



## Making the most of materials computations

Thygesen, Kristian Sommer; Jacobsen, Karsten Wedel

*Published in:*  
Science

*Link to article, DOI:*  
[10.1126/science.aah4776](https://doi.org/10.1126/science.aah4776)

*Publication date:*  
2016

*Document Version*  
Early version, also known as pre-print

[Link back to DTU Orbit](#)

*Citation (APA):*  
Thygesen, K. S., & Jacobsen, K. W. (2016). Making the most of materials computations. *Science*, 354(6309), 180-181. <https://doi.org/10.1126/science.aah4776>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Making the most of materials computations

- *Databases of theoretical structures and properties of materials can speed real-world discovery*

Kristian S. Thygesen

CAMD, Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark; [thygesen@fysik.dtu.dk](mailto:thygesen@fysik.dtu.dk)

Karsten W. Jacobsen

CAMD, Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark; [kwj@fysik.dtu.dk](mailto:kwj@fysik.dtu.dk)

For more than a century, materials scientists have accumulated experimental data about materials like their crystal structure or thermal, electronic and mechanical properties. These data have been a corner stone in the development, selection, and design of materials [1].

Over the past decade more than a handful of computational databases storing materials properties obtained by quantum mechanical calculations have been established. They are mostly based on Density Functional Theory (DFT), which calculates the electronic structure of a material and thereby provides atomic-scale information about its properties. The calculations are much faster and cheaper to perform than experiments, and in some cases with comparable accuracy. Some of the larger repositories containing together more than a million material entries are the Materials Project [2], the Automatic Flow for Materials Discovery [3], the Open Quantum Materials Database [4], the Novel Materials Discovery Repository [5], the CatApp database [6], and the Computational Materials Repository [7]. All of these have web-interfaces providing free access to the data.

The computational databases supplement the experimental ones (as for example the Inorganic Crystal Structure Database [8]) by providing additional systematic information about materials. Maybe even more interestingly, they provide information about materials which have never been synthesized by Nature or in a laboratory, but where the properties can still be calculated. The databases therefore provide a playground for exploration of new materials.

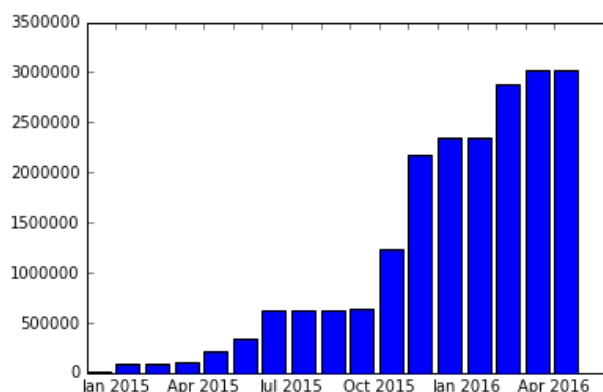


Figure 1: The number of DFT calculations in the NOMAD repository from December 2014 until April 2016. The large jumps in October/November 2015 and February 2016 are due to inclusion of data from the AFLOWLIB and OQMD repositories, respectively.

A popular application of DFT databases is thermodynamic stability analysis, where for a given combination of chemical elements stable compositions and structures at low temperatures are identified. This feature is available in most of the major databases and is for example used as a first check of the stability of a given hypothetical material composition. Combining the computed data with experimental data about the free energy of dissolution makes it possible to extend this analysis to Pourbaix diagrams, where the stability is mapped out as a function of pH and applied potential [9, 10].

The databases are steadily growing in size both because of systematic inclusion of new crystal structures and compositions, and because computational screening studies aimed at identifying useful materials within for example batteries or solar energy, contribute with new materials. In the future many materials screening projects can be expected to take their starting point directly in the computational materials databases.

The further development of computational materials databases in order to improve their quality and their applicability for materials science and design involves a number of opportunities and challenges. Here we discuss a few of the important ones.

#### **Data representation:**

Computed materials data have the advantage compared to experimental data that they are very well defined and easy to reproduce. A DFT calculation is specified by the atomic numbers and positions of all the atoms plus a few additional parameters to describe the approximations being made. In practice it may be a challenge though to label the computed materials in a way, which is both useful and sufficiently unique for the user. For example, when should two calculations be considered describing the same material in a particular phase when the atomic positions are not exactly the same?

#### **Data quality:**

An important concern is the accuracy of the computed data. Computational materials scientists apply a range of different codes all solving the same fundamental equations of DFT but applying diverse numerical approaches. In a large community effort it has recently been shown that in simple situations, like the calculation of the equation-of-state of elemental crystals, agreement between different codes can be established [11], but more complex material properties still represent a challenge and accuracies need to be estimated. One practical aspect of code comparison is that different codes produce output in different formats. This issue is currently addressed by the NOMAD project [12], where parsers for the 40 most popular codes are developed to obtain a common file format.

At a more fundamental level, DFT calculations require physical approximations for the so-called exchange-correlation energy of the electrons. Further refinements of these approximations will be necessary to increase the accuracy of the calculations, which might be necessary depending on the application.

#### **Complex systems:**

So far the computational databases mainly contain bulk solids in simple crystal structures. One reason for this is simply that the calculation time grows quickly with the number of atoms in the unit cell. However, many material properties are in fact determined by defects like vacancies/interstitials, impurities, or grain boundaries, or by surfaces. This means that for many applications, the materials stored in the databases are not directly relevant. To broaden the applicability of computational databases, more complex structures and materials therefore have to be included, but as complexity increases, the number of possible materials also grows tremendously. An example of a first step to systematically address the more complicated situation of molecules reacting at surfaces is a DFT-database of adsorption energies [6] of relevance to heterogeneous catalysis.

### **Beyond DFT:**

There are several materials properties or phenomena that lie beyond the present capabilities of DFT. These include the description of electronic excitations, strongly correlated systems, and certain transport phenomena where DFT can provide qualitative estimates at best. There is thus a need to move beyond DFT to study for example light absorption of materials for photovoltaics or photocatalysis. Different corrections to the standard DFT approaches can be applied, but the best calculations today of properties like electronic band structures or optical spectra are based on many-body perturbation theory applying the so-called GW or Bethe-Salpeter approximations. However, these calculations are computationally much more expensive and numerically involved than DFT calculations making the issue of data quality and consistency even greater. The first steps towards establishing GW databases for code benchmarking [13] and for the electronic structure of two-dimensional materials [14] have recently been taken. However, much further work is needed to establish consistent and accurate beyond-DFT data and thereby open up new opportunities for the computational databases.

### **Machine learning:**

It might be debatable whether the computed materials databases have the characteristics (volume, variety, velocity) to qualify as “Big Data” [15]. However, as the computational databases grow, new possibilities for investigating the materials space will certainly arise. It will be necessary to develop or adapt tools from machine learning for searching and analyzing many millions of computed materials, and it will be possible to apply these tools to learn new correlations or structure-property relations of materials. Simple relations of this type are an integral part of materials science and chemistry. Examples include valence rules and geometrical tolerance factors for predicting chemical and structural stability. These rules have emerged from empirical observations performed over decades of research, but when considering large numbers of computed materials, these rules can be systematically deduced as correlations in the data. The use of modern statistical learning may potentially allow researchers to identify deeper and more complex patterns in the materials space, and to use correlations to accelerate the prediction of new materials or to design materials with specific tailored properties.

### **Acknowledgments**

The research in the groups of the authors is supported by The Danish National Research Foundation's Center for Nanostructured Graphene (CNG), The Villum Center for the Science of Sustainable Fuels and Chemicals (V-SUSTAIN), and The EU Center of Excellence, Novel Materials Discovery Laboratory (NOMAD).

## References

- [1] M. F. Ashby, *Materials Selection in Mechanical Design*, 4 ed., Butterworth-Heinemann, Oxford, 2010.
- [2] 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, *APL Mater.* **1**, 011002 (2013), <http://materialsproject.org>.
- [3] 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, *Computational Materials Science* **58**, 218 (2012), <http://aflowlib.org>.
- [4] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *JOM* **65**, 1501 (2013), <http://oqmd.org>.
- [5] <http://nomad-repository.eu>; <http://nomad-coe.eu>.
- [6] J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard, and J. K. Nørskov, *Angewandte Chemie International Edition*, **51**, 272 (2011).
- [7] D. D. Landis, J. S. Hummelshøj, S. Nestorov, J. Greeley, M. Dulak, T. Bligaard, J. K. Nørskov, and K. W. Jacobsen, *Computing in Science and Engineering* **14**, 51 (2012), <http://cmr.fysik.dtu.dk>.
- [8] ICSD; <https://www.fiz-karlsruhe.de/de/leistungen/kristallographie/icsd.html>.
- [9] K. A. Persson, B. Waldrick, P. Lazic, and G. Ceder, *Phys. Rev. B*, **85**, 235438 (2012).
- [10] I. E. Castelli, K. S. Thygesen, and K. W. Jacobsen, *Topics in Catalysis*, **57**, 265 (2013).
- [11] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Du ak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Granas, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iu an, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Kucukbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordstrom, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K.

Schwarz, S. Sharma, F. Tavazza, P. Thunstrom, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G. X. Zhang, and S. Cottenier, *Science* **351**, 1415 (2016)

[12] L. M. Ghiringhelli, C. Carbogno, S. Levchenko, F. Mohamed, G. Huhs, M. Lueders, M. Oliveira, and M. Scheffler, Psi-k Highlight July 2016;  
[http://psi-k.net/download/highlights/Highlight\\_131.pdf](http://psi-k.net/download/highlights/Highlight_131.pdf).

[13] M. J. van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, *Journal of Chemical Theory and Computation*, **11**, 5665 (2015).

[14] F. Rasmussen and K. S. Thygesen, *J. Phys. Chem. C* **119**, 13169 (2015)

[15] J. Hill, G. Mulholland, K. Persson, R. Seshadri, C. Wolverton, and B. Meredig, *MRS Bulletin*, **41**, 399 (2016).

Data for the figure:

year_month	Number of DFT calculations
2014-12	20116
2015-01	83278
2015-02	92113
2015-03	112404
2015-04	221405
2015-05	345387
2015-06	630393
2015-07	631439
2015-08	631635
2015-09	634001
2015-10	1235389
2015-11	2171584
2015-12	2341854
2016-01	2346157
2016-02	2880312
2016-03	3026496
2016-04	3026797