An extension of the Hirsch index: Indexing scientific topics and compounds

M. G. Banks*

Max-Planck-Institute für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany
(Dated: September 14, 2006)

An interesting twist of the Hirsch index is given, in terms of an index for topics and compounds. By comparing both the hb index and m for a number of compounds and topics, it can be used to differentiate between a new so-called hot topic with older topics. This quick method is shown to help new comers to identify how much interest and work has already been achieved in their chosen area of research.

The Hirsch index "h index,"[1,2] has recently been of much scientific discussion for its use in ranking the output by a given scientist. Using the Science Citation Index (SCI) under the Web of Science (WoS) offered by Thomson ISI (available at http://isiknowledge.com), and by sorting the authors papers by ascending order in terms of citations. The h index of a given author is when h of his or hers N_p papers have at least h citations each. The corresponding remaining papers have thus $\leq h$ citations each. In the case where the author has a unique name, this procedure is a very easy and fast way in order to compare scientists. More importantly, it does not fully depend on the total number of papers, N_p, which could be bolstered by conference proceedings, or the total number of citations or significant papers. The Hirsch index measures the cumulative impact of a scientist, which comes mainly from the quality of the research and also from the size of the community in which the scientist publishes.

In this paper I give an interesting twist on the h index and apply it further to the case of interesting topics and compounds. It is common practice today that many so called buzzwords are in use, in which a new person to the field hears the word often but is left wondering, even what is meant by it, without much explanation about the meaning being given. It is getting increasingly difficult for a newcomer, i.e. PhD student or undergraduate student to unravel all this information in order to choose a masters or PhD topic that could be interesting and relevant for their interest and future development. Especially in solid state physics, there is much use of "compound culture" where a given student works on a single compound during a masters or PhD course. A student or for that matter also a post-doc could have the following questions: How much has already been done in this topic or compound? Is it a so called "hot topic"? Will many people be interested in the results that I obtain from my thesis or research? In today's wealth of information this could be a difficult question to answer without doing a huge amount of initial searching in the literature. A simple and quick, but by no means full answer could be easily obtained by comparing the h index for topics or compounds, which will from now on be called the "h-b index".

No rigorous mathematical treatment will be given here, but it will follow from the procedures given from ref (1), in which the h index is defined, which I now assign as the h-b index. The following linear relation of the h index with the number of years from the first paper published (eq. 6 from ref (1)) is given as

$$h \sim mn \qquad (1)$$

With h, substituted for h-b. In this case, it is assumed that h-b increases linearly with the number of years, n. The gradient is thus given by m, which will vary from topic to topic or compound to compound. This linear relation may or may not be the case, but it is an approximation for which in a large number of cases will hold. In the case of a scientist's name, this can be shown that from the first paper published until a given maximum, in which afterwards the h index is constant. A plot of an authors h index against number of years (n) will give an approximate straight line[5]. This also holds for a compound or topic search where it is generally found that the hb index increases linearly with the number of years since the initial discovery, until a given maximum when it is approximately constant. Even for new expanding topics this initial linear behavior is found.

The h-b index is found by entering a topic or compound in the Thomson ISI Web of Science database and then ordering the results in terms of citations, by largest first. The h-b index is defined as above in the same manner as the h index.

The number of years, n, is found by the year of the first paper published on this topic or compound until 2005. In this manner, m is easily calculated as the ratio of h to n. Before any results or observations are given this procedure will be discussed in terms of its legitimacy. The database used, will in the case of topics or compounds, search the title, keywords (incl. keywords plus) and the abstract[6]. For example, for a given compound, the paper could be about a high T_c superconductor, but include MgB_2 in its abstract, in which case this paper would also be included for an MgB_2 search. Although this hinders the results for an MgB_2 search if the paper is highly referenced, this difficulty would occur in a limited number of cases and would not affect the result in a large way. As it is assumed that the title and keywords would contain what the paper is written for. The abstract is the

only way a result could be ambiguous, but for the large majority of cases, the results will hold. In the case of compounds, searching a compound by its chemical name or formula should in a wide range produce sound results, as the compound name is largely unique, as in the same way the author's name. However, in some limited cases another compound formula which contains the searched formula string could be present. This is analogous to the case of an author name that may not be unique, but for the most cases is (for example, about 76% of the Nobel prize winners in the last 20 years had unique names, from ref (1)). In the case of topics, the search is more difficult and care must be taken. In the case, where I specifically mention a topic (or compound), the exact search string will be presented e.g. "Kondo AND lattice". The topic itself is not as concise as searching for compounds, due to the more chance of the occurrence of the word rather than a compound formula. As stated above there are some pitfalls, however the results are nevertheless a good ballpark estimate, of the current interest and the amount of people working in the area, now and over the years. Next, I will give some results of searches I have undertaken, and then discuss possible implications in respect of the above questions and searches.

First is the discussion of compounds. The results I will present are to compounds that were mostly unambiguously identified, with no or very little "washing" out of the h-b index. In some cases, e.g., H_2O, this string exists in many other chemical formulas which contains "H_2O", and thus appears many times, so it is impossible to know what the number is for only "H_2O". Most compounds do not exhibit this feature, and some of them are shown in table I. It should be noted that in calculating m, the year of the first paper about the compound was used, which in most cases is the paper of synthesis and/or describing the crystal structure. This I believe is nevertheless the best way of representing m, of course the h-b index is regardless of this point.

It could be argued that using m is not a good estimate of how much interest has gone into a given compound, due to the fact that the starting year I chose is probably the year of synthesis or structure determination. There could be an initial time lag where no further experiments were carried out on the compound until a much later date. An interesting case could be MgB_2, after the discovery of superconductivity, the work into this compound exploded, but the time between the first paper in 1954, gives a 40 year difference, this naturally affects its m number a great deal. In such cases a comparison of the h-b index is necessary. However, it is not the scope of the paper to argue what is a good starting point for the determination of m. Some simple conclusions can be pointed out by looking at table I. The superconductors (e.g., V_3Si, Nb_3Sn, MgB_2) all have large values of h-b and m, a clear indication of the enthusiasm and large number of scientists working in the field of superconductivity. A comparison with research in rare earth systems (e.g., $PrPb_3$, $TmGa_3$) shows that the number of people

TABLE I: h-b index for specific compounds given by a name or chemical formula, sorted by ascending m.

Compound	h-b index	m
$PrPb_3$	6	0.26
$TmGa_3$	6	0.30
$Si28$	17	0.31
CeB_6	32	0.76
V_3Si	39	0.77
Ni_2MnGa	37	0.82
Nb_3Sn	48	0.94
MgB_2	67	1.31
$CeCu_2Si_2$	39	1.44
$SrTiO_3$	94	1.96
GaN	144	2.12
C-60	182	5.20

working in these compounds is much less than superconductivity (exceptions do exist, e.g., $CeCu_2Si_2$). Work in C-60 bucky balls, is the largest that I have found to date, with an m of 5.2 which represents this as a unique compound, in which there has been a large amount of work done.

Therefore I find from looking at all my results from searches carried out on compounds the following conclusions.

i) $0 < m \leq 0.5$, i.e. a maximum h-b of 20 after 40 years. This represents a compound which is likely to be of interest to the researchers in that particular field of research, where the field is a smaller community.

ii) $0.5 \leq m \leq 2$, i.e a maximum h-b of 40 after 20 years. This represents a compound that is likely to be a hot topic area of research, where the community is very large. Or a compound with very interesting features.

iii) $m \geq 2$, i.e a minimum of h-b of 40 after 20 years. This reflects a unique compound, which has far reaching consequences rather than just in its own area of research. It is likely to be a compound with application purposes, or unique characteristic features.

Next, I will discuss the larger area of topics, which of course extends much further than just the scope of solid state physics. The searching of topics was performed in the same manner as with the compounds above. There are a few points of interest which should be addressed first. Searching a topic would give a hit, when the string is represented in the title, keywords or the abstract. This is expected in the majority of cases to be legitimate. If a false hit is present, this would occur in most cases from the abstract, but its effect is neglected here. Some searches of certain topics is meaningless, e.g. "magnetism" or "specific heat", and would give answers in which not much information could be extracted. However for today's so called "hot topics" such searches could be used. Table II shows some of the search results.

In the same way as with specific compounds, the use

TABLE II: h-b index for specific topics, search strings are given, sorted by ascending m.

Topic	h-b index	m
Borides	46	0.44
pyrochlore	61	0.62
Spin flop	34	0.83
Optical lattice	43	0.90
Antiferroquadrupolar	18	1.00
amorphous silicon	116	1.10
Spin frustration	30	1.36
ferroelectricity	78	1.39
Spin liquid	45	1.55
kondo AND lattice	63	1.97
perovskites	103	2.10
spin ice	17	2.13
magnetoresistance	172	2.39
quantum information	65	2.41
geometrical frustration	21	2.63
quantum critical point	42	2.63
porous silicon	104	3.25
spin glass	108	3.38
Spin valve	48	3.43
heavy fermion	97	3.73
superstrings	99	3.96
Teleportation	61	5.08
quantum computation	73	5.21
M-theory	79	6.58
giant magnetoresistance	116	6.82
fullerenes	140	7.78
quantum dots	149	7.84
Nanowires	105	8.75
carbon nanotubes	167	12.85

of certain keywords describing the topic, may take some time after the initial discovery to surface in common use, so there may also be some initial time lag, between the true discovery, and the naming of it. However in most cases this would only be a year at the most. Table II shows the results of some searches for specific topics in many areas of physics. Some of the older topics are most likely to be found at the top of the table (as the table is m ascending), which is likely also caused by the increasing number of researchers in all fields of science over the years. Therefore, most of the new hot topics are represented at the bottom of the table, e.g., nanowires. In some cases it may be more useful in order to compare by h-b number, which of course is still modified by the total number of people or groups in that area, but not dependant on n, the number of years. Some more general conclusions are now given from the searches of topics:

i) A large m ($m \geq 3$) number is likely to come from a topic which is a so-called "hot topic", as a large m number comes from the fact that the number of years is small, while maintaining a large enough h-b number.

ii) A large h-b number (h-$b \geq 100$) and a large m number ($m \geq 3$) represents a topic in which there has been a lot of research already, which was and still is a hot topic.

iii) A small m ($m \leq 2$) but a large h-b number (h-$b \geq 100$) probably reflects this as a older topic, which has a good contribution throughout the years, which is why the m number is low.

Of course these rules are by no means complete, but kind of a first approach, that one could follow, but not adhere too. There are some topics which are revolutionary in some way, like carbon nanotubes or nanowires, which have had a huge amount of work done, in a very short amount of time.

I have presented here mainly the results for the case of physics, and with a little more emphasis on solid state physics. I am sure that this would also work for any area of science, where topics can be clearly separated from each other. I have discussed a lot of topics and compounds, without talking about the physics itself. My results are only based on a few numbers. As for the case of a scientist, a number does not reflect the potential of a person. In this case, also the number does not reflect the science or the interesting physics. Rather it does reflect the interest in the immediate community and further, and the amount of work which has been done. It should be noted that using this method in order to join a particular area of research is another question and based on many other variables. However, this method can be used to get a feeling of what topic or compound is mainstream research at this present time. Therefore, there can be some conclusions drawn without going into the details of the specific research area, this I find itself is a success of this technique. With regard to the questions that I posed at the start of the article. I would now like to answer or attempt to reason why this technique could help a student or researcher find his way through topics. Regarding the question of finding out how much research has gone into the topic already, I think this is a useful way of finding it out, and probably represents the best way in order to get a feeling of how much research has already been done. The question of whether this is a "hot topic" or not is difficult to answer in a number, as "hot topics" come from interesting physics, or interesting features. However, what can be immediately concluded is that a large number of people work in the area, which in itself is a way to identify it as a "hot topic". In a similar manner working in a "hot topic" area will make your work more visible by a larger community.

Concerning Hirsch's comments regarding membership of a national academy or institution by comparing a given researchers h index. An interesting point could be made by looking at the h-b index when allocating funding money from grants. This method could be used to compare different researchers applying for a given research grant in the same field. By searching for a researchers

name and topic of proposed research, then comparing the h-b index, could give interesting results and comparative features of already published work in the area.

In summary, I have proposed a tool to compare different topics and compounds based on the result obtained by J. E. Hirsch[1,2]. I have found interesting correlations, which can be used to give some conclusions which point to whether this is an "older topic" or a new "hot topic" by comparing the h-b index and m. It has been shown that this method could help new comers to the field, to identify how much interest and work has already been achieved in their chosen area of research.

[*] Electronic address: m.banks@fkf.mpg.de
[1] J. E. Hirsch, Proc. Nat. Acad. Sci **102**, 16569 (2005).
[2] J. E. Hirsch, available from http://arxiv.org/abs/physics/0508025.
[3] Nature **441**, 265 (2006).
[4] Physics World **19**, 7 (2006).
[5] W. Marx, *to be published.*
[6] Using Thomson ISI WoS, the abstracts are only given for papers published after 1990.