The ab initio materials project (AIMP) v4.0 database

Introduction

The AIMP compound database v4.0 contains data for 120'600 solid phases that were calculated by various groups using *ab initio* methods in the context of the Materials Project (www.materialsproject.org) [1-3] as of 20th of January 2021. Derived from these 120'600 structures, a total of 86'999 compounds are introduced to the AIMP compound database, with a range of compounds having multiple calculated crystallographic structures that are introduced as different phases into the database. The Materials Project repository contains results from *ab initio* calculations at 0K. The models used to extrapolate thermodynamic properties to temperatures above 298K are described below after remarks on validity and possible application areas. The information which model is used is encoded in the phase name for each individual phase as detailed below.

The AIMP solutions database v4.0 contains data for metallic FCC, BCC, HCP solid solutions. Phase stability of these solid solutions has been calculated from *ab initio* calculations by GTT-Technologies. The models used are described below.

AIMP is developed by GTT-Technologies, using data from www.materialsproject.org as well as own data as input.

Validity and applicability

Unlike all other databases available in FactSage, AIMP contains uncurated data.

Therefore, AIMP cannot be expected to lead to as accurate results as it is the case when using other FactSage databases.

Using data analytics, all formation enthalpies, vibrational entropies as well as heat capacities have been checked to be generally reasonable and acceptable given the inaccuracies of the first principles methods that were used. As mentioned below most formation enthalpies have been corrected based on data in existing FactSage databases. The heat capacity and entropy of all elements have been adjusted to mostly reproduce the melting point when combining with the liquid phase from SGTE Solutions or SGPS databases.

There are three major applications for this database:

- Benchmarking the accuracy of *ab initio* calculations extrapolated to relevant temperatures.
- Using as starting point for a CALPHAD assessment.
- Combining standard FactSage databases with AIMP to estimate thermochemical properties in parts of chemical compound space where otherwise no data is available to describe the behavior of minor elements.

Heat capacity

Constant-pressure heat capacity C_p is obtained by using the Neumann-Kopp approach with elemental C_p data that has been fitted between 298K and 3000K under the constraint that elemental melting points are reproduced when combining AIMP with the SGTE Solutions database.

For each of the elements H, B, C, N, O, F, Cl and Br, the elemental Neumann-Kopp contribution to heat capacity was derived from the analysis of ~20 compounds in existing FactSage databases. The same was done for the functional groups OH, CO_3, NO_3 and SO_4. Compounds containing these functional groups were identified based on the crystal structure information in Materials Project [1-3] to take into account that the functional groups have characteristic vibrational properties distinct from the constituting elements.

Entropy at 298K

A machine-learning (ML) model for entropy at 298K S^{298K} has been developed. The training set consists of about 1'500 compounds for which phonon calculated values of S^{298K} have been reported [4,5] and additional 3'400 compounds from Calphad databases, i.e. values based on evaluations of experiments. The ML model allows calculation of S^{298K} based on three descriptors: composition, enthalpy of formation, unit cell volume, making it very robust. E.g. the mean absolute error with respect to a test set (= data set not used for training) of 151 compounds is 2.11 J/(K mol atoms).

For the compounds for which phonon calculated values of S^{298K} are available, these have been used, rather than the ML model values. A preference was given to data from Materials Project [4,5] over data from the Kyoto Phonon DB [6].

Formation enthalpies and corrections

The formation enthalpies for all compounds at room temperature are assumed to be the same as stored in the Materials Project at 0K ($\Delta H_f^{298K} = \Delta H_f^{0K}$) if all constituting elements' ground states are the same at 0K and 298K and the crystals do not contain functional groups.

If a constituting element's ground state changes between 0K and 298K or when crystals contain functional groups, formation enthalpies ΔH_f^{298K} are corrected by element-specific corrections based on observed differences in formation enthalpies between *ab initio* calculations and FactSage databases. The element-specific correction terms have been determined based on compounds that are both in the Materials Project database and in FactSage databases. This procedure applies to the following elements and groups:

H, B, C, N, O, F, P, S, Cl, Br, I, Na, Ti, Sn, Hg, OH, H_2O, CO_3, NO_3 and SO_4

Density and elastic properties

For all compounds the density is calculated from the relaxed unit cell volume given in Materials Project. It should be noted that the Generalized Gradient Approximation (GGA) used in the *ab initio* calculations results in a systematic error, overestimating the unit cell volume, i.e. underestimating density [7,8].

For some of the phases the elastic constants are given in Materials Project.

Magnetic properties

All compounds in AIMP are treated as non-magnetic. Therefore, larger errors must be expected in systems that contain typical magnetic elements such as Iron or Cobalt.

Discarded phases

A total of 3569 compounds in the Materials Project database have been discarded due to one of the following reasons:

- Phases containing noble gas elements.
- Phases containing Pa, Ac or Np due to the lack of data to benchmark and train the ML model on.
- Amorphous structures.
- Structures containing surfaces.
- Molecules (H_2, O_2…).
- Phases that are very slightly more stable than commonly accepted ground states, e.g., trigonal-Ag mp-989737 which is 345 J/mol more stable than FCC-Ag mp-124.
- For some compounds (e.g. SiO_2), Materials Project contains data for a very large number of different crystal structures. Only the 49 most stable phases are included.

Phase name scheme

The compound database follows a naming scheme where each calculated structure is added as a 'phase' for the composition of the formula-unit, scaling structures with multiple formula units per unit cell to the respective formula-unit.

The unique Materials Project ID as used on www.materialsproject.org is used as the phase name. The mode of calculation of the enthalpy of formation is appended as either 'HM', if it is the unmodified formation enthalpy from the Materials Project Database, or 'HC', if the enthalpy of formation correction is applied as detailed above. The appendix SR indicates that the entropy is derived using the ML regression model while SP indicates that entropy is obtained from the phonon calculations in Materials Project [4,5] and SK from phonon calculations in the Kyoto Phonon DB [6]. Elemental entropies of FCC, BCC and HCP structures have been adjusted to match the (hypothetical) transition temperature between the ground state at 298 K and the other states accepted in Calphad assessments [9], these entropies are indicated by SC. Here, all heat capacities are modelled using the Neumann-Kopp approach which is indicated by CN at the end of the phase name. See Table 1 for reference. Examples are given below:

Al_2SiO_5: *mp-5065_HC_SR_CN* - ΔH_f^{298K} has been corrected, S^{298K} is obtained from the ML regression model and C_p is calculated using the Neumann-Kopp approach.

$MoSe_2$: *mp-7581_HM_SP_CN* - ΔH_f^{298K} has not been corrected, S^{298K} is taken from phonon calculations and C_p is calculated using the Neumann-Kopp approach. Please note that the naming convention ensures that the data stored in Materials Project can always be easily retrieved following the link: e.g. https://materialsproject.org/materials/mp-5065/ will display the results for Al_2SiO_5.

Table 1: AiMP 4.0 database naming scheme

	Compound Identifier	***Enthalpy***		***Entropy***				***Heat capacity***
Concept	Materials Project ID	Unmodified	Corrected	ML-Regression	Phonon calculated		Calphad adjusted	Neumann-Kopp-Rule
					[4,5]	[6]		
Name	mp-XXYY	HM	HC	SR	SP	SK	SC	CN

For a few elements, *ab initio* calculations of the FCC, BCC or HCP have been performed by GTT-Technologies which are included as well. The naming convention is the same as described above. GTT-IDs starting with 1 indicate BCC structures, with 2 HCP structures and with 3 or 4 FCC structures. Example:

Au(S3): *GTT-1824-79_HM_SC_CN* – BCC-Au, ΔH_f^{298K} as calculated by GTT-Technologies, S^{298K} adjusted to reproduce the hypothetical FCC → BCC transition temperature of 3864 K obtained from the description used in Calphad assessments [9] and C_p is calculated using the Neumann-Kopp approach.

Solid Solutions

Enthalpies of mixing at 0 K have been systematically calculated by GTT-Technologies for the FCC_A1, BCC_A2 and HCP_A3 solutions. Based on these, 1900 binary interaction parameters have been derived.

In the phase **FCC_A1**, 1146 interaction parameters have been derived for binary systems combining any metal with atomic number between 3 (Li) and 83 (Bi) with one of the following elements:
Al, Ca, Ni, Cu, Sr, Rh, Pd, Ag, Ir, Pt, Au, Pb

In the phase **BCC_A2**, 608 interaction parameters have been derived for binary systems combining any metal with atomic number between 3 (Li) and 83 (Bi) with one of the following elements:
Li, Na, K, V, Fe, Nb, Mo, Ta, W

In the phase **HCP_A3**, 146 interaction parameters have been derived for binary systems combining any metal with atomic number between 3 (Li) and 83 (Bi) with one of the following elements:
Mg, Ti, Zr

Lattice stabilities of the solution endmembers are described using the same methodology as described for the compounds above.

Further information

Please contact us via info@gtt-technologies.de if you need further information.

References

[1]: 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 (*=equal contributions). *The Materials Project: A materials genome approach to accelerating materials innovation.* APL Materials, 2013, 1(1), 011002.

[2]: S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. Persson, G. Ceder. *Python Materials Genomics (pymatgen) : A Robust, Open-Source Python Library for Materials Analysis.* Computational Materials Science, 2013, 68, 314–319.

[3]: S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson. *The Materials Application Programming Interface (API): A simple, flexible and efficient API for*

materials data based on REpresentational State Transfer (REST) principles. Computational Materials Science, 2015, 97, 209–215.

[4]: G. Petretto, S. Dwaraknath, H.P.C. Miranda, D. Winston, M. Giantomassi, M.J. van Setten, X.Gonze, K.A. Persson, G. Hautier and G.-M. Rignanese. *High-throughput density-functional perturbation theory phonons for inorganic materials.* Scientific Data, 2018, 5, 180065.

[5]: https://doi.org/10.6084/m9.figshare.c.3938023.v1

[6]: http://phonondb.mtl.kyoto-u.ac.jp/ph20180417/index.html

[7]: 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*, Computational Materials Science, vol. 50, 2011, pp. 2295-2310.

[8]: https://www.materialsproject.org/wiki/index.php/Volume_Change_Error_manual

[9]: A. Dinsdale, *SGTE data for Pure Elements*, CALPHAD, 1991, 15, 317-425.