

Automated Bonding Analysis with Crystal Orbital Hamilton Populations

Janine George,^{*[a,b,c]} Guido Petretto,^[c] Aakash Naik,^[a,b] Marco Esters,^[d] Adam J. Jackson,^[e] Ryky Nelson,^[f] Richard Dronskowski,^[f,g] Gian-Marco Rignanesi,^[c] Geoffroy Hautier^{*[c,h]}

Dedication ((optional))

-
- [a] Dr. Janine George, Aakash Naik
Federal Institute for Materials Research and Testing
Department Materials Chemistry
Unter den Eichen 87, 12205 Berlin, Germany.
E-mail: Janine.george@bam.de
- [b] Dr. Janine George, Aakash Naik
Friedrich Schiller University Jena,
Institute of Condensed Matter Theory and Solid-State Optics
Max-Wien-Platz 1, 07743 Jena, Germany
- [c] Dr. Janine George (former address), Dr. Guido Petretto (former address), Prof. Dr. Gian-Marco Rignanesi, Prof. Dr. Geoffroy Hautier (former address)
Université catholique de Louvain,
Institute of Condensed Matter and Nanosciences,
Chemin des Étoiles 8, 1348 Louvain-la-Neuve, Belgium
- [d] Dr. Marco Esters (former address)
Department of Chemistry
University of Oregon
Eugene, Oregon 97403, USA
and
Dr. Marco Esters
Department of Mechanical Engineering and Materials Science and Center for Autonomous Materials Design
Duke University
Durham, North Carolina 27708, USA
- [e] Dr. Adam J. Jackson
Scientific Computing Department,
Science & Technology Facilities Council,
Rutherford Appleton Laboratory,
Didcot OX11 0QX, UK
- [f] Dr. Ryky Nelson, Prof. Dr. Richard Dronskowski
RWTH Aachen University
Institute of Inorganic Chemistry,
D-52056 Aachen, Germany.
- [g] Prof. Dr. Richard Dronskowski
RWTH Aachen University
Jülich-Aachen Research Alliance (JARA-CSD),
D-52056 Aachen, Germany
And
Prof. Dr. Richard Dronskowski
Shenzhen Polytechnic
Hoffmann Institute of Advanced Materials,
7098 Liuxian Blvd, Nanshan District, Shenzhen 518055, China
- [h] Prof. Dr. Geoffroy Hautier
Dartmouth College
Thayer School of Engineering,
Hanover, New Hampshire 03755, USA
E-mail: geoffroy.hautier@dartmouth.edu

Supporting information for this article is given via a link at the end of the document.

Abstract: Understanding crystalline structures based on their chemical bonding is growing in importance. In this context, chemical bonding can be studied with the Crystal Orbital Hamilton Population (COHP), allowing to quantify interatomic bond strength. Here we present a new set of tools to automate the calculation of COHP and analyze the results. We use the program packages *VASP* and *LOBSTER* and the Python packages *atomate* and *pymatgen*. The analysis produced by our tools includes plots, a textual description, and key data in machine-readable format. To illustrate those capabilities, we have selected simple test compounds (NaCl, GaN), the oxynitrides BaTaO₂N, CaTaO₂N, and SrTaO₂N, and the thermoelectric material Yb₁₄MnSb₁₁. We show correlations between bond strengths and stabilities in the oxynitrides, as well as the influence of the Mn-Sb bonds on the magnetism in Yb₁₄MnSb₁₁. Our contribution enables high-throughput bonding analysis and will facilitate the use of bonding information for machine learning studies.

Introduction

The concept of chemical bonding plays a major role in understanding both molecules and crystals and their properties.^[1,2] Chemical bonding in crystals can be studied using various techniques based on electronic structure theory methods. These techniques are usually divided into density-based and orbital-based methods. For example, a widely used density-based method is Bader analysis, which is based on electron density topology analysis.^[3] Bader analysis has also been used in a high-throughput manner.^[4–6] In this paper, however, we will focus on a more quantum-chemical orbital-based bonding analysis, which has been shown to be successful in understanding crystal structures in the past,^[2,7–9] and its automation. These methods have the advantage that bonding and antibonding contributions directly fall off from the orbital phases such that one arrives at a picture that better relates to the chemist’s orbital-based understanding of atoms and molecules.^[8] An important technique in this context is Crystal Orbital Hamilton Population (COHP),^[10] a bond-weighted density of states that provides information about bond energy and covalency in crystals.^[7] There are other bond indicators not based on the Hamilton matrix but on the overlap matrix or bond order (Crystal Orbital Overlap Populations, COOP and Crystal Orbital Bond Index, COBI).^[11] The first one relates to bond strengths, the second one to bond orders. In addition, there are also orbital-based indicators for the ionicity of compounds.^[11,12]

COHP has been used for about three decades to study all kinds of compounds, for example, phase change materials,^[13] intermetallics,^[14] magnetic materials,^[15] and many other classes of materials. Nowadays, these populations can also be derived according to electronic structure theory calculations using the projector augmented wave method. This is done by projecting the original delocalized basis based on plane waves onto atomic orbitals (e.g., Slater orbitals), as implemented, for example, in the computer program *LOBSTER*.^[16–19] *LOBSTER* is also capable of computing COHPs, COOPs, COBIs, and many other similar properties based on the information resulting from the projection. In addition, Mulliken and Löwdin charges and corresponding Madelung energies can be calculated.^[11,20] This procedure is illustrated in Figure 1. Here we use *VASP* as the code for our electronic structure theory calculations.

To date, this type of projection has not been implemented in any of the standard codes for electronic structure theory based on plane waves (e.g., *VASP*,^[21–23] *Abinit*,^[24] or *Quantum Espresso*^[25]). Therefore, this type of bonding analysis is a multi-step process involving at least two programs. New users not only have to use

and learn non-standard settings for the electronic structure theory codes, they must also learn to use *LOBSTER*. Naturally, the users should also be aware of fundamental quantum-chemical concepts not necessarily taught in courses of quantum mechanics or computational materials science. Hence, this usually involves a lot of training for first-time users of the programs, a lot of manual work, possible errors and problems in interpretation. This limits the use of this bonding analysis tool for high-throughput calculations, although there have been some studies where *LOBSTER* has been used for a larger number of compounds.^[26] Workflow-managing codes nowadays provide ways to automate such complex calculations.^[6,27,28] This has been done in the past for several types of standard calculations (e.g., magnetic ground-state calculations, defect calculations,^[6,29–32]). This automation is also directly related to the need for high-throughput calculations and accurate computational data in materials science for machine learning and data-driven material discovery.^[33–35]

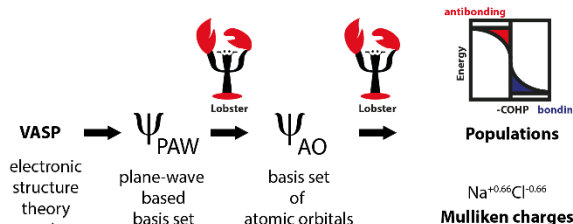


Figure 1. Illustration of the process of bonding analysis using the *VASP* and *LOBSTER* tools. *VASP* is used to calculate a wave function using the projector-augmented wave method. This calculation depends on the basis function chosen in the later *LOBSTER* run. *LOBSTER* then projects this wave function onto a basis set of atomic orbitals that must be adapted to the system at hand, corresponding to a unitary transformation between a totally delocalized and a localized representation. Then, *LOBSTER* calculates various populations and charges based on these populations, which allow to assess the bonding situation in the compound. This is a multi-step process whose steps are interdependent. We have now developed automated tools to perform all the steps of this procedure and to ensure the compatibility of the different steps.

For example, high-throughput bonding analysis could provide further insight into important interactions within crystal structures and provide opportunities to determine coordination environments in crystal structures not only based on geometric constraints but also on electronic structure.^[36,37] This has the advantage that the relevance of neighbors can be evaluated based on the integrated COHP as a measure of a bond energy and hence a covalent bond strength. In purely geometric approaches, the strength of covalent bonds can only be evaluated indirectly using limits on bond lengths. Moreover, some of these bond properties could serve as descriptors in machine learning studies since the underlying bonding situation is expected to be causative for many materials properties. Recently, two such studies have been published.^[26,38]

Here, we present an implementation of bonding analysis using the programs *VASP* and *LOBSTER* in a Python-based workflow. We have automated this process using the programs *VASP*, *LOBSTER*,^[7,16–19] *pymatgen*,^[39] *custodian*,^[39] and *atomate*.^[28] This means that the entire bonding analysis workflow, leading to an analysis text and a summary of key properties, can be started with very little Python code. In addition to the computational workflow, we have also automated the analysis of *LOBSTER* results. Without this automation, users must do this by hand, and there is currently no unified strategy for doing so, unless the user is properly trained. Therefore, automating this process will also help in standardizing the analysis of *LOBSTER* results. This automation of the whole process will hopefully lead to fewer errors

in the future and less training being required. We also expect that the automated analysis will contribute to a much wider use of the tools and serve as a starting point for more detailed and individual bonding analysis by the user. In addition, high-throughput studies should now be in reach for non-expert *LOBSTER* users.

To illustrate the capabilities of the code, we have tested it on a range of binary and more complex materials: we use the binaries NaCl, GaN, then the oxynitrides CaTaO₂N, BaTaO₂N, SrTaO₂N, and finally Yb₁₄Mn₁Sb₁₁, where the bonding analysis is performed fully automatically. In the case of the oxynitrides, we also show how easily correlations with other material properties can be derived based on the developed tools. We test the correlation between the total energy of the systems and the strongest integrated COHP upon varying the anion order in these perovskite systems. In addition, Yb₁₄Mn₁Sb₁₁ illustrates the potential of the code to identify important interactions for magnetic properties and potential bonding-based driving forces for properties in materials.

Results and Discussion

The workflow

The implemented workflow is shown in Figure 2.

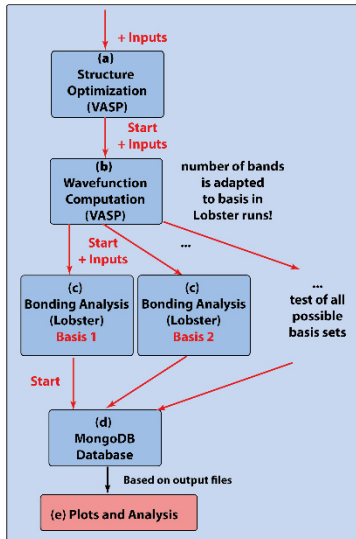


Figure 2. This graphic illustrates the progress that has been made in this work. Instead of creating all input files through custom scripts or by hand and starting and evaluating calculations by hand, there is now a fully automated workflow for the entire bonding analysis process using the *VASP* and *LOBSTER* programs. This workflow can automatically create all input files, start and monitor calculations. We have also developed tools to evaluate all the results and create automatic output texts and dictionaries of the most important bonding properties, which can be used to populate databases and machine learn the data.

The crystal structure is optimized in a first optional step (a). This step is optional to ensure that structures of any origin can be used as input for the subsequent COHP analysis.

Then, the wave functions are calculated based on DFT (b). In this step, special requisites for the DFT calculation are considered. Due to the required projection from plane waves to atomic orbitals, as many bands must be included in the DFT calculation as are needed for the projection to atomic orbitals. The number of bands actually depends on the structure at hand and on the basis chosen for the projection in *LOBSTER*. COHPs can only be computed if this number is correctly set by sheer mathematics. In

Section 2 of SI, it is demonstrated how the minimum number of bands is computed for GaN. This detail also illustrates the need for automation, since the input in this earlier step depends on an input that is chosen in a later step of the computational procedure, and *VASP* and *LOBSTER* runs are highly interdependent. This is also one of the most labor-intensive steps in the entire bonding analysis workflow and requires a lot of training for new users. Our workflow ensures that users always select the correct minimum number of bands according to the proposed basis. Currently, two basis sets (*Koga* and *pbeVASPfit2015*) can be used for the entire periodic table but only *pbeVASPfit2015* also includes additional orbitals relevant for the solid state (e.g., 2p in metallic Be^[17]). Some more trivial settings for the DFT run must also be chosen. For example, DFT runs must be performed without most of the symmetries to be compatible with *LOBSTER* (*VASP* input: *ISYM*=0) as only time-reversal symmetry is implemented in *LOBSTER* so far.

The next step is to test several atomic orbital bases for the projection of the wave function in *LOBSTER* (c). Some *LOBSTER* beginners trained in plane waves but lacking atomic-orbital knowledge may overlook the need to thoroughly test the basis for the projection. Here this is done systematically and all basis functions possible in *LOBSTER* are exploited. We start with a minimal basis consisting only of occupied valence orbitals in the atomic ground state of each atom. This is determined using the pseudopotential file of *VASP*. We then systematically add unoccupied orbitals for each element and test all possible combinations. This test is further constrained by the basis functions currently available in *LOBSTER*.

Particular attention has been paid to the handling of the large wavefunction files that are generated as they can become a burden on storage on typical supercomputers, especially for high-throughput studies. They can be automatically deleted. The most important outputs of the *VASP* and *LOBSTER* runs will be stored in a MongoDB database (d). For *LOBSTER*, this is the information from the *lobsterout* file. Additionally, other output files can be stored in the database as well.

In the last step (e), an automatic analysis of the *LOBSTER* calculations typically with the lowest absolute charge spilling is performed. This charge spilling indicates how much charge was lost in the projection from the plane-wave wave function to the atomic-orbital wave function and it is one of the most important quality criteria for the projection; in principle, the transformation is exact but small deviations may occur, just like in DFT. If used in the “cation-anion” mode, the automated analysis first detects cations and anions based on Mulliken charges calculated using *LOBSTER*^[40] and then identifies the strongest cation-anion bond within the crystal structure based on the integrated Crystal Orbital Hamiltonian populations (ICOHPs). For the test systems shown below, only bonds with a strength of at least 10% of the strongest cation-anion bond were considered. This cutoff can be adjusted. In case all bonds (not only cation-anion bonds) are considered in the analysis, the ICOHP cutoff will typically be set to 10% of the strongest bond.

To implement this workflow and simplify the bonding analysis, we extended and created several Python packages.

Two new workflows are available in the *atomate*^[28] workflow library (see Figure S1 in the supporting information). The first one is the workflow above. The second one relies only on a predefined basis. This workflow should only be used if the quality of the predefined basis has already been tested. The concrete DFT workflows are implemented in *atomate* using the workflow manager *fireworks*, the error handler *custodian*, and the *pymatgen* library for material analysis. With *atomate* and the other underlying Python packages, it is possible to perform complex calculations mostly with *VASP* (e.g., band structure calculations) on high-performance computers. All input files are generated automatically.

To handle the input and output files, we have implemented classes for almost every input and output file of *LOBSTER* in *pymatgen*. Plotting classes for COHPs, COOPs, COBIs are now

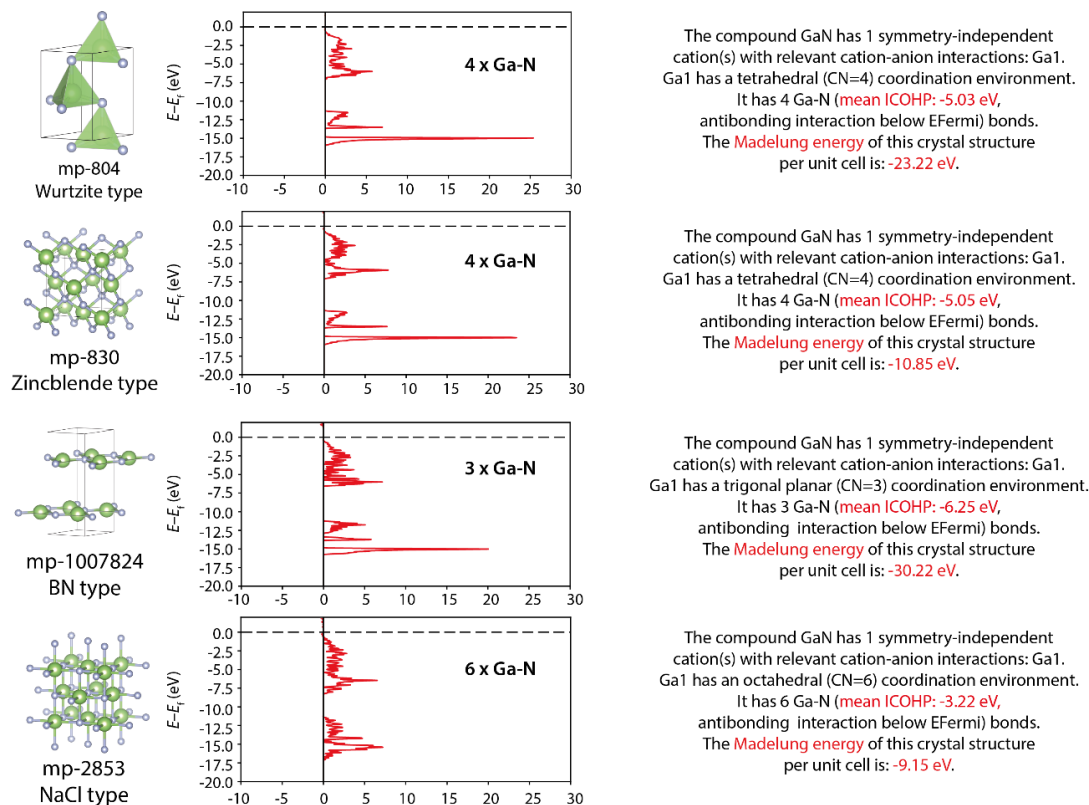


Figure 3. Illustration of the automatic output for four different GaN phases. The structures of the phases are depicted next to the automatic COHP plots and the text description. All Ga-N COHPs are nearly fully bonding except for small traces of anti-bonding interactions close to E_{Fermi}. The algorithm correctly recognizes all coordination environments. In addition, the Madelung energies per unit cell are given. Table 1 shows the Madelung energies per formula unit of GaN. For the COHP plots, a Gaussian smearing was used as the VASP version determined the band gap at the bottom of the band gap with the tetrahedron smearing but not the Gaussian smearing. We expect that this will be fixed again in future versions of VASP. Furthermore, we manually adapted the energy range here and restarted the Lobster run from the projection step. Very often, only states close to the Fermi energy E_{Fermi} are relevant for our bonding analysis and we therefore have a different standard setting for the energy range in our automatic workflow.

available and there is an interface to the existing plotting classes for densities of states and *fatbands* in *pymatgen*. New classes have been developed to allow for the automatic analysis of local environments based on orbital-based bonding analysis. They connect to available tools for determining and analyzing atomic environments in *pymatgen* (*ChemEnv*,^[36] *localenv*,^[37]). The bonds between atoms are determined based on the magnitude of the ICOHP values. Comparisons with ideal polyhedra are then made based on, for example, *ChemEnv* and its implementation of continuous symmetry measures describing a distance to a shape^[36]. In combination with *localenv*,^[37] structure graph objects and further information about the neighboring atoms can be retrieved.

In order to perform hundreds to thousands of calculations in parallel, error handling and validation are critical for high-throughput studies. We have therefore also added new implementations for validators in *custodian*, which must be successfully passed. For example, this system checks whether the main LOBSTER output file (*lobsterout*) has been created and whether enough bands have been selected in the DFT run. Also, a validator was added to check if the charge spilling is reasonably low (<5%). We use the latter validator only when exactly one set of basis functions is used for the LOBSTER calculations. If different basis functions are tested in the workflow, this check is not performed. Also, a job class has been introduced to run LOBSTER jobs with *custodian*.

Two tutorials showing how to use these classes and functions are available online.^[41,42]

Parts of this implementation have been used in other publications, e.g., to understand defects and their stability in photovoltaic materials or to benchmark a new implementation in the LOBSTER program.^[16,43] In the latter study,^[16] we identified the computational settings used in this study that lead to very well converged LOBSTER results. These and other details on the DFT

and LOBSTER settings can be found in the methods section of this paper.

The tools for automatically analyzing LOBSTER outputs are implemented in the *lobsterpy* package. An overview of the package can be found in Figure S1 in the supporting information. The *Analysis* class takes care of the automatic analysis of LOBSTER output and is the starting point for another class called *Description*, which will provide automated text and tools for automatic plotting. Detailed examples of COHP data analysis are attached to the code repository corresponding to this paper. By default, only cation-anion bonds are included in the analysis. Especially for large compounds such as $\text{Yb}_{14}\text{Mn}_1\text{Sb}_{11}$, the automatic output otherwise provides too much information for the user, since weak cation-cation interactions are also present. An analysis of all bonds can be enforced since there are compounds where cation-cation interactions are responsible for the stability of a material and the particular structure (e.g., $\text{Ge}_4\text{Se}_3\text{Te}^{[44]}$). For very large coordination environments (coordination number >13), only the coordination number is determined, but not the environment, which is based on a limitation of *ChemEnv*. *ChemEnv* only has reference polyhedra for coordination numbers ≤ 13 .

Lobsterpy also offers a command-line interface that connects and extends the plotting tools available in *pymatgen*. The user can provide customized style sheets for the plots. This is also possible for automatic COHP plots. The automatic description can also be printed to the screen.

Automatic Test of Best Basis and Projections

The *Lobster* workflow in *atomate* was run for several test systems (NaCl, GaN, MTaO_2N ($M = \text{Ba}, \text{Ca}, \text{Sr}$), $\text{Yb}_{14}\text{Sb}_1\text{Mn}_{11}$). First, we will discuss the resulting projections.

For most systems, only one basis is available (see Table 1) since their constituting ions are in noble gas configuration (full octet).

For $\text{Yb}_{14}\text{Sb}_1\text{Mn}_{11}$, we will test the influence of the 5d orbital for Yb. This orbital would be unoccupied in a Yb atom according to the corresponding pseudopotential file. However, its use reduces the absolute charge spilling by more than 1%. Therefore, we will use the larger basis in the following. For all systems tested here, the charge spilling is less than 2%, indicating a very good projection. Note that a subsequent Löwdin orthogonalization of the local basis automatically carried out in *LOBSTER* assures that the entire Hilbert space is recovered, so no electron density gets lost.

Table 1. Possible bases for our test systems.

Composition	Basis 1	Basis 2
GaN	Ga (4s, 4p, 3d), N (2s 2p)	-
NaCl	Na (3s 3p), Cl (3s 3p)	
CaTaO ₂ N	Ca (3s 3p 4s), Ta (6s, 5p, 5d), O (2s 2p), N (2s 2p)	
SrTaO ₂ N	Sr (4s 4p 5s), Ta (6s, 5p, 5d), O (2s 2p), N (2s 2p)	
BaTaO ₂ N	Ba (5s 5p 6s), Ta (6s, 5p, 5d), O (2s 2p), N (2s 2p)	
Yb ₁₄ Mn ₁ Sb ₁₁	Yb (6s 5p), Mn (3p 3d 4s), Sb (Ss, 5p)	Yb (6s 5p 5d), Mn (3p 3d 4s), Sb (Ss, 5p)

We assume that this automatic checking of the basis will indeed ensure that users test the projections, rather than simply choosing a basis that gives reasonable spilling under a certain limit without further checking. We also hope that high-throughput studies will allow us to develop further criteria for the quality of a basis besides the absolute charge spilling. We now discuss the results of the automatic analysis of the output files.

Two illustrative examples: NaCl and GaN

We start with the analysis of two rather simple binary systems: GaN and NaCl. We usually describe chemical bonds in crystalline materials in terms of their metallic, covalent, and ionic contributions. The Van Arkel-Ketelaar triangle represents the different contributions for materials and allows us to map materials according to the nature of their bonds.^[1,45–47] COHP provides a way to quantify covalent contributions. Madelung energies are commonly used to quantify ionic contributions. *LOBSTER* calculates these based on quantum-chemical Mulliken or Löwdin charges rather than just based on formal oxidation states. In the following, we will show that such information about the bonding situation in crystals can be easily extracted using our automation. We will also use this information on covalency and ionicity to discuss the stability of the different GaN and NaCl phases.

We start with the four most stable GaN phases available in the Materials Project database^[29] (see Table 2 and Figure 3). These four phases have been investigated by experimental and theoretical studies.^[48] GaN is a well-known semiconducting compound for optoelectronic applications and is therefore expected to show significant covalent interactions.^[49] Based on COHPs, our code correctly recognizes all coordination environments. This is consistent with geometric evaluations based on *ChemEnv* and its standard settings. *ChemEnv* compares coordination environments to ideal shapes.^[36]

We now proceed to discuss the stability of the compounds and what is effectively causing this. GaN in the wurtzite and zincblende structures is predicted to be more stable than GaN in NaCl structure type based on the DFT total energies, which is consistent with our expectations. GaN in the BN structure type lies between these extremes. We expected that the stabilization of the

zincblende and wurtzite structure type compared to the NaCl structure type is due to a higher covalency as GaN is semiconducting.^[1] Indeed, we find that the more stable compounds crystallizing in the wurtzite and zincblende types have more covalent interactions than GaN crystallizing in the NaCl structure type. This is indicated by the sum of Ga-N ICOHPs per Ga atom (Table 2 and Figure 3) and it is also connected to the Ga-N distances in the compounds. Ga in wurtzite and zincblende structure type show much stronger COHPs, especially for lower-lying states (mostly with contributions of N (2s)). This is, of course, directly correlated to the smaller distances in these compounds as, for example, the more localized 2s orbitals (around -15 eV below the Fermi energy E_{Fermi}) cannot overlap that easily with the Ga orbitals anymore. The dispersion of the bands seems to be very similar in all four compounds, which is a bit surprising. The total energy of GaN in the BN structure, however, cannot be explained that easily based on covalent contributions alone. The calculated Madelung energies suggest that there may be electrostatic stabilization that could explain this energetic ordering. GaN in the BN structure exhibits the most negative Madelung energy per formula unit based on Mulliken charges of all four compounds. The same tendency can be found based on Löwdin charges.

Table 2. Comparison of total DFT energies per formula unit, Madelung energies calculated based on Mulliken charges, and sum of Ga-N ICOHPs. The phases are referenced by the material project IDs (mp-...). The lowest, i.e., most stabilizing values are highlighted in bold.

Com- positi on	Phas e (MP- ID)	Total energies/ formula unit (eV)	Sum of Ga-N ICOHPs per Ga atom (Covalency) (eV)	Shortest Ga-N distance (Å)	Madelung energy per formula unit (Ionicity) (eV)
GaN	Wurt zite (mp- 804)	-12.16	-20.12	1.976	-11.61
GaN	Zinc- blend e (mp- 830)	-12.15	-20.20	1.970	-10.85
GaN	BN (mp- 1007 824)	-11.46	-18.75	1.854	-15.11
GaN	NaCl (mp- 2853)	-11.21	-19.23	2.137	-9.16

As shown in this simple example, correlations of DFT total energies with covalent bond strengths or Madelung energies can now be easily tested on a much larger scale to potentially understand phase transitions based on bonding properties and electrostatics.

We now turn to the analysis of simple but very ionic systems - two NaCl phases. Here we do not expect covalency to be the driving force for the stability of these compounds. Nevertheless, it is possible to identify the correct coordination environments in these phases based on the bond energy given by COHP (Fig. 4).

For the binary ground-state structure of NaCl (mp-22862, Figure 4 a), the code correctly recognizes the octahedral coordination environment for Na^+ and summarizes the mean values for the integrated COHPs. For the high-pressure phase of NaCl (mp-22851, Figure 4 b), the cubic environment is correctly recognized.

The differences from the ground-state structure of NaCl are visible in the sum of ICOHP values (*ICOHP_sum*) and the mean value of ICOHPs per bond (*ICOHP_mean*). The ground-state structure of NaCl has a more negative value (i.e., it has stronger covalent bonds). However, this value should be taken with caution because NaCl is a comparatively ionic compound and exhibits very small ICOHP values in general. In this case, the electrostatics and calculated Madelung energies should lead to a much better understanding of the compound stability: Indeed, the ground-state structure of NaCl also shows a more negative Madelung energy than the high-pressure phase by more than 0.3 eV per cell.

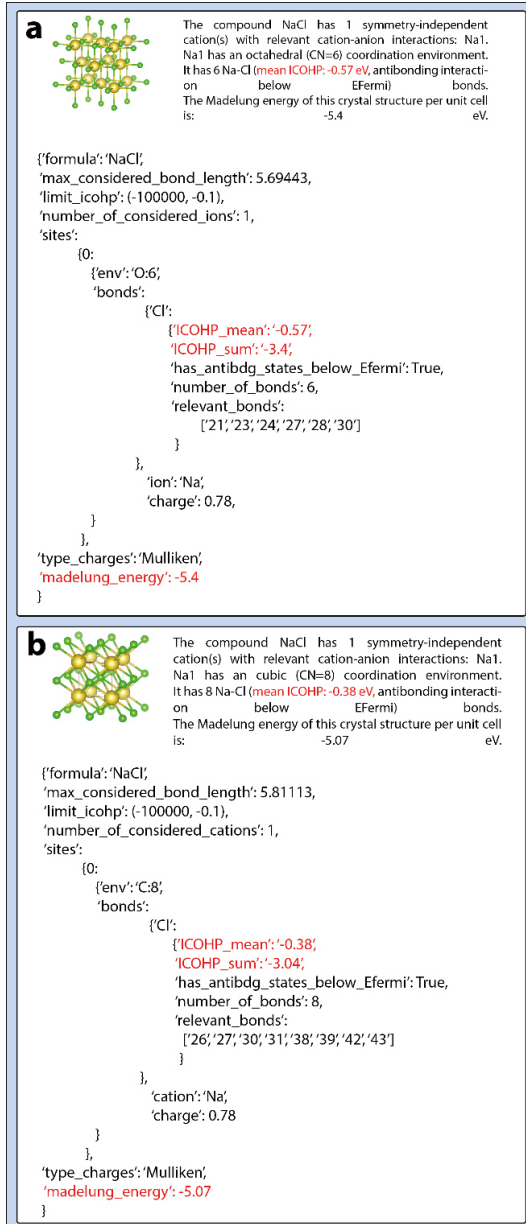


Figure 4. Automatic analysis and structure of NaCl in the ground-state structure (a) and the high-pressure CsCl-type structure (b). Additionally, the automatic output dictionaries for both phases are shown. Note that the automatic output provides a counting scheme for the atoms that starts at 1 (similar to *LOBSTER* outputs), but that the digits for sites within the dictionaries start at 0 (standard in Python).

Two advanced examples: oxynitrides and Yb₁₄Mg₁Sb₁₁

Now, two more advanced examples will be discussed.

We first focus on the oxynitride compounds CaTaO₂N, SrTaO₂N, and BaTaO₂N, which crystallize in a perovskite-related structure. The anion order in these compounds have been under some

debate in the past, and possible driving forces for certain anion orders have been discussed.^[50–52] We now automatically analyze the importance of certain bonds for the anion order in these compounds (Figure 5).

For this purpose, we vary the anion order of the systems in a supercell with up to 2 formula units. This results in seven different structural models per compound with different anion orders. The energetically least and most favorable models are shown in Figure 5 a and b. Based on our automated bonding analysis, we can then easily correlate the total energy of the systems with the strongest covalent interaction (Ta-N bond) in the systems (Figure 5 c). We also see an anti-correlation with the bond energies of the Ta-O bonds because the more covalent Ta-N bonds strengthen at the expense of the Ta-O bonds. The covalency of the Ta-N interactions is therefore crucial for the anionic ordering of these oxynitrides.

The Ca/Sr/Ba-O bonds are too weak to be detected with our automated analysis and the corresponding relative ICOHP cutoff (10% of the absolute largest ICOHP). The ICOHPs help to focus on the most important covalent interactions in our system, which could also be an advantage for determining coordination environments compared to purely geometric determinations of the latter.

Thanks to our automation tools, the whole bonding analysis process can be performed with only a few lines of Python code, as shown in the repository accompanying this paper. This correlation between the strongest covalent interaction and the total energy of the system has already been observed for the oxynitride CaTaO₂N^[50,51]. This correlation has now been confirmed for SrTaO₂N and BaTaO₂N, which was to be expected given the chemical similarity of the compounds. One now arrives at such an analysis almost fully automatically.

As a final example, we show the analysis of a very large and complex structure, namely that of the well-known thermoelectric material Yb₁₄Mn₁Sb₁₁.^[53] It has 104 atoms in the unit cell, and such a bonding analysis by hand would require a lot of manual work. Up to a distance limit of 6 Å (maximum considered distance in our current implementation of the workflow), more than 1400 bonds are found in this structure. The band structure of the compound in the nonmagnetic state has been recently studied.^[54] Similarly, an analysis of the band structure of Yb₁₄Mg₁Sb₁₁ crystallizing in the same type of structure, has been carried out.^[55] The magnetism in Yb₁₄Mn₁Sb₁₁ has also already been studied both experimentally and theoretically.^[56] From previous studies, we expect an influence of the Mn-Sb bonds on the magnetism.

To investigate the influence of chemical bonding on magnetism in this compound, we decided to compare COHPs from non-magnetic and ferromagnetic runs. This type of investigation was inspired by a study by Landrum and Dronskowski on magnetism and driving forces for itinerant magnetism based on chemical bonding.^[15] We then performed two static *LOBSTER* runs that differ in the magnetic model after structural optimizations in the ferromagnetic setting. We introduce magnetic moments on Mn only.

Similar to Perez et al.^[54] in the nonmagnetic state, Yb, Mn, and Sb states play an important role in the valence band of Yb₁₄Mn₁Sb₁₁. We find that both Yb-Sb and Mn-Sb bonds are relevant cation-anion bonds in the structures, the Yb-Sb bonds being on average stronger than the Mn-Sb bonds. The bonds in the "Sb₃⁷⁻" are not considered in the standard automatic analysis of *LOBSTER* output files. By default, this analysis focuses only on cation-anion bonds to obtain a readable result even for very large structures such as Yb₁₄Mn₁Sb₁₁. However, it is possible to enforce their analysis. The automatic text output and the structure for the ferromagnetic setting are shown in Fig. 6.

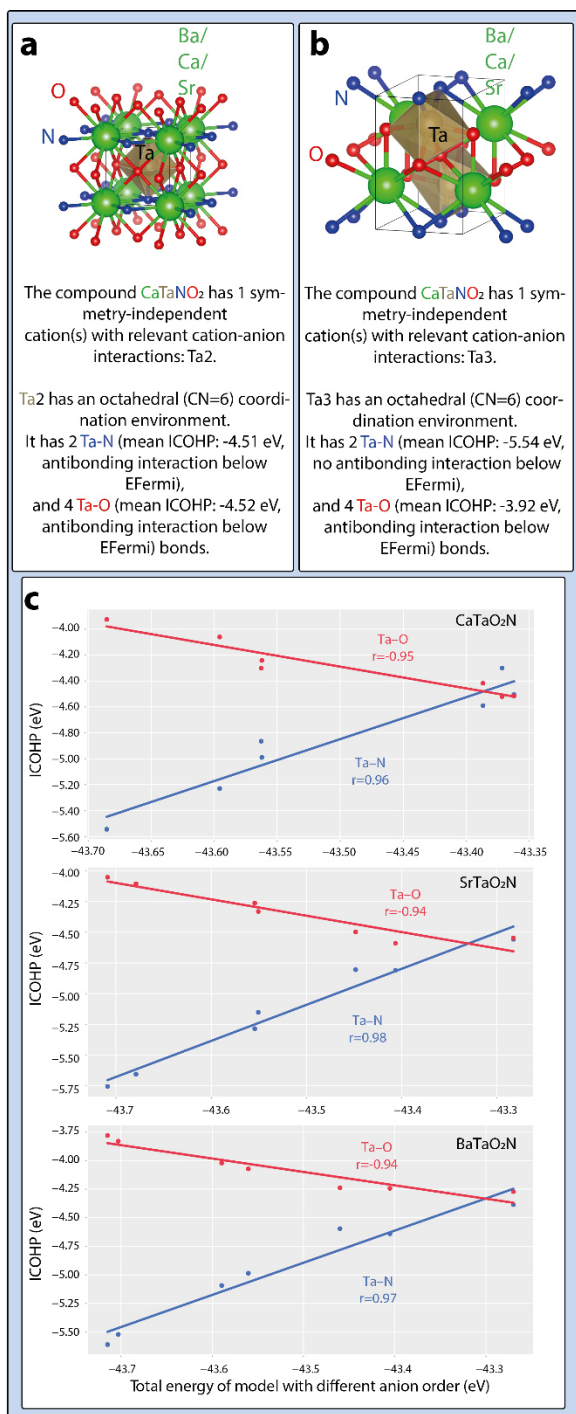


Figure 5. a) Structure as depicted with VESTA^[57] and automatic analysis of the conventional cell for CaTaO_2N . The code is able to automatically determine relevant bonds, coordination environments and average bond strengths (average ICOHPs); note that more negative ICOHPs indicate stronger bonds. In this case, only the Ta-O and Ta-N bonds were detected. Potential Ca/Sr/Ba-O and Ca/Sr/Ba-N bonds are much less covalent and play a less important direct role for the material. This information can then be easily correlated with other material properties (e.g., total energies) to understand these properties based on bond strengths. b) Structure of the most stable Ca/Sr/Br TaNO_2 model including up to 2 formula units. c) A correlation plot of the most important ICOHPs for this compound and the total energy of the systems with different anion orders is shown. The correlation coefficient r for the strongest covalent interaction (Ta-N) is always higher than 0.95, indicating a significant correlation.

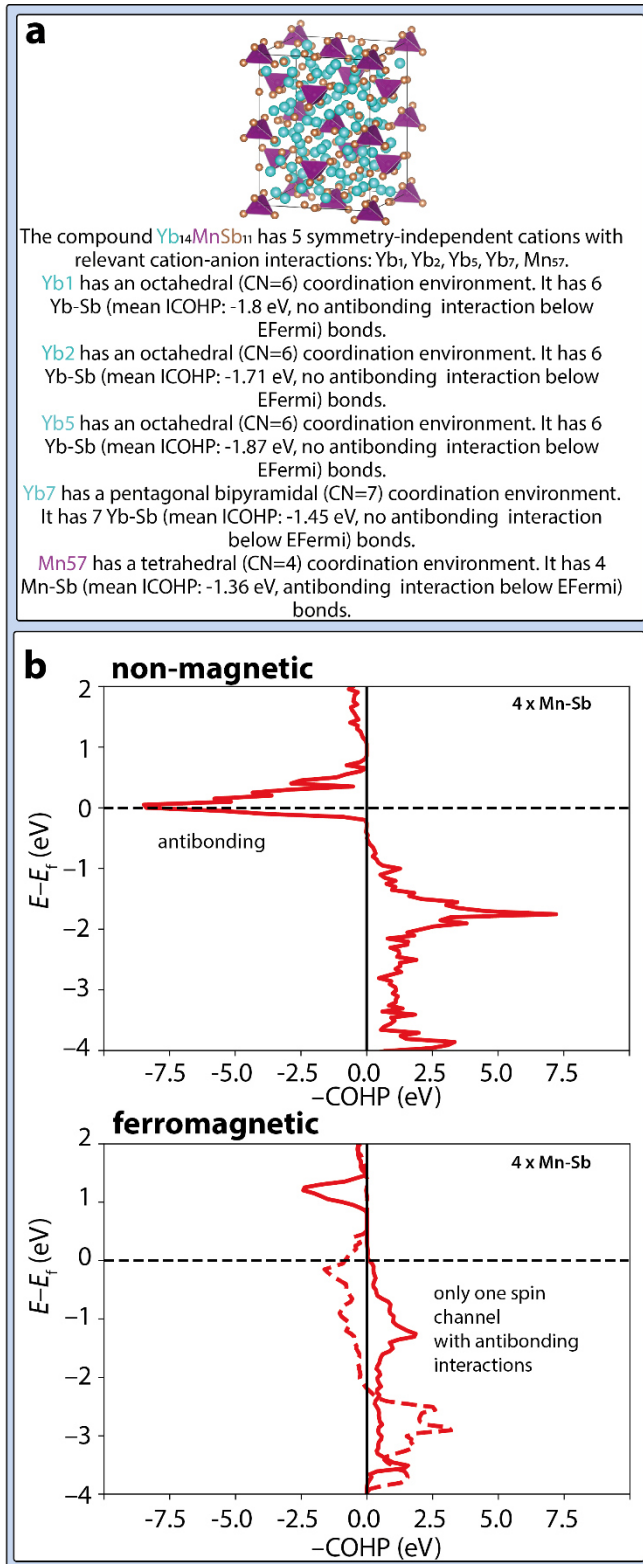


Figure 6. a) Structure of $\text{Yb}_{14}\text{MnSb}_{11}$ above the automatic text analysis of the bonding situation in the compound. Yb shows octahedral and pentagonal bipyramidal environments and Mn shows a tetrahedral environment. Ferromagnetism was considered in these calculations. b) Here, the structure was optimized in a ferromagnetic environment and the COHPs were calculated in a nonmagnetic and in a ferromagnetic setting.

In the ferromagnetic setting, the octahedral environments are discovered for three of four crystallographically nonequivalent Yb atoms. The other is determined to be a pentagonal bipyramid. In contrast, the purely geometric tool *ChemEnv*,^[36] based on Voronoi analysis with distance and solid angle cutoffs and determination of continuous symmetry measures, determines octahedral environments for all four. *SimplestChemEnvStrategy* and a standard relative solid angle cutoff of 0.3 were used. This lower cutoff is relative to the largest solid angle in the Voronoi construction. One of the identified octahedral environments is, however, highly distorted, as indicated by a continuous symmetry measure greater than 7. This is the environment that our orbital-based analysis determines to be a pentagonal bipyramid. If we change the angle cutoff to 0.2, the geometric algorithm also arrives at the pentagonal bipyramidal environment. The determination of the tetrahedral coordination environment for Mn agrees well with the geometric-based determination. Overall, the agreement between the geometric and orbital-based determination of the coordination environments is very good.

Upon analyzing the COHP diagrams with and without spin polarization for Yb₁₄Mn₁Sb₁₁ (Figure 6b), we indeed find an influence of the Mn-Sb bonds on the magnetism as it has been seen before^[56]: strong antibonding Mn-Sb interactions are discovered below the Fermi level for the calculation without spin polarization. When spin polarization is turned on, the antibonding interactions in one spin channel disappear completely and this spin channel now has fully bonding Mn-Sb bonds. In the other spin channel, small antibonding Mn-Sb interactions remain around the Fermi level. This finding agrees well with the antibonding fingerprint for other ferromagnetic compounds such as the transition metals, FeNi₃, FePd₃, and MnSb, demonstrated by Landrum and Dronskowski.^[15] In contrast to the results of Landrum and Dronskowski, we do not detect stronger Mn-Sb bonds for the ferromagnetic setting, however. These bonds are slightly weaker than in the non-magnetic setting. In our gedankenexperiment, however, we find an "oxidation" of Mn when spin polarization is considered. The Mulliken charges of Mn change dramatically from 0.69 (nonmagnetic) to 0.91 (ferromagnetic). This makes the Madelung energy much more negative with a change from -104.21 to -118.82 eV. This could indicate additional electrostatic stabilization for this compound in the ferromagnetic setting, rather than the typical covalent stabilization observed for other compounds such as α -Fe. These differences to the situation in α -Fe with homopolar Fe-Fe bonds are not surprising as we now observe this change in a heteropolar bond where Sb is more electronegative than Mn and thereby captures some of the electrons of Mn when spin-polarization is switched on.

Conclusion

This work simplifies bonding analysis by providing workflow tools and tools for automated analysis. Prior to this study, chemical bonding analysis based on orbitals was a task with many manual steps and therefore error prone. We have demonstrated our automation here using several test systems, including the study of the anionic order of several oxynitrides and the complex crystal structure of Yb₁₄Mn₁Sb₁₁. Thanks to automation, these calculations can now be performed in a high-throughput manner. In addition, we have shown how the tools allow correlating important bonding properties with other material properties. We expect that these tools will facilitate the search for new descriptors for machine learning of material properties. Bader charges^[58] or other information based on electronic structure theory have already helped in such studies. In addition, we hope that the tools will help to provide further chemical understanding of materials. The study also illustrates how much work is needed to automate just one task in computational chemistry. Future work will aim to include orbital-by-orbital analysis of bonds, various indicators of covalency (ICOOP, ICOBI), and k-dependent covalency in our

automated analysis, which could be particularly useful for compounds with very steep bands. In addition, pre-convergence steps with a smaller number of bands could be included in our *LOBSTER* workflow to speed up calculations and convergence of calculations, which is currently a problem for compounds requiring a very large number of bands (such as Yb₁₄Mn₁Sb₁₁). Interfaces to other DFT codes could be developed.

For all examples shown, we provide the analysis scripts in our Github repository (<https://github.com/JaGeo/LobsterAutomation>, link to zenodo.org will be provided in the final publication).

Methods

For all materials considered in this study, DFT calculations based on the PAW method^[59,60] and the PBE approximation of the exchange correlation functional^[61] were performed. For optimization, we relied on many of the default settings of the *atomate* package imported via *pymatgen* from the MPRelaxSet. However, we chose more precise optimization settings. We use an energy difference criterion for the convergence of the electronic structure of 10⁻⁶ eV, and the structure is considered relaxed if the energy difference between two consecutive steps is below a setting of 10⁻⁵ eV. We relied on the standard cutoff for the plane-wave energy of 520 eV. In contrast to the MPRelaxSet in *pymatgen*, we used the VASP pseudopotentials with version number 5.4, with the additional difference that s-electrons are included in the valence of W. We used 6000 k-points per reciprocal atom to ensure that the energies for the defect phases of the oxynitrides converged well. We also chose a Gaussian smearing for the optimization.

When calculating the wave function, the energy is converged until it is less than 10⁻⁶ eV. This stricter criterion, as opposed to the default setting in the materials project, was necessary to ensure that the wave functions were of sufficient quality. Otherwise, we would sometimes have obtained different charge spillings for the two spin channels of actually not magnetic compounds, which is contrary to expectation and indicates problems with convergence of the wave function. In addition, we considered time-reversal symmetry in the VASP calculation and used 6000 k points per reciprocal atom to ensure convergence of the *LOBSTER* results. The number of bands is adjusted according to the largest basis available in *LOBSTER* for the compound in question. The other parameters are the established defaults for the Materials Project PBE calculations as implemented in MPRelaxSet. The entire implementation of a generator for the input can be found in *pymatgen* and in *pymatgen.io.vasp.sets.LobsterSet*. These settings were previously determined to produce well-matched results when testing a new implementation in *LOBSTER* for several hundred *LOBSTER* calculations.^[16]

The *LOBSTER* run itself was then based on the *pbeVASPfit2015* basis and fits the atomic basis to the selected POTCAR files. All standard outputs of a *LOBSTER* run (i.e., information on Crystal Orbital Hamilton populations, Crystal Orbital Overlap Populations, Crystal Orbital Bond Indices, projected density of states, Mulliken charges, Löwdin charges, etc.) were calculated.

To generate the different anion configurations of the perovskite structures, we again relied on *pymatgen*, *enumlib*^[62] and also on symmetry identification based on spglib^[63].

The following program versions have been used: *pymatgen* 2022.2.1, *atomate* 1.0.3, *LobsterPy* 0.1.0, *Lobster* 4.1.0, VASP 6.2.1.

Data and code availability:

All data is available on zenodo.org. All code is available under an open-source license. We also have a github repository to reproduce the whole publication (<https://github.com/JaGeo/LobsterAutomation>, link to zenodo.org will be provided in the final publication).

Acknowledgements

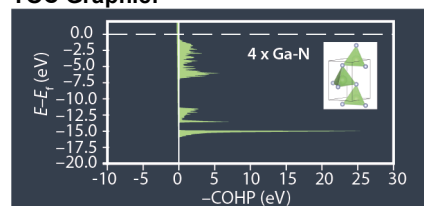
J.G. acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 837910. J.G. and A.N. gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre (www.lrz.de) (project *pn73da*). J. G. thanks Kirill Yuseenko for feedback on the manuscript in BAM's MatChIngCamp. Furthermore, we thank David Waroquiers for help with the *ChemEnv* code, Matthew Horton and Alex Ganose for feedback and discussions with regard to *pymatgen*'s and *atomate*'s extension.

Keywords: bonding analysis • automation • density-functional theory • oxynitrides • machine learning

- [1] J. K. Burdett, J. K. Burdett, *Chemical Bonding in Solids*, Oxford University Press, **1995**.
- [2] R. Hoffmann, *Angew. Chem. Int. Ed.* **1987**, *26*, 846–878.
- [3] R. F. W. Bader, T. T. Nguyen-Dang, in *Advances in Quantum Chemistry*, Elsevier, **1981**, pp. 63–124.
- [4] O. Andriuc, M. Siron, J. H. Montoya, M. Horton, K. A. Persson, *J. Chem. Inf. Model.* **2021**, *61*, 3908.
- [5] L. A. Burton, F. Ricci, W. Chen, G.-M. Rignanese, G. Hautier, *Chemistry of Materials* **2018**, *30*, 7521–7526.
- [6] 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, O. Levy, *Computational Materials Science* **2012**, *58*, 227–235.
- [7] R. Dronskowski, P. E. Blöchl, *J. Phys. Chem.* **1993**, *97*, 8617–8624.
- [8] R. Hoffmann, *Solids and Surfaces: A Chemist's View of Bonding in Extended Structures*, **1988**.
- [9] R. Dronskowski, *Computational Chemistry of Solid State Materials: A Guide for Materials Scientists, Chemists, Physicists and Others*, Wiley-VCH, Weinheim, **2005**.
- [10] M.-H. Whangbo, R. Hoffmann, *J. Am. Chem. Soc.* **1978**, *100*, 6093.
- [11] P. C. Müller, C. Ertural, J. Hempelmann, R. Dronskowski, *J. Phys. Chem. C* **2021**, *125*, 7959–7970.
- [12] M. Esser, S. Maintz, R. Dronskowski, *J. Comput. Chem.* **2017**, *38*, 620–628.
- [13] P. M. Konze, R. Dronskowski, V. L. Deringer, *Phys. Physica Status Solidi* **2019**, *13*, 1800579.
- [14] S. Steinberg, R. Dronskowski, *Crystals* **2018**, *8*, 225.
- [15] G. A. Landrum, R. Dronskowski, *Angew. Chem. Int. Ed.* **2000**, *39*, 1560–1585.
- [16] R. Nelson, C. Ertural, J. George, V. L. Deringer, G. Hautier, R. Dronskowski, *J. Comput. Chem.* **2020**, *41*, 1931–1940.
- [17] S. Maintz, V. L. Deringer, A. L. Tchougréeff, R. Dronskowski, *J. Comput. Chem.* **2016**, *37*, 1030.
- [18] S. Maintz, V. L. Deringer, A. L. Tchougréeff, R. Dronskowski, *J. Comput. Chem.* **2013**, *34*, 2557–2567.
- [19] V. L. Deringer, A. L. Tchougréeff, R. Dronskowski, *J. Phys. Chem. A* **2011**, *115*, 5461–5466.
- [20] C. Ertural, S. Steinberg, R. Dronskowski, *RSC Advances* **2019**, *9*, 29821–29830.
- [21] G. Kresse, J. Hafner, *Phys. Rev. B* **1993**, *47*, 558.
- [22] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169.
- [23] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- [24] X. Gonze, B. Amadon, G. Antonius, F. Arnardi, L. Baguet, J.-M. Beuken, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, N. Brouwer, F. Bruneval, G. Brunin, T. Cavignac, J.-B. Charraud, W. Chen, M. Côté, S. Cottenier, J. Denier, G. Geneste, P. Ghosez, M. Giantomassi, Y. Gillet, O. Gingras, D. R. Hamann, G. Hautier, X. He, N. Helbig, N. Holzwarth, Y. Jia, F. Jollet, W. Lafargue-Dit-Hauret, K. Lejaeghere, M. A. L. Marques, A. Martin, C. Martins, H. P. C. Miranda, F. Naccarato, K. Persson, G. Petretto, V. Planes, Y. Pouillon, S. Prokhorenko, F. Ricci, G.-M. Rignanese, A. H. Romero, M. M. Schmitt, M. Torrent, M. J. van Setten, B. Van Troeye, M. J. Verstraete, G. Zerah, J. W. Zwanziger, *Comput. Phys. Commun.* **2020**, *248*, 107042.
- [25] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, *J. Phys. Condens. Matter* **2017**, *29*, 465901.
- [26] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, Z. Ulissi, *ACS Catal.* **2021**, *11*, 6059–6072.
- [27] J. George, *Trends Chem.* **2021**, *3*, 697–699.
- [28] K. Mathew, J. H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I. Chu, T. Smidt, B. Bocklund, M. Horton, J. Dagdelen, B. Wood, Z.-K. Liu, J. Neaton, S. P. Ong, K. Persson, A. Jain, *Comput. Mater. Sci.* **2017**, *139*, 140–152.
- [29] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, *1*, 011002.
- [30] C. W. Andersen, R. Armiento, E. Blokhin, G. J. Conduit, S. Dwarakanath, M. L. Evans, Á. Fekete, A. Gopakumar, S. Gražulis, A. Merkys, F. Mohamed, C. Oses, G. Pizzi, G.-M. Rignanese, M. Scheidgen, L. Talirz, C. Toher, D. Winston, R. Aversa, K. Choudhary, P. Colinet, S. Curtarolo, D. Di Stefano, C. Draxl, S. Er, M. Esters, M. Fornari, M. Giantomassi, M. Govoni, G. Hautier, V. Hegde, M. K. Horton, P. Huck, G. Huhs, J. Hummelshøj, A. Karirya, B. Kozinsky, S. Kumbhar, M. Liu, N. Marzari, A. J. Morris, A. A. Mostofi, K. A. Persson, G. Petretto, T. Purcell, F. Ricci, F. Rose, M. Scheffler, D. Speckhard, M. Uhrin, A. Vaitkus, P. Villars, D. Waroquiers, C. Wolverton, M. Wu, X. Yang, *Sci Data* **2021**, *8*, 217.
- [31] D. Broberg, B. Medasani, N. E. R. Zimmermann, G. Yu, A. Canning, M. Haranczyk, M. Asta, G. Hautier, *Comput. Phys. Commun.* **2018**, *226*, 165–179.
- [32] M. K. Horton, J. H. Montoya, M. Liu, K. A. Persson, *npj Comput Mater* **2019**, *5*, 64.
- [33] G. Hautier, *Comput. Mater. Sci.* **2019**, *163*, 108–116.
- [34] J. George, G. Hautier, *Trends in Chemistry* **2021**, *3*, 86–95.
- [35] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547–555.
- [36] D. Waroquiers, J. George, M. Horton, S. Schenk, K. A. Persson, G.-M. Rignanese, X. Gonze, G. Hautier, *Acta Cryst B* **2020**, *76*, 683–695.
- [37] H. Pan, A. M. Ganose, M. Horton, M. Aykol, K. A. Persson, N. E. R. Zimmermann, A. Jain, *Inorg. Chem.* **2021**, *60*, 1590–1603.
- [38] A. R. Ferreira, *Phys. Rev. Materials* **2020**, *4*, 113603.
- [39] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* **2013**, *68*, 314.
- [40] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- [41] J. George, A. Naik, “Tutorial on Automated Bonding Analysis,” can be found under https://jageo.github.io/sites/Tutorial_LobsterAtomate-Update.html (Archived Version: https://web.archive.org/web/20220222081328/https://jageo.github.io/sites/Tutorial_LobsterAtomate-Update.html), **2022**.
- [42] J. George, “Tutorial on Lobster classes in pymatgen,” can be found under <http://web.archive.org/web/20220222081116/http://matgenb.materialsvirtuallab.org/2019/01/11/How-to-plot-and-evaluate-output-files-from-Lobster.html> (Archived version: <http://web.archive.org/web/20220222081116/http://matgenb.materialsvirtuallab.org/2019/01/11/How-to-plot-and-evaluate-output-files-from-Lobster.html>), **2022**.

- [43] D. Dahliah, G. Brunin, J. George, V.-A. Ha, G.-M. Rignanese, G. Hautier, *Energy Environ. Sci.* **2021**, *14*, 5057–5073.
- [44] M. Küpers, P. M. Konze, S. Maintz, S. Steinberg, A. M. Mio, O. Cojocaru-Mirédin, M. Zhu, M. Müller, M. Luysberg, J. Mayer, M. Wuttig, R. Dronskowski, *Angew. Chem. Int. Ed.* **2017**, *56*, 10204–10208.
- [45] L. C. Allen, J. F. Capitani, G. A. Kolks, G. D. Sproul, *J. Mol. Struct.* **1993**, *300*, 647–655.
- [46] A. E. van Arkel, *Molecules and Crystals in Inorganic Chemistry*, Butterworth Scientific Publications, **1949**.
- [47] J. A. A. Ketelaar, *Chemical Constitution. An Introduction to the Theory of the Chemical Bond*, **1958**.
- [48] S. Strite, H. Morkoç, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena* **1992**, *10*, 1237–1266.
- [49] “The Nobel Prize in Physics 2014,” can be found under <https://www.nobelprize.org/prizes/physics/2014/prize-announcement/>, **2022**.
- [50] H. Wolff, R. Dronskowski, *J. Comput. Chem.* **2008**, *29*, 2260–2267.
- [51] Z. Ma, A. Jaworski, J. George, A. Rokicinska, T. Thersleff, T. M. Budnyak, G. Hautier, A. J. Pell, R. Dronskowski, P. Kuśtrowski, A. Slabon, *J. Phys. Chem. C* **2020**, *124*, 152–160.
- [52] H.-C. Wang, J. Schmidt, S. Botti, M. A. L. Marques, *J. Mater. Chem. A* **2021**, *9*, 8501–8513.
- [53] S. R. Brown, S. M. Kauzlarich, F. Gascoin, G. J. Snyder, *Chem. Mater.* **2006**, *18*, 1873–1877.
- [54] C. J. Perez, M. Wood, F. Ricci, G. Yu, T. Vo, S. K. Bux, G. Hautier, G.-M. Rignanese, G. J. Snyder, S. M. Kauzlarich, *Sci. Adv.* **2021**, *7*, eabe9439.
- [55] Y. Liu, M. Y. Toriyama, Z. Cai, M. Zhao, F. Liu, G. Jeffrey Snyder, *Appl. Phys. Lett.* **2021**, *119*, 213902.
- [56] A. He, E. L. K. Wille, L. M. Moreau, S. M. Thomas, J. M. Lawrence, E. D. Bauer, C. H. Booth, S. M. Kauzlarich, *Phys. Rev. Materials* **2020**, *4*, 114407.
- [57] K. Momma, F. Izumi, *J. Appl. Crystallogr.* **2011**, *44*, 1272–1276.
- [58] S. B. Torrisi, M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram, L. Hung, *npj Comput Mater* **2020**, *6*, 109.
- [59] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953–17979.
- [60] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758–1775.
- [61] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [62] G. L. W. Hart, L. J. Nelson, R. W. Forcade, *Computational Materials Science* **2012**, *59*, 101–107.
- [63] A. Togo, I. Tanaka, *arXiv:1808.01590 [cond-mat]* **2018**.

TOC Graphic:



The compound GaN has 1 symmetry-independent cation(s) with relevant cation-anion interactions: Ga1.

Ga1 has a tetrahedral (CN=4) coordination environment.

It has 4 Ga-N (mean ICOHP: -5.03 eV, antibonding interaction below E_{Fermi}) bonds.

Supporting Information for Automated Bonding Analysis with Crystal Orbital Hamilton Populations

Janine George,^{*,[a,b,c]} Guido Petretto,^[c] Aakash Naik,^[a,b] Marco Esters,^[d] Adam J. Jackson,^[e] Ryky Nelson,^[f]
Richard Dronskowski,^[f,g] Gian-Marco Rignanese,^[c] Geoffroy Hautier^{*,[c,h]}

- [a] Dr. Janine George, Aakash Naik
Federal Institute for Materials Research and Testing
Department Materials Chemistry
Unter den Eichen 87, 12205 Berlin, Germany.
E-mail: Janine.george@bam.de
- [b] Dr. Janine George, Aakash Naik
Friedrich Schiller University Jena,
Institute of Condensed Matter Theory and Solid-State Optics
Max-Wien-Platz 1, 07743 Jena, Germany
- [c] Dr. Janine George (former address), Dr. Guido Petretto (former address), Prof. Dr. Gian-Marco Rignanese, Prof. Dr. Geoffroy Hautier (former address)
Université catholique de Louvain,
Institute of Condensed Matter and Nanosciences,
Chemin des Étoiles 8, 1348 Louvain-la-Neuve, Belgium
- [d] Dr. Marco Esters (former address)
Department of Chemistry
University of Oregon
Eugene, Oregon 97403, USA
and
Dr. Marco Esters
Department of Mechanical Engineering and Materials Science and Center for Autonomous Materials Design
Duke University
Durham, North Carolina 27708, USA
- [e] Dr. Adam J. Jackson
Scientific Computing Department,
Science & Technology Facilities Council,
Rutherford Appleton Laboratory,
Didcot OX11 0QX, UK
- [f] Dr. Ryky Nelson, Prof. Dr. Richard Dronskowski
RWTH Aachen University
Institute of Inorganic Chemistry,
D-52056 Aachen, Germany.
- [g] Prof. Dr. Richard Dronskowski
RWTH Aachen University
Jülich-Aachen Research Alliance (JARA-CSD),
D-52056 Aachen, Germany
And
Prof. Dr. Richard Dronskowski
Shenzhen Polytechnic
Hoffmann Institute of Advanced Materials,
7098 Liuxian Blvd, Nanshan District, Shenzhen 518055, China
- [h] Prof. Dr. Geoffroy Hautier
Dartmouth College
Thayer School of Engineering,
Hanover, New Hampshire 03755, USA
E-mail: geoffroy.hautier@dartmouth.edu

Section 1. Overview on code developments

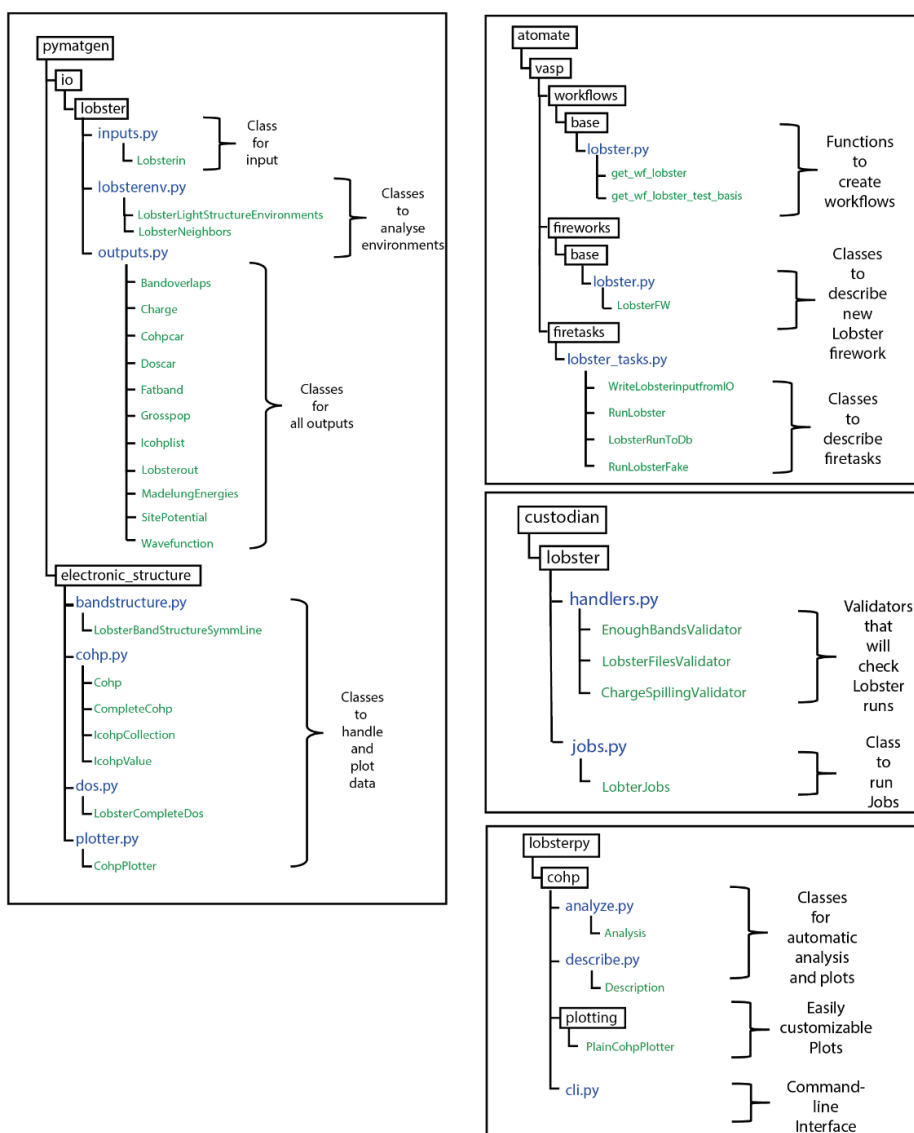


Figure S1. This graphic shows the new classes and functions that have been included in the packages *atomate*, *custodian* and *pymatgen* to develop the workflows discussed in this publication. Furthermore, new classes have been implemented in the package *lobsterpy* to automatically analyze the outputs.

Section 2. How to calculate the number bands in Lobster

Example: GaN

With the following basis: Ga (4s, 4p, 3d), N (2s 2p)

For each formula unit of Ga, we need the following number of bands:

$$1 (\text{Ga}(4s)) + 3 (\text{Ga}(4p)) + 5 (\text{Ga}(3d)) + 1 (2s) + 3 (2p) = 13$$

The available basis functions in Lobster for each pseudopotential can be found in *pymatgen*.

pymatgen/io/lobster/lobster_basis/BASIS_PBE_54_max.yaml includes the largest basis that can be used in Lobster with the pbeVASPfit2015 basis set,

pymatgen/io/lobster/lobster_basis/BASIS_PBE_54_min.yaml the smallest basis (occupied orbitals according to pseudopotential). In between these two extrema, all combinations will be tested with our algorithm.