.4	10.9	9.1 [1]	-1.31	-0.44
Br	1.6	25.6	13.3 [24]	-11.80	-0.51
Kr	0.7	4.01	3.61 [27]		-0.40
Rb	2.8	3.6	3.1 [1]	-0.43	-0.03
Sr	11.1	12.0	11.6 [1]	-0.29	-0.09
Y	41.4	37.3	36.6 [1]	-0.38	-0.33
Zr	93.5	84.3	83.3 [1]	-0.55	-0.43
Nb	171.7	173.2	170.2 [1]	-2.24	-0.73
Mo	262.9	276.2	272.5 [1]	-2.34	-1.37
Tc	298.4	303.1	297 [1]	-4.28	-1.78
Ru	310.9	335.5	320.8 [1]	-6.11	-8.61
Rh	254.3	277.1	270.4 [1]	-4.49	-2.24

Pd	168.4	187.2	180.8 [1]	-4.80	-1.59
Ag	89.3	105.7	100.7 [1]	-4.05	-0.96
Cd	36.7	50.7	46.7 [1]	-3.17	-0.79
In	35.7	44.7	41.1 [1]	-3.18	-0.43
Sn	35.8	42.8	42.5 [28]	-0.04	-0.28
Sb	30.7	39.5	38.3 [1]	-0.82	-0.40
Te	17.8	26.2	23 [1]	-1.46	-1.75
I	1.9	23.9	13.6 [24]	-9.76	-0.59
Xe	0.5	3.83	3.64 [29]		-0.19
Cs	2.0	2.3	2 [1]	-0.33	-0.01
Ba	8.9	10.6	10.3 [1]	-0.23	-0.06
Lu	47.3	42.0	41.1 [1]	-0.64	-0.27
Hf	107.2	110.7	109 [1]	-1.14	-0.52
Ta	193.0	202.7	200 [1]	-2.13	-0.56
W	298.0	327.5	323.2 [1]	-2.83	-1.45
Re	366.4	380.8	372 [1]	-5.61	-3.15
Os	386.7	424.6	418 [1]	-4.32	-2.29
Ir	337.0	362.2	355 [1]	-4.94	-2.30
Pt	241.7	285.5	278.3 [1]	-5.71	-1.50
Au	141.3	182.0	173.2 [1]	-7.08	-1.72
Hg	12.9				
Tl	26.7	37.4	35.9 [1]	-1.45	-0.09
Pb	36.2	46.3	43 [1]	-2.98	-0.36
Bi	22.5	32.1	31.5 [1]	-0.46	-0.12
Po	33.9	27.4	26 [1]	-1.28	-0.17
Rn	0.6				

Tab. 2.4 Pressure derivative of the bulk modulus B_1 (dimensionless) by means of VASP-PBE and from experiment

	$B_{1,VASP}$	$B_{1,exp}$
H	3.94	
He	7.52	
Li	3.16	3.51 [30]
Be	3.24	
B	3.99	
C	9.89	8.9 [30]
N	6.17	
O	5.62	
F	7.14	
Ne	8.13	9.23 [30]
Na	3.27	4.13 [30]
Mg	3.92	4.8 [30]
Al	4.62	4.45 [30]
Si	4.29	4.43 [30]
P	13.56	4.5 [30]
S	20.11	6.75 [30]
Cl	11.01	5.2 [30]
Ar	6.81	7.2 [30]
K	3.70	4.089 [30]
Ca	3.22	3.1 [30]
Sc	3.41	2.8 [31]
Ti	3.60	3.4 [30]
V	3.81	4.135 [30]
Cr	7.40	6.895 [30]
Mn	6.40	6.6 [30]
Fe	4.93	4.6 [30]
Co	4.58	4.26 [32]
Ni	4.97	4 [30]
Cu	5.05	4.88 [30]
Zn	5.77	4.4 [30]
Ga	4.66	
Ge	4.85	4.76 [33]
As	12.58	3.3 [34]
Se	9.54	5.8 [30]
Br	16.58	5.2 [30]
Kr	8.49	7.2 [30]
Rb	3.73	3.885 [30]
Sr	5.02	2.485 [30]
Y	3.21	2.2 [30]
Zr	3.55	2.575 [30]
Nb	3.99	4.015 [30]
Mo	4.34	3.98 [30]
Tc	4.55	
Ru	4.96	6.61 [32]
Rh	5.24	4.5 [32]

Pd	5.54	5 [30]
Ag	5.89	4.725 [30]
Cd	6.21	4.9 [30]
In	5.69	5.35 [30]
Sn	4.83	4 [30]
Sb	7.70	4.3 [30]
Te	10.28	8.4 [30]
I	13.57	6 [30]
Xe	7.43	6.34 [30]
Cs	3.32	3.79 [30]
Ba	3.13	2.43 [30]
Lu	3.55	
Hf	3.46	3.95 [32]
Ta	3.98	3.75 [30]
W	4.26	4.32 [30]
Re	4.57	5.41 [32]
Os	5.00	4.5 [35]
Ir	5.23	4.83 [32]
Pt	5.45	5.18 [32]
Au	6.07	6.4 [30]
Hg	8.72	
Tl	5.55	5.8 [36]
Pb	5.11	5.335 [30]
Bi	7.64	2.4 [30]
Po	5.35	
Rn	8.57	

Tab. 2.5 Elastic constants (in GPa) by means of VASP-PW91 and from experiment (both taken from [37])

	$C_{11,VASP}$	$C_{11,exp}$	$C_{12,VASP}$	$C_{12,exp}$	$C_{33,VASP}$	$C_{33,exp}$
Li	14.6	14.8	13.8	12.5	14.6	14.8
Be	312.2	299.4	24.6	27.6	377.9	342.2
Na	9.8	9.45	7.9	7.79	9.8	9.45
Mg	58.1	63.48	27.6	25.94	64.7	66.45
Al	101.0	114.3	61.0	61.92	101.0	114.3
K	3.7	4.16	3.2	3.41	3.7	4.16
Ca	22.1	22.8	15.3	16	22.1	22.8
Sc	104.8	99.3	37.7	39.7	105.2	107
Ti	174.8	176.1	87.5	86.9	189.5	190.5
V	272.0	232.4	144.8	119.36	272.0	232.4
Cr	247.6	391	73.4	89.6	247.6	391
Fe	279.2	243.1	148.8	138.1	279.2	243.1
Co	363.4	319.5	168.2	166.1	411.7	373.6
Ni	275.5	261.2	160.1	150.8	275.5	261.2
Cu	174.8	176.2	122.8	124.94	174.8	176.2
Zn	159.5	179.09	56.0	37.5	57.0	68.8
Rb	3.1	3.42	2.7	2.88	3.1	3.42
Sr	15.0	15.3	10.5	10.3	15.0	15.3
Y	78.0	83.4	24.8	29.1	82.4	80.1
Zr	145.3	155.4	67.3	67.2	166.1	172.5
Nb	247.2	252.7	140.0	133.2	247.2	252.7
Mo	466.0	450.02	165.2	172.92	466.0	450.02
Tc	491.4	433	222.4	199	546.8	470
Ru	558.9	576.3	188.5	187.2	631.8	640.5
Rh	405.3	413	185.5	194	405.3	413
Pd	198.0	234.12	155.7	176.14	198.0	234.12
Ag	115.9	131.44	85.1	97.3	115.9	131.44
Cd	93.2	129.23	40.7	39.99	43.9	56.68
Cs	2.1	2.59	2.0	2.17	2.1	2.59
Ba	12.3	12.6 [38,39]	7.6	11.1 [38,39]	12.3	12.6 [38,39]
Hf	183.6	190.1	71.1	74.5	197.9	204.4
Ta	260.9	266.32	165.2	158.16	260.9	266.32
W	517.8	532.55	201.7	204.95	517.8	532.55
Re	613.1	644.6	279.9	277	674.5	717
Os	729.6	763.3	230.0	227.9	816.9	843.2
Ir	580.8	599.47	232.0	255.82	580.8	599.47
Pt	296.4	358	225.6	253.6	296.4	358
Au	159.1	201.63	136.7	169.67	159.1	201.63
Tl	34.9	44.4	29.2	37.6	48.7	60.2
Pb	75.2	55.54	60.2	45.42	75.2	55.54

	$C_{13,VASP}$	$C_{13,exp}$	$C_{44,VASP}$	$C_{44,exp}$
Li	13.8	12.5	11.5	10.8
Be	10.9	11	165.4	166.2
Na	7.9	7.79	6.6	6.18
Mg	21.6	21.7	14.2	18.42
Al	61	61.92	25.4	31.62
K	3.2	3.41	2.6	2.86
Ca	15.3	16	13.2	14
Sc	29.3	29.4	31.6	27.7
Ti	80	68.3	40.9	50.8
V	144.8	119.36	17.6	45.95
Cr	73.4	89.6	48.3	103.2
Fe	148.8	138.1	93	121.9
Co	116.3	102.1	91.3	82.4
Ni	160.1	150.8	126.3	131.7
Cu	122.8	124.94	76.3	81.77
Zn	51.8	55.4	23.2	45.95
Rb	2.7	2.88	2	2.21
Sr	10.5	10.3	13.9	9.9
Y	22.7	19	26.4	26.9
Zr	69.5	64.6	24.3	36.3
Nb	140	133.2	14.2	30.97
Mo	165.2	172.92	99.5	125.03
Tc	183.4	199	129.9	177
Ru	176.4	167.3	179.9	189.1
Rh	185.5	194	176.5	184
Pd	155.7	176.14	69.7	71.17
Ag	85.1	97.3	42.1	51.08
Cd	34.5	40.95	11.8	24.2
Cs	2	2.17	1.4	1.6
Ba	7.6	11.1 [38,39]	10.5	9.5 [38,39]
Hf	71.1	65.5	51.1	60
Ta	165.2	158.16	70.4	87.36
W	201.7	204.95	139.4	163.13
Re	223	195.9	142.5	168.5
Os	229	218	251.5	269.3
Ir	232	255.82	249.8	268.82
Pt	225.6	253.6	50.7	77.4
Au	136.7	169.67	27.6	45.44
Tl	24.8	30	4.7	8.8
Pb	60.2	45.42	25.8	19.42

3. Difference between codes

For each element the crystal structure is the one defined by the corresponding CIF file in CIFs.tar.gz.

Tab. 3.1 Equilibrium volume V_0 (in Å^3/atom), bulk modulus B_0 (in GPa) and its pressure derivative B_1 (dimensionless) by means of WIEN2k-PBE for fixed geometries. These data are available in WIEN2k.txt as well, as an unformatted text file

	$V_{0,WIEN2k}$	$B_{0,WIEN2k}$	$B_{1,WIEN2k}$
H	17.39	10.3	3.03
He	17.78	0.8	6.53
Li	20.22	13.9	3.75
Be	7.92	123.0	3.15
B	7.25	237.6	3.48
C	11.65	209.6	3.57
N	28.90	54.3	3.75
O	18.56	51.2	3.93
F	19.35	35.0	4.25
Ne	24.35	1.2	8.66
Na	37.09	7.7	3.74
Mg	22.93	35.7	4.26
Al	16.50	77.5	4.67
Si	20.55	88.7	4.23
P	21.61	68.9	4.42
S	17.35	85.5	4.44
Cl	39.18	19.3	4.49
Ar	52.21	0.7	7.84
K	73.71	3.6	3.73
Ca	42.21	17.3	3.17
Sc	24.62	54.6	3.40
Ti	17.41	112.7	3.59
V	13.52	185.2	3.73
Cr	11.91	183.8	7.37
Mn	11.61	131.2	0.85
Fe	11.45	196.1	6.14
Co	10.95	216.2	5.08
Ni	10.99	204.7	4.85
Cu	12.02	143.7	5.20
Zn	15.27	75.6	5.38
Ga	20.38	49.1	5.33
Ge	24.09	59.5	5.08
As	22.57	69.7	4.29
Se	29.97	47.6	4.55
Br	39.78	22.6	5.15
Kr	65.23	0.9	22.01
Rb	91.13	2.8	2.32
Sr	54.56	11.3	4.17
Y	32.86	41.3	2.95

Zr	23.40	94.1	3.25
Nb	18.16	168.7	3.50
Mo	15.83	260.4	4.26
Tc	14.47	301.4	4.56
Ru	13.81	315.4	4.96
Rh	14.08	260.6	5.44
Pd	15.33	170.4	5.85
Ag	17.86	91.3	5.80
Cd	22.87	44.2	7.09
In	27.50	34.8	4.87
Sn	36.88	36.0	5.04
Sb	31.79	50.7	4.50
Te	35.01	44.8	4.70
I	50.34	18.7	5.22
Xe	87.32	0.6	-0.72
Cs	117.75	2.0	3.59
Ba	63.20	8.4	3.30
Lu	29.06	47.7	4.05
Hf	22.55	108.1	3.23
Ta	18.30	193.7	4.77
W	16.17	302.6	4.30
Re	14.99	364.6	4.47
Os	14.32	402.2	4.35
Ir	14.53	341.7	6.96
Pt	15.68	251.8	5.34
Au	17.99	139.9	5.99
Hg	29.72	8.2	8.08
Tl	31.46	26.7	4.57
Pb	32.00	39.6	5.95
Bi	36.94	42.6	4.63
Po	37.57	45.5	5.02
Rn	92.76	0.5	13.13

Tab. 3.2 Equilibrium volume V_0 (in Å^3/atom), bulk modulus B_0 (in GPa) and its pressure derivative B_1 (dimensionless) by means of VASP-PBE and GPAW-PBE for fixed geometries

	$V_{0,VASP}$	$B_{0,VASP}$	$B_{1,VASP}$	$V_{0,GPAW}$	$B_{0,GPAW}$	$B_{1,GPAW}$
H	17.44	10.1	3.03	17.46	10.3	2.93
He	17.72	0.9	7.42	17.69	0.8	6.73
Li	20.29	13.8	3.16	20.28	14.0	3.41
Be	7.92	123.3	3.29	8.01	123.3	3.37
B	7.25	237.0	3.47	7.24	237.1	3.53
C	11.66	208.4	3.56	11.65	208.7	3.62
N	29.65	56.4	3.86	28.91	54.0	3.67
O	19.19	51.4	4.06	19.28	53.6	3.96
F	19.54	34.2	4.20	19.21	33.7	4.50
Ne	24.61	1.1	14.48	23.75	2.1	20.67
Na	37.07	7.7	3.04	37.04	7.8	3.91
Mg	22.85	36.5	3.92	22.97	36.3	4.18
Al	16.49	77.3	4.65	16.51	78.2	4.61
Si	20.45	88.8	4.30	20.52	88.8	4.36
P	21.36	68.5	4.34	21.53	68.1	4.44
S	17.17	83.5	4.14	17.22	83.7	4.32
Cl	38.21	19.2	4.33	38.95	19.0	4.59
Ar	52.65	0.8	7.35	52.66	0.8	3.27
K	73.84	3.6	3.82	73.78	3.6	2.49
Ca	42.17	17.5	3.02	42.46	17.4	3.19
Sc	24.66	54.4	3.41	24.66	54.2	3.10
Ti	17.37	112.5	3.60	17.26	114.4	3.71
V	13.48	181.6	4.02	13.73	185.5	4.38
Cr	11.83	176.8	7.17	11.87	161.8	6.67
Mn	11.57	115.4	0.35	11.75	126.6	1.12
Fe	11.38	185.7	4.91	11.48	196.7	5.03
Co	10.88	210.8	4.99	10.92	212.4	5.00
Ni	10.94	193.7	4.92	10.92	203.6	4.60
Cu	12.03	136.8	5.10	12.10	139.4	4.59
Zn	15.29	74.2	5.62	15.25	75.1	4.80
Ga	20.36	48.9	5.09	20.52	50.3	5.24
Ge	23.91	58.8	4.78	23.98	60.1	4.84
As	22.69	68.7	4.29	22.64	68.9	4.43
Se	29.83	47.2	4.48	29.75	47.2	4.61
Br	39.47	22.5	4.85	40.38	20.8	4.66
Kr	66.56	0.7	4.82	65.58	0.8	13.16
Rb	91.24	2.8	3.73	91.17	2.7	3.06
Sr	54.54	11.1	4.59	55.15	11.2	3.68
Y	32.92	41.4	3.20			
Zr	23.53	93.8	3.58	23.46	94.9	3.38
Nb	18.34	171.8	3.97	18.23	174.2	3.78
Mo	15.92	262.9	4.34	15.89	263.7	4.38
Tc	14.60	298.5	4.55			
Ru	13.84	311.5	4.95	14.09	310.9	4.87
Rh	14.18	254.2	5.25	14.32	255.2	5.42

Pd	15.44	168.2	5.64	15.41	167.3	5.98
Ag	18.05	89.1	5.78	18.07	90.7	5.52
Cd	23.01	43.8	7.23	22.87	43.3	6.28
In	27.54	35.4	5.34	27.53	35.1	5.60
Sn	36.86	35.8	4.85	36.90	36.0	4.90
Sb	31.79	50.9	4.54			
Te	34.96	45.0	4.72	34.17	46.1	4.81
I	50.12	18.7	5.09	51.15	17.6	4.88
Xe	87.77	0.5	7.76			
Cs	116.96	2.0	3.48	117.01	2.1	3.59
Ba	63.55	8.9	3.11	63.51	8.9	2.95
Lu	29.47	47.2	3.44			
Hf	22.50	107.8	3.48			
Ta	18.33	194.3	3.97	18.52	198.4	3.99
W	16.23	304.4	4.30	16.44	307.5	4.35
Re	14.93	372.5	4.55			
Os	14.36	400.9	4.95	14.52	397.9	4.98
Ir	14.57	346.0	5.17	14.70	351.6	5.39
Pt	15.72	248.1	5.53	15.79	247.6	5.67
Au	18.18	136.4	5.91	18.21	137.8	5.77
Hg	29.94	7.6	12.94			
Tl	31.52	26.5	5.43			
Pb	32.08	39.8	4.40	32.10	39.1	5.56
Bi	36.99	42.5	4.55	36.97	42.9	4.76
Po	37.53	45.4	4.63			
Rn	93.11	0.6	7.24			

[1] C. Kittel, *Introduction to Solid State Physics* , 8th ed. (John Wiley & Sons, Inc, 2005).
[2] N. Greenwood and A. Earnshaw, *Chemistry of the Elements* , 2nd ed. (Butterworth-Heinemann, 1997).
[3] T. Tohei, A. Kuwabara, F. Oba, and I. Tanaka, Phys. Rev. B **73**, 064304 (2006).
[4] V. Moruzzi, J. Janak, and K. Schwarz, Phys. Rev. B **37**, 790 (1988).
[5] P. Villars and J. Daams, J. Alloys Compd. **197**, 177 (1993).
[6] L. Bolz, M. Boyd, F. Mauer, and H. Peiser, Acta Crystallogr. **12**, 247 (1959).
[7] L. Meyer, C. Barrett, and S. Greer, J. Chem. Phys. **49**, 1902 (1968).
[8] D. Batchelder, D. Losee, and R. Simmons, Phys. Rev. **162**, 767 (1967).
[9] B. Powell, K. Heal, and B. Torrie, Mol. Phys. **53**, 929 (1984).
[10] O. Peterson, D. Batchelder, and R. Simmons, Phys. Rev. **150**, 703 (1966).
[11] D. Losee and R. Simmons, Phys. Rev. **172**, 944 (1968).
[12] D. Sears and H. Klug, J. Chem. Phys. **37**, 3002 (1962).
[13] C. Swenson, Phys. Rev. **111**, 82 (1958).
[14] D. Lide (ed.), *CRC Handbook of Chemistry and Physics* , 89th ed. (CRC Press, Boca Raton, Florida, 2008-2009).
[15] J. Nelson and D. Riley, Proc. Phys. Soc. London **57**, 477 (1945).
[16] R. Keyes, Phys. Rev. **92**, 580 (1953).
[17] J. Wallis, I. Sigalas, and S. Hart, J. Appl. Crystallogr. **19**, 273 (1986).
[18] R. Pawar and V. Deshpande, J. Mater. Sci. **5**, 1061 (1970).
[19] M. Anderson and C. Swenson, Phys. Rev. B **28**, 5395 (1983).
[20] Y. Tsuru, Y. Shinzato, Y. Saito, M. Shimazu, M. Shiono, and M. Morinaga, J. Ceram. Soc. Jpn. **118**, 241 (2010).
[21] E. Kim and C. Chen, Phys. Lett. A **326**, 442 (2004).
[22] T. Scott, Phys. Rep. **27C**, 89 (1976).
[23] Y. Endoh, G. Shirane, and J. Skalyo, Phys. Rev. B **11**, 1681 (1975).
[24] E.-F. Düsing, W. Grosshans, and W. Holzapfel, J. Phys. Colloques (Paris) **45**, 203 (1984).
[25] G. Keeler and D. Batchelder, J. Phys. C **3**, 510 (1970).
[26] H. Fujihisa and K. Takemura, Phys. Rev. B **52**, 13257 (1995).
[27] J. Skalyo, Y. Endoh, and G. Shirane, Phys. Rev. B **9**, 1797 (1974).
[28] D. Price, J. Rowe, and R. Nicklow, Phys. Rev. B **3**, 1268 (1971).
[29] N. Lurie, G. Shirane, and J. Skalyo, Phys. Rev. B **9**, 2661 (1974).
[30] E. Knittle, in *Mineral Physics and Crystallography: A Handbook of Physical Constants* , edited by T. Ahrens (American Geophysical Union Press, 1995) pp. 98-142.
[31] S. Batsanov, in *Effects of Explosions on Materials: Modification and Synthesis under High Pressure Shock Compression* (Springer, New York, 1994).
[32] M. Guinan and D. Steinberg, J. Phys. Chem. Solids **35**, 1501 (1974).
[33] C. Menoni, J. Hu, and I. Spain, Phys. Rev. B **34**, 362 (1986).
[34] H. Beister, K. Strössner, and K. Syassen, Phys. Rev. B **41**, 5535 (1990).
[35] T. Kenichi, Phys. Rev. B **70**, 012101 (2004).
[36] J. Li, S. Liang, H. Guo and B. Liu, J. Alloys Compd. **431**, 23 (2007).
[37] S. Shang, A. Saengdeejing, Z. Mei, D. Kim, H. Zhang, S. Ganeshan, Y. Wang, and Z. Liu, Comput. Mater. Sci. **48**, 813 (2010).
[38] L. Hector, J. Herbst, W. Wolf, P. Saxe, and G. Kresse, Phys. Rev. B **76**, 014121 (2007).
[39] U. Buchenau, M. Heiroth, H. Schober, J. Evers, and G. Oehlinger, Phys. Rev. B **30**, 3502 (1984).

README.txt

```
README FILE
-----------

This file provides a summary of the necessary information to calculate a
code's Delta factor. The full procedure and background are described in
the article by K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S.
Cottenier: "Error estimates for solid-state density-functional theory
predictions: an overview by means of the ground-state elemental
crystals", Crit. Rev. Solid State (2014).

In addition to this README.txt, the following files should have been
included with the Supplementary Material:
- calcDelta.py
- CIFs.tar.gz
- eosfit.py
- WIEN2k.txt
In order to be able to run calcDelta.py and eosfit.py, python and numpy
are needed.

CONTENTS OF THE SUPPLEMENTARY MATERIAL
--------------------------------------

CIFs.tar.gz contains the CIF files of 71 elemental crystals in their
ground state structure (except for sulfur and manganese, where a simpler
structure that appears at higher temperatures or pressures is taken). For
each of these structures, the CIF file presents the crystal at its
calculated equilibrium (PBE functional). Spin-orbit coupling has not been
incorporated in these calculations.

calcDelta.py is a python script, that calculates the Delta factor between
a given code and WIEN2k. Results are saved in Delta-out.txt. The script
is called by "python calcDelta.py filename", where 'filename' refers to a
file containing V0, B0, and B1 information from the code under study for
each elemental crystal. Other options allow to compare 'filename' to
another code, or to print output to the screen. By typing "python
calcDelta.py --help" a more extensive help can be displayed.

WIEN2k.txt is an input file for calcDelta.py. It is not called in the
program arguments by default, however. It consists of similar information
as 'filename', but for WIEN2k. These data are used as a reference. The
WIEN2k.txt file should be in the work directory at all times in order for
calcDelta.py to work.

eosfit.py is a python script, that calculates the Birch-Murnaghan
equation of state for a single set of E(V) data points. It is called by
"python eosfit.py filename2", where 'filename2' refers to a file
containing the crystal volumes (in A^3/atom) and the corresponding
energies (in eV/atom). By typing "python eosfit.py --help" a more
extensive help can be printed on screen.

CALCULATING THE DELTA FACTOR
----------------------------

Unpack the CIFs.tar.gz file and translate every CIF file into a structure
input file of the code under investigation. Make similar input files for
```

```
94%, 96%, 98%, 102%, 104%, and 106% of the equilibrium volume. For each
elemental crystal this yields seven structures.

Calculate the energy for each of the 71x7 input files. Include spin
polarization for O, Cr and Mn (antiferromagnetic) and Fe, Co, and Ni
(ferromagnetic). The antiferromagnetism in the oxygen crystal should take
place between the O2 molecules. Make sure the energy has converged in
function of the computational parameters. Force optimizations should not
be employed, since the computation of Delta requires fixed cell
geometries, both with respect to the cell shape and the internal
coordinates. Put the energies next to the volumes in a separate text file
for each element (expressed per atom) and use eosfit.py to determine the
Birch-Murnaghan equation of state parameters. If the predicted
equilibrium volume is smaller than 96% or larger than 104% of the WIEN2k
equilibrium volume, calculate an additional number of volumes so again 7
volume points are obtained, approximately symmetrical around V0. This
ensures that for the pressure derivative of the bulk modulus an accurate
value is obtained.

Assemble all sets of Birch-Murnaghan parameters in one text file and put
the corresponding element symbol in front of the data. Use calcDelta.py
to determine the Delta factor of the code under consideration, compared
to WIEN2k.

More elaborate comparisons can be performed as well. A Linux-friendly
output scheme has been provided, which allows to execute such analyses.
To scale the elementwise Delta (from filename) to the difference with
experiment, for example, type:
"python calcDelta.py VASP-relaxed.txt exp.txt --stdout | grep -E -v
"\#|np|-" > test.txt"
"python calcDelta.py filename WIEN2k.txt --stdout | grep -E -v "\#|np|-"
| join test.txt - | grep -v "N/A" | awk '{print $1, $3/$2*100}' | sort -n
-k 2"

To compare the performance of the code from filename to that of VASP
(fixed-geometry), type:
"python calcDelta.py VASP.txt WIEN2k.txt --stdout | grep -E -v "\#|np|-"
> test.txt"
"python calcDelta.py filename WIEN2k.txt --stdout | grep -E -v "\#|np|-"
| join test.txt - | grep -v "N/A" | awk '{print $1, $3/$2}' | sort -n -k
2"

FURTHER INFORMATION
-------------------

K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier: "Error
estimates for solid-state density-functional theory predictions: an
overview by means of the ground-state elemental crystals", Crit. Rev.
Solid State (2014). (also available at <http://arxiv.org/abs/1204.2733>)

Corresponding author: <Stefaan.Cottenier@UGent.be>
```

eosfit.py

```
# eosfit.py fits E(V) data to a Birch-Murnaghan equation of state.
# Current version: 2.0
#
# Copyright (C) 2012 Kurt Lejaeghere <Kurt.Lejaeghere@UGent.be>, Center for
# Molecular Modeling (CMM), Ghent University, Ghent, Belgium
#
# eosfit.py is free software; you can redistribute it and/or modify it under
# the terms of the GNU Lesser General Public License as published by the Free
# Software Foundation; either version 2.1 of the License, or (at your option)
# any later version.
#
# eosfit.py is distributed in the hope that it will be useful, but WITHOUT ANY
# WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS
# FOR A PARTICULAR PURPOSE.  See the GNU Lesser General Public License for
# more details.
#
# You should have received a copy of the GNU Lesser General Public License
# along with eosfit.py; if not, see <http://www.gnu.org/licenses/>.

# The following code is based on the source code of eos.py from the Atomic
# Simulation Environment (ASE) <https://wiki.fysik.dtu.dk/ase/>.

# Python and numpy are required to use this script.

import numpy as np
from sys import argv

def BM(energies):

    fitdata = np.polyfit(energies[:,0]**(-2./3.), energies[:,1], 3, full=True)
    ssr = fitdata[1]
    sst = np.sum((energies[:,1] - np.average(energies[:,1]))**2.)
    residuals0 = ssr/sst
    deriv0 = np.poly1d(fitdata[0])
    deriv1 = np.polyder(deriv0, 1)
    deriv2 = np.polyder(deriv1, 1)
    deriv3 = np.polyder(deriv2, 1)

    volume0 = 0
    x = 0
    for x in np.roots(deriv1):
        if x > 0 and deriv2(x) > 0:
            volume0 = x**(-3./2.)
            break

    if volume0 == 0:
        print('Error: No minimum could be found')
        exit()

    derivV2 = 4./9. * x**5. * deriv2(x)
    derivV3 = (-20./9. * x**(13./2.) * deriv2(x) -
        8./27. * x**(15./2.) * deriv3(x))
    bulk_modulus0 = derivV2 / x**(3./2.)
    bulk_deriv0 = -1 - x**(-3./2.) * derivV3 / derivV2

    return volume0, bulk_modulus0, bulk_deriv0, residuals0

usage = '''\
Use: python eosfit.py filename
    calculates the Birch-Murnaghan equation of state from a given file,
    containing in its columns the volumes in A^3/atom and energies in eV/atom,
    respectively
    output is printed in filename.eosout
--help gives an overview of all options
'''
```

```
if len(argv) != 2:
    print 'Error: Wrong number of arguments'
    exit()
else:
    if argv[1] == '--help':
        print usage
        exit()
infile = argv[1]

data = np.loadtxt(infile)
volume, bulk_modulus, bulk_deriv, residuals = BM(data)

echarge = 1.60217733e-19

outstr = '''\
Equation Of State parameters - least squares fit of a Birch Murnaghan curve

%.5f \t %.5f \t %.3f

   V0 \t \t  B0 \t \t  BP
[A^3/at] \t [GPa] \t \t [--]

1-R^2: %f

''' % (volume, (bulk_modulus * echarge * 1.0e21), bulk_deriv, residuals[0])

outfile = open(infile+'.eosout', 'w')
outfile.write(outstr)

outfile.close()
```

calcDelta.py

```
# calcDelta.py determines the Delta factor of a code. Current version: 2.0
# Copyright (C) 2012 Kurt Lejaeghere <Kurt.Lejaeghere@UGent.be>, Center for
# Molecular Modeling (CMM), Ghent University, Ghent, Belgium
#
# calcDelta.py is free software; you can redistribute it and/or modify it
# under the terms of the GNU Lesser General Public License as published by the
# Free Software Foundation; either version 2.1 of the License, or (at your
# option) any later version.
#
# In addition to the regulations of the GNU Lesser General Public License,
# publications and communications based in parts on this program or on
# parts of this program are required to cite the following articles:
#
# K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, "Error
# estimates for solid-state density-functional theory predictions: an overview
# by means of the ground-state elemental crystals", Crit. Rev. Solid State
# (2014).
#
# calcDelta.py is distributed in the hope that it will be useful, but WITHOUT
# ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or
# FITNESS FOR A PARTICULAR PURPOSE.  See the GNU Lesser General Public License
# for more details.
#
# You should have received a copy of the GNU Lesser General Public License
# along with calcDelta.py; if not, see <http://www.gnu.org/licenses/>.

# Python and numpy are required to use this script.

import numpy as np
from sys import argv, stdout, exit

usage = '''\
calcDelta.py -- Support script to calculate the Delta factor of a code

Use: python calcDelta.py infile [reffile] [--stdout]
     where reffile and infile refer to files containing the element name, V0,
     B0, and B1 information (V0 in A^3/atom, B0 in GPa, and B1 dimensionless)
     in columns
     This command calculates the Delta factor of the code in infile compared
     to the one in reffile. When reffile is not explicitly given, WIEN2k.txt
     is used by default.
     Additional output is printed in Delta-out.txt. The option --stdout can be
     used to explicitly print all elements to standard output (on screen)
     instead.
     Attention: the presence of WIEN2k.txt in the same folder is required for
     full functionality!
python calcDelta.py --help displays the current instructions
'''

if len(argv) not in [2, 3, 4]:
    print 'Error: Wrong number of arguments'
    exit()
if argv[1] == '--help':
    print usage
    exit()

reffile = 'WIEN2k.txt'
if len(argv) > 2 and argv[2] != '--stdout':
    reffile = argv[2]

data_f = np.loadtxt(argv[1], dtype={'names':('element', 'V0', 'B0', 'BP'),
    'formats':('S2', np.float, np.float, np.float)})
data_w = np.loadtxt(reffile, dtype={'names':('element', 'V0', 'B0', 'BP'),
    'formats':('S2', np.float, np.float, np.float)})
```

```
try:
    len(data_f['element'])
except TypeError:
    print 'Error: ' + argv[1] + ': at least two elements required'
    exit()
eloverlap = list(set(data_f['element']) & set(data_w['element']))

v0w = np.zeros(len(eloverlap))
b0w = np.zeros(len(eloverlap))
b1w = np.zeros(len(eloverlap))

v0f = np.zeros(len(eloverlap))
b0f = np.zeros(len(eloverlap))
b1f = np.zeros(len(eloverlap))

elw = list(data_w['element'])
elf = list(data_f['element'])

for i in range(len(eloverlap)):
    searchnr = elw.index(eloverlap[i])
    v0w[i] = data_w['V0'][searchnr]
    b0w[i] = data_w['B0'][searchnr] * 10.**9. / 1.602176565e-19 / 10.**30.
    b1w[i] = data_w['BP'][searchnr]

    searchnr = elf.index(eloverlap[i])
    v0f[i] = data_f['V0'][searchnr]
    b0f[i] = data_f['B0'][searchnr] * 10.**9. / 1.602176565e-19 / 10.**30.
    b1f[i] = data_f['BP'][searchnr]

Vi = 0.94 * v0w
Vf = 1.06 * v0w

a3f = 9. * v0f**3. * b0f / 16. * (b1f - 4.)
a2f = 9. * v0f**(7./3.) * b0f / 16. * (14. - 3. * b1f)
a1f = 9. * v0f**(5./3.) * b0f / 16. * (3. * b1f - 16.)
a0f = 9. * v0f * b0f / 16. * (6. - b1f)

a3w = 9. * v0w**3. * b0w / 16. * (b1w - 4.)
a2w = 9. * v0w**(7./3.) * b0w / 16. * (14. - 3. * b1w)
a1w = 9. * v0w**(5./3.) * b0w / 16. * (3. * b1w - 16.)
a0w = 9. * v0w * b0w / 16. * (6. - b1w)

x = [0, 0, 0, 0, 0, 0, 0]

x[0] = (a0f - a0w)**2
x[1] = 6. * (a1f - a1w) * (a0f - a0w)
x[2] = -3. * (2. * (a2f - a2w) * (a0f - a0w) + (a1f - a1w)**2.)
x[3] = -2. * (a3f - a3w) * (a0f - a0w) - 2. * (a2f - a2w) * (a1f - a1w)
x[4] = -3./5. * (2. * (a3f - a3w) * (a1f - a1w) + (a2f - a2w)**2.)
x[5] = -6./7. * (a3f - a3w) * (a2f - a2w)
x[6] = -1./3. * (a3f - a3w)**2.

Fi = np.zeros_like(Vi)
Ff = np.zeros_like(Vf)

for n in range(7):
    Fi = Fi + x[n] * Vi**(-(2.*n-3.)/3.)
    Ff = Ff + x[n] * Vf**(-(2.*n-3.)/3.)

Delta1 = 1000. * np.sqrt((Ff - Fi) / (0.12 * v0w))
Dmax = Delta1.argmax()
Dmin = Delta1.argmin()
Delta = Delta1.mean()
Dstdev = Delta1.std()
total = len(v0w)
```

```
outfile = stdout
if argv[-1] != '--stdout':
    outfile = open('Delta-out.txt', 'w')
elementlist = ['H', 'He', 'Li', 'Be', 'B', 'C', 'N', 'O', 'F', 'Ne', 'Na',
               'Mg', 'Al', 'Si', 'P', 'S', 'Cl', 'Ar', 'K', 'Ca', 'Sc', 'Ti',
               'V', 'Cr', 'Mn', 'Fe', 'Co', 'Ni', 'Cu', 'Zn', 'Ga', 'Ge',
               'As', 'Se', 'Br', 'Kr', 'Rb', 'Sr', 'Y', 'Zr', 'Nb', 'Mo',
               'Tc', 'Ru', 'Rh', 'Pd', 'Ag', 'Cd', 'In', 'Sn', 'Sb', 'Te',
               'I', 'Xe', 'Cs', 'Ba', 'Lu', 'Hf', 'Ta', 'W', 'Re', 'Os', 'Ir',
               'Pt', 'Au', 'Hg', 'Tl', 'Pb', 'Bi', 'Po', 'Rn']

outfile.write('--------------------\n')
outfile.write('# Delta-factor of ' + argv[1] + ' with respect to ' +
    reffile + ' (in meV/atom)\n')
outfile.write('# (%i elements of %i included)\n' % (total, len(elementlist)))
outfile.write('# calculated with calcDelta.py version 2.0 \n')
outfile.write('--------------------\n')

for el in elementlist:
    while True:
        try:            i = eloverlap.index(el)
            outfile.write(eloverlap[i] + '\t %.3f \n' % Delta1[i])
            break
        except ValueError:
            outfile.write(el + '\t N/A \n')
           break

outfile.write('--------------------\n')
outfile.write('np.mean  %.3f\n' % Delta)
outfile.write('np.std   %.3f\n' % Dstdev)
outfile.write('np.max   %.3f  (%.2s)\n' % (Delta1[Dmax], eloverlap[Dmax]))
outfile.write('np.min   %.3f  (%.2s)\n' % (Delta1[Dmin], eloverlap[Dmin]))
outfile.write('--------------------\n')

outfile.close()
```

WIEN2k.txt

```
# WIEN2k version 11.1

H     17.387     10.315     3.025
He    17.778     0.847      6.534
Li    20.216     13.879     3.754
Be    7.915      123.039    3.154
B     7.245      237.599    3.476
C     11.654     209.648    3.565
N     28.896     54.288     3.753
O     18.564     51.231     3.927
F     19.347     35.015     4.246
Ne    24.349     1.242      8.664
Na    37.089     7.728      3.736
Mg    22.934     35.748     4.263
Al    16.503     77.512     4.667
Si    20.553     88.673     4.227
P     21.614     68.863     4.416
S     17.346     85.495     4.440
Cl    39.175     19.324     4.489
Ar    52.209     0.704      7.838
K     73.710     3.586      3.734
Ca    42.208     17.300     3.167
Sc    24.621     54.559     3.402
Ti    17.407     112.712    3.591
V     13.517     185.231    3.731
Cr    11.910     183.841    7.374
Mn    11.611     131.159    0.852
Fe    11.451     196.127    6.137
Co    10.948     216.191    5.076
Ni    10.986     204.685    4.852
Cu    12.017     143.702    5.198
Zn    15.270     75.617     5.377
Ga    20.377     49.087     5.330
Ge    24.094     59.486     5.084
As    22.571     69.732     4.294
Se    29.973     47.637     4.548
Br    39.780     22.629     5.154
Kr    65.226     0.851      22.009
Rb    91.130     2.842      2.321
Sr    54.561     11.333     4.172
Y     32.864     41.320     2.952
Zr    23.402     94.061     3.247
Nb    18.155     168.691    3.504
Mo    15.825     260.410    4.264
Tc    14.472     301.391    4.555
Ru    13.810     315.359    4.955
Rh    14.082     260.576    5.440
Pd    15.326     170.442    5.854
Ag    17.855     91.328     5.799
Cd    22.866     44.247     7.089
In    27.497     34.759     4.870
Sn    36.883     36.020     5.038
Sb    31.786     50.718     4.495
Te    35.009     44.806     4.703
I     50.339     18.699     5.222
```

Xe	87.318	0.569	-0.724
Cs	117.748	1.964	3.592
Ba	63.200	8.431	3.299
Lu	29.060	47.725	4.048
Hf	22.551	108.082	3.225
Ta	18.296	193.730	4.774
W	16.175	302.590	4.300
Re	14.987	364.610	4.466
Os	14.320	402.199	4.352
Ir	14.533	341.743	6.958
Pt	15.678	251.750	5.342
Au	17.993	139.863	5.985
Hg	29.716	8.155	8.077
Tl	31.456	26.683	4.565
Pb	32.000	39.551	5.954
Bi	36.939	42.556	4.626
Po	37.570	45.523	5.018
Rn	92.765	0.544	13.133

Supplementary Material of Paper 2

Ab-initio-based thermal property predictions at a low cost: an error analysis

K. Lejaeghere, J. Jaeken, V. Van Speybroeck and S. Cottenier

Supplementary Material

All citations in this document refer to the list that is given in the end, and not to the bibliography of the main paper.

1. Computational parameters

cut-off energy
400 eV for most elements
600 eV for He, B, C, N, O, F, and Ne

self-consistent cycle
10^{-4} eV energy convergence criterion

reciprocal space representation
6750 / N k-points for N-atom cells

FFT grid
default (2 G_{cut} x 2 G_{cut} x 2 G_{cut} ; G_{cut} being the cut-off electronic wave vector)

augmentation charge grid
default (4 G_{cut} x 4 G_{cut} x 4 G_{cut})

potential files
see Tab. 1.1

Tab. 1.1 VASP potentials used for the elemental crystals

	potential file
H	PAW_PBE H 15Jun2001
He	PAW_PBE He 05Jan2001
Li	PAW_PBE Li_sv 23Jan2001
Be	PAW_PBE Be 06Sep2000
B	PAW_PBE B 06Sep2000
C	PAW_PBE C 08Apr2002
N	PAW_PBE N 08Apr2002
O	PAW_PBE O 08Apr2002
F	PAW_PBE F 08Apr2002
Ne	PAW_PBE Ne 05Jan2001
Na	PAW_PBE Na_pv 05Jan2001
Mg	PAW_PBE Mg 05Jan2001
Al	PAW_PBE Al 04Jan2001
Si	PAW_PBE Si 05Jan2001
P	PAW_PBE P 17Jan2003
S	PAW_PBE S 17Jan2003
Cl	PAW_PBE Cl 17Jan2003
Ar	PAW_PBE Ar 07Sep2000
K	PAW_PBE K_sv 06Sep2000
Ca	PAW_PBE Ca_pv 06Sep2000
Sc	PAW_PBE Sc_sv 07Sep2000
Ti	PAW_PBE Ti_pv 07Sep2000
V	PAW_PBE V_pv 07Sep2000
Cr	PAW_PBE Cr_pv 07Sep2000
Mn	PAW_PBE Mn_pv 07Sep2000
Fe	PAW_PBE Fe 06Sep2000
Co	PAW_PBE Co 06Sep2000
Ni	PAW_PBE Ni 06Sep2000
Cu	PAW_PBE Cu 05Jan2001
Zn	PAW_PBE Zn 06Sep2000
Ga	PAW_PBE Ga_d 06Sep2000
Ge	PAW_PBE Ge_d 06Sep2000
As	PAW_PBE As 06Sep2000
Se	PAW_PBE Se 06Sep2000
Br	PAW_PBE Br 06Sep2000
Kr	PAW_PBE Kr 07Sep2000
Rb	PAW_PBE Rb_sv 06Sep2000
Sr	PAW_PBE Sr_sv 07Sep2000
Y	PAW_PBE Y_sv 06Sep2000
Zr	PAW_PBE Zr_sv 07Sep2000
Nb	PAW_PBE Nb_pv 08Apr2002
Mo	PAW_PBE Mo_pv 08Apr2002
Tc	PAW_PBE Tc_pv 06Sep2000
Ru	PAW_PBE Ru 06Sep2000
Rh	PAW_PBE Rh 06Sep2000
Pd	PAW_PBE Pd 05Jan2001
Ag	PAW_PBE Ag 06Sep2000

Cd	PAW_PBE Cd 06Sep2000
In	PAW_PBE In_d 06Sep2000
Sn	PAW_PBE Sn_d 06Sep2000
Sb	PAW_PBE Sb 06Sep2000
Te	PAW_PBE Te 08Apr2002
I	PAW_PBE I 08Apr2002
Xe	PAW_PBE Xe 07Sep2000
Cs	PAW_PBE Cs_sv 08Apr2002
Ba	PAW_PBE Ba_sv 06Sep2000
Lu	PAW_PBE Lu_3 06Sep2000
Hf	PAW_PBE Hf_pv 06Sep2000
Ta	PAW_PBE Ta_pv 07Sep2000
W	PAW_PBE W_pv 06Sep2000
Re	PAW_PBE Re 17Jan2003
Os	PAW_PBE Os_pv 20Jan2003
Ir	PAW_PBE Ir 06Sep2000
Pt	PAW_PBE Pt 05Jan2001
Au	PAW_PBE Au 06Sep2000
Hg	PAW_PBE Hg 06Sep2000
Tl	PAW_PBE Tl_d 06Sep2000
Pb	PAW_PBE Pb_d 06Sep2000
Bi	PAW_PBE Bi_d 06Sep2000
Po	PAW_PBE Po_d 25May2007
Rn	PAW_PBE Rn 28Aug2006

2. Computational and experimental data

Tab. 2.1 Moleculization energy ΔE_{mol} (in kJ/mol), equilibrium volume V_0 (in Å^3/atom), bulk modulus B_0 (in GPa) and pressure derivative of the bulk modulus B_1 (dimensionless) by means of VASP-PBE for the ground-state elemental crystals.

	$\Delta E_{mol,VASP}$	$V_{0,VASP}$	$B_{0,VASP}$	$B_{1,VASP}$
H	0.76	17.41	0.416	3.9
He	1.10	17.73	0.905	7.5
Li	155	20.28	13.83	3.2
Be	360	7.922	122.8	3.2
B	600	7.252	210.5	4.0
C	758	11.66	1.554	9.9
N	1.31	29.65	0.770	6.2
O	2.12	19.18	1.422	5.6
F	2.20	19.52	1.562	7.1
Ne	1.80	24.69	1.067	8.1
Na	104	37.07	7.674	3.3
Mg	145	22.85	36.48	3.9
Al	331	16.49	77.30	4.6
Si	439	20.45	88.80	4.3
P	336	21.38	7.265	14
S	294	38.48	0.355	20
Cl	4.48	38.18	1.334	11
Ar	9.16	52.67	0.788	6.8
K	83.9	73.85	3.598	3.7
Ca	185	42.17	17.54	3.2
Sc	400	24.66	54.30	3.4
Ti	528	17.37	112.5	3.6
V	571	13.49	181.4	3.8
Cr	390	11.83	177.0	7.4
Mn	372	11.10	212.3	6.4
Fe	468	11.37	185.9	4.9
Co	496	10.88	208.4	4.6
Ni	466	10.94	193.7	5.0
Cu	335	12.03	136.7	5.0
Zn	106	15.35	57.92	5.8
Ga	254	20.36	48.27	4.7
Ge	361	23.91	58.77	4.9
As	286	22.73	25.49	13
Se	251	29.84	4.413	9.5
Br	12.3	39.29	1.558	17
Kr	2.34	66.44	0.664	8.5
Rb	74.7	91.24	2.790	3.7
Sr	156	54.53	11.11	5.0
Y	402	32.92	41.38	3.2
Zr	604	23.53	93.50	3.6
Nb	669	18.34	171.7	4.0
Mo	603	15.92	262.9	4.3

Tc	658	14.60	298.4	4.5
Ru	644	13.84	310.9	5.0
Rh	567	14.18	254.3	5.2
Pd	357	15.44	168.4	5.5
Ag	240	18.05	89.26	5.9
Cd	71.5	23.05	36.73	6.2
In	223	27.54	35.68	5.7
Sn	306	36.86	35.82	4.8
Sb	261	31.79	30.75	7.7
Te	233	34.98	17.78	10
I	17.5	49.78	1.928	14
Xe	2.69	87.78	0.531	7.4
Cs	69.0	117.0	1.979	3.3
Ba	182	63.55	8.862	3.1
Lu	395	29.32	47.32	3.5
Hf	606	22.42	107.2	3.5
Ta	766	18.30	193.0	4.0
W	852	16.28	298.0	4.3
Re	760	14.96	366.4	4.6
Os	770	14.44	386.7	5.0
Ir	676	14.61	337.0	5.2
Pt	514	15.78	241.7	5.5
Au	300	18.05	141.3	6.1
Hg	20	28.27	12.91	8.7
Tl	146	31.17	26.70	5.5
Pb	189	32.27	36.18	5.1
Bi	200	38.27	22.50	7.6
Po	157	39.62	33.90	5.4
Rn	3.78	89.60	0.640	8.6

Tab. 2.2 VASP-PBE-based estimates for the volumetric thermal expansion coefficients $\alpha_{V,rt}$ of the ground-state elemental crystals at room temperature (in 10^{-5}/K), using the approximations by Slater (S) [1], Dugdale-MacDonald (DM) [2], Vashchenko-Zubarev (VZ) [3], Guinea *et al.* (G) [4] and Tsuru *et al.* (Ts) [5], as well as the experimental values (exp).

	$\alpha_{V,rt}$(S)	$\alpha_{V,rt}$(DM)	$\alpha_{V,rt}$(VZ)	$\alpha_{V,rt}$(G)	$\alpha_{V,rt}$(Ts)	$\alpha_{V,rt}$(exp)
H	1032	842	651	1574	1829	
He	928	841	755	879	1265	
Li	20.9	16.0	11.0	17.7	8.99	13.8 [6]
Be	6.18	4.76	3.34	6.25	3.88	3.39 [6]
B	4.96	4.06	3.15	3.86	2.32	2.49 [6]
C	1092	1016	940	31.5	1.84	2.4 [7]
N	529	469	408	675	1062	
O	402	351	300	486	656	
F	462	417	372	451	632	
Ne	613	561	508	536	772	
Na	21.4	16.5	11.7	21.4	13.3	21.3 [6]
Mg	8.92	7.27	5.61	10.6	9.60	7.44 [6]
Al	6.97	5.89	4.80	5.69	4.21	6.93 [6]
Si	4.52	3.75	2.99	4.14	3.17	0.78 [6]
P	176	167	159	16.2	4.14	9.9 [8]
S	2998	2897	2796	58.4	4.74	25.3 [9]
Cl	434	407	380	245	311	
Ar	323	290	257	190	152	
K	26.3	21.1	15.9	24.8	16.6	25.0 [6]
Ca	8.09	6.22	4.36	9.99	7.53	6.69 [6]
Sc	4.76	3.73	2.70	5.05	3.49	3.06 [6]
Ti	3.46	2.75	2.05	3.64	2.64	2.58 [6]
V	2.94	2.38	1.81	3.13	2.44	2.52 [6]
Cr	6.99	6.33	5.67	4.09	3.57	1.47 [6]
Mn	5.33	4.74	4.16	3.95	3.75	6.51 [6]
Fe	4.50	3.85	3.19	3.72	2.98	3.54 [6]
Co	3.87	3.27	2.66	3.48	2.81	3.90 [6]
Ni	4.53	3.88	3.23	3.72	2.99	4.02 [6]
Cu	5.94	5.10	4.26	4.98	4.15	4.95 [6]
Zn	12.7	11.1	9.55	12.0	13.1	9.06 [6]
Ga	9.11	7.71	6.30	7.40	5.48	5.4 [6]
Ge	6.66	5.68	4.70	5.19	3.86	1.74 [6]
As	43.8	41.4	39.0	9.07	4.87	5.05 [10]
Se	145	134	124	20.3	5.55	16.61 [6]
Br	550	527	504	135	114	
Kr	383	351	320	364	595	
Rb	27.6	22.2	16.8	26.8	18.6	23.6 [11]
Sr	16.0	13.7	11.5	12.0	8.95	6.75 [6]
Y	4.37	3.36	2.34	4.99	3.46	3.18 [6]
Zr	3.03	2.40	1.78	3.21	2.31	1.71 [6]
Nb	2.41	1.97	1.53	2.55	2.08	2.19 [6]
Mo	1.98	1.65	1.32	2.33	2.31	1.44 [6]
Tc	2.00	1.69	1.37	2.18	2.12	2.1 [12]

Ru	2.22	1.90	1.58	2.22	2.16	1.92 [6]
Rh	2.82	2.44	2.05	2.58	2.46	2.46 [6]
Pd	4.14	3.61	3.08	3.83	3.90	3.54 [6]
Ag	7.14	6.28	5.42	5.94	5.80	5.67 [6]
Cd	14.4	12.7	11.1	15.0	19.5	9.24 [6]
In	11.3	9.87	8.47	7.89	6.24	9.63 [6]
Sn	7.05	6.00	4.96	5.81	4.55	1.62 [6]
Sb	15.6	14.2	12.8	7.32	5.34	3.30 [6]
Te	33.1	30.9	28.7	9.72	5.99	5.04 [6]
I	286	271	257	90.3	79.7	
Xe	315	286	256	330	518	
Cs	26.8	20.8	14.8	29.3	20.2	29.1 [6]
Ba	10.3	7.81	5.36	11.5	7.65	6.18 [6]
Lu	4.80	3.80	2.81	4.99	3.52	2.97 [6]
Hf	2.69	2.12	1.54	3.06	2.30	1.77 [6]
Ta	2.14	1.75	1.36	2.25	1.82	1.89 [6]
W	1.68	1.39	1.11	1.82	1.64	1.35 [6]
Re	1.60	1.35	1.10	1.81	1.83	1.86 [6]
Os	1.73	1.48	1.24	1.78	1.81	1.53 [6]
Ir	2.06	1.78	1.50	2.03	2.06	1.92 [6]
Pt	2.78	2.42	2.06	2.64	2.71	2.64 [6]
Au	4.65	4.11	3.57	4.22	4.64	4.26 [6]
Hg	47.6	43.8	40.0	42.8	68.1	
Tl	13.0	11.3	9.65	10.6	9.54	8.97 [6]
Pb	8.47	7.29	6.11	7.88	7.39	8.67 [6]
Bi	17.6	16.0	14.4	8.90	6.95	4.02 [6]
Po	7.74	6.71	5.68	8.06	8.90	7.05 [6]
Rn	297	273	249	251	368	

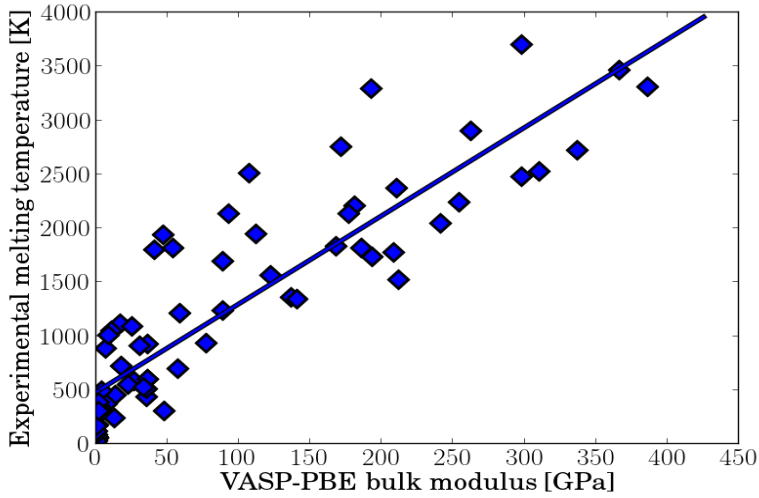


Fig. 2.3 Least-squares linear regression between the experimental melting temperature T_m and the VASP-PBE bulk modulus B_0 for the ground-state elemental crystals: $T_m = 482.8 + 8.172\ B_0$.

Tab. 2.4 VASP-PBE-based estimates for the melting temperatures T_m of the ground-state elemental crystals (in K), based on the bulk modulus B_0 (Eq. (17) of the article) or the moleculization energy ΔE_{mol} [4], as well as the experimental values.

	$T_m(B_0)$	$T_m(\Delta E_{mol})$	T_m(exp)
H	486.2	2.932	14.2 [13]
He	490.2	4.238	
Li	595.8	596.8	453.7 [14]
Be	1486	1384	1562 [14]
B	2203	2309	2365 [14]
C	495.5	2916	
N	489.1	5.050	63.15 [14]
O	494.4	8.170	54.36 [14]
F	495.6	8.486	53.48 [14]
Ne	491.5	6.944	24.56 [14]
Na	545.5	401.9	371.0 [14]
Mg	780.9	558.6	922 [14]
Al	1115	1274	933.5 [14]
Si	1208	1691	1687 [14]
P	542.2	1295	883 [15]
S	485.7	1130	388.4 [14]
Cl	493.7	17.25	172.2 [14]
Ar	489.3	35.26	83.81 [14]
K	512.2	323.0	336.3 [14]
Ca	626.1	712.6	1113 [14]
Sc	926.5	1538	1814 [14]
Ti	1402	2031	1946 [14]
V	1965	2196	2202 [14]
Cr	1929	1502	2133 [14]
Mn	2218	1430	1520 [14]
Fe	2002	1802	1811 [14]
Co	2186	1910	1770 [14]
Ni	2066	1793	1728 [14]
Cu	1600	1291	1358 [14]
Zn	956.1	409.2	692.7 [14]
Ga	877.3	978.0	302.9 [14]
Ge	963.1	1388	1211 [14]
As	691.1	1102	1089 [14]
Se	518.9	966.9	494 [14]
Br	495.5	47.15	265.9 [14]
Kr	488.2	9.007	115.8 [14]
Rb	505.6	287.7	312.6 [14]
Sr	573.6	599.3	1042 [14]
Y	820.9	1549	1801 [14]
Zr	1247	2325	2128 [14]
Nb	1886	2575	2750 [14]
Mo	2631	2319	2895 [14]
Tc	2921	2534	2477 [14]
Ru	3024	2478	2527 [14]

Rh	2561	2184	2236 [14]
Pd	1859	1375	1827 [14]
Ag	1212	923.9	1235 [14]
Cd	782.9	275.3	594.3 [14]
In	774.4	859.0	429.8 [14]
Sn	775.5	1178	505.1 [14]
Sb	734.1	1005	903.9 [14]
Te	628.1	895.4	722.7 [14]
I	498.6	67.27	386.7 [14]
Xe	487.2	10.36	161.4 [14]
Cs	499.0	265.5	301.6 [14]
Ba	555.2	701.3	1002 [14]
Lu	869.5	1522	1938 [14]
Hf	1359	2333	2504 [14]
Ta	2060	2947	3293 [14]
W	2918	3280	3695 [14]
Re	3477	2927	3459 [14]
Os	3643	2964	3306 [14]
Ir	3236	2602	2720 [14]
Pt	2458	1977	2045 [14]
Au	1638	1155	1338 [14]
Hg	588.3	78.75	234.3 [14]
Tl	701.0	562.2	577 [14]
Pb	778.5	725.6	600.7 [14]
Bi	666.7	771.4	544.6 [14]
Po	759.9	602.4	527 [14]
Rn	488.0	14.56	

Tab. 2.5 VASP- and GPAW-PBE-based estimates for the room-temperature volumetric thermal expansion coefficient $\alpha_{V,rt}$ in the Slater approximation (S, in 10^{-5}/K) and the melting temperature based on the bulk modulus B_0 (Eq. (17) of the article, in K). These numbers are based on E(V) properties of the ground-state elemental crystals, where the cell shape and the atomic positions are kept fixed. More computational details about the VASP and GPAW calculations can be found in [16]. Error estimates from a linear regression $Y = \beta_1 X + \beta_0$, with Y the experimental numbers (Tabs. 2.2 and 2.4) and X the raw theoretical values, are given at the bottom.

	VASP - fixed geometry		**GPAW - fixed geometry**	
	$\alpha_{V,rt}$(S)	**$T_m(B_0)$**	**$\alpha_{V,rt}$(S)**	**$T_m(B_0)$**
H		565.4		567.2
He				
Li	20.9	595.6	22.4	597.4
Be	6.28	1490	6.37	1491
B	3.78	2420	3.85	2421
C	2.75		2.80	
N		943.8		924.1
O		903.1		920.9
F		762.2		758.4
Ne		491.6		499.6
Na	19.7	545.6	25.6	546.7
Mg	8.91	780.8	9.57	779.1
Al	7.02	1114	6.86	1122
Si	4.52	1208	4.58	1209
P	5.67	1043	5.80	1039
S				
Cl		639.3		638.2
Ar		489.3		489.7
K	27.2	512.2	16.8	512.3
Ca	7.52	626.0	8.02	625.0
Sc	4.75	927.1	4.28	925.6
Ti	3.46	1402	3.55	1418
V	3.12	1967	3.29	1998
Cr	6.77	1927	6.83	1805
Mn				
Fe	4.49	2001	4.31	2090
Co	4.21	2205	4.16	2219
Ni	4.48	2066	3.98	2147
Cu	6.00	1600	5.23	1622
Zn	9.66	1089	8.07	1096
Ga	9.88	882.6	9.85	893.7
Ge	6.54	963.5	6.48	973.6
As	5.26	1044	5.45	1046
Se	6.10	868.8	6.31	868.3
Br		666.4		652.8
Kr		488.4		489.4
Rb	27.6	505.6	23.0	504.8
Sr	14.5	573.8	11.3	574.2
Y				

Zr	3.05	1249	2.84	1258
Nb	2.39	1886	2.25	1906
Mo	1.98	2631	2.00	2638
Tc				
Ru	2.22	3028	2.15	3024
Rh	2.82	2560	2.89	2568
Pd	4.23	1857	4.53	1850
Ag	7.01	1211	6.56	1224
Cd	14.2	840.6	12.4	836.4
In	10.6	772.4	11.3	770.0
Sn	7.08	775.7	7.12	776.9
Sb				
Te	5.78	850.3	5.89	859.6
I		635.8		626.2
Xe				
Cs	28.3	498.9	27.2	500.1
Ba	10.2	555.2	9.54	555.8
Lu				
Hf				
Ta	2.12	2070	2.06	2104
W	1.66	2970	1.65	2996
Re				
Os	1.66	3759	1.66	3735
Ir	1.99	3310	2.02	3356
Pt	2.76	2510	2.82	2506
Au	4.66	1598	4.49	1609
Hg				
Tl				
Pb	6.60	807.9	8.63	802.0
Bi	5.55	830.4	5.79	833.0
Po				
Rn				

β_1	0.894	1.049	0.894	1.043
β_0	-0.67	-154.4	-0.64	-151.6
SER	2.22	457.8	2.14	457.8

Tab. 2.6 Zero-point- and thermally corrected [16] experimental moleculization energy ΔE_{mol} (in kJ/mol), equilibrium volume V_0 (in $Å^3$/atom) and bulk modulus B_0 (in GPa), and non-corrected pressure derivative of the bulk modulus B_1 (dimensionless) for the ground-state elemental crystals.

	$\Delta E_{mol,exp}$	$V_{0,exp}$	$B_{0,exp}$	$B_{1,exp}$
H		6.58 [17]	0.5 [14]	
He		9.40 [17]	0.42 [14]	
Li	161 [14]	20.41 [17]	13.1 [14]	3.51 [35]
Be	333 [14]	7.82 [17]	108.1 [14]	
B	569 [14]	7.16 [17]	185.9 [14]	
C	719 [15]	8.06 [17]	55.6 [26]	8.9 [35]
N		21.09 [18]	2.69 [27]	
O		15.98 [17]		
F		14.47 [19]		
Ne	2.62 [14]	17.85 [20]	2.8 [28]	9.23 [35]
Na	108 [14]	37.18 [17]	7.9 [14]	4.13 [35]
Mg	149 [14]	22.65 [17]	38.6 [14]	4.8 [35]
Al	331 [14]	16.27 [17]	77.1 [14]	4.45 [35]
Si	452 [14]	19.82 [17]	101.3 [14]	4.43 [35]
P	333 [14]	18.50 [17]	33.2 [14]	4.5 [35]
S	277 [14]	24.09 [17]	23.8 [14]	6.75 [35]
Cl		26.95 [21]	40.8 [29]	5.2 [35]
Ar	8.60 [14]	35.55 [22]	3.30 [30]	7.2 [35]
K	91.0 [14]	72.20 [17]	3.8 [14]	4.089 [35]
Ca	180 [14]	42.95 [17]	15.9 [14]	3.1 [35]
Sc	379 [14]	24.77 [17]	44.5 [14]	2.8 [36]
Ti	472 [14]	17.50 [17]	107.4 [14]	3.4 [35]
V	516 [14]	13.81 [17]	165.8 [14]	4.135 [35]
Cr	401 [14]	11.82 [17]	204.6 [14]	6.895 [35]
Mn	286 [14]	11.97 [17]	174.7 [31]	6.6 [35]
Fe	417 [14]	11.64 [17]	175.1 [14]	4.6 [35]
Co	428 [14]	10.96 [17]	198.4 [14]	4.26 [37]
Ni	432 [14]	10.81 [17]	192.5 [14]	4 [35]
Cu	339 [14]	11.65 [17]	144.3 [14]	4.88 [35]
Zn	133 [14]	14.86 [17]	64.7 [14]	4.4 [35]
Ga	274 [14]	19.15 [17]	60.2 [14]	
Ge	375 [14]	22.44 [17]	79.4 [14]	4.76 [38]
As	287.9 [14]	21.23 [17]	40.8 [14]	3.3 [39]
Se	238 [14]	26.22 [17]	10.9 [14]	5.8 [35]
Br		31.47 [21]	25.6 [29]	5.2 [35]
Kr	11.9 [14]	44.04 [23]	4.01 [32]	7.2 [35]
Rb	82.7 [14]	89.16 [17]	3.6 [14]	3.885 [35]
Sr	167 [14]	55.60 [17]	12.0 [14]	2.485 [35]
Y	425 [14]	32.95 [17]	37.3 [14]	2.2 [35]
Zr	606 [14]	23.18 [17]	84.3 [14]	2.575 [35]
Nb	733 [14]	17.97 [17]	173.2 [14]	4.015 [35]
Mo	662 [14]	15.51 [17]	276.2 [14]	3.98 [35]
Tc	665 [14]	14.22 [17]	303.1 [14]	

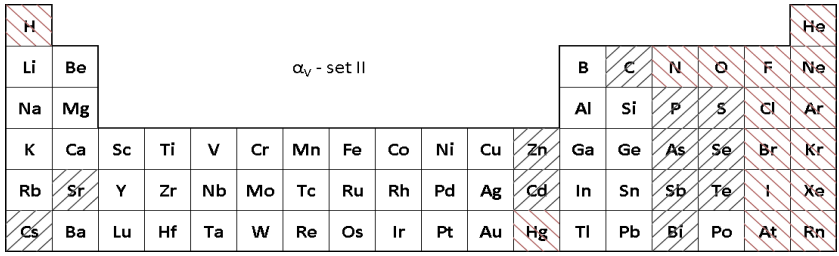
Ru	656 [14]	13.45 [17]	335.5 [14]	6.61 [37]
Rh	558 [14]	13.57 [17]	277.1 [14]	4.5 [37]
Pd	379 [14]	14.56 [17]	187.2 [14]	5 [35]
Ag	286 [14]	16.85 [17]	105.7 [14]	4.725 [35]
Cd	114 [14]	21.01 [17]	50.7 [14]	4.9 [35]
In	244 [14]	25.70 [17]	44.7 [14]	5.35 [35]
Sn	305 [14]	33.97 [17]	42.8 [33]	4 [35]
Sb	267 [14]	29.68 [17]	39.5 [14]	4.3 [35]
Te	212 [14]	33.30 [17]	26.2 [14]	8.4 [35]
I		37.04 [17]	23.9 [29]	6 [35]
Xe	16.5 [14]	56.87 [24]	3.83 [34]	6.34 [35]
Cs	78.0 [14]	110.32 [17]	2.3 [14]	3.79 [35]
Ba	184 [14]	62.29 [17]	10.6 [14]	2.43 [35]
Lu	430 [14]	29.27 [17]	42.0 [14]	
Hf	623 [14]	22.30 [17]	110.7 [14]	3.95 [37]
Ta	784 [14]	17.93 [17]	202.7 [14]	3.75 [35]
W	863 [14]	15.80 [17]	327.5 [14]	4.32 [35]
Re	779 [14]	14.62 [17]	380.8 [14]	5.41 [37]
Os	793 [14]	13.85 [17]	424.6 [14]	4.5 [40]
Ir	674 [14]	14.06 [17]	362.2 [14]	4.83 [37]
Pt	566 [14]	15.02 [17]	285.5 [14]	5.18 [37]
Au	370 [14]	16.82 [17]	182.0 [14]	6.4 [35]
Hg	66 [14]	22.25 [25]		
Tl	183 [14]	27.99 [17]	37.4 [14]	5.8 [41]
Pb	197 [14]	29.86 [17]	46.3 [14]	5.335 [35]
Bi	211 [14]	35.13 [17]	32.1 [14]	2.4 [35]
Po	145 [14]	36.93 [17]	27.4 [14]	
Rn	19.6 [14]			

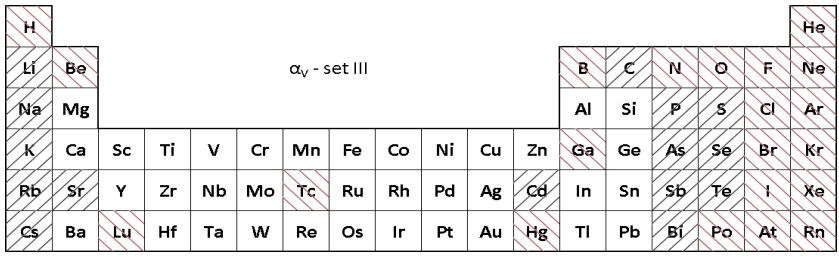
Tab. 2.7 Experiment-based estimates for the volumetric thermal expansion coefficients $\alpha_{V,rt}$ and melting temperatures T_m of the ground-state elemental crystals at room temperature (in 10^{-5}/K), using the approximations by Slater (S) [1], Dugdale-MacDonald (DM) [2], Vashchenko-Zubarev (VZ) [3], Guinea *et al.* (G) [4], Tsuru *et al.* (Ts) [5] and based on the bulk modulus B_0 (Eq. (17) of the article).

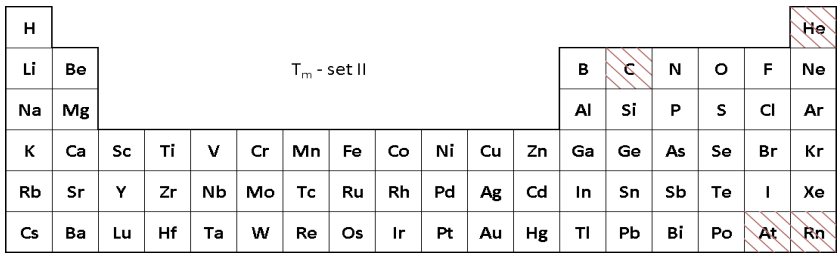
	$\alpha_{V,rt}$(S)	$\alpha_{V,rt}$(DM)	$\alpha_{V,rt}$(VZ)	$\alpha_{V,rt}$(G)	$\alpha_{V,rt}$(Ts)	T_m(B0)	T_m(G)
H						486.7	
He						486.3	
Li	24.7	19.5	14.3	17.8	8.64	589.6	620.5
Be				6.96	4.18	1367	1283
B				4.25	2.45	2002	2188
C	39.6	36.5	33.4	6.51	1.94	937.1	2766
N						504.8	
O							
F							
Ne	369	341	314	323	532	505.7	10.09
Na	26.7	22.0	17.3	20.7	12.8	547.5	417.5
Mg	10.6	9.00	7.42	10.3	9.37	798.3	572.5
Al	6.79	5.69	4.59	5.74	4.21	1113	1274
Si	4.23	3.54	2.85	3.88	3.08	1310	1740
P	14.0	11.8	9.6	8.17	4.18	754.2	1282
S	25.3	22.7	20.1	9.70	5.04	656.8	1065
Cl	9.17	7.91	6.66			816.2	
Ar	121	110	97.7	116	162	509.8	33.10
K	28.7	23.6	18.5	23.6	15.3	513.5	350.0
Ca	8.42	6.39	4.36	10.6	7.73	612.4	693.4
Sc	4.63	3.38	2.13	5.71	3.67	846.8	1460
Ti	3.38	2.65	1.91	3.93	2.95	1360	1816
V	3.44	2.83	2.23	3.40	2.70	1838	1984
Cr	5.62	5.05	4.48	3.75	3.48	2154	1543
Mn	6.21	5.55	4.89	4.78	4.87	1910	1100
Fe	4.34	3.66	2.98	4.01	3.34	1914	1606
Co	3.74	3.11	2.47	3.83	3.25	2104	1648
Ni	3.65	2.99	2.32	3.90	3.22	2056	1663
Cu	5.60	4.78	3.96	4.90	4.11	1662	1306
Zn	8.76	7.33	5.89	10.3	10.5	1011	512.1
Ga				6.58	5.09	974.8	1055
Ge	5.15	4.37	3.60	4.52	3.71	1132	1445
As	7.08	5.49	3.90	7.40	4.84	816.6	1108
Se	39.8	34.9	30.1	14.2	5.86	571.5	915.4
Br	12.5	10.8	9.1			692.1	
Kr	80.5	72.7	64.9	80.7	117	515.6	45.70
Rb	23.2	18.9	14.5	22.8	16.8	511.8	318.4
Sr	6.69	4.62	2.54	11.1	8.33	580.7	644.2
Y	3.14	2.02	0.90	5.12	3.28	787.7	1634
Zr	2.38	1.67	0.96	3.40	2.30	1172	2331
Nb	2.45	2.01	1.56	2.45	1.90	1898	2819
Mo	1.76	1.44	1.12	2.19	2.10	2740	2549

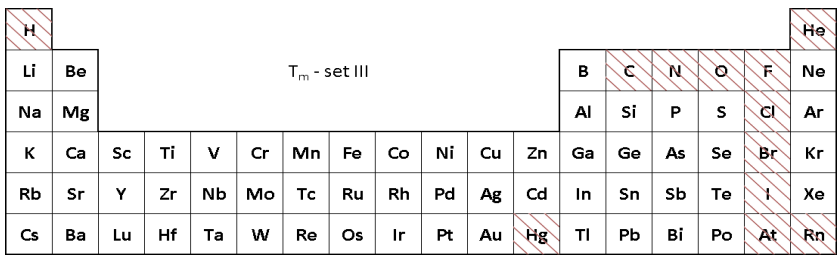
Tc				2.18	2.10	2959	2558
Ru	2.88	2.57	2.27	2.15	2.13	3225	2523
Rh	2.29	1.93	1.56	2.55	2.50	2747	2149
Pd	3.54	3.04	2.53	3.64	3.68	2013	1457
Ag	5.11	4.33	3.56	5.18	4.87	1347	1101
Cd	8.89	7.59	6.29	10.6	12.2	896.8	438.6
In	9.04	7.84	6.64	6.98	5.71	848.1	939.1
Sn	5.22	4.27	3.32	5.55	4.57	832.7	1173
Sb	7.00	5.83	4.65	6.61	5.22	805.7	1028
Te	19.1	17.6	16.0	8.58	6.56	697.0	817.6
I	13.2	11.7	10.1			678.5	
Xe	57.1	50.8	44.4	61.6	84.5	514.1	63.50
Cs	27.7	22.3	17.0	26.0	17.9	502.0	300.0
Ba	6.58	4.49	2.40	10.6	7.57	569.4	708.3
Lu				5.08	3.24	826.1	1655
Hf	3.04	2.48	1.92	2.98	2.24	1387	2399
Ta	1.95	1.57	1.19	2.19	1.78	2139	3018
W	1.60	1.33	1.06	1.75	1.62	3159	3320
Re	1.89	1.64	1.39	1.78	1.79	3594	2998
Os	1.47	1.23	1.00	1.71	1.76	3953	3051
Ir	1.83	1.56	1.29	1.99	2.07	3443	2594
Pt	2.34	2.02	1.70	2.37	2.46	2816	2179
Au	4.10	3.65	3.20	3.48	3.77	1970	1422
Hg					21.2		252.8
Tl	10.8	9.48	8.17	8.45	7.63	788.8	703.3
Pb	7.48	6.49	5.49	7.08	7.07	861.5	758.1
Bi	3.80	2.57	1.35	7.58	6.60	744.9	812.5
Po				9.65	9.62	707.1	557.5
Rn					70.9		75.60

Tab. 2.8 Common inliers between all PBE-based predictions (set II) and between all PBE- and experiment-based predictions (set III) for both α_V and T_m. These are the notations used in Tab. IV of the main article. The legend is given below.









X1 outlier in one of the data sets

X2 not enough experimental data available

Tab. 2.9 Elastic constants C_{ij} (in GPa) by means of VASP-PBE for some ground-state elemental crystals. These values were obtained from the stress tensors of 2% deformed unit cells.

	C_{11}	C_{12}	C_{44}	C_{13}	C_{33}
Be	301.2	15.6	161.6	13.2	368.6
Ne	1.1	1.0	0.4	1.0	1.1
Mg	64.8	18.7	18.9	22.6	67.1
Al	118.3	56.9	23.7	56.9	118.3
Si	154.9	55.8	98.4	55.8	154.9
Ar	1.1	0.7	0.4	0.7	1.1
K	3.9	3.4	2.9	3.4	3.9
Ca	23.3	14.6	14.3	14.6	23.3
Sc	109.2	34.2	30.0	38.4	114.3
Ti	185.8	58.0	42.8	72.8	180.1
V	269.3	137.5	20.7	137.5	269.3
Cr	415.3	58.2	95.2	58.2	415.3
Fe	256.0	151.1	87.7	151.1	256.0
Co	350.6	139.7	97.4	126.2	401.9
Ni	274.0	153.7	128.9	153.7	274.0
Cu	158.9	125.7	75.4	125.7	158.9
Zn	157.3	36.4	32.9	51.1	64.1
Ge	105.1	35.8	71.9	35.8	105.1
Kr	0.9	0.6	0.3	0.6	0.9
Rb	3.0	2.7	2.1	2.7	3.0
Sr	15.0	9.2	13.0	9.2	15.0
Y	77.8	19.1	25.4	20.7	77.5
Zr	148.6	52.5	25.8	69.0	165.9
Nb	239.6	137.7	14.7	137.7	239.6
Mo	469.2	159.8	105.5	159.8	469.2
Tc	461.3	169.3	135.3	158.7	498.2
Ru	515.8	139.0	184.4	161.6	590.7
Rh	404.0	179.7	176.9	179.7	404.0
Pd	199.0	152.8	57.1	152.8	199.0
Ag	96.6	85.7	40.6	85.7	96.6
Cd	101.1	41.1	12.6	56.3	79.7
Sn	56.3	25.6	43.0	25.6	56.3
Xe	0.7	0.5	0.3	0.5	0.7
Cs	2.1	1.9	1.5	1.9	2.1
Ba	11.8	7.4	10.2	7.4	11.8
Lu	90.7	24.3	30.5	22.4	88.5
Hf	179.7	52.0	53.1	62.9	184.6
Ta	263.8	159.5	71.4	159.5	263.8
W	512.2	201.0	144.8	201.0	512.2
Re	560.3	206.6	163.1	192.1	623.9
Os	657.2	163.9	255.8	200.2	775.9
Ir	581.0	228.9	250.3	228.9	581.0
Pt	295.9	224.6	50.5	224.6	295.9
Au	146.2	131.7	33.2	131.7	146.2
Tl	32.4	21.1	7.5	23.9	51.1

Pb	52.3	33.7	19.4	33.7	52.3
Po	82.4	26.8	3.0	26.8	82.4

3. Phonon calculations

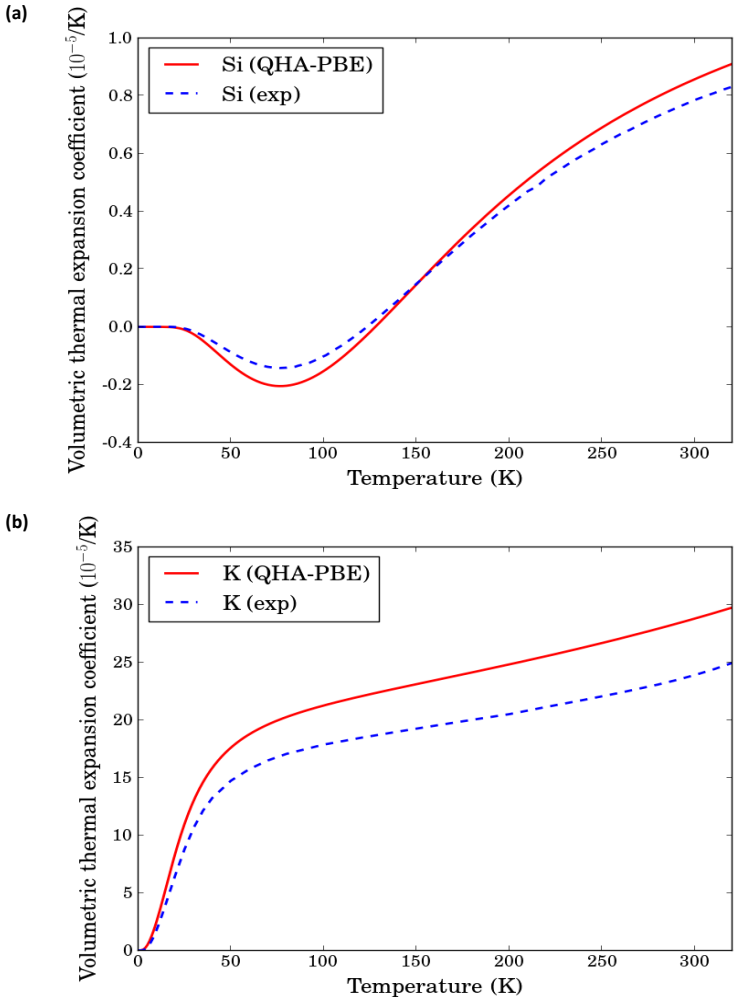


Fig. 3.1 Volumetric thermal expansion coefficient $\alpha_{V,rt}(T)$ of (a) Si and (b) K using the quasiharmonic approximation (based on PBE-calculated phonons, [42]) and from experiment [43, 44]. Phonon frequencies were extracted from the dynamical matrix, which was constructed from the forces in 3x3x3 supercells corresponding with atomic displacements of 0.02 Å. We used the same computational settings as for the rest of the calculations (see Sec. 1 of this Supplementary Material), except for the convergence criterion of the electronic cycle (10^{-9} eV) and a more accurate determination of the augmentation charges (ENAUG equal to 1600 eV, support grid enabled). A 7x7x7 Monkhorst-Pack k-grid was employed.

[1] J. Slater, Phys. Rev. **57**, 744 (1940).
[2] J. Dugdale and D. MacDonald, Phys. Rev. **89**, 832 (1953).
[3] V. Vashchenko and V. Zubarev, Sov. Phys. Solid State **5**, 653 (1963).
[4] F. Guinea, J. Rose, J. Smith, and J. Ferrante, Appl. Phys. Lett. **44**, 53 (1984).
[5] Y. Tsuru, Y. Shinzato, Y. Saito, M. Shimazu, M. Shiono, and M. Morinaga, J. Ceram Soc. Jpn. **118**, 241 (2010).
[6] D. Lide (ed.), *CRC Handbook of Chemistry and Physics* , 89th ed. (CRC Press, Boca Raton, Florida, 2008-2009).
[7] J. Nelson and D. Riley, Proc. Phys. Soc. London **57**, 477 (1945).
[8] R. Keyes, Phys. Rev. **92**, 580 (1953).
[9] J. Wallis, I. Sigalas, and S. Hart, J. Appl. Crystallogr. **19**, 273 (1986).
[10] R. Pawar and V. Deshpande, J. Mater. Sci. **5**, 1061 (1970).
[11] M. Anderson and C. Swenson, Phys. Rev. B **28**, 5395 (1983).
[12] J. Marples and C. Koch, Phys. Lett. A **41**, 307 (1972).
[13] R.Mills and E. Grilly, Phys. Rev. **101**, 1246 (1956)
[14] C. Kittel, Introduction to Solid State Physics, 8th ed. (John Wiley & Sons, Inc, 2005).
[15] N. Greenwood and A. Earnshaw, Chemistry of the Elements, 2nd ed. (Butterworth-Heinemann, 1997).
[16] K. Lejaeghere, V. Van Speybroeck, G. Van Oost and S. Cottenier, Crit. Rev. Solid State **39**, 22 (2014).
[17] P. Villars and J. Daams, J. Alloys Compd. **197**, 177 (1993).
[18] L. Bolz, M. Boyd, F. Mauer, and H. Peiser, Acta Crystallogr. **12**, 247 (1959).
[19] L. Meyer, C. Barrett, and S. Greer, J. Chem. Phys. **49**, 1902 (1968).
[20] D. Batchelder, D. Losee, and R. Simmons, Phys. Rev. **162**, 767 (1967).
[21] B. Powell, K. Heal, and B. Torrie, Mol. Phys. **53**, 929 (1984).
[22] O. Peterson, D. Batchelder, and R. Simmons, Phys. Rev. **150**, 703 (1966).
[23] D. Losee and R. Simmons, Phys. Rev. **172**, 944 (1968).
[24] D. Sears and H. Klug, J. Chem. Phys. **37**, 3002 (1962).
[25] C. Swenson, Phys. Rev. **111**, 82 (1958).
[26] E. Kim and C. Chen, Phys. Lett. A **326**, 442 (2004).
[27] T. Scott, Phys. Rep. **27C**, 89 (1976).
[28] Y. Endoh, G. Shirane, and J. Skalyo, Phys. Rev. B **11**, 1681 (1975).
[29] E.-F. Düsing, W. Grosshans, and W. Holzapfel, J. Phys. Colloques (Paris) **45**, 203 (1984).
[30] G. Keeler and D. Batchelder, J. Phys. C **3**, 510 (1970).
[31] H. Fujihisa and K. Takemura, Phys. Rev. B **52**, 13257 (1995).
[32] J. Skalyo, Y. Endoh, and G. Shirane, Phys. Rev. B **9**, 1797 (1974).
[33] D. Price, J. Rowe, and R. Nicklow, Phys. Rev. B **3**, 1268 (1971).
[34] N. Lurie, G. Shirane, and J. Skalyo, Phys. Rev. B **9**, 2661 (1974).
[35] E. Knittle, in *Mineral Physics and Crystallography: A Handbook of Physical Constants* , edited by T. Ahrens (American Geophysical Union Press, 1995) pp. 98-142.
[36] S. Batsanov, in *Effects of Explosions on Materials: Modification and Synthesis under High Pressure Shock Compression* (Springer, New York, 1994).
[37] M. Guinan and D. Steinberg, J. Phys. Chem. Solids **35**, 1501 (1974).
[38] C. Menoni, J. Hu, and I. Spain, Phys. Rev. B **34**, 362 (1986).
[39] H. Beister, K. Strössner, and K. Syassen, Phys. Rev. B **41**, 5535 (1990).
[40] T. Kenichi, Phys. Rev. B **70**, 012101 (2004).
[41] J. Li, S. Liang, H. Guo and B. Liu, J. Alloys Compd. **431**, 23 (2007).
[42] A. Togo, F. Oba and I. Tanaka, Phys. Rev. B **78**, 134106 (2008); http://phonopy.sourceforge.net.
[43] K. Lyon, G. Salinger, C. Swenson and G. White, J. Appl. Phys. **48**, 865 (1977).
[44] D. Schouten and C. Swenson, Phys. Rev. B **10**, 2175 (1974).

Supplementary Material of Paper 3

Ranking the stars: A refined Pareto approach to computational materials design

K. Lejaeghere, S. Cottenier and V. Van Speybroeck, *Phys. Rev. Lett.* (2013)
Supplemental Material

1. Computational parameters

cut-off energy
400 eV for most elements
600 eV for He, B, C, N, O, F, and Ne

self-consistent cycle
10^{-4} eV energy convergence criterion

reciprocal space representation
6750 / N k-points for N-atom cells (Monkhorst-Pack grid)

FFT grid
default (2 G_{cut} x 2 G_{cut} x 2 G_{cut} ; G_{cut} being the cut-off electronic wave vector)

augmentation charge grid
default (4 G_{cut} x 4 G_{cut} x 4 G_{cut})

spin polarization
included for cells containing Cr, Mn, Fe, Co or Ni

spin-orbit coupling
included for cells containing Tl or heavier elements

potential files
see Tab. 1.1

alloy structures
built from a supercell approach
m x m x m conventional bcc unit cell with a cubic sublattice of impurities superimposed

Tab. 1.1 VASP potentials used for the different elements

	potential file
H	PAW_PBE H 15Jun2001
He	PAW_PBE He 05Jan2001
Li	PAW_PBE Li_sv 23Jan2001
Be	PAW_PBE Be 06Sep2000
B	PAW_PBE B 06Sep2000
C	PAW_PBE C 08Apr2002
N	PAW_PBE N 08Apr2002
O	PAW_PBE O 08Apr2002
F	PAW_PBE F 08Apr2002
Ne	PAW_PBE Ne 05Jan2001
Na	PAW_PBE Na_pv 05Jan2001
Mg	PAW_PBE Mg 05Jan2001
Al	PAW_PBE Al 04Jan2001
Si	PAW_PBE Si 05Jan2001
P	PAW_PBE P 17Jan2003
S	PAW_PBE S 17Jan2003
Cl	PAW_PBE Cl 17Jan2003
Ar	PAW_PBE Ar 07Sep2000
K	PAW_PBE K_sv 06Sep2000
Ca	PAW_PBE Ca_pv 06Sep2000
Sc	PAW_PBE Sc_sv 07Sep2000
Ti	PAW_PBE Ti_pv 07Sep2000
V	PAW_PBE V_pv 07Sep2000
Cr	PAW_PBE Cr_pv 07Sep2000
Mn	PAW_PBE Mn_pv 07Sep2000
Fe	PAW_PBE Fe 06Sep2000
Co	PAW_PBE Co 06Sep2000
Ni	PAW_PBE Ni 06Sep2000
Cu	PAW_PBE Cu 05Jan2001
Zn	PAW_PBE Zn 06Sep2000
Ga	PAW_PBE Ga_d 06Sep2000
Ge	PAW_PBE Ge_d 06Sep2000
As	PAW_PBE As 06Sep2000
Se	PAW_PBE Se 06Sep2000
Br	PAW_PBE Br 06Sep2000
Kr	PAW_PBE Kr 07Sep2000
Rb	PAW_PBE Rb_sv 06Sep2000
Sr	PAW_PBE Sr_sv 07Sep2000
Y	PAW_PBE Y_sv 06Sep2000
Zr	PAW_PBE Zr_sv 07Sep2000
Nb	PAW_PBE Nb_pv 08Apr2002
Mo	PAW_PBE Mo_pv 08Apr2002
Tc	PAW_PBE Tc_pv 06Sep2000
Ru	PAW_PBE Ru 06Sep2000
Rh	PAW_PBE Rh 06Sep2000
Pd	PAW_PBE Pd 05Jan2001
Ag	PAW_PBE Ag 06Sep2000

Cd	PAW_PBE Cd 06Sep2000
In	PAW_PBE In_d 06Sep2000
Sn	PAW_PBE Sn_d 06Sep2000
Sb	PAW_PBE Sb 06Sep2000
Te	PAW_PBE Te 08Apr2002
I	PAW_PBE I 08Apr2002
Xe	PAW_PBE Xe 07Sep2000
Cs	PAW_PBE Cs_sv 08Apr2002
Ba	PAW_PBE Ba_sv 06Sep2000
Lu	PAW_PBE Lu_3 06Sep2000
Hf	PAW_PBE Hf_pv 06Sep2000
Ta	PAW_PBE Ta_pv 07Sep2000
W	PAW_PBE W_pv 06Sep2000
Re	PAW_PBE Re 17Jan2003
Os	PAW_PBE Os_pv 20Jan2003
Ir	PAW_PBE Ir 06Sep2000
Pt	PAW_PBE Pt 05Jan2001
Au	PAW_PBE Au 06Sep2000
Hg	PAW_PBE Hg 06Sep2000
Tl	PAW_PBE Tl_d 06Sep2000
Pb	PAW_PBE Pb_d 06Sep2000
Bi	PAW_PBE Bi_d 06Sep2000
Po	PAW_PBE Po_d 25May2007
Rn	PAW_PBE Rn 28Aug2006

2. Computational results

Tab. 2.1 Pareto set for the bcc binary tungsten alloys with respect to the mass density ρ_m (kg/m^3) and molar price p (USD/mol). The Pareto compounds are ranked according to the minimum win fraction (mwf)

	mwf	ρ_m	p
W	72.51%	18815	5.42
WPb	27.49%	16233	2.73
W_3Pb	25.46%	17623	4.08
W_7Pb	24.40%	18229	4.75
$W_{15}Pb$	21.14%	18474	5.09
W_7Hg	19.88%	18553	5.18
WHg	15.20%	17720	4.46
$W_{15}Hg$	14.56%	18626	5.30
WIr	10.44%	19816	2213.24
WOs	10.11%	20008	7325.41
WRe	9.50%	19657	1492.39
W_3Ir	6.80%	19369	1109.33
W_3Re	5.76%	19245	748.91
W_7Ir	4.30%	19112	557.38
W_7Re	3.49%	19034	377.17
$W_{15}Ir$	2.47%	18958	281.40
$W_{15}Re$	2.16%	18927	191.29
WO	0.63%	14640	2.71
$W_{15}Fe$	0.38%	18333	5.09
W_3Fe	0.27%	16643	4.07

Tab. 2.2 Pareto set for the bcc binary tungsten alloys and elemental crystals with respect to the mass density ρ_m (kg/m^3) and molar price p (USD/mol). The Pareto compounds are ranked according to the minimum win fraction (mwf)

	mwf	**ρ_m**	**p**
W	53.94%	18815	5.42
Pb	46.06%	10664	0.04
Fe	42.71%	8153	0.01
WPb	27.49%	16233	2.73
W_3Pb	25.46%	17623	4.08
Ir	24.46%	21913	4421.06
W_7Pb	24.40%	18229	4.75
$W_{15}Pb$	21.14%	18474	5.09
O	20.87%	1385	0.003
W_7Hg	19.88%	18553	5.18
Re	17.36%	20707	2979.36
WHg	15.20%	17720	4.46
$W_{15}Hg$	14.56%	18626	5.30
WIr	10.44%	19816	2213.24
WRe	9.50%	19657	1492.39
W_3Ir	6.80%	19369	1109.33
W_3Re	5.76%	19245	748.91
Os	4.47%	21993	14645.40
W_7Ir	4.30%	19112	557.38
W_7Re	3.49%	19034	377.17
$W_{15}Ir$	2.47%	18958	281.40
$W_{15}Re$	2.16%	18927	191.29
WO	0.63%	14640	2.71
$W_{15}Fe$	0.38%	18333	5.09
W_3Fe	0.27%	16643	4.07

Tab. 2.3 Pareto set for the bcc binary tungsten alloys with respect to the bulk modulus B_0 (GPa), cohesive energy ΔE_{coh} (kJ/mol) and molar price p (USD/mol). The Pareto compounds are ranked according to the minimum win fraction (mwf)

	mwf	B_0	ΔE_{coh}	p
WC	56.14%	365	737	2.86
W	43.86%	304	867	5.42
WRe	13.34%	336	801	1492.39
W_3Re	9.53%	320	834	748.91
W_3V	8.36%	276	801	4.41
W_7V	8.09%	290	834	4.92
W_7Cr	8.04%	295	803	4.76
$W_{15}V$	7.95%	297	851	5.17
$W_{15}Cr$	7.88%	300	835	5.09
W_3Fe	5.81%	265	751	4.07
W_7Re	5.76%	312	851	377.17
W_7Fe	4.42%	282	809	4.75
$W_{15}Si$	4.18%	294	837	5.09
W_3Cr	2.64%	288	742	4.10
$W_{15}Re$	2.59%	307	858	191.29
W_3Si	1.75%	266	742	4.08
WN	1.38%	342	618	2.74
W_7Ni	1.08%	288	805	4.89
$W_{15}Fe$	1.08%	296	834	5.09
WFe	0.81%	246	650	2.72
W_7Si	0.56%	283	801	4.75
WCr	0.48%	281	625	2.78
$W_{15}Os$	0.16%	307	858	920.42
WO	0.07%	242	542	2.71

Tab. 2.4 Pareto set for the bcc binary tungsten alloys and elemental crystals with respect to the bulk modulus B_0 (GPa), cohesive energy ΔE_{coh} (kJ/mol) and molar price p (USD/mol). The Pareto compounds are ranked according to the minimum win fraction (mwf)

	mwf	B_0	ΔE_{coh}	p
WC	56.14%	365	737	2.86
W	43.86%	304	867	5.42
Fe	32.31%	186	468	0.01
Os	22.66%	400	801	14645.40
Mn	18.95%	212	372	0.15
C	16.05%	2	758	0.29
WRe	13.34%	336	801	1492.39
O	10.95%	1	294	0.003
W_3Re	9.53%	320	834	748.91
W_3V	8.36%	276	801	4.41
Re	8.17%	373	754	2979.36
W_7V	8.09%	290	834	4.92
W_7Cr	8.04%	295	803	4.76
$W_{15}V$	7.95%	297	851	5.17
$W_{15}Cr$	7.88%	300	835	5.09
V	6.83%	181	571	1.38
N	6.54%	1	503	0.06
Ni	6.03%	194	466	1.12
W_3Fe	5.81%	265	751	4.07
W_7Re	5.76%	312	851	377.17
W_7Fe	4.42%	282	809	4.75
$W_{15}Si$	4.18%	294	837	5.09
W_3Cr	2.64%	288	742	4.10
$W_{15}Re$	2.59%	307	858	191.29
W_3Si	1.75%	266	742	4.08
WN	1.38%	342	618	2.74
W_7Ni	1.08%	288	805	4.89
$W_{15}Fe$	1.08%	296	834	5.09
WFe	0.81%	246	650	2.72
Co	0.75%	208	496	2.59
W_7Si	0.56%	283	801	4.75
WCr	0.48%	281	625	2.78
$W_{15}Os$	0.16%	307	858	920.42
WO	0.07%	242	542	2.71

Tab. 2.5 Pareto set for the mechanically stable bcc binary tungsten alloys with respect to the Cauchy pressure P_C (GPa), cohesive energy ΔE_{coh} (kJ/mol) and molar price p (USD/mol). The formation energy ΔE_f (eV/atom) is given as an additional criterion. All Pareto compounds are ranked according to the minimum win fraction (mwf)

	mwf	P_C	ΔE_{coh}	p	ΔE_f
W_7C	58.06%	189	778	4.78	0.78
W_3P	41.94%	213	694	6.39	0.41
WTa	22.10%	128	841	10.04	-0.10
W	19.00%	64	867	5.42	0.00
W_7Ta	18.93%	72	863	6.58	-0.06
W_3Ta	16.41%	74	858	7.73	-0.09
$W_{15}Co$	14.78%	79	835	5.25	0.08
$W_{15}Ni$	14.61%	79	834	5.15	0.08
$W_{15}C$	14.21%	95	824	5.10	0.37
W_7Si	13.94%	140	801	4.75	0.13
WSi	10.67%	146	641	2.73	0.13
WMn	10.57%	160	605	2.79	0.15
$W_{15}V$	10.16%	61	851	5.17	-0.03
W_7V	8.97%	70	834	4.92	-0.05
WV	8.62%	138	726	3.40	-0.08
WFe	7.97%	114	650	2.72	0.18
$W_{15}Si$	7.08%	64	837	5.09	0.03
W_3Si	3.57%	119	742	4.08	0.18
W_3Sb	3.50%	162	656	4.20	0.62
$W_{15}Fe$	3.43%	69	834	5.09	0.07
W_7Fe	3.43%	62	809	4.75	0.08
WC	3.09%	-77	737	2.86	0.78
$W_{15}P$	2.81%	85	825	5.67	0.08
W_3Ni	2.53%	131	741	4.35	0.26
W_7Ni	2.51%	83	805	4.89	0.11
W_7Ge	2.43%	158	782	15.63	0.22
W_3Fe	2.42%	75	751	4.07	0.16
W_3V	2.14%	76	801	4.41	-0.09
W_7Os	2.03%	78	850	1835.42	0.08
$W_{15}Cr$	1.75%	65	835	5.09	0.02
W_7Cr	0.79%	63	803	4.76	0.04
$W_{15}O$	0.70%	91	810	5.08	0.22
$W_{15}Rh$	0.45%	75	843	455.27	0.05
W_3Co	0.36%	125	750	4.72	0.25
$W_{15}N$	0.33%	84	811	5.09	0.34
W_7O	0.04%	108	753	4.75	0.43

3. Error estimates

Error estimates were determined from a linear regression between the experimental and computational properties of the elemental crystals, as done in [18]. For the mass density and the Cauchy pressure, this procedure yielded error estimates of 300 kg/m^3 and 17 GPa, respectively. The error bars for the bulk modulus (15 GPa) and the cohesive energy (30 kJ/mol) were taken from [18]. For the (logarithmic) price, we applied an error estimate of 0.12 (25-30 %) to account for price fluctuations. Its value was chosen as to encompass an approximately equal number of compounds as the other error bars.

4. Energetic and mechanical stability

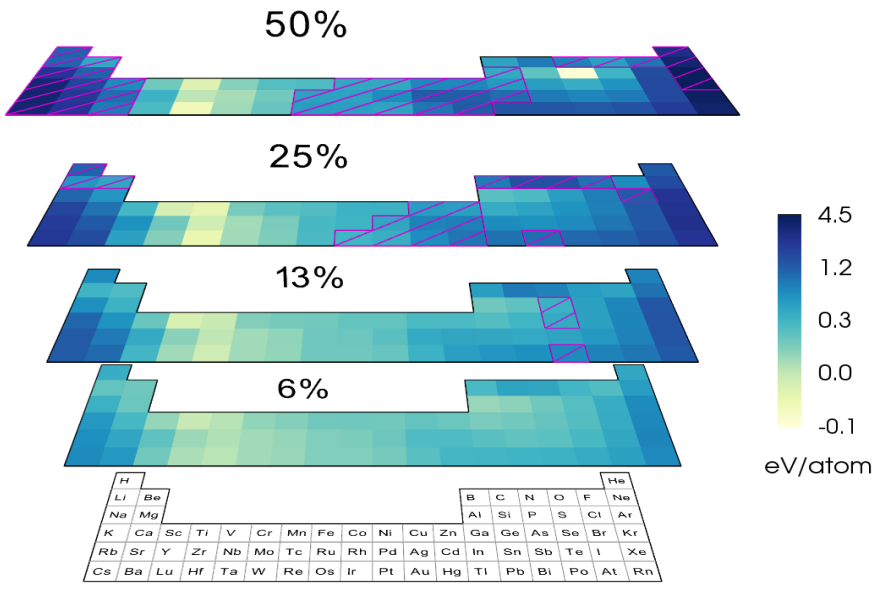


Fig. 4.1 Energetic and mechanical stability for the binary tungsten alloys. Each alloy is represented by a square at the corresponding impurity element and concentration. The colour code depicts the formation energy, while the hatched areas indicate mechanically unstable structures [33]

D
List of symbols

α	thermal expansion coefficient
α_l	linear thermal expansion coefficient
α_V	volumetric thermal expansion coefficient
β	regression coefficient
β_0	regression intercept
β_1	regression slope
γ	Grüneisen parameter
δ	deformation parameter
Δ	code-comparison criterion
Δ_1	modified code-comparison criterion
$\Delta B^{(1)}$	correction to B_0 due to thermal expansion
$\Delta B^{(2)}$	correction to B_0 due to zero-point effects
ΔE_{bcc}	formation energy from the pure bcc phases
ΔE_{coh}	cohesive energy
ΔE_{evol}	energy difference in evolutionary algorithms
ΔE_f	formation energy
ΔE_{form}	formation energy
ΔE_{mol}	moleculization energy
ΔE_{react}	reaction energy
ΔN	difference in electron number
$\Delta V^{(1)}$	correction to V_0 due to thermal expansion
$\Delta V^{(2)}$	correction to V_0 due to zero-point effects
ϵ	one-particle energy / strain / random DFT error
ϵ_0	vacuum permittivity

ζ	zero-point energy
η	random error on experimental values
θ	angle
Θ_D	Debye temperature
κ	thermal conductivity / inverse interaction length
λ	wavelength
ν	Poisson modulus
ρ	electron density
σ	stress / standard deviation
ϕ	basis function
ψ	one-particle wave function
$\tilde{\psi}$	pseudized wave function
Ψ	many-body wave function
ω	angular frequency
Ω	unit cell volume

a, b	unspecified prefactor
A, B, C	constant
b	Burgers vector
B_0	bulk modulus
B_1	pressure derivative of the bulk modulus
B_2	second-order pressure derivative of the bulk modulus
c	expansion coefficient
C	expansion coefficient matrix / stiffness constant
C_V	volumetric heat capacity
$\mathcal{C}_V$	electronic part of the volumetric heat capacity
e	electron
E	(electronic) energy (matrix) / Young's modulus
$\mathcal{E}$	prelogarithmic energy
E_0	equilibrium energy
E_{at}	atomic energy
E_{cut}	cutoff energy
E_F	Fermi energy
f	objective function
$\tilde{f}$	objective function evaluated at a Pareto solution $\tilde{\boldsymbol{x}}$
G	shear modulus
$\boldsymbol{G}$	reciprocal lattice vector
h	bond integral
H	Hamiltonian (matrix) / enthalpy
$\mathcal{H}$	Heaviside step function
I	interstitial space
$\boldsymbol{k}$	3D reciprocal-space coordinate

k_B	Boltzmann constant
K_Q	fracture toughness
ℓ	orbital momentum quantum number
l_{TF}	Thomas-Fermi screening length
m, M	mass
m	magnetic quantum number
M	surface heat capability / number of inequivalent atoms / Mendeleev number
N	number of electrons / objectives / particles
$\bar{N}$	average number of electrons
N_d	number of d electrons
n	neutron / band index / density of states
p	p-value / molar price
P	pressure / probability
$\mathcal{P}$	Pareto set
P_C	Cauchy pressure
r	Pearson product-moment correlation coefficient
$\boldsymbol{r}$	3D real-space coordinate
r_{WS}	Wigner-Seitz radius
$\boldsymbol{R}$	3D real-space ionic coordinate
R	interatomic distance
R^2	coefficient of determination
S	overlap matrix / muffin tin sphere
T	kinetic energy / temperature / theoretical stochastic variable
T_m	melting temperature
$\mathcal{T}$	pseudization transformation
u	function with the same periodicity as the real-space lattice / atomic radial wave function
U	internal energy
v	potential
$\tilde{v}$	pseudized potential
V	potential energy / volume
V_0	equilibrium volume
w	orbital occupation number
W	band width
$\boldsymbol{x}$	compound represented as a materials property vector
$\tilde{\boldsymbol{x}}$	Pareto compound represented as a materials property vector
X, Y	objective function / stochastic variable / transition metal
Y	spherical harmonic
z	coordination number
Z	atomic number

E

List of abbreviations

AIMD	*ab initio* molecular dynamics
APW	augmented plane-wave method
ASE	Atomic Simulation Environment
at	atom
bcc	body-centred cubic
BM	Birch-Murnaghan
BZ	first Brillouin zone
C	dominant correlation
CIF	crystallographic information file
CMM	Center for Molecular Modeling
COD	Crystallographic Open Database
CPU	central processing unit
D	London dispersion
DBTT	ductile-to-brittle transition temperature
D-D	deuterium-deuterium
DDD	discrete dislocation dynamics
DDR	double data rate
DEMO	demonstration reactor
DFT	density-functional theory
DFT-D	density-functional theory with dispersion corrections
DM	Dugdale and MacDonald
DOS	density of states
dpa	displacements per atom
D-T	deuterium-tritium

EAM	embedded-atom method
EFDA	European Fusion Development Agreement
eff	effective
ELM	edge-localized mode
EOS	equation of state
EWI	Economie, Wetenschap en Innovatie
exp	experimental
FDR	fourteen data rate
FEM	finite-elements method
fcc	face-centred cubic
FF	force field
FFT	fast Fourier transform
F/M	ferritic-martensitic steel
FWO	Fonds Wetenschappelijk Onderzoek
G	Guinea *et al.*
geom	geometry
GGA	generalized-gradient approximation
H	Hartree
HF	Hartree-Fock
HPC	high-performance computing
IB	InfiniBand
ICSD	Inorganic Crystal Structure Database
IFMIF	International Fusion Material Irradiation Facility
IPM	independent-particle model
ITER	(Lat.) the road
KMC	kinetic Monte Carlo
KS	Kohn-Sham
L	low coordination number
LAPW	linearized augmented plane-wave method
LCAO	linear combination of atomic orbitals
LDA	local-density approximation
lo	local orbital
LO	local orbital
max	maximum
MCDM	multicriterion decision making
min	minimum
mwf	minimum win fraction
N/A	not applicable
n-SO	without spin-orbit coupling
PAW	projector augmented-wave method
PBE	functional of Perdew, Burke and Ernzerhof
PBE+U	Hubbard-corrected PBE functional

PBS	portable batch system
PDOS	partial density of states
PRACE	Partnership for Advanced Computing in Europe
pos	position
PW91	functional of Perdew and Wang (1991)
QDR	quad data rate
QHA	quasiharmonic approximation
QSPR	quantitative structure-property relationship
RAM	random-access memory
RANSAC	random sample consensus
reg	regression
rep	repulsion
RF	radio frequency
RPA	random-phase approximation
rt	room temperature
S	Slater
SCF	self-consistent field
SCOOP	Scalable Concurrent Operations in Python
SER	standard error of the regression
SO	spin-orbit coupling
TB	tight binding
th	theory
tot	total
Ts	Tsuru *et al.*
tokamak	toroidal chamber with magnetic coils
uts	ultimate tensile strength
VASP	Vienna Ab initio Simulation Package
vdW-DF2	van der Waals density functional version 2
VSC	Vlaams Supercomputer Centrum
VZ	Vashchenko and Zubarev
wf	win fraction
WVM	Tungsten Vacuum Metallizing
xc	exchange-correlation

F

List of publications

Publications in international peer-reviewed journals

1. *Accuracy of structure optimizations, phonons and elasticity of metal-organic frameworks*
   A. Ghysels, D. Vanpoucke, K. Lejaeghere, and V. Van Speybroeck
   to be submitted to J. Chem. Theory Comp.

2. *The 245 keV level of ^{111}Cd: nuclear quadrupole moment and its error bar*
   L. Errico, K. Lejaeghere, J. Runco, M. Renteria, and S. Cottenier
   to be submitted to Phys. Rev. B

3. *Ab initio based thermal property predictions at a low cost: An error analysis*
   K. Lejaeghere, J. Jaeken, V. Van Speybroeck, and S. Cottenier
   Phys. Rev. B **89**, 014304 (2014)

4. *Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals*
   K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier
   Crit. Rev. Solid State **39**, 1-24 (2014)

5. *Ranking the Stars: A Refined Pareto Approach to Computational Materials Design*
   K. Lejaeghere, S. Cottenier, and V. Van Speybroeck
   Phys. Rev. Lett. **111**, 075501 (2013)

6. *Assessment of a low-cost protocol for an ab initio based prediction of the mixing enthalpy at elevated temperatures: The Fe-Mo system*
   K. Lejaeghere, S. Cottenier, S. Claessens, M. Waroquier, and V. Van Speybroeck
   Phys. Rev. B **83**, 184201 (2011)

Invited oral contributions to national and international conferences

The presenting author has been indicated with an asterisk (*).

1. *How to validate pseudopotentials? Methods of comparison*
   K. Lejaeghere*, S. Cottenier, and V. Van Speybroeck
   CFCAM Workshop 'Pseudopotentials and PAW atomic data: beyond a "black art"?', Paris, France (January 28-29 2014)

2. *The Delta parameter: which precision has a meaning on physical results?*
   S. Cottenier*, K. Lejaeghere, and V. Van Speybroeck
   CFCAM Workshop 'Pseudopotentials and PAW atomic data: beyond a "black art"?', Paris, France (January 28-29 2014)

3. *DFT vs. reality: the error budget*
   S. Cottenier*, K. Lejaeghere, J. Jaeken, and V. Van Speybroeck
   International workshop on computational physics and materials science ("Total energy and force methods"), Lausanne, Switzerland (January 9-11 2014)

4. *DFT-based Thermal Properties: Three Levels of Error Management*
   K. Lejaeghere*, J. Jaeken, G. Van Oost, V. Van Speybroeck, and S. Cottenier
   IMA Workshop on Uncertainty Quantification in Materials Modeling, Minneapolis, MN, USA (December 16-17 2013)

5. *The ground state elemental crystals as a guideline for the assessment of solid state DFT accuracy*
   K. Lejaeghere*, V. Van Speybroeck, G. Van Oost, and S. Cottenier
   NIST workshop on Atomistic Simulations for Industrial Needs, Gaithersburg, MD, USA (August 13-14 2013)

6. *Inverse Materials Design for Tungsten Alloys*
   K. Lejaeghere, S. Cottenier*, and V. Van Speybroeck
   International Workshop on High-Throughput Materials Development, Ghent, Belgium (June 11-12 2013)

7. *Finding candidate materials for the first wall of fusion reactors by an inverse strategy*
   K. Lejaeghere, S. Cottenier*, M. Waroquier, and V. Van Speybroeck
   MRS Fall Meeting, Boston, MA, USA (November 25-30 2012)

8. *Error bars for solid-state density-functional theory predictions*
   K. Lejaeghere, G. Van Oost, V. Van Speybroeck, and S. Cottenier*
   17th ETSF Workshop on Electronic Excitations, Coimbra, Portugal (October 1-5 2012)

9. *The ground state elemental crystals as a guideline for the assessment of solid state DFT accuracy*
   K. Lejaeghere*, V. Van Speybroeck, G. Van Oost, and S. Cottenier
   CECAM workshop on Validation and Verification in Electronic-Structure calculations, Lausanne, Switzerland (September 5-7 2012)

10. *Error bars on DFT predictions: sense or nonsense?*
    K. Lejaeghere, G. Van Oost, V. Van Speybroeck, and S. Cottenier*
    19th WIEN2k Workshop, Tokyo, Japan (September 3-7 2012)

11. *Computational materials design for electricity generation in the 22nd century*
    S. Cottenier*, K. Lejaeghere, and V. Van Speybroeck
    4th Workshop on novel methods for electronic structure calculations and 1st South American congress on materials, La Plata, Argentina (November 2-4 2011)

Other oral contributions to national and international conferences

The presenting author has been indicated with an asterisk (*).

1. *An extended Pareto approach to computational materials design: tungsten alloys for nuclear fusion reactors*
   K. Lejaeghere*, S. Cottenier, and V. Van Speybroeck
   EMRS Spring Meeting, Lille, France (May 26-30 2014)

2. *Checking the engines: quantitative error bar assessment for DFT-based property predictions*
   K. Lejaeghere, J. Jaeken, V. Van Speybroeck, and S. Cottenier*
   EMRS Spring Meeting, Lille, France (May 26-30 2014)

3. *Tailoring Metal-organic frameworks for adsorption applications*
   D.E.P. Vanpoucke*, T. Verstraelen, M. Vandichel, A. Ghysels, K. Lejaeghere, and V. Van Speybroeck
   EMRS Spring Meeting, Lille, France (May 26-30 2014)

4. *An extended Pareto approach to computational materials design*
   K. Lejaeghere*, S. Cottenier, and V. Van Speybroeck
   DPG Spring Meeting, Regensburg, Germany (March 11-15 2013)

5. *The ground state elemental crystals as a benchmark set for solid state DFT: intrinsic accuracy and code comparison*
   K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier*
   ICAMM 2012, Nantes, France (June 11-16 2012)

6. *The ground state elemental crystals as a benchmark set for solid state DFT: intrinsic accuracy and code comparison*
   K. Lejaeghere, G. Van Oost, V. Van Speybroeck, and S. Cottenier*
   ADIS 2012, Tegernsee, Germany (April 29 - May 4 2012)

7. *The ground state elemental crystals as a benchmark set for solid state DFT: intrinsic accuracy and code comparison*
   K. Lejaeghere*, V. Van Speybroeck, G. Van Oost, L. Malerba, and S. Cottenier
   TEC PhD meeting, Mol, Belgium (January 26 2012)

8. *Ab initio screening of suitable tungsten alloys as first wall material in nuclear fusion reactors*
   K. Lejaeghere*, V. Van Speybroeck, G. Van Oost, L. Malerba, and S. Cottenier
   TEC PhD meeting, Jülich, Germany (January 20 2011)

Poster contributions to national and international conferences

The presenting author has been indicated with an asterisk (*).

1. *A high-throughput approach to computational materials design: tungsten alloys for nuclear fusion reactors*
   K. Lejaeghere*, S. Cottenier, and V. Van Speybroeck
   14th FEA PhD Symposium, Ghent, Belgium (December 6 2013)
   poster award 2013 by Atlas Copco

2. *Benchmark Set for Accuracy Assessment of DFT Codes and Methods*
   K. Lejaeghere, V. Van Speybroeck, and S. Cottenier*
   20th WIEN2k Workshop, State College, PA, USA (August 12-15 2013)

3. *Materials for Energy: The Role of Computational Materials Design*
   K. Lejaeghere*, K. Rijpstra, M. Sluydts, J. Jaeken, M. Waroquier, V. Van Speybroeck, and S. Cottenier
   Colloquium on Physics and the Energy Challenge, Brussels, Belgium (April 20 2013)

4. *The ground state elemental crystals as a benchmark set for solid state DFT properties*
   K. Lejaeghere*, V. Van Speybroeck, G. Van Oost, and S. Cottenier
   DFTM-2012, Ghent, Belgium (April 1-6 2012)

5. *The ground state elemental crystals as a benchmark set for solid state DFT: intrinsic accuracy and code comparison*
   K. Lejaeghere*, V. Van Speybroeck, and S. Cottenier
   QCB10, Brussels, Belgium (February 10 2012)

6. *Modelling of the enthalpy of mixing for the development of new "bulk metallic glasses" (BMG)*
   K. Lejaeghere*, S. Cottenier, M. Waroquier, and V. Van Speybroeck
   Umicore Scientific Awards, Brussels, Belgium (March 31 2011)

7. *Assessment of a low-cost protocol for an ab initio based prediction of the mixing enthalpy at elevated temperatures: the Fe-Mo system*
   K. Lejaeghere*, S. Cottenier, M. Waroquier, and V. Van Speybroeck
   11th FirW PhD Symposium, Ghent, Belgium (December 1 2010)

8. *Computationele zoektocht naar een waardig alternatief voor staal*
   K. Lejaeghere*, S. Cottenier, and V. Van Speybroeck
   ie-prijzen 2010, Brussels, Belgium (November 25 2010)

9. *Modelling of the enthalpy of mixing for the development of new "bulk metallic glasses" (BMG)*
   K. Lejaeghere*, S. Cottenier, M. Waroquier, and V. Van Speybroeck
   CAMD Summer School 2010, Lyngby, Denmark (August 14-20 2010)

10. *Modelling of the enthalpy of mixing for the development of new "bulk metallic glasses" (BMG)*
    K. Lejaeghere, S. Cottenier, M. Waroquier, and V. Van Speybroeck
    OCAS PhD Day, Zelzate, Belgium (June 10, 2010)

Master's thesis

Modelling of the enthalpy of mixing for the development of new "bulk metallic glasses" (BMG)
K. Lejaeghere
Master's thesis at the Center for Molecular Modeling (CMM), Ghent University, 2009-2010
Promoters: prof. dr. ir. V. Van Speybroeck and dr. S. Cottenier

Bibliography

[1] United Nations Population Division, *World Population Prospects: The 2010 Revision*, United Nations, New York, 2011.

[2] International Energy Agency, *World Energy Outlook*, IEA Publications, 2012.

[3] J. Wesson, *The science of JET*, JET, March 2000.

[4] G. Van Oost, *Plasma Physics (partim hot plasmas)*, UGent course taught during the academic year 2008-2009.

[5] G. Van Oost, *Plasma Technology and Fusion Technology (partim fusion technology)*, UGent course taught during the academic year 2009-2010.

[6] Linda Hall Library, "It's a questions of physics: what's the difference between fission & fusion?," `http://atomic.lindahall.org/what-is-fission-and-fusion.html` (last consulted March 2014).

[7] HyperPhysics, "Nuclear Binding Energy," `http://hyperphysics.phy-astr.gsu.edu/hbase/nucene/nucbin.html` (last consulted December 2013).

[8] O.A. Hurricane, D.A. Callahan, D.T. Casey, P.M. Celliers, C. Cerjan, E.L. Dewald, T.R. Dittrich, T. Döppner, D.E. Hinkel, L.F. Berzak Hopkins, J.L. Kline, S. Le Pape, T. Ma, A.G. MacPhee, J.L. Milovich, A. Pak, H.-S. Park, P.K. Patel, B.A. Remington, J.D. Salmonson, P.T. Springer, and R. Tommasini, "Fuel gain exceeding unity in an inertially confined fusion implosion," *Nature*, 2014, doi:10.1038/nature13008.

[9] EFDA, "European Fusion Development Agreement," `http://www.efda.org` (last consulted December 2013).

[10] M. Rieth, S.L. Dudarev, S.M. Gonzalez de Vicente, J. Aktaa, T. Ahlgren, S. Antusch, D.E.J. Armstrong, M. Balden, N. Baluc, M.-F. Barthe, W.W. Basuki, M. Battabyal, C.S. Becquart, D. Blagoeva, H. Boldyryeva, J. Brinkmann, M. Celino, L. Ciupinski, J.B. Correia, A. De Backer,

C. Domain, E. Gaganidze, C. García-Rosales, J. Gibson, M.R. Gilbert, S. Giusepponi, B. Gludovatz, H. Greuner, K. Heinola, T. Höschen, A. Hoffmann, N. Holstein, F. Koch, W. Krauss, H. Li, S. Lindig, J. Linke, Ch. Linsmeier, P. López-Ruiz, H. Maier, J. Matejicek, T.P. Mishra, M. Muhammed, A. Muñoz, M. Muzyk, K. Nordlund, D. Nguyen-Manh, J. Opschoor, N. Ordás, T. Palacios, G. Pintsuk, R. Pippan, J. Reiser, J. Riesch, S.G. Roberts, L. Romaner, M. Rosiński, M. Sanchez, W. Schulmeyer, H. Traxler, A. Ureña, J.G. van der Laan, L. Veleva, S. Wahlberg, M. Walter, T. Weber, T. Weitkamp, S. Wurster, M.A. Yar, J.H. You, and A. Ziverlonghi, "Recent progress in research on tungsten materials for nuclear fusion applications in Europe," *J. Nucl. Mater.*, vol. 432, pp. 482–500, 2013.

[11] D. King, A. Airaghi, H. Bolt, J. Calvo, B. Frois, M. Gaube, L. Högberg, G. Marbach, and S. Walsgrove, "Conclusions of the Fusion Fast Track Experts Meeting," November 2001.

[12] EFDA, *Fusion Electricity: A roadmap to the realisation of fusion energy*, November 2012.

[13] V. Barabash, The ITER International Team, A. Peacock, S. Fabritsiev, G. Kalinin, S. Zinkle, A. Rowcliffe, J.-W. Rensman, A.A. Tavassoli, P. Marmy, P.J. Karditsas, F. Gillemot, and M. Akiba, "Materials challenges for ITER – Current status and future activities," *J. Nucl. Mater.*, vol. 367-370, pp. 21–32, 2007.

[14] R. Lässer, N. Baluc, J.-L. Boutard, E. Diegele, S. Dudarev, M. Gasparotto, A. Möslang, R. Pippan, B. Riccardi, and B. van der Schaaf, "Structural materials for DEMO: The EU development, strategy, testing and modelling," *Fusion Eng. Des.*, vol. 82, pp. 511–520, 2007.

[15] S.J. Zinkle, "Advanced materials for fusion technology," *Fusion Eng. Des.*, vol. 74, pp. 31–40, 2005.

[16] D.L. Smith, S. Majumdar, M. Billone, and R. Mattas, "Performance limits for fusion first-wall structural materials," *J. Nucl. Mater.*, vol. 283-287, pp. 716–720, 2000.

[17] N.P. Taylor and R. Pampin, "Activation properties of tungsten as a first wall protection in fusion power plants," *Fusion Eng. Des.*, vol. 81, pp. 1333–1338, 2006.

[18] G.A. Cottrell, R. Pampin, and N.P. Taylor, "Transmutation and phase stability of tungsten armor in fusion power plants," *Fusion Sci. Technol.*, vol. 50, pp. 89–98, 2006.

[19] A.-A.F. Tavassoli, "Present limits and improvements of structural materials for fusion reactors – a review," *J. Nucl. Mater.*, vol. 302, pp. 73–88, 2002.

[20] Y. Shimomura, "ITER and plasma surface interaction issues in a fusion reactor," *J. Nucl. Mater.*, vol. 363-365, pp. 467–475, 2007.

[21] A. Loarte, B. Lipschultz, A.S. Kukushkin, G.F. Matthews, P.C. Stangeby, N. Asakura, G.F. Counsell, G. Federici, A. Kallenbach, K. Krieger, A. Mahdavi, V. Philipps, D. Reiter, J. Roth, J. Strachan, D. Whyte, R. Doerner, T. Eich, W. Fundamenski, A. Herrmann, M. Fenstermacher, P. Ghendrih, M. Groth, A. Kirschner, S. Konoshima, B. LaBombard, P. Lang, A.W. Leonard, P. Monier-Garbet, R. Neu, H. Pacher, B. Pegourie, R.A. Pitts, S. Takamura, J. Terry, E. Tsitrone, and the ITPA Scrape-off Layer and Divertor Physics Topical Group, "Chapter 4: Power and particle control," *Nucl. Fuison*, vol. 47, pp. S203–S263, 2007.

[22] D.E. Post, "The role of atomic collisions in fusion," in *Physics of Ion-Ion and Electron-Ion Collisions*, F. Brouillard and J.W. McGowan, Eds., 1983, vol. 83 of *NATO Advanced Study Institutes*, pp. 37–99.

[23] S.J. Zinkle and N.M. Ghoniem, "Operating temperature windows for fusion reactor structural materials," *Fusion Eng. Des.*, vol. 51-52, pp. 55–71, 2000.

[24] S.J. Zinkle, J.P. Robertson, and R.L. Klueh, "Thermophysical and mechanical properties of Fe-(8-9%)Cr reduced activation steels," Fusion Materials Semiann. Prog. Report, Oak Ridge National Lab, June 1998.

[25] D.S. Gelles and M.A. Sokolov, "Fracture toughness variability in F82H," Fusion Materials Semiann. Prog. Report, Oak Ridge National Lab, June 2003.

[26] C.L. Younger and G.N. Wrights, "Effect of reactor irradiation on ductile-brittle transition and stress-strain behavior of tungsten," NASA technical note, September 1970.

[27] Plansee SE, "Tungsten," `http://www.plansee.com/en/Materials-Tungsten-403.htm` (last consulted January 2014).

[28] G.A. Geach and J.E. Hughes, "The Alloys of Rhenium with Molybdenum or with Tungsten and Having Good High-Temperature Properties," in *Plansee Seminar Proceedings*, F. Benesovsky, Ed., London, 1955, Pergamon Press.

[29] G. Pfennig, H. Klewe-Nebenius, and W. Seelmann-Eggebert, *Karlsruher Nuklidkarte*, Druckhaus Haberbeck, Lage (Lippe), Germany, 6th edition, 1998.

[30] Element Materials Technology, "Finite Element Analysis (FEA)," `http://www.element.com/services-index/finiteelementanalysis-fea` (last consulted March 2014).

[31] Laboratoire d'Etude des Microstructures, "3-D simulation of dislocation dynamics," `http://zig.onera.fr/DisGallery/3D.html` (last consulted March 2014).

[32] C.-C. Fu, J. Dalla Torre, F. Willaime, J.-L. Bocquet, and A. Barbu, "Multiscale modelling of defect kinetics in irradiated iron," *Nature Mater.*, vol. 4, pp. 68–74, 2005.

[33] ESA, "IMPRESS, Supporting Education Across Europe – Glossary (G)," `http://www.spaceflight.esa.int/impress/text/education/Glossary/Glossary_G.html` (last consulted March 2014).

[34] W.K. Liu, E.G. Karpov, S. Zhang, and H.S. Park, "An introduction to computational nanomechanics and materials," *Comp. Methods Appl. Mech. Eng.*, vol. 193, pp. 1529–1578, 2004.

[35] S.L. Dudarev, J.-L. Boutard, R. Lässer, M.J. Caturla, P.M. Derlet, M. Fivel, C.-C. Fu, M.Y. Lavrentiev, L. Malerba, M. Mrovec, D. Nguyen-Manh, K. Nordlund, M. Perlado, R. Schäublin, H. Van Swygenhoven, D. Terentyev, J. Wallenius, D. Weygand, and F. Willaime, "The EU programme for modelling radiation effects in fusion reactor materials: An overview of recent advances and future goals," *J. Nucl. Mater.*, vol. 386-388, pp. 1–7, 2009.

[36] D. Duffy, "Modelling materials for fusion power," *Int. Mater. Rev.*, vol. 56, pp. 324–340, 2011.

[37] T.Diaz de la Rubia, H.M. Zbib, T.A. Khraishi, B.D. Wirth, M. Victoria, and M.J. Caturia, "Multiscale modelling of plastic flow localization in irradiated materials," *Nature*, vol. 406, pp. 871–874, 2000.

[38] G.S. Smith, E.B. Tadmor, and E. Kaxiras, "Multiscale Simulation of Loading and Electrical Resistance in Silicon Nanoindentation," *Phys. Rev. Lett.*, vol. 84, pp. 1260–1263, 2000.

[39] S.S. Rao, *The Finite Element Method in Engineering*, Butterworth-Heinemann, Boston, 5th edition, 2011.

[40] A. Chatterjee and D.G. Vlachos, "An overview of spatial microscopic and accelerated kinetic monte carlo methods," *J. Comp.-Aided Mater. Des.*, vol. 14, pp. 253–308, 2007.

[41] L.K. Mansur, "Theory and experimental background on dimensional changes in irradiated alloys," *J. Nucl. Mater.*, vol. 216, pp. 97–123, 1994.

[42] I. Steinbach, "Phase-field models in materials science," *Model. Simul. Mater. Sci. Eng.*, vol. 17, pp. 073001, 2009.

[43] H.M. Zbib, M. Rhee, and J.P. Hirth, "On plastic deformation and the dynamics of 3D dislocations," *Int. J. Mech. Sci.*, vol. 40, pp. 113–127, 1998.

[44] D.J. Bacon, D.M. Barnett, and R.O. Scattergood, "Anisotropic continuum theory of lattice defects," *Prog. Mater. Sci.*, vol. 23, pp. 51–262, 1980.

[45] W.H. Dickhoff and D. Van Neck, *Many-Body Theory Exposed!*, World Scientific, 2nd edition, 2008.

[46] R.G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, vol. 16 of *International series of monographs on chemistry*, Oxford University Press, New York, 1989.

[47] R. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, 2004.

[48] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.*, vol. 136, pp. B864–B871, 1964.

[49] W. Kohn and L.J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.*, vol. 140, pp. A1133–A1138, 1965.

[50] Nobel Media AB, "The Nobel Prize in Chemistry 1998," `http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/` (last consulted March 2014).

[51] J.P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," *AIP Conf. Proc.*, vol. 577, pp. 1–20, 2001.

[52] J.P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865, 1996.

[53] G.I. Csonka, J.P. Perdew, A. Ruzsinszky, P.H.T. Philipsen, S. Lebègue, J. Paier, O.A. Vydrov, and J.G. Ángyán, "Assessing the performance of recent density functionals for bulk solids," *Phys. Rev. B*, vol. 79, pp. 155107, 2009.

[54] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Phys. Rev. B*, vol. 46, pp. 6671, 1992.

[55] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais, "Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Phys. Rev. B*, vol. 48, pp. 4978, 1993.

[56] A.E. Mattsson, R. Armiento, P.A. Schultz, and T.R. Mattsson, "Nonequivalence of the generalized gradient approximations PBE and PW91," *Phys. Rev. B*, vol. 73, pp. 195123, 2006.

[57] M. Swart, "Density Functional Theory poll," `http://www.marcelswart.eu/`.

[58] V. Ozoliņš and M. Körling, "Full-potential calculations using the generalized gradient approximation: Structural properties of transition metals," *Phys. Rev. B*, vol. 48, no. 24, pp. 18304–18307, 1993.

[59] J.P. Perdew, A. Ruzsinszky, G.I. Csonka, O.A. Vydrov, G.E. Scuseria, L.A. Constantin, X. Zhou, and K. Burke, "Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces," *Phys. Rev. Lett.*, vol. 100, pp. 136406, 2008.

[60] Z. Wu and R.E. Cohen, "More accurate generalized gradient approximation for solids," *Phys. Rev. B*, vol. 73, pp. 235116, 2006.

[61] Y. Zhang and W. Yang, "Comment on "Generalized Gradient Approximation Made Simple"," *Phys. Rev. Lett.*, vol. 80, pp. 890, 1998.

[62] B. Hammer, L.B. Hansen, and J.K. Nørskov, "Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals," *Phys. Rev. B*, vol. 59, pp. 7413–7421, 1999.

[63] P. Haas, F. Tran, P. Blaha, and K. Schwarz, "Construction of an optimal GGA functional for molecules and solids," *Phys. Rev. B*, vol. 83, pp. 205117, 2011.

[64] L. He, F. Liu, G. Hautier, M.J.T. Oliveira, M.A.L. Marques, F.D. Vila, J.J. Rehr, G.-M. Rignanese, and A. Zhou, "Accuracy of generalized gradient approximation functionals for density-functional perturbation theory calculations," *Phys. Rev. B*, vol. 89, pp. 064305, 2014.

[65] F. Tran, R. Laskowski, P. Blaha, and K. Schwarz, "Performance on molecules, surfaces, and solids of the Wu-Cohen GGA exchange-correlation energy functional," *Phys. Rev. B*, vol. 75, pp. 115131, 2007.

[66] V.N. Staroverov, G.E. Scuseria, J. Tao, and J.P. Perdew, "Tests of a ladder of density functionals for bulk solids and surfaces," *Phys. Rev. B*, vol. 69, pp. 075102, 2004.

[67] V.N. Staroverov, G.E. Scuseria, J. Tao, and J.P. Perdew, "Erratum: Tests of a ladder of density functionals for bulk solids and surfaces," *Phys. Rev. B*, vol. 78, pp. 239907(E), 2008.

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

[69] J.J. Mortensen, L.B. Hansen, and K.W. Jacobsen, "Real-space grid implementation of the projector augmented wave method," *Phys. Rev. B*, vol. 71, pp. 035109, 2005.

[70] L. Vitos, "Total-energy method based on the exact muffin-tin orbitals theory," *Phys. Rev. B*, vol. 64, pp. 014107, 2001.

[71] S. Cottenier, *Density Functional Theory and the family of (L)APW-methods: a step-by-step introduction*, 2nd edition, (freely available at `http://www.wien2k.at/reg_user/textbooks`).

[72] G.H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, 3rd edition, 1996.

[73] M.C. Gibson, *Implementation and Application of Advanced Density Functionals*, Doctoraatsthesis, University of Durham, 2006.

[74] G. Kresse, M. Marsman, and J. Furthmüller, *VASP the guide*, Universität Wien, september 9 2013 edition, (freely available at `http://www.vasp.at/`).

[75] H.J. Monkhorst and J.D. Pack, "Special points for Brillouin-zone integrations," *Phys. Rev. B*, vol. 13, no. 12, pp. 5188–5192, 1976.

[76] P.E. Blöchl, O. Jepsen, and O.K. Andersen, "Improved tetrahedron method for Brillouin-zone integrations," *Phys. Rev. B*, vol. 49, pp. 16223–16233, 1994.

[77] Wikipedia, "Pseudopotential," `http://en.wikipedia.org/wiki/Pseudopotential` (last consulted March 2014).

[78] U. von Barth and C.D. Gelatt, "Validity of the frozen-core approximation and pseudopotential theory for cohesive energy calculations," *Phys. Rev. B*, vol. 21, pp. 2222–2228, 1980.

[79] M.D. Segall, P.J.D. Lindan, M.J. Probert, C.J. Pickard, P.J. Hasnip, S.J. Clark, and M.C. Payne, "First-principles simulation: ideas, illustrations and the CASTEP code," *J. Phys.: Condens. Matter*, vol. 14, pp. 2717, 2002.

[80] L. Kleinman and D.M. Bylander, "Efficacious Form for Model Pseudopotentials," *Phys. Rev. Lett.*, vol. 48, pp. 1425–1428, 1982.

[81] D.R. Hamann, M. Schlüter, and C. Chiang, "Norm-Conserving Pseudopotentials," *Phys. Rev. Lett.*, vol. 43, pp. 1494–1497, 1979.

[82] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism," *Phys. Rev. B*, vol. 41, pp. 7892–7895, 1990.

[83] P.E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B*, vol. 50, pp. 17953–17979, 1994.

[84] S. Cottenier, "Density Functional Theory and the family of (L)APW-methods: a step-by-step introduction," 2002-2013 (2nd edition), freely available at `http://www.wien2k.at/reg_user/textbooks`.

[85] J.C. Slater, "An Augmented Plane Wave Method for the Periodic Potential Problem," *Phys. Rev.*, vol. 92, pp. 603–608, 1953.

[86] M.M. Saffren and J.C. Slater, "An Augmented Plane-Wave Method for the Periodic Potential Problem. II," *Phys. Rev.*, vol. 92, pp. 1126–1128, 1953.

[87] O.K. Andersen, "Linear methods in band theory," *Phys. Rev. B*, vol. 12, pp. 3060–3083, 1975.

[88] D.D. Koelling and G.O. Arbman, "Use of energy derivative of the radial solution in an augmented plane wave method: application to copper," *J. Phys. F: Metal Phys.*, vol. 5, pp. 2041–2054, 1975.

[89] E. Sjöstedt, L. Nordström, and D.J. Singh, "An alternative way of linearizing the augmented plane-wave method," *Solid State Comm.*, vol. 114, pp. 15–20, 2000.

[90] G.K.H. Madsen, P. Blaha, K. Schwarz, E. Sjöstedt, and L. Nordström, "Efficient linearization of the augmented plane-wave method," *Phys. Rev. B*, vol. 64, pp. 195134, 2001.

[91] R.A. Potyrailo and E.J. Amis (Ed.), *High-throughput analysis: a tool for combinatorial materials science*, Kluwer Academic Publishers, New York, 2003.

[92] H.M. Berman, W.K. Olson, D.L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A.R. Srinivasan, and B. Schneider, "The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids," *Biophys. J.*, vol. 63, pp. 751–759, 1992.

[93] G.J. Ferland, K.T. Korista, D.A. Verner, J.W. Ferguson, J.B. Kingdon, and E.M. Verner, "CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra," *Publ. Astron. Soc. Pac.*, vol. 110, pp. 761–778, 1998.

[94] C.E. Wilmer, M. Leaf, C.Y. Lee, O.K. Farha, B.G. Hauser, J.T. Hupp, and R.Q. Snurr, "Large-scale screening of hypothetical metal-organic frameworks," *Nature Chem.*, vol. 4, pp. 83–89, 2011.

[95] US National Science and Technology Council, "Materials Genome Initiative for Global Competitiveness," June 2011, `http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf` (last consulted December 2013).

[96] D. Morgan, G. Ceder, and S. Curtarolo, "High-throughput and data mining with *ab initio* methods," *Meas. Sci. Technol.*, vol. 16, pp. 296–301, 2005.

[97] S. Curtarolo, G.L.W. Hart, M. Buongiorno Nardelli, N. Mingo, S. Sanvito, and O. Levy, "The high-throughput highway to computational materials design," *Nature Mater.*, vol. 12, pp. 191–201, 2013.

[98] Wim Dewitte, "The computational conveyer belt," February 2014.

[99] A. Belsky, M. Hellenbrandt, V.L. Karen, and P. Luksch, "New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design," *Acta. Crystallogr. B*, vol. 58, pp. 364–369, 2002.

[100] S. Gražulis, D. Chateigner, R.T. Downs, A.F.T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, "Crystallography Open Database – an open-access collection of crystal structures," *J. Appl. Crystallogr.*, vol. 42, no. 4, pp. 726–729, 2009.

[101] A. Franceschetti and A. Zunger, "The inverse band-structure problem of finding an atomic configuration with given electronic properties," *Nature*, vol. 402, pp. 60–63, 1999.

[102] Q. Zhu, A.R. Oganov, M.A. Salvadó, P. Pertierra, and A.O. Lyakhov, "Denser than diamond: *Ab initio* search for superdense carbon allotropes," *Phys. Rev. B*, vol. 83, pp. 193410, 2011.

[103] A.O. Lyakhov and A.R. Oganov, "Evolutionary search for superhard materials: Methodology and applications to forms of carbon and TiO_2," *Phys. Rev. B*, vol. 84, pp. 092103, 2011.

[104] C.J. Pickard and R.J. Needs, "Structures at high pressure from random searching," *Phys. Status Solidi B*, vol. 246, pp. 536–540, 2009.

[105] S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, pp. 671–680, 1983.

[106] V. Černý, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *J. Optimiz. Theory App.*, vol. 45, pp. 41–51, 1985.

[107] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, vol. 21, pp. 1087, 1953.

[108] N.A. Barricelli, "Esempi numerici di processi di evoluzione," *Methodos*, vol. 6-7, pp. 45–68, 1954.

[109] N. Chakraborti, "Genetic algorithms in materials design and processing," *Int. Mater. Rev.*, vol. 49, pp. 246–260, 2004.

[110] D.M. Deaven and K.M. Ho, "Molecular Geometry Optimization with a Genetic Algorithm," *Phys. Rev. Lett.*, vol. 75, pp. 288–291, 1995.

[111] G.H. Jóhannesson, T. Bligaard, A.V. Ruban, H.L. Skriver, K.W. Jacobsen, and J.K. Nørskov, "Combined Electronic Structure and Evolutionary Search Approach to Materials Design," *Phys. Rev. Lett.*, vol. 88, pp. 255506, 2002.

[112] A.R. Oganov and C.W. Glass, "Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications," *J. Chem. Phys.*, vol. 124, pp. 244704, 2006.

[113] N.L. Abraham and M.I.J. Probert, "A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction," *Phys. Rev. B*, vol. 73, pp. 224104, 2006.

[114] L.B. Vilhelmsen and B. Hammer, "Systematic Study of Au_6 to Au_{12} Gold Clusters on MgO(100) F Centers Using Density-Functional Theory," *Phys. Rev. Lett.*, vol. 108, pp. 126101, 2012.

[115] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philos. Mag.*, vol. 2, pp. 559–572, 1901.

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

[117] A. Togo and I. Tanaka, "Evolution of crystal structures in metallic elements," *Phys. Rev. B*, vol. 87, pp. 184104, 2013.

[118] Y. Wang, J. Lv, L. Zhu, and Y. Ma, "Crystal structure prediction via particle-swarm optimization," *Phys. Rev. B*, vol. 82, pp. 094116, 2010.

[119] S.R. Bahn and K.W. Jacobsen, "An object-oriented scripting interface to a legacy electronic structure code," *Comp. Sci. Eng.*, vol. 4, pp. 56–66, 2002.

[120] A. van de Walle, M. Asta, and G. Ceder, "The Alloy Theoretic Automated Toolkit: A User Guide," *Calphad*, vol. 26, pp. 539–553, 2002.

[121] B. Kozinsky, "Two Key Challenges in Computational Materials Screening: Theory and Practice," presentation at the CECAM workshop on Validation and Verification in Electronic-Structure Calculations, September 2012.

[122] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, and D. Morgan, "AFLOW: An automatic framework for high-throughput materials discovery," *Comp. Mater. Sci.*, vol. 58, pp. 218–226, 2012.

[123] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, and G. Ceder, "Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis," *Comp. Mater. Sci.*, vol. 68, pp. 314–319, 2013.

[124] D.D. Landis, J.S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J.K. Nørskov, and K.W. Jacobsen, "The Computational Materials Repository," *Comp. Sci. Eng.*, vol. 14, pp. 51–57, 2012.

[125] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, pp. 011002, 2013.

[126] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, "AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations," *Comp. Mater. Sci.*, vol. 58, pp. 227–235, 2012.

[127] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)," *JOM*, vol. 65, pp. 1501–1509, 2013.

[128] J.S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard, and J.K. Nørskov, "CatApp: A Web Application for Surface Chemistry and Heterogeneous Catalysis," *Angew. Chem. Int. Ed.*, vol. 51, pp. 272–274, 2012.

[129] G. Yuan and F. Gygi, "ESTEST: a framework for the validation and verification of electronic structure codes," *Comp. Sci. Disc.*, vol. 3, pp. 015004, 2010.

[130] S. Adams, P. de Castro, P. Echenique, J. Estrada, M.D. Hanwell, P. Murray-Rust, P. Sherwood, J. Thomas, and J. Townsend, "The Quixote project: Collaborative and Open Quantum Chemistry data management in the Internet age," *J. Cheminform.*, vol. 3, pp. 38, 2011.

[131] NIST, "Physical Reference Data," `http://www.nist.gov/pml/data/` (last consulted February 2014).

[132] G. Hautier, S.P. Ong, A. Jain, C.J. Moore, and G. Ceder, "Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability," *Phys. Rev. B*, vol. 85, pp. 155208, 2012.

[133] M. Sluydts, D.M. Tomecka, K. Lejaeghere *et al.*, "PBE-based band gap predictions and their error bars," in preparation.

[134] F. Guinea, J.H. Rose, J.R. Smith, and J. Ferrante, "Scaling relations in the equations of state, thermal expansion, and melting of metals," *Appl. Phys. Lett.*, vol. 44, no. 1, pp. 53–55, 1984.

[135] B.G. Sumpter and D.W. Noid, "On the Design, Analysis, and Characterization of Materials Using Computatinal Neural Networks," *Annu. Rev. Mater. Sci.*, vol. 26, pp. 223–277, 1996.

[136] J.M. Serra, L.A. Baumes, M. Moliner, P. Serna, and A. Corma, "Zeolite Synthesis Modelling with Support Vector Machines: A Combinatorial Approach," *Comb. Chem. High Throughput Screen.*, vol. 10, pp. 13–24, 2007.

[137] D.G. Pettifor, "The structures of binary compounds. I. Phenomenological structure maps," *J. Phys. C: Solid State Phys.*, vol. 19, pp. 285–313, 1986.

[138] M. Ashby, *Materials Selection in Mechanical Design*, Butterworth-Heinemann, Burlington, Massachusetts, 3rd edition, 1999.

[139] K.M. Miettinen, *Nonlinear Multiobjective Optimization*, International Series in Operations Research and Management Science. Kluwer Academic Publishers, Boston, 1999.

[140] R.L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*, Cambridge University Press, Cambridge, 1993.

[141] V. Pareto, *Manuale di economia politica con una introduzione alla scienza sociale*, Società Editrice Libraria, Milano, 1906.

[142] S. Börzsönyi, D. Kossmann, and K. Stocker, "The Skyline Operator," in *Proceedings of the 17th International Conference on Data Engineering*, Heidelberg, 2001, pp. 421–430.

[143] P. Godfrey, R. Shipley, and J. Gryz, "Algorithms and analyses for maximal vector computation," *VLDB J.*, vol. 16, pp. 5–28, 2007.

[144] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*, vol. 16 of *Wiley Interscience Series in Systems and Optimization*, Wiley, New York, 2001.

[145] T. Bligaard, G.H. Jóhannesson, A.V. Ruban, H.L. Skriver, K.W. Jacobsen, and J.K. Nørskov, "Pareto-optimal alloys," *Appl. Phys. Lett.*, vol. 83, pp. 4527–4529, 2003.

[146] M.D. Halls and K. Tasaki, "High-throughput quantum chemistry and virtual screening for lithium ion battery electrolyte additives," *J. Power Sources*, vol. 195, pp. 1472–1478, 2010.

[147] J.T. Buchanan, "A naïve approach for solving MCDM problems: the GUESS method," *J. Oper. Res. Soc.*, vol. 48, pp. 202–206, 1997.

[148] V. Venkat, S.H. Jacobson, and J.A. Stori, "A Post-Optimality Analysis Algorithm for Multi-Objective Optimization," *Comput. Optim. Appl.*, vol. 28, pp. 357–372, 2004.

[149] H.A. Taboada, F. Baheranwala, D.W. Coit, and N. Wattanapongsakorn, "Practical solutions for multi-objective optimization: An application to system reliability design problems," *Reliab. Eng. Syst. Safe.*, vol. 92, pp. 314–322, 2007.

[150] C. O'Mahony and N. Wilson, "Sorted-Pareto Dominance: an extension to the Pareto Dominance relation and its application in Soft Constraints," in *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, Athens, 2012, pp. 798–805.

[151] I. Das, "A preference ordering among various Pareto optimal alternatives," *Struct. Optim.*, vol. 18, pp. 30–35, 1999.

[152] N. Lopez, O. Aguirre, J.F. Espiritu, and H.A. Taboada, "Using game theory as a post-Pareto analysis for renewable energy integration problems considering multiple objectives," in *Proceedings of the 41st International Conference on Computers and Industrial Engineering*, Los Angeles, 2011, pp. 678–683.

[153] S.L. Dudarev, "Density Functional Theory Models for Radiation Damage," *Annu. Rev. Mater. Sci.*, vol. 43, pp. 35–61, 2013.

[154] A. Bakaev, D. Terentyev, G. Bonny, T.P.C. Klaver, P. Olsson, and D. Van Neck, "Interaction of minor alloying elements of high-Cr ferritic steels with lattice defects: An *ab initio* study," *J. Nucl. Mater.*, vol. 444, pp. 237–246, 2014.

[155] X. Li, C. Zhang, J. Zhao, and B. Johansson, "Mechanical properties and defective effects of bcc V-4Cr-4Ti and V-5Cr-5Ti alloys by first-principles simulations," *Comp. Mater. Sci.*, vol. 50, pp. 2727–2731, 2011.

[156] C.J. Ortiz, M.J. Caturla, C.C. Fu, and F. Willaime, "Influence of carbon on the kinetics of He migration and clustering in α-Fe from first principles," *Phys. Rev. B*, vol. 80, pp. 134109, 2009.

[157] F. Gao, H. Xiao, X. Zu, M. Posselt, and W.J. Weber, "Defect-Enhanced Charge Transfer by Ion-Solid Interactions in SiC using Large-Scale *Ab Initio* Molecular Dynamics Simulations," *Phys. Rev. Lett.*, vol. 103, pp. 027405, 2009.

[158] M. Muzyk, D. Nguyen-Manh, K.J. Kurzydłowski, N.L. Baluc, and S.L. Dudarev, "Phase stability, point defects, and elastic properties of W-V and W-Ta alloys," *Phys. Rev. B*, vol. 84, pp. 104115, 2011.

[159] L. Romaner, C. Ambrosch-Draxl, and R. Pippan, "Effect of Rhenium on the Dislocation Core Structure in Tungsten," *Phys. Rev. Lett.*, vol. 104, pp. 195503, 2010.

[160] Center for Molecular Modeling, "Comparing Solid State DFT Codes, Basis Sets and Potentials," `https://molmod.ugent.be/deltacodesdft` (last consulted March 2014).

[161] M. Parizeau, Y. Hold, and O. Gagnon, "scoop – Scalable COncurrent Operations in Python," `http://www.pyscoop.org/` (last consulted March 2014).

[162] A. Bengtson, K. Persson, and D. Morgan, "*Ab initio* study of the composition dependence of the pressure-induced spin crossover in perovskite $(Mg_{1-x},Fe_x)SiO_3$," *Earth Planet. Sci. Lett.*, vol. 265, pp. 535–545, 2008.

[163] J. He, M.-X. Chen, X.-Q. Chen, and C. Franchini, "Structural transitions and transport-half-metallic ferromagnetism in $LaMnO_3$ at elevated pressure," *Phys. Rev. B*, vol. 85, pp. 195135, 2012.

[164] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Comput. Mater. Sci.*, vol. 6, pp. 15, 1996.

[165] J. Hafner, "Ab-Initio Simulations of Materials Using VASP: Density-Functional Theory and Beyond," *J. Comput. Chem.*, vol. 29, pp. 2044–2078, 2008.

[166] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Phys. Rev. B*, vol. 59, pp. 1758, 1999.

[167] F. Birch, "Finite Elastic Strain of Cubic Crystals," *Phys. Rev.*, vol. 71, pp. 809–824, 1947.

[168] Wikipedia, "Brillouin-Zone," `http://de.wikipedia.org/wiki/Brillouin-Zone` (last consulted March 2014).

[169] C. Kittel, *Introduction to Solid State Physics*, John Wiley & Sons, Inc, 8th edition, 2005.

[170] M.J. Mehl, J.E. Osburn, D.A. Papaconstantopoulos, and B.M. Klein, "Structural properties of ordered high-melting-temperature intermetallic alloys from first-principles total-energy calculations," *Phys. Rev. B*, vol. 41, pp. 10311–10323, 1990.

[171] M.J. Mehl, B.M. Klein, and D.A. Papaconstantopoulos, "First principles calculations of elastic properties of metals," in *Intermetallic Compounds – Principles and Practice*, J.H. Westbrook and R.L. Fleischer, Eds., London, 1994, vol. 1, pp. 195–210, John Wiley and Sons.

[172] Y. Le Page and P. Saxe, "Symmetry-general least-squares extraction of elastic coefficients from *ab initio* total energy calculations," *Phys. Rev. B*, vol. 63, pp. 174103, 2001.

[173] Y. Le Page and P. Saxe, "Symmetry-general least-squares extraction of elastic data for strained materials from *ab initio* calculations of stress," *Phys. Rev. B*, vol. 65, pp. 104104, 2002.

[174] S. Shang, Y. Wang, and Z.-K. Liu, "First-principles elastic constants of α- and θ-Al_2O_3," *Appl. Phys. Lett.*, vol. 90, pp. 101909, 2007.

[175] D.G. Pettifor, "A chemical scale for crystal-structure maps," *Solid State Comm.*, vol. 51, pp. 31–34, 1984.

[176] C.D. Gelatt, Jr., H. Ehrenreich, and R.E. Watson, "Renormalized atoms: Cohesion in transition metals," *Phys. Rev. B*, vol. 15, pp. 1613–1628, 1977.

[177] E.C. Snow and J.T. Waber, "The APW energy bands for the body centered and face centered cubic modifications of the 3d transition metals," *Acta Metall.*, vol. 17, pp. 623–635, 1969.

[178] D.G. Pettifor, *Bonding and Structure of Molecules and Solids*, Oxford Science Publications. Oxford University Press, Oxford, 1995.

[179] D.G. Pettifor, "Theoretical predictions of structure and related properties of intermetallics," *Mater. Sci. Technol.*, vol. 8, pp. 345–349, 1992.

[180] S.F. Pugh, "Relation between the Elastic Moduli and the Plastic Properties of Polycrystalline Pure Metals," *Philos. Mag.*, vol. 45, pp. 823–843, 1954.

[181] C.R. Weinberger, G.J. Tucker, and S.M. Foiles, "Peierls potential of screw dislocations in bcc transition metals: Predictions from density functional theory," *Phys. Rev. B*, vol. 87, pp. 054114, 2013.

[182] J.J. Gilman, *Electronic basis of the strength of materials*, Cambridge University Press, Cambridge, 2003.

[183] G. Gottstein, *Physical Foundations of Materials Science*, Springer-Verlag, Berlin Heidelberg, 2004.

[184] P.E.A. Turchi, M. Sluiter, F.J. Pinski, D.D. Johnson, D.M. Nicholson, G.M. Stocks, and J.B. Staunton, "First-principles study of phase stability in Cu-Zn substitutional alloys," *Phys. Rev. Lett.*, vol. 67, pp. 1779–1782, 1991.

[185] L. Nordheim, "Zur Elektronentheorie der Metalle. I," *Ann. Phys. (Berlin)*, vol. 401, pp. 607–640, 1931.

[186] A. Zunger, S.-H. Wei, L.G. Ferreira, and J.E. Bernard, "Special Quasirandom Structures," *Phys. Rev. Lett.*, vol. 65, pp. 353–356, 1990.

[187] D.W. Taylor, "Vibrational Properties of Imperfect Crystals with Large Defect Concentrations," *Phys. Rev.*, vol. 156, pp. 1017–1029, 1967.

[188] J.M. Sanchez, F. Ducastelle, and D. Gratias, "Generalized cluster description of multicomponent systems," *Physica A*, vol. 128, pp. 334–350, 1984.

[189] S. Liu, R. Tao, and C.M. Tam, "Optimizing cost and CO_2 emission for construction projects using particle swarm optimization," *Habitat Int.*, vol. 37, pp. 155–162, 2013.

[190] M. Choudhury, V. Jalan, S. Sarkar, and A. Basu, "Evolution, optimization, and language change: the case of the Bengali verb inflections," *Proc. ACL SIGMORPHON*, vol. 9, pp. 65–74, 2007.

[191] N. Assadian and S.H. Pourtakdoust, "Multiobjective genetic optimization of Earth-Moon trajectories in the restricted four-body problem," *Adv. Space Res.*, vol. 45, pp. 398–409, 2010.

[192] F. Jollet, M. Torrent, and N. Holzwarth, "Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format," *Comp. Phys. Comm.*, vol. 185, pp. 1246–1254, 2014.

[193] HPC UGent, "High Performance Computing Infrastructure," `http://www.ugent.be/hpc/en/` (last consulted March 2014).

[194] "Tier1 supercomputer," http://www.ugent.be/hpc/en/infrastructure/tier1 (last consulted January 2014).

[195] PRACE, "Partnership for Advanced Computing in Europe," http://www.prace-ri.eu/ (last consulted March 2014).

[196] K. Hoste and E. Pauwels, "Introduction to HPC-UGent and VSC," FOSDEM'14, Brussels, February 2014.

[197] J.C. Slater and G.F. Koster, "Simplified LCAO Method for the Periodic Potential Problem," *Phys. Rev.*, vol. 94, pp. 1498–1524, 1954.

[198] F. Cyrot-Lackmann, "Sur le calcul de la cohésion et de la tension superficielle des métaux de transition par une méthode de liaisons fortes," *J. Phys. Chem. Solids*, vol. 29, pp. 1235–1243, 1968.

[199] F. Ducastelle and F. Cyrot-Lackmann, "Moments developments and their application to the electronic charge distribution of d bands," *J. Phys. Chem. Solids*, vol. 31, pp. 1295–1306, 1970.

[200] J. Friedel, "Transition metals. Electronic structure of the d-band. Its role in the crystalline and magnetic structures," in *The Physics of Metals*, J.M. Ziman, Ed., New York, 1969, vol. 1, pp. 340–408, Cambridge University Press.

[201] F. Ducastelle, "Modules élastiques des métaux de transition," *J. Phys. (Paris)*, vol. 31, pp. 1055–1062, 1970.

[202] D.G. Pettifor, "Theory of the Heats of Formation of Transition-Metal Alloys," *Phys. Rev. Lett.*, vol. 42, pp. 846–850, 1979.

[203] A.J. Skinner and D.G. Pettifor, "Transferability and the pair potential within the tight-binding bond model: an analytical study for hydrogen," *J. Phys.: Condens. Matter*, vol. 3, pp. 2029–2047, 1991.

[204] R.A. Buckingham, "The Classical Equation of State of Gaseous Helium, Neon and Argon," *Proc. Royal Soc. London A*, vol. 168, pp. 264–283, 1938.

This research was supported by the Research Foundation Flanders.

The computational resources and services used in this work were provided by Ghent University (Stevin), the Hercules Foundation (Tier-1 Flemish Supercomputer Infrastructure) and the Flemish Government – department EWI.