Ψ_k Scientific Highlight Of The Month

No. 135 April 2017

Statistical Data about Density Functional Calculations

Phivos Mavropoulos*, and Peter Dederichs†

Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich, Germany

Abstract

We present statistical data on the publications in the field of Density Functional Theory. Our focus is on the European landscape, but we also include the most significant non-European states. Data comparing the number of publications per state and per million of inhabitants, as well as the average number of citations in the last 10 years, are shown.

1 Introduction

Computational Sciences are outstanding growth areas of research, profiting from both the tremendous progress in hardware development as well as the continuous improvements of computer codes and the invention of new computational methods. In the future an increasingly larger part of our technology development and our gross domestic product will depend on computer applications, in particular in Materials and Nano Sciences. Ab-initio calculations based on density functional theory (DFT), mathematically founded by Walter Kohn and coworkers, are making a big progress in these fields, since they allow parameter-free and element specific calculations of the material properties.

In this paper we present the results of a literature search for papers on ab-initio calculations, using data available from a search in the ISI Web of Knowledge. We report on the total number of publications in the field, their growth as a function of time, the importance of different regions in the world, e.g. USA versus Europe, the number of papers of different European states, the average number of citations per paper, etc. In total we find huge numbers of publications, e.g. in 2016 more than 23000 worldwide, showing the strong importance of density functional theory.

We note that the field is very well suited for such a literature search, since it can be characterized by rather few key-words. For each publication these key-words are searched in Topic, i.e., in the title, abstract, and among the key-words of the publication. As listed in the Appendix, the survey includes Physics, Chemistry, Materials Science, and Biology-oriented fields, and DFT-based as well as non-DFT ab-initio methods.

*ph.mavropoulos@fz-juelich.de

†p.h.dederichs@fz-juelich.de

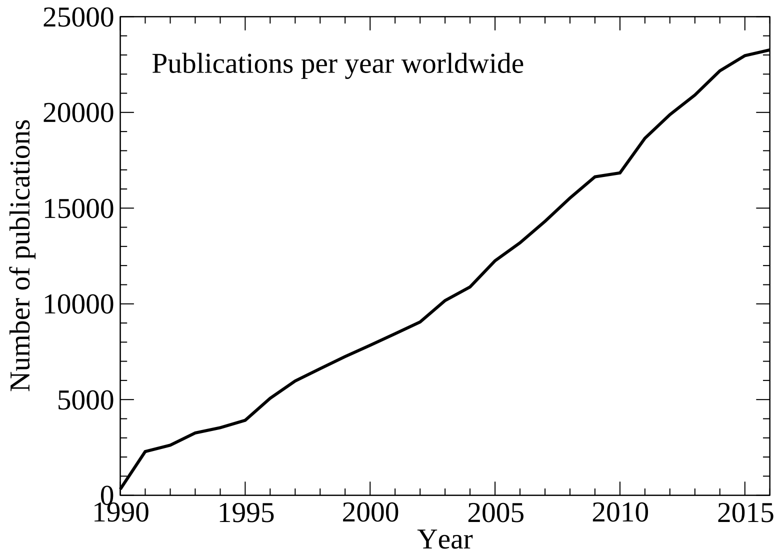

Figure 1: Number of publications per year worldwide. Before 1990, we find less than 500 publications per year. Data collected in March 2017.

2 Size and Growth of the Field of ab-initio Calculations

Fig. 1 shows the total number of publications in the field, published worldwide and obtained in the years 1990-2016. The figure shows a strong growth of about 1000 publications per year. The increase is more or less linear, and clearly not exponential. For the last 15 years it roughly amounts to a growth of 2.6 from 9000 in 2002 to 23000 in 2016. Note that the data were collected in March 2017, and that the 2016 values are therefore expected to increase by the end of 2017 as more papers are added in the database.

The more than 23000 publications are a conservative estimate, since there might be a large number of "dark" papers which do not contain the key-words in the Topic section, but show e.g. a short neutral sentence like "We have calculated the energy..." or similar, and the real information can only be found in the main text.

The large number of publications in the field is also frightening, since nobody can read all these papers, not even the ones in the different subfields. Thus it is important that Psi-k organises every year workshops on many topical problems and subfields. Another big problem is that annually more and more people perform DFT calculations and use user-friendly codes, but do not have the necessary basic knowhow. Thus training DFT skills, promoted by Psi-k tutorials, become ever more important.

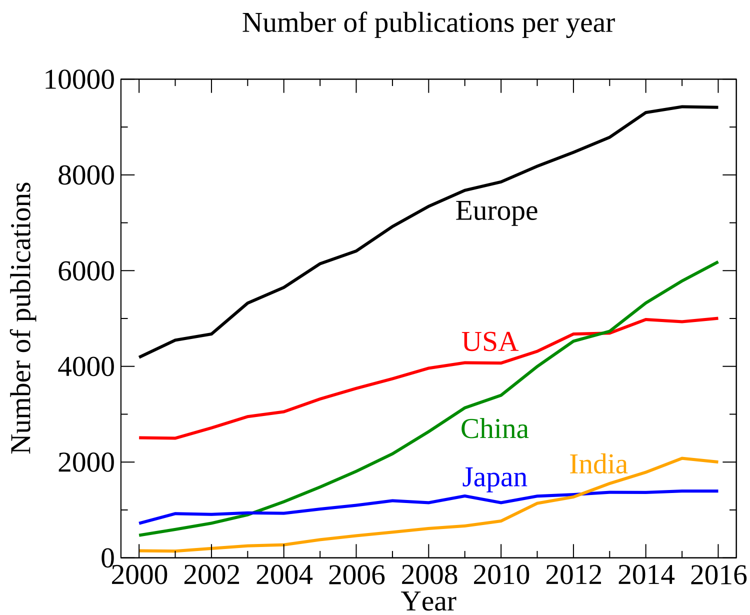

Figure 2: Regional growth of publications, 2000-2016. Europe is world-leading, but China shows the fastest growth rate. Japan shows a saturation. Finally, India has entered the field very dynamically, showing an exponential increase similar to that of China 10 years ago. Note that publications with authors from different states are counted for each state given in the address.

3 Regional Strength and local Growth

In addition to the total number of publications worldwide we would like to discuss here the number of publications from the most important scientific areas in the world, i.e. of authors from Europe, including all European states, from USA and from Asia, in particular China, India and Japan. Note that publications with authors from different states are counted for each state or region given in the address.

Firstly, one sees in Fig. 2 that the publications from Europe as well as from the USA increase more or less linearly with time, with the numbers from Europe being for all years roughly 50% larger than the US-numbers. Most remarkable is the strong super-linear increase of the publications from China, which have overtaken the USA in 2014. This is a tremendous and unmatched increase of the number of publications in refereed journals, something which one should not underestimate. On the other hand, the count is a pure quantitative measure, and does not say anything about the quality of the publications. At present, publications from USA or from Europe achieve on average a higher citation count than publications from China (see Sec. 6), but this might very well change in the future.

The figure also includes publications from Japan and India. While the publications from Japan show a plateau since about 2011, publications from India show a strong increase by a factor 5 over the last ten years.

In total the data imply that Europe is presently the world leader in ab-initio calculations. This

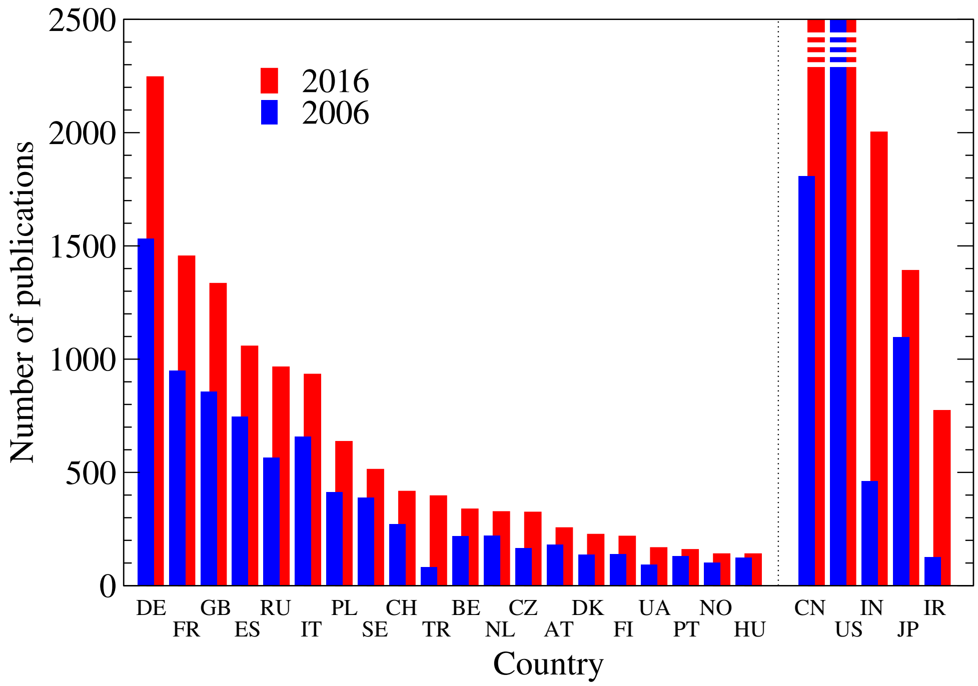

Figure 3: Publications in 2006 and 2016 per country. Shown are the results for the top 20 European states, plus the top 5 non-European countries, with respect to the number of publications in 2016. For the US, we find 3540 publications in 2006 and 5004 in 2016. For China we find 6184 publications in 2016. Note that publications with authors from different states are counted for each state given in the address.

conclusion is also supported by the fact that many European Electronic Structure codes are used worldwide.

Independently of this it would be important for our field, if in the future the contributions from USA would increase further, which presumable would need a greater funding for more positions. Without doubt the importance of our field in particular for materials science as well as the high quality of the US-contributions (see Fig. 5) would justify this.

4 The European ab-initio Landscape

Here we want to discuss the number of publication in the ab-initio field for different European states. Fig. 3 shows the number of publications for 20 European states with the largest 2016 publication numbers: Germany (DE), France (FR), United Kingdom (GB), Spain (ES), Russia (RU), Italy (IT), Poland (PL), Sweden (SE), Switzerland (CH), Turkey (TR), Belgium (BE), Netherlands (NL), Czech Republic (CZ), Austria (AT), Denmark (DK), Finland (FI), Ukraine (UA), Portugal (PT), Norway (NO), Hungary (HU). Also shown are results for the top 5 non-European states in 2016 publications: China (CN), United States (US), India (IN), Japan (JP), Iran (IR). This set of the top 20 European and top 5 non-European states forms the basis of our analysis in the remainder of the paper. Compared are the publications for the year 2006 (in blue) with those for the year 2016 (in red).

The state with the largest number of publications in Europe is Germany, in agreement with its large population. The next states in the sequence are France, United Kingdom, Spain, Russia and Italy. Clearly, the trend of publications does not follow the trend of population. We will come back to this in the next section.

Comparing 2006 with 2016, all states show considerable growth. However, the strongest growth within Europe is witnessed in Turkey (4.8 growth rate in 10 years). Outside Europe, the strongest growth is witnessed in Iran (6.1 growth rate), followed by India (4.3) and China (3.4). Here, we consider the strong growth of the "newcomers" India, Iran and Turkey, as most surprising. In our search we also found other non-European states that show such strong growth. The prospective power of these states in the field cannot be overestimated. Of course there is always the question of quality, that cannot be judged merely by the number of publications or growth. This we will discuss in Sec. 6. Still, newcomers in the field will produce high-quality papers, if there is enough funding and attractive working conditions. Compared to this strong growth, the near stagnation of the ab-initio field in Japan is rather disheartening (see also Fig. 2).

5 Publications per Million of Inhabitants

What additional information can one obtain from the publication search? Here we try to get some information about the funding of our field in different states. The number of publications for a specific state clearly depends on the quality of the researchers in this state, but also on the number of positions for research in the field. Both these factors are related to the funding, which ideally should scale proportionally to the population. This motivates us to examine the number of publications per million inhabitants of each state. To clarify, the latter quantity is just a normalised output of research papers, which, however, we interpret as being proportional to the normalised funding.

Fig. 4 shows the number of publications per million inhabitants for the top 20 European and top 5 non-European states in 2016 publications. For instance for France the total number of publications in Fig. 3, i.e. 1455, is divided by 67 due to the population of 67 million French citizens, resulting to a value of about 22. The values for the UK and Spain are rather similar. As a reference we take the average of the four first European states with respect to the number of publications (Germany, France, UK, and Spain) which is indicated by a horizontal line. Thus we would consider the funding in France, UK and Spain as adequate and compared to this the funding in Germany as somewhat better. However, the funding of our field in Italy seems not to be as good. Moreover, poor results are found for Russia, in spite of its great tradition in physical sciences.

Looking at the European standard we also find that the funding of the electronic structure field is not adequate in the USA and that more research positions are needed. As we had pointed out already in Sec. 3 it would be most important for the development of our field that the funding in the USA would be increased, in particular also due to the high quality of US-research (see Sec. 6). The data for Japan show also another, even bigger deficit of funding, according to our interpretation.

We will now discuss the ratio of publications per million inhabitants for the smaller European

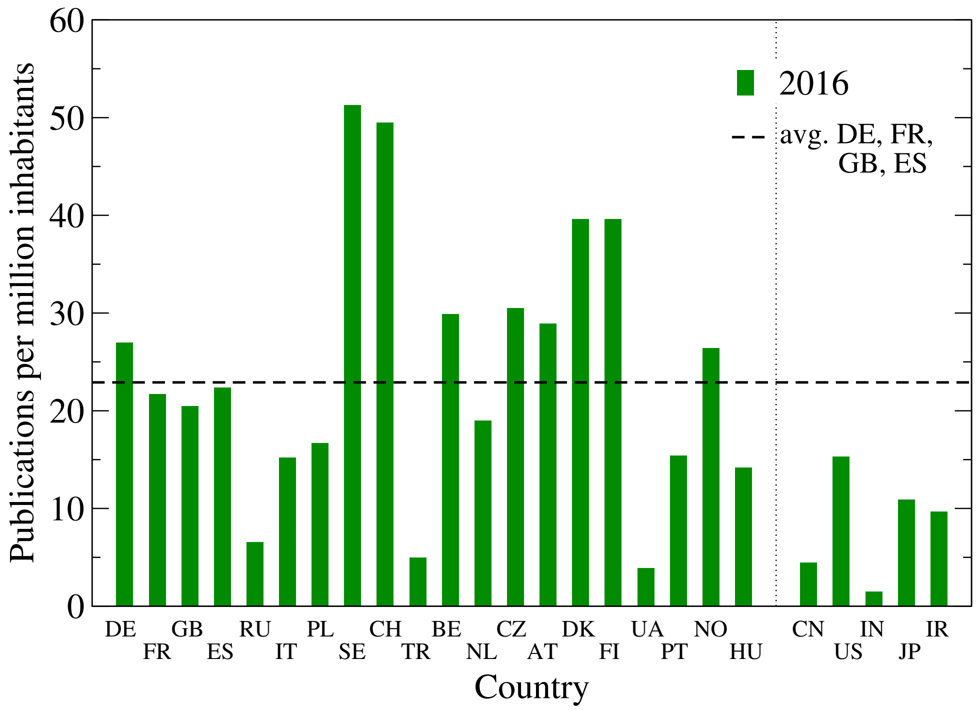

Figure 4: Total number of publications in 2016 (from Fig. 3) divided by the population in millions for the different states. The horizontal line denotes the average of Germany, France, UK and Spain.

states. The results in Fig. 4 are surprising, since in many smaller states the ratio is much higher than for the larger European states. For the two top values, *i.e.* Sweden and Switzerland, the ratio is nearly double as large as the ratio for Germany. Also Denmark and Finland show extraordinary large values. Thus the real leaders in our field are the smaller states, presumably due to considerably better funding. Also Belgium, the Czech Republic, and Austria are well above the average of the four most-publishing states (DE, FR, GB, ES) represented by a horizontal line. Only the Netherlands, Hungary and Portugal (among the less populous states) fall below this average.

6 Average Citation Count for different Countries

A large quantity of publications does not necessarily reflect quality. Of course it is difficult to average the quality of hundreds, or thousands, of publications, but one of many measures of quality is the average number of citations per publication. This we show in Fig. 5, for publications averaged over the years 2005, 2006, and 2007, and their citations sampled in March 2017.

Denmark hits the top value of more than 51 citations per publication, followed by Switzerland with 46, then the Netherlands and Germany. Impressive is also the high score of Italian publications, certifying the high quality of Italian contributions to our field. Also impressive is the high citation rate of Czech scientists, higher than many Western-European states. The poor rate for Russia is in line with the low number of publications per million inhabitants discussed above.

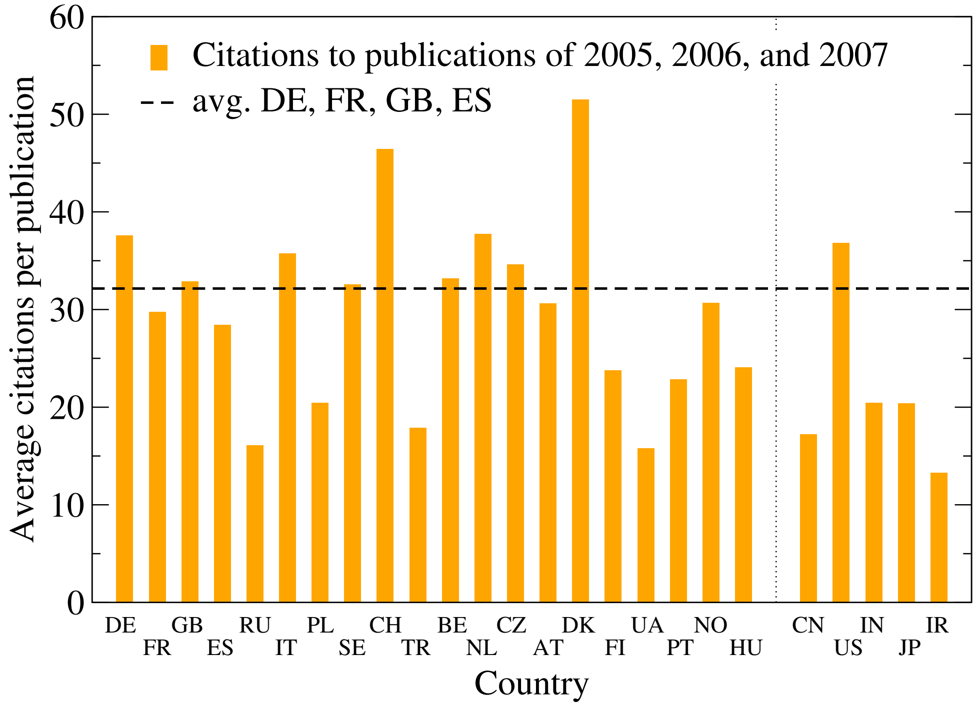

Figure 5: Number citations per publication averaged over publications in 2005, 2006, and 2007 and sampled in March 2017. Note that publications with authors from different states are counted for each state given in the address.

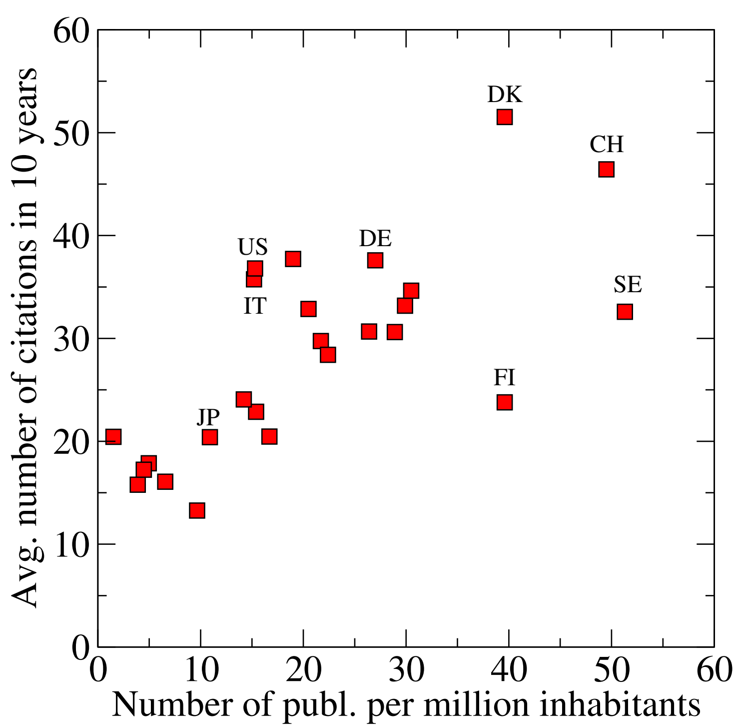

Figure 6: Fig. 6. Number of the average number of citations (from Fig. 5) versus number the of publications per million inhabitants (from Fig. 4). In spite of the fluctuations, a correlation of the two quantities is evident.

Comparing Fig. 4 and 5, we see that the number of publications per million inhabitants does not correlate one-to-one with the citation average. Nevertheless, there is a degree of correlation, which we depict in Fig. 6. Thus we might say that the higher the number of publications per million inhabitants, in short the better the funding, the higher is on average the citation rate, i.e., the better is the quality. Of course this is an oversimplification, but in rough agreement with the trend of Fig. 6.

7 Conclusions and Outlook

The continuous growth of publications in the field of ab-initio electronic structure calculations is undisputed, on-going, and expected to reach almost 30000 papers per year by 2020. This large count includes primarily the fields of Quantum Chemistry, Materials Science, and Physics, but also extends all the way to Biology, Energy, and Engineering. From our personal experience in Physics and Materials Science we see that a relatively small number of codes is responsible for the bulk of these calculations, at least when one compares the number of codes to the amount of publications. This is of course understandable, since the development of an ab-initio code needs normally more than 10 years of work by experienced people, and since the existing codes are continually being augmented by new features, *e.g.* new functionals or by a connection to non-density-functional techniques. We also know that a significant portion of these codes is being developed in Europe.

The growth of the ab-initio field is governed by two basic facts. For the first time in history we have the chance of understanding materials properties on the most elementary level, and due to the advance of computational methods and high-performance computing, these materials can be extremely complex. Secondly, and extremely importantly for applications, we will be able to design novel materials with improved or even required properties and performance. For our society this will mean in the future an enormous shift away from tests and experiments to computer-aided design and discovery. Such materials will be crucial for scientific and technological advances. Both facts will assure that also in the foreseeable future our field will not stop growing.

The Psi-k European Network offers a great service to the scientific community by sponsoring workshops and tutorials, about 30 every year in many European states. These events give the chance to impart a proper philosophy of performing ab-initio calculations to the many new or inexperienced users. In addition, the success of the self-funding system of Psi-k gives excellent arguments for applying for extra funding by the states, which is, as we believe, less than optimal in many cases. The need for supporting ab-initio theory, successfuly reaching far from its initial disciplines of physics and chemistry to computer aided design and discovery in many scientific and technological fields, cannot be overstressed. We hope that the present statistical survey can help for improving the funding in our field.

Appendix A: Details on the search and on the choice of keywords

The search was performed in the ISI Web of Science Core Collection (www.webofknowledge.com).

The following keywords were used in the Topic section: “Ab-initio” or “Density Functional” or “First Principles” or “First Principle” or “DFT” or “LDA” not “Discrete Fourier” not “Discriminant Analysis”.

The addition of the keywords DFT and LDA increases the record count by approximately 10-15% with respect to the other keywords. Of course, not all hits correspond to condensed matter calculations, but the vast majority of them do. For example, there are publications in electromagnetic wave propagation that use the term ab-initio meaning that only the dielectric constant is given. Also, the term Density Functional Theory occurs in classical liquid theories and in Nuclear Physics, while the acronyms DFT and LDA have occasionally a different meaning than Density Functional Theory and Local Density Approximation.

On the other hand, there exist papers in the ab-initio electronic structure field that do not match these keywords in their Topic section, merely referring to calculations or theory. Since a full-text search is at present not possible, these dark papers are missing from our statistics.

We restricted the search to the following Web of Science Categories: Chemistry Physical, Physics Atomic Molecular Chemical, Chemistry Multidisciplinary, Physics Condensed Matter, Materials Science Multidisciplinary, Physics Applied, Chemistry Inorganic Nuclear, Nanoscience Nanotechnology, Chemistry Organic, Physics Multidisciplinary, Engineering Electrical Electronic, Spectroscopy, Optics, Biochemistry Molecular Biology, Crystallography, Metallurgy Metallurgical Engineering. The restriction results in a small underestimation of the publication counts. However, relaxing the restriction may lead to an overproportional increase of “wrong papers,” not directly related to our topics.

According to our previous experience, we expect the results for the year 2016 to be updated to slightly higher values, perhaps by 3-4%, by the middle or the end of 2017, because not all 2016 publications have been incorporated yet in the databases. This results in an apparent slowdown of the growth, or a plateau formation, in the 2016 results of all figures.

Here we should also make reference to two papers which include publication counts in our field. The review paper on DFT by R. O. Jones, Rev. Mod. Phys. **87**, 897 (2015) includes a publication count up on Web of Science up to February 2015. The trend is the same as the one of our Fig. 1, but our count is somewhat larger because of the use of more keywords here. The second review by A. Pribram-Jones, D. A. Gross, and K. Burke, Annu. Rev. Phys. Chem. **66**, 283 (2015), presents a citation count (not publication count) to DFT papers resulting in a total number of almost 30000 citations in 2013. Our count for 2013 gives 20900 publications.

Appendix B: Country codes and population

The country codes appearing in the figures follow the ISO-3166 system, see https://www.iso.org/iso-3166-country-codes.html . The population of the states was retrieved from the Wikipedia site https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population in March 2017.