Jointly published by Akadémiai Kiadó, Budapest *Scientometrics,*
and Kluwer Academic Publishers, Dordrecht *Vol. 61, No. 3 (2004) 339–359*

A comparative study of research performance in computer science

JIANCHENG GUAN,[a,b] NAN MA[a]

[a] *School of Management, Beijing University of Aeronautics and Astronautics, Beijing (P. R. China)*
[b] *Research Center for Technological Innovation, Tsinghua University, Beijing (P. R. China)*

The paper compares the research performance in computer science of four major Western countries, India and China, based on the data abstracted from INSPEC database during the period 1993–2002. A total of 9,632 computer science papers recorded in INSPEC database were used for the comparison. The findings indicate that, on the one hand, the number of papers produced in China has considerably increased in the past few years. Particularly, in recent years, China occupies a remarkable high position in terms of counts of papers indexed by the INSPEC database. On the other hand, Chinese scientists preferred to publish in domestic journals and proceedings and shares of SCI-papers to the total journal papers for China have still remained the lowest. This indicates that the research activities of Chinese scientists in computer science are still rather "local" and suffer from a low international visibility. Various scientometric indicators, such as Normalized Impact Factor, ratio of papers in high quality journals are further adopted to analyze research performance and diverse finding are obtained. Nevertheless, for these surrogate indicators, China has optimistically achieved great progress, characterized with "low level of beginning and high speed of developing". The policy implication of the findings lies in that China, as well as other less developed countries in science, can earn relative competitive advantages in some new emerging or younger disciplines such as computer science by properly using catch-up strategy.

Introduction

The assessment of scientific research is an extremely delicate and sophisticated venture (BRAUN et al., 1985). The scientific position of a given country in the international context usually could be analyzed from both qualitative and quantitative points of view. First, the number of publications of a country and its contribution to the total world can be used. Second, the impact of its research outputs, preferably by scientific disciplines, can be measured through citation or some other surrogate Impact Factor measures (BORDONS et al., 2002).

Scientometric analysis plays an important role in the assessment of the performance of scientific research, for it could solve some structural problems such as the

Received June 28, 2004

Address for correspondence:
JIANCHENG GUAN
School of Management, Beijing University of Aeronautics and Astronautics
100083 Beijing, P. R. China
E-mail: guanjianch@sina.com, guanjianch@buaa.edu.cn

0138–9130/2004/US $ 20.00

comparison between a country and some other countries on the impact of research outputs of several scientific fields, the scale and characteristics of the international comparison, the structure of several fields, and the relationship which exits between them, as well (VAN RAAN, 1990).

Computer science has being widely regarded as one of key research fields for several decades in the world. In China, computer science, as well as other sub-disciplines of information science, is an area of highest priority, as has been identified by NSFC and Chinese Science and Technology Ministry (GUAN & WANG, 2004). The development of computer science and technology also listed as a research area of top priority both in Chinese National Key Basic Research Program (973 program) and Chinese High Technology Development Program (863 program).

Furthermore, computer science and technology is playing an increasingly important role in the S&T development as well as in economic growth in China (Annual Report of Science and Technology Development of China, 2002). The evidences show that the most notable successes have occurred in upgrading of China's industrial competitiveness in the information industry, particularly in computer industry (LU & LAZONICK, 2001).

The present study attempts to evaluate the output performance and the trend of development on the basic research of computer science in China, through the analysis of some bibliometric indicators of the scientific publications by international comparison. Similar studies have been made by GARG (2002) and GARG & PADHI (2001, 2002) in a series of studies on laser S&T between India and China. In addition, there are some studies that deal with the scientific performance of India and China as individual countries by GARG & DUTT (1992), FRAME & NARIN (1987), and ARUNACHALAM et al., (1987).

Research objectives

The objective of the study is to present a comparative assessment of the status of computer science and technology (S&T) research in China and other five representative countries, namely, USA, Germany, UK, Japan, and India, by using different scientometric indicators and analytical techniques. USA, Germany, UK and Japan are all science-giants, representing highest academic level in the world. India is one of the most populous developing countries as well as one of rapid S&T-growing countries, like China.

The comparison has been made on the following features of the research output:

I. To study the quantitative distribution and growth characteristics of the scientific publications in the field of computer science in recent 10 years, and compare the relative research effort and performance by the scientists of all six countries considered in this study.

II. To study the communication behavior and mainstream connectivity through analyzing distribution profiles of scientific papers on "country of publication".
III. To identify the patterns of collaboration for computer science of six countries.
IV. To study the numbers and proportion of the scientific papers appeared in journals covered by SCI, as to identify the mainstream connectivity and mainstream readership of the research outputs.
V. To compare the research outputs' impact (i.e. relative research qualities) in computer science in the six countries using some proxy measures constructed by the Impact Factor provided by the *Journal Citation Report* (JCR) published by Institute of Science Information (ISI).

Data and methodology

The data source for this study was the INSPEC database, which is published by the Institute of Electrical Engineers (IEE). The world famous database supplier–Ovid, which is the subsidiary company of the Wolters Kluwer cooperation, provides this database on line. INSPEC is the world's leading English abstracting and indexing service providing information on all aspects of physics, electronics, electrical engineering, computer and its control and information technology.

After in-depth searching into the classification codes in INSPEC, we find that computer science in INSPEC covers almost 200 sub-categories. Considering that one of the major objectives of this research is to evaluate the relative research qualities of the countries concerned in computer science, we need to use impact factor (IF), which is only provided by JCR, and normalized impact factor (NIF) that is based on IF to evaluate the research quality of the countries considered. Referring to the knowledge structure of computer science and JCR's thesaurus on computer science, we considered computer science is composed of 8 sub-fields, including Computation Theory, Programming, Software Engineering, Computer Architecture, Artificial Intelligence, Computer Networks, Information Systems and Interdisciplinary Application. However, it is extremely difficult to distinguish and assign so large number of papers of the 200 sub-categories classified by INSPEC into the 8 subfields classified by JCR's thesaurus on computer science.

GARG (2002) and GARG & PADHI (2001, 2002) used 'laser' and 'lasers' in INSPEC database as their search key words to collect the publications and to evaluate them by NIF. Largely following their approaches, we applied the search technique in computer science. We believe that using "key phrase" as searching strategy in INSPEC database is more precise than using all the classification codes. Moreover, the subject categories and terminology provided by ISI are widely recognized by many researchers and scientometricians in their studies. It is relatively simple and easy to use. Thus, through combining "Key Phrase Identifier" with "Subject Headings", we firstly searched the above 8 sub-fields as independent subject, and then they were combined together. By using and combining these two search strategies, we got abstracting information of all articles published in the field of computer science. We filled in the authors' affiliation and publication year that we need to study. Thus, records incorporating USA, UK, Germany, Japan, India and China in the affiliation and 8 sub-fields of computer science in the subject field were downloaded for the time period 1993–2002. By the way, these data are all in text (.txt) format. The bibliometric details for each record included document type, title of the article, author(s), affiliation, and name of the source where it is published along with its country of publication, language of the article, and treatment of the subject.

The data downloaded from the INSPEC need to be transferred into an analyzable form. Therefore, we developed a program specifically used for the data transformation. This program was compiled on the platform of Microsoft Visual Basic 6.0, along with Microsoft SQL Server 2000 as its database tool. Through running this program, we could transfer the original data form (.txt) into the form of table in SQL Server 2000, and then it has been exported into the sheet form in Microsoft Excel 2000 and saved.

In addition, we have downloaded journals' Impact Factor in 1997 and 2002 from the web of Institute of Science Information (ISI). It has been directly input into Microsoft Excel 2000. In this way, we could get one more bibliometric detail for the scientific paper, which is the Impact Factor of the journal in which the paper has been published.

Results and discussion

Absolute outputs

The most direct and common used indicators that describe the research outputs are number of papers published in academic journals and conference proceedings and citation counts of the published papers. To evaluate numerical characteristics of research performance, many studies have used the number of publications and the number of citations received by those publications as measures of productivity, either at the discipline level or in the regions. Many important conclusions are first deduced from counts of publications.

First of all, we would like to study the characteristics of the number of publications published by six countries in the field of computer science. The counts of publication indexed by INSPEC database during the period 1993–2002 are given in Table 1.

An analysis of the total output, presented in Table 1, indicates that The United States took the absolute predominance among all six countries; it constituted about one third of the world output, while two developed European countries UK and Germany took about one fourth and one sixth of the USA's. However, except for China, the other five countries have all a relative – to different extent – pessimistic declined trend in publication volume.

Table 1. Publication output during 1993–2002

Year	USA	UK	Germany	Japan	India	China	**World**
1993	3567	915	622	484	120	162	**10457**
1994	3887	1035	673	541	95	256	**11329**
1995	3519	1000	602	616	132	252	**10997**
1996	3535	946	705	660	85	323	**11878**
1997	3212	869	709	588	100	349	**11178**
1998	2874	772	629	603	96	423	**10571**
1999	2666	816	567	603	111	459	**9915**
2000	2442	686	506	558	90	664	**9153**
2001	2256	606	494	545	96	854	**9344**
2002	1833	441	328	380	64	759	**7254**
Total	**29791**	**8086**	**5835**	**5578**	**989**	**4501**	**102076**

Among three Asian countries, Japan occupies highest ranking in the number of publication, for it is roughly equal to that of Germany. India is the least productive country in the field of computer science, while China is the most remarkable country of all six countries, for its number of publications in the field of computer science has been sharply increasing in recent 10 years. The share of China's output in the whole world has considerably increased from 1.5% in 1993 to 9% in 2001 and then over 10% in 2002. This was largely due to the scientific policies made by Chinese government in the recent years, on which computer science and technology has been listed as an area with top priority (GUAN & WANG, 2004). This also indicates that developing countries, like China, can earn some comparative advantages during their catch-up scientific activities in some new emerging science discipline.

Figure 1 shows the general tendency of growth rates of the publications in the field of computer science for the six countries during the period 1993–2002.

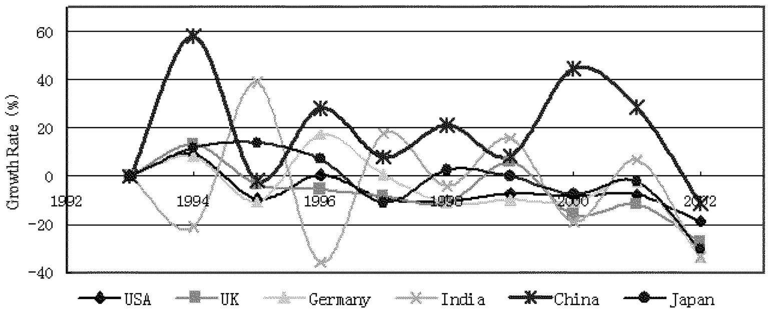

Figure 1. General tendency of growth rates of the publications

The growth rates for the different years are calculated with respect to the previous year and not to the starting year 1993. It is obvious from the Figure 1 that China's growth rate of publications is also the highest, for it has been kept positive for 9 years out of the 10 years period considered.

Relative research effort

Using the number of publications to measure the research outcome is generally based on a hypothesis that all the papers carry the same influence and quality. However, counts of scientific papers do not reveal anything about the impact of the research (NARIN & FRAME, 1989). In most cases, publications differ considerably in value and authority. Therefore, counts of number of papers make less sense, which in turn calls for using bibliometric indicators as the additional indicators to measure research outcome. In doing so, we first use the concept of Activity Index.

AI was first suggested by FRAME (1977), and elaborated by SCHUBERT & BRAUN (1986). NAGPAUL (1995) and GARG & PADHI (1998) have used AI in their early studies either.

Activity Index could evaluate the relative research efforts a country devoted to a given sub-field, with considering the effect of the size of the country and also the size of the sub-field.

Mathematically, AI has been defined as follows:

$$\mathrm{AI} = \frac{\text{the country's share of the world's publication output in the given field}}{\text{the country's share of the world's publication output in all science fields}} \times 100$$

In order to study the development of research activities among the six countries during the period 1993–2002, we use a Transformative expression to calculate Activity Index, which was first suggested by PRICE (1981), and have application in the Bibliometrics of Alkaloid Chemistry research in India (KARKI & GARG, 1997). We call this Transformative Activity Index (TAI), and the expression of TAI is given below:

$$\mathrm{TAI} = \{(C_i/C_o)/(W_i/W_o)\} \times 100$$

where: C_i denotes the number of publications in a given field of a specific country in the i-th year,

C_o denotes the number of publications in a given field of a specific country during research period,

W_i denotes the number of publications in a given field of the whole world in the i-th year,

W_o denotes the number of publications in a given field of the whole world during research period.

TAI=100 indicates that a country's research effort in the given field corresponds precisely to the world average, TAI>100 reflects higher than average, and TAI<100 indicates lower than the average.

Exactly speaking, the Transformative Activity Index (TAI) defined in our article is a relative indicator. Referred to EGGHE & ROUSSEAU (2002), it can be expressed as: The share of articles published by authors of country *c* in the year *j* with respect to all articles published by authors of country *c*, divided by the share of all articles published in the year *j* (over the whole world) with respect to all articles published over the whole period (over the whole world). Thus, Transformative Activity Index (TAI) can be explained as an equivalent measure for relative research effort. Compared with the usual expression for Activity Index (AI), our TAI is more concentrated on the relative change on research effort during the given period of time. In total, it is a variation of the original AI and can also reflect the relative changes of research efforts.

The values of TAI during the period 1993–2002 for the countries considered are calculated and shown as Figure 2.

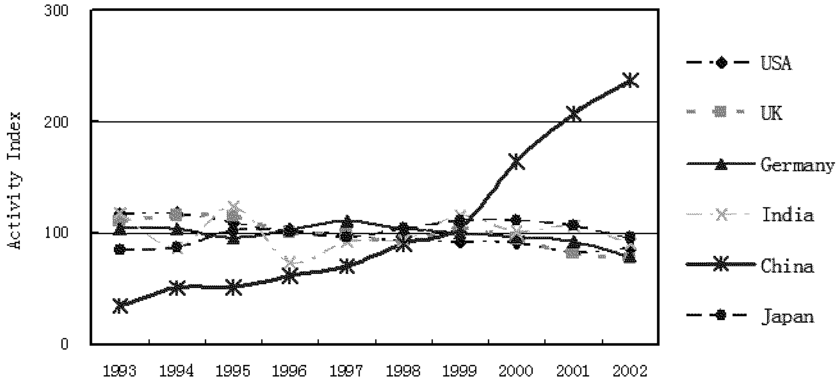

Figure 2. Trends of TAI during 1993–2002

Figure 2 indicates that except for China, the relative research efforts of other five countries are all roughly equal to the world average during 1993–2002, with slight fluctuations. Among them, USA, UK and Germany represent a tiny declined trend, while Japan and India keep fluctuating around 100 above and below.

Referring to the meaning of Transformative Activity Index, we can deduce that the relative research effort China devotes to its computer science has got really remarkable improvement in recent 10 years. The Activity Index of China in 2002 is 7 times as many as that in 1993.

Communication behavior and mainstream connectivity

There is no border between research activities of different countries or regions. In fact, the scientists, including the scientists from developing countries and non-English countries, are not willing to totally publish their research outputs in domestic journals. In order to promote communicating with other fellow scientists, many scientists prefer to publish their results in international journals and proceedings. To publish papers in international journals is not only to strengthen the impact of research outputs, but also help improving the development and advancement of all the research activities.

One of the research objectives is to analyze the publication distribution profiles for the compared countries, in order to evaluate the mainstream connectivity of research outputs in each country. The searching results show that in total more than 90% of publications in computer science are published in English language, either for English native-speaker countries and India or other occident countries, except for China. Therefore, we use the indicator of "country of publication" to evaluate mainstream connectivity of research output for the countries considered in the study, particularly for the comparison between India and China. At least in China, the domestic journals have been performing poor on mainstream connectivity and lack of international visibility (REN & ROUSSEAU, 2002)

The distribution data for 1997 and 2002 are given in Table 2. Table 2 indicates that, USA, UK as well as Germany are the main countries where international journals are numerously published and international conferences are more frequently held. This gave them great advantages of publishing their research outputs in international journals and conferences.

But for India and China, their domestic journals and conferences haven't been – to a greater extent – recognized by the international academia, yet. Thus, it makes more sense to study publication profiles of these two countries' in international journals and proceedings for assessing the mainstream connectivity of the research outputs.

From Table 2, we can conclude that, Indian scientists preferred to publish their research output in international journals and proceedings abroad, while Chinese scientists are more likely to publish in domestic ones. For India in 1997, the share of papers in domestic proceedings and journals accounted for 20 per cent and 80 percent of their total papers were appeared in the journals and proceedings published from abroad. The trend to publish abroad for India is further increased. Only 6.25 per cent publish domestically and 93.75% internationally in 2002. On the other hand, Chinese scientists have published 55 per cent articles domestically in 1997, and in 2002 the trend to publish domestically for China is further increased: the share takes 67 per cent. Even for those papers published in China, only 15 per cent were published in English in 1997, and further deceased to 8.1 percent in 2002.

Table 2. Distribution profiles of research output

Country of publication / Authors' affiliation	USA	UK	Germany	Japan	India	China
1997						
Domestic	2361	381	228	199	20	193
USA	–	251	242	222	38	96
UK	365	–	83	38	26	26
Germany	140	97	–	52	1	3
Netherlands	270	91	92	40	10	8
Others	76	49	64	37	5	23
Total	3212	869	709	588	100	349
2002						
Domestic	1248	154	127	131	4	506
USA	–	154	104	128	23	129
UK	198	–	33	30	16	55
Germany	125	56	–	43	5	15
Netherlands	226	51	44	34	11	36
Others	36	26	20	14	5	18
Total	1833	441	328	380	64	759

One possible reason for fewer shares of papers in international journals and proceedings published by Chinese scientists may be their inability to write articles in English that are of an acceptable standard in terms of English expression. In India, where English is an official language, its scientists have more advantages on writing and expressing in English than Chinese. This was also in accordance with the findings made by GARG (2002) in his research on laser S&T in India and China.

As was indicated by GARG (2002) and GARG & PADHI (2002), papers appeared in international journals and proceedings are better connected to mainstream science than those published in domestic journals and proceedings. For this reason, we can conclude that although the number of papers published by Indian scientists was less than that published by Chinese scientists, papers by Indian scientists on computer science are better connected to the mainstream science than those by Chinese scientists.

From this point, we suggest that Chinese scientists should pay more attentions to the communication channel used for publication, and strengthen the capability of writing and expressing in English. None but this, could they improve the mainstream connectivity and impact of the research outputs.

China's preference to publish research outputs in domestic journals and proceedings and in Chinese language has boomed the publication of its own domestic journals in the field of computer science, such as: Chinese Journal of Computers, Chinese Journal of

Advanced Software Research, *Ruan jian xue bao*, *Dian zi qi jian*, and many other transactions published by the universities, colleges, and research institutions. These publications in the field of computer science, however, still need improving their quality and raise their international impact.

Collaboration patterns

In modern society, science is no longer a matter of the individuals. A large proportion of big projects is completed by groups. Thus, it is necessary to study collaboration patterns of authors.

In the following context, we present the collaboration profiles of six countries in the field of computer science in 1997 and 2002 using different indicators like Co-authorship Index and Collaborative Coefficient.

Co-authorship Index (CAI)

Co-authorship Index has been used by GARG & PADHI (2001), and is obtained by calculating proportional output of single, two, multi and mega-authored papers in different periods and for different sub-fields of computer science. This methodology is very similar to the one suggested by PRICE (1981) and used to calculate Activity Index. Mathematically, Co-authorship Index can be described as follows:

$$CAI = \{(N_{ij}/N_{io})/(N_{oj}/N_{oo})\} \times 100$$

where: N_{ij} denotes the number of papers co-authored by j authors in the i-th country,
N_{io} denotes the total number of papers in the i-th country,
N_{oj} denotes the number of papers co-authored by j authors in all countries,
N_{oo} denotes the total number of papers in all countries.

CAI=100 indicates that the number of publications corresponds precisely to the average of all countries under one co-authorship pattern. CAI>100 reflects higher than the average, and CAI<100 indicates lower than the average.

Here, the papers have been divided into four categories according to their number of authors. They are: single authored papers, two authored papers, multi-authored and mega-authored papers. Multi-authored papers included papers with three and four authors, while mega authored papers included papers with five or more authors. Table 3 indicates the profiles of CAI for the compared countries.

An analysis of data indicates that the Co-authorship Index of single-authored and two-authored papers in three occident countries is more than the average level of six countries. That is to say, papers published by scientists in occident countries preferred to work separately or only with one collaborator.

Table 3. The profiles of CAI

Country \ Indicators	Co-Authorship Index (CAI)			
	Single Author	Two authors	Multi-author	Mega-author
1997				
USA	110.1	103.1	87.2	90.7
UK	108.3	107.0	88.0	76.0
Germany	107.6	92.1	101.1	97.1
India	54.9	113.2	152.0	143.0
Japan	60.4	81.0	141.5	205.0
China	50.9	98.1	160.7	86.3
2002				
USA	117	105.3	86.2	97.2
UK	117.1	104.1	94.3	65.6
Germany	145.2	90.1	81.6	98.7
India	80.8	76.0	118.5	155.2
Japan	48.7	185.6	53.8	126.5
China	43.4	93.9	139.3	97.5

Asian scientists preferred "multi-authors" and "mega-authors" these two collaboration patterns, that is to say they are more likely to work together with more than two collaborators. Japan's Co-authorship Index in the column of "mega-author" in 1997 has even reached 205, which was far more than the average number. This indicates that scientists in Japan are more willing to work in big groups. It is also in accordance with the impression that Japan usually gives us.

Table 3 also reflects that, surprisingly, Chinese scientists preferred to collaborate with 2~3 collaborators. For the Co-authorship Index under "multi-author" is 160.7 in 1997 and 139.3 in 2002. This contradicts to the traditional impression, on which Chinese prefer to do all work by oneself. Our finding concludes that Chinese scientists are more willing to work in small groups which usually have 3~4 scientists.

Collaborative Coefficient (CC)

Following from GARG (2002), this measure has been first suggested by AJIFERUKE et al. (1988) and is based on the fractional productivity originally defined by PRICE & BEAVER (1966).

Collaborative Coefficient can be shown as follows:

$$CC = 1 - \left\{ \sum_{j=1}^{k} (1/j) F_j / N \right\}$$

where: F_j denotes the number of papers finished by j authors,
N denotes the total number of papers.

According to the description made by AJIFERUKE et al. (1988), CC tends to zero as single authored papers dominate and to 1-1/j as j-authored papers dominate. This implies that the higher the value of CC, the higher is the probability of multi or mega authored papers.

From Collaborative Coefficient CC, we can derive the average weighted number of authors as follows

$$\bar{j} = 1/(1 - CC) = \frac{N}{\sum_{j=1}^{k} F_j/j}$$

This indicator may represent the distribution of the different collaboration patterns as a whole. Figure 3 delineates profiles of the average number of authors for the concerned countries.

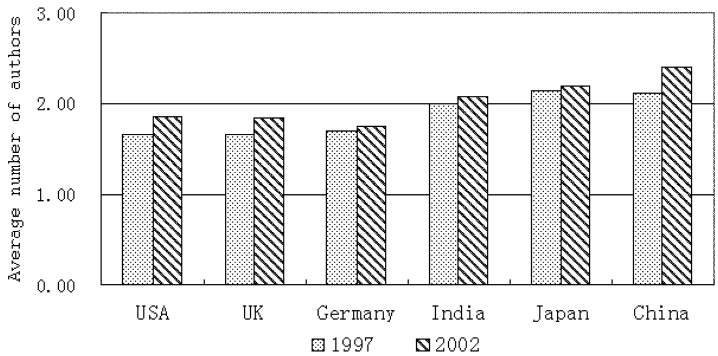

Figure 3. Profiles of the average number of authors in 1997 and 2002

As we can see in the above picture, the average number of authors for a scientific paper in the Asian countries is slightly higher than those of occident countries. We can primarily deduce that scientists in the field of computer science in Asian countries are preferred to work in groups, while those in occident countries are more willing to work separately or only with one collaborator. Again surprisingly, the average number of Chinese authors holds a highest position, particularly in 2002. This result just coincides with the finding mentioned before.

Impact of research output

It is quite clear that *Science Citation Index* is of the most importance as an international database with a unique and powerful method for its citations, which can provide a system of scientometric indicators based on number of publications, authors, references and citations (ZHANG & ZHANG, 1997). Thus, papers appeared in journals covered by the SCI and in journals with high impact of SCI can often indicate mainstream readership and mainstream connectivity.

Although SCI has many disadvantages and limitations (BORDONS et al., 2002), for example, it shows a considerable bias in favor of periodicals published in the USA and other English-language periodicals, and against periodicals with non-Roman alphabets, it is still the most comprehensive and frequently used source database for a broad review of international science.

SCI and non-SCI covered papers

1) The papers counts indexed by SCI. Table 4 describes the counts of papers in computer science indexed by SCI for the compared countries and for 1997 and 2002. It indicates that, the number of SCI-papers published by the scientists of USA is far more than other five countries either in 1997 or 2002.

Table 4. Counts distribution of SCI-papers in computer science

Country / Year	USA	UK	Germany	Japan	India	China
1997	795	172	129	79	33	39
2002	635	135	82	85	38	118

This result certainly relates to the limitations of SCI, but it was also to a great extent due to the high level of research outputs of USA. The number of SCI-papers by UK is simply equal to that of Germany's, and Japan consecutively follows behind. India and China both lie in the last two positions. Noticeably, China's number of SCI-papers has been growing fast in the recent 5 years. Until 2002, it has increased to 118, which is near to UK's and already beyond Germany and Japan. In other words, the research outputs in the field of computer science of China has gradually gained more and more mainstream readership, and been more and more recognized by the mainstream science of the world in basic research of computer science.

2) The ratio of SCI-papers to total journal papers. First, we delete all papers of conference proceedings from data base because almost no paper appeared in conference proceedings in computer science could be indexed by SCI. Table 5 delineates the ratio distribution of SCI-papers to total journal papers. It indicates that, in the field of computer science, the ratio of papers appeared in SCI covered journals to total journal papers in three occident countries are all rather high, which are on average above 50 per cent.

Table 5. The ratio distribution of SCI-papers to total journal papers

Year \ Country	USA	UK	Germany	Japan	India	China
1997	54.16	48.18	54.66	26.87	50.00	15.98
2002	62.38	54.00	58.57	37.61	76.00	17.91

Among three Asian countries, in 1997, about 50 per cent of journal papers published by Indian scientists appeared in SCI-covered journals, while China's share in SCI covered journals took only 16 per cent. The difference was even larger in 2002. In 2002, India's share of SCI-papers increase to 76 per cent, whereas China's share still remains 18 per cent, not even exceeded 20 per cent . That is, China still holds the lowest ranking among the six countries after five years past since 1997.

Generally speaking, we cannot conclude from the absolute number of papers appeared in SCI covered journals that, the impact of China's research outputs has already been comparable to or over that of Germany's. In the ratios of SCI-papers to total journal papers, however, China ranks the last in six countries, and has no notable growth in the period 1997–2002. From analyzing Chinese journals covered by ISI, REN & ROUSSEAU (2002) found that Chinese scientific journals, even covered by ISI, are still lack of international visibility. Our finding from analyzing the ratio of SCI-papers to total journal papers supports their conclusion. This also indicates that the research activities of Chinese scientists in computer science are also still rather "local" and suffer from a low international visibility, as Chinese scientific journals (REN & ROUSSEAU, 2002). The papers published by Chinese scientists are still less connected to the mainstream science in the field of basic research of computer science. Therefore, the research of computer science of China needs to get more mainstream readership through publishing more papers in international journals covered by SCI. In addition, an important reason for less proportion of papers appeared in SCI-covered journals might be China's preference to publish in domestic journals, many of which are not covered by SCI.

Impact Factor and Normalized Impact Factor

Impact Factor (IF) is an important indicator describing the quality of journals. It is determined by the average citation rate of each papers received by this journal. The Impact Factor of a journal was used in the literature as a measure of expected citations for each paper published in it, which is as an indirect measure of its quality and impact (BORDONS et al., 2002).

There are several objections raised against the use of impact factors. First, impact factor patterns vary among fields. For example, IFs are far higher in molecular biology and genetics than in material science and mathematics. As VAN LEEUWEN et al. (1999) suggested in their study, a critical comment about Impact Factor is that the distribution of citations within a journal is much skewed. This shows that a large portion of papers in a journal received no citations at all. The second aspect comes from the coverage of ISI database. It does not cover all significant scientific journals, particularly for those non-central countries (BORDONS et al., 2002). Third, journals from non-English-speaking countries are not as well represented as those from English-speaking countries in ISI database (MAY, 1997). Therefore, it remains an open question whether the comparison, which is based on NIF that is deducted by IF, introduces biases.

However, Impact Factor's validity was supported by the strong negative correlation between the value of the IF and the degree of uncitedness. Despite the limitations and shortcomings of citation analysis, the evaluation method relied on IF is still widely used with certain improvement and advantages (BORDONS et al., 2002). BORDONS et al. (2002) stress the need to avoid inter-field comparisons of using impact factor, which is most commonly used by the bibliometric community. Since our evaluation is focused on computer science, the limitation that IF varies sharply across sub-disciplines can be to a great extent avoided. Particularly, in order to eliminate the effect of different sub-disciplines, we will use the Normalized Impact Factor as a surrogate measure to evaluate research quality of the countries considered, following Garg's approach (GARG, 2002). Thus, we could use Impact Factor of the journal as a proxy measure to evaluate the impact and quality of each paper published in it.

In the *Journal Citation Report* published by ISI, journals are categorized into different large disciplines. Journals' Impact Factor differentiates largely according to their discipline. Since the papers on computer science are published in a variety of journals dealing with different subject fields like mathematics, electronics, electronic engineering and so on, so we have to consider the effect of the disciplines. For example, the journals in computer science usually have Impact Factors ranged from 1 to 3, and highest to 5 at most. But some journals, such as *Nature* and *Science*, have Impact Factor that could be more than 20 or 25. The value two as impact factor may refer to a top journal in one field, and to a mediocre one in another domain (EGGHE & ROUSSEAU, 2002).

In order to eliminate the effect of different disciplines, we use the Normalized Impact Factor as a surrogate measure, which seems to be a plausible indicator to reflect relative importance degree within the discipline. The procedure was first suggested by SEN (1992) and used by GARG & PADHI (1999) in their earlier study on laser S&T. Normalized Impact Factor can be described as follows:

$$NIF_{ij} = \{GIF_{ij} / MAX(GIF_j)\} \times 10$$

where: NIF_{ij} denotes the Normalized Impact Factor of the journal i in sub-field j,
GIF_{ij} denotes the Garfield's Impact Factor of the journal i in sub-field j,
MAX (GIF_j) denotes the highest Impact Factor in the j-th set of journals used for publication.

Similar to Garg's study on laser S&T (GARG, 2002), we extend to divide NIF into five categories. They are: NIF≤0.100(very low), 0.100<NIF≤1.000(low), 1.000<NIF≤2.000(medium), 2.000<NIF≤3.000(high), NIF>3.000(very high).

According to the above regulations, we count the number of papers indexed by SCI in each NIF category respectively, and calculate the papers' percentage of each category. Table 6 describes the distributions.

Based on the category distributions of papers according to the Normalized Impact Factor of the journals, it is observed that about one-third of the Chinese papers appeared in the journals with low Normalized Impact Factor, both in 1997 and 2002, while in case of USA about one-fifth of the papers were published in low quality journals.

Table 6. The NIF distributions of papers indexed by SCI

Country NIF	USA (%)	UK (%)	Germany (%)	Japan (%)	India (%)	China (%)
1997						
≤ 0.100 (very low)	20 (2.52)	5 (2.91)	7 (5.43)	15 (18.99)	1 (3.03)	0 (0)
> 0.100 ≤ 1.000(low)	138 (17.36)	48 (27.91)	27 (20.93)	30 (37.97)	13 (39.39)	14 (35.90)
> 1.000 ≤ 2.000(medium)	227 (28.55)	61 (35.47)	49 (37.98)	15 (18.99)	9 (27.27)	13 (33.33)
> 2.000 ≤ 3.000(high)	170 (21.38)	29 (16.86)	33 (25.58)	8 (10.13)	7 (21.21)	9 (23.08)
> 3.000 (very high)	240 (30.19)	29 (16.86)	13 (10.08)	11 (13.92)	3 (9.09)	3 (7.69)
Total	**795 (100)**	**172 (100)**	**129 (100)**	**79 (100)**	**33 (100)**	**39 (100)**
2002						
≤ 0.100 (very low)	3 (0.47)	2 (1.48)	1 (1.22)	0 (0)	2 (5.26)	0 (0)
> 0.100 ≤ 1.000(low)	151 (23.78)	31 (22.96)	23 (28.05)	41 (48.24)	16 (42.11)	41 (34.75)
> 1.000 ≤ 2.000(medium)	131 (20.63)	42 (31.11)	25 (30.49)	19 (22.35)	9 (23.68)	37 (31.36)
> 2.000 ≤ 3.000(high)	142 (22.36)	37 (27.41)	13 (15.85)	7 (8.24)	3 (7.89)	15 (12.71)
> 3.000 (very high)	208 (32.76)	23 (17.04)	20 (24.39)	18 (21.18)	8 (21.05)	25 (21.19)
Total	**635 (100)**	**135 (100)**	**82 (100)**	**85 (100)**	**38 (100)**	**118 (100)**

The share of papers in very high quality Normalized Impact Factor journals for USA is also the highest in all six countries. However, the share of Chinese papers in high quality Normalized Impact Factor journals grows fast in the research period, from 7.69 per cent in 1997 to 21.19 percent in 2002. This indicates an optimistic trend.

As suggested by GARG (2002), there are some other surrogate measures used for evaluating the impact of research outputs. They are respectively as follows:

Normalized Impact Factor per Paper. The indicator of Normalized Impact Factor per Paper was based on Normalized Impact Factor, and is calculated as follows:

$$\left(\sum P_i \times \sum F_i\right)/N$$

where: P_i denotes the number of papers published in journal i,
F_i denotes the Normalized Impact Factor in journal i, where paper was published,
N denotes the total number of papers.

Proportion of Papers in High Quality Journals (PHQ). According to Garg's suggestion in his study, papers with a Normalized Impact Factor>3 have been considered as high quality ones.

Publication Effective Index (PEI). This measure indicates whether the impact of publications of a country in a research field is commensurate with the publication effort devoted to it. The value of PEI>1 for a country indicates that the impact of publications is more than the research effort devoted to it for that particular country and vice versa. This indicator is the ratio of the proportion of impact (TNIMP%) to the proportion of publications (TNP%). Formula can be shown as follows:

$$\mathrm{PEI} = \frac{\mathrm{TNIMP}_i / \mathrm{TNIMP}_t}{\mathrm{TNP}_i / \mathrm{TNP}_t}$$

where: TNIMP_i denotes the total Normalized Impact Factor of country i,
TNIMP_t denotes the total Normalized Impact Factor of all countries,
TNP_i denotes the total number of papers of country i,
TNP_t denotes the total number of papers of all countries.

The values of different surrogate measures used for comparing the impact of research outputs for the six nations are given in Table 7.

Based on different surrogate measures of quality, it is observed that the value of NIMP/Paper for USA was the highest either in 1997 or in 2002, and UK has been kept above average in both years. The other four countries have all achieved – to different extents – progress on this indicator, where Germany has increased from 1.721 (below average) to 2.085 (above average), and China has already been near to the average value as well.

For the indicator PHQ, USA also achieved the highest share, and has been kept above 30 per cent in both years. However, the other five countries have considerably developed in this indicator all the same. Among them, China is the most exceptional one, which has grown from 7.69 per cent to 21.19 per cent and already exceeded the values of Japan's and India's. This reflects that the research quality of China in computer science has considerably increased, although Chinese scientists in computer science still strongly prefer to publish domestically.

Table 7. Impact indicators of the six countries

Indicator / Country	TNP	TNIMP	NIMP/Paper	PHQ	PEI
1997					
USA	795	2071	2.605	30.19	1.140
UK	172	334	1.942	16.86	0.850
Germany	129	222	1.721	10.8	0.753
Japan	79	108	1.367	13.92	0.598
India	33	53	1.606	9.09	0.703
China	39	62	1.590	7.69	0.696
Total	**1247**	**2850**	**1.805**	**14.758**	
2002					
USA	635	1572	2.476	32.76	1.099
UK	135	276	2.050	17.04	0.910
Germany	82	171	2.085	24.39	0.926
Japan	85	149	1.753	21.18	0.778
India	38	65	1.724	21.05	0.765
China	118	228	1.932	21.19	0.858
Total	**1093**	**2462**	**2.003**	**22.93**	

Based on the value of PEI, it is observed that only USA's research effort was commensurate with its publication efforts. But it is observed that, up to 2002, the value of UK and Germany's have very approximated to 1. China has also got great improvement on this indicator, and it indicates China's research effort has gradually increased to approximate to its publication effort.

Conclusions and implications

Based on the foregoing analysis, the following conclusions are derived:

1. Optimistically, the changes that have occurred in recent years on the Chinese computer science research are having noticeable effects on world's research output profile. China's output of research in the field of computer science has

been growing fast in the recent 10 years, and kept positive growth rate for nine out of the 10 years considered, while the other five countries all have a relative pessimistic declined trend.

2. The Activity Index of China in 2002 has been 7 times as many as that in 1993. This indicates China has really got great improvement on either relative research effort or absolute number of publications in the field of computer science. The Activity Indexes for the other five countries during the 10 years almost follow the same trend. That is, the changing increments of Activity Indexes for the other five countries fluctuate between positive and negative in alternative years.
3. Scientists in China preferred to publish a large proportion of research outputs in domestic journals and proceedings, while scientists of India, a similar science-developing country to China, preferred to publish a major portion in the scientific journals and proceedings from developed countries of the west. To some extent, this could indicate that the research output from Indian scientists is better connected to the mainstream science and more recognized by international academia.
4. With respect to collaboration profile, the scientists in occident countries are more likely to work independently or only prefer to work with one collaborator, while Asian scientists preferred to work in big groups that usually have more than three scientists.
5. Based on the *Science Citation Index*, we found that China's number of papers from journals covered by SCI has considerably increased in the past 5 years. This implies China is being approximated to the mainstream science and getting more mainstream readership. But, the proportion of SCI-papers to the total journal papers indicates that China ranked the last in all six countries. This indicates that the research activities of Chinese scientists in computer science are still rather "local" and suffer from a low international visibility. Thus, China was still in a relative low level in the field of computer science in the international context.

Based on different surrogate measures of impact, it is observed that, China has showed characteristics of "low level of beginning and high speed of developing" on all aspects.

As a whole, the research outputs from scientists in China have generally lacked the original innovation, and still remains in a relatively low level. Although some of the research outputs by Chinese scientists have been already accepted by the mainstream international academia, they are still way behind the major occident's countries as well as some Asian countries such as Japan and even India.

The policy implication of the findings lies in that China, as well as other less developed countries in science, can earn relative competitive advantages in some new emerging or younger disciplines such as computer science by properly using catch-up strategy.

*

This research was funded by National Natural Science Foundation of China (Project No.70172026) and Department of Information Science of NSFC. The authors wish to express their appreciation to the anonymous reviewer and Prof. Tibor Braun for their helpful comments and suggestions of this article. The authors also wish to gratefully thank Prof. Ronald Rousseau of KHBO, Industrial Sciences and Technology, Belgium, for his constructive suggestions and many comments to the paper, which made us significantly improve this paper. The authors bear certainly responsibility for any remaining shortcomings.

References

AJIFERUKE, I., BURRELL, Q., TAGUE, J. (1988), Collaborative coefficient: A single measure of the degree of collaboration in research, *Scientometrics*, 14 : 421–433.

Annual Report of Science and Technology Development of China (By Research Group on Development and Strategy of S&T of China, 2002), Beijing: Economic Management Publishing House (in Chinese), 148–152.

ARUNACHALAM, S., SINGH, U. N., SINHA, R. (1993), The sleeping dragon wakes up: A scientometric analysis of the growth of science and the usage of journals in China, *Current Science*, 65 : 809–822.

BORDONS, M., FERNANDEZ, M. T., GOMEZ, I. (2002), Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country, *Scientometrics*, 53 (2) : 195–206.

BRAUN, T., SCHUBERT, A., GLÄNZEL, W. (1985), *Scientometric Indicators: A 32 Countries Comparative Evaluation of Publishing Performance and Citation Impact.* World Scientific Publishing, Singapore.

BRAUN, T., GLÄNZEL, W., SCHUBERT, A. (1990), Publication productivity: From frequency distributions to scientometric indicators, *Journal of Information Science*, 16 : 37–44.

CARPENTER, M. P., NARIN, F. (1981), The adequacy of the Science Citation Index (SCI) as an indicator of international scientific activity, *Journal of American Society for Information Science*, 32 : 430–439.

EGGHE, L., ROUSSEAU, R. (2002) A general frame-work for relative impact indicators, *Canadian Journal of Information and Library Science*, 27 (1) : 29–48.

FRAME, J. D. (1977) Mainstream research in Latin America and Caribbean, *Interciencia*, 2 : 143–198.

FRAME, J. D., NARIN, F. (1987), The growth of Chinese Scientific research 1973-84, *Scientometrics*, 12 : 135–144.

GARG, K. C. (2002), Scientometrics of laser research in India and China, *Scientometrics*, 55 (1) : 71–85.

GARG, K. C., DUTT, B. (1992), Bibliometrics of Indian Science as reflected through Science Citation Index, *Journal of Scientific and Industrial Research,* 51 : 329–340.

GARG, K. C., PADHI, P. (1998), Scientometrics study of laser patent literature, *Scientometrics*, 43 : 443–454.

GARG, K. C., PADHI, P. (1999), Scientometrics of laser research literature as viewed through the Journal of Current Laser Abstract, *Scientometrics*, 45 : 251–268.

GARG, K. C., PADHI, P. (2001), A study of collaboration in laser science and technology, *Scientometrics*, 50 (2) : 415–427.

GARG, K. C., PADHI, P. (2002), Scientometrics of laser research in India during 1970–1994, *Scientometrics*, 55 (2) : 215–241.

GUAN, J. C., WANG, J. X. (2004), Evaluation and interpretation of knowledge production efficiency, *Scientometrics*, 59 (1) : 131–155.

IRVINE, J., MARTIN, B., PEACOCK, T., TURNER, R. (1985), Charting the decline in British science, *Nature*, 316 : 587–590.

International Information Service in Physics, Electro-technology, Computer Control [DB/OL], http://www.ovid.com

Journal Citation Report, ISI, http://www.webofkonwledge.com

KING, J. (1989), A review of bibliometric and other science indicators and their role in research evaluation, *Journal of Information Science*, 13 : 261–276.

KARKI, M. M. S., GARG, K. C. (1997), Bibliometrics of Alkaloid Chemistry research in India, *Journal of Chemical Information and Computer Science*, 37 : 157–161.

LU, Q., LAZONICK, W. (2001), The organisation of innovation in a transitional economy: business and government in Chinese electronic publishing, *Research Policy*, 30 : 55–77.

MAY, R. M., (1997), The scientific wealth of nations, *Science*, 275 (5301) : 793–796.

NAGPAUL, P. S. (1995), Contribution of Indian university to the mainstream scientific literature: A bibliometric assessment, *Scientometrics*, 32 : 11–36.

NARIN, F., FRAME, J.D. (1989), The growth of Japanese science and technology, *Science*, 245 : 600–605.

PRICE, D. DE. S. (1981), The analysis of scientometric metrics for policy implications, *Scientometrics*, 3 : 47–54.

PRICE, D. DE. S., BEAVER, D. DE. B. (1966), Collaboration in an invisible college, *American Psychologist*, 3 : 1011.

VAN RAAN, A. F. J., Bibliometric indicators as research performance evaluation tools, In: *Proceedings of the European University Institute Conference on 'Research Management in Europe Today*', Florence, 13-15 December 1990.

REN, S., ROUSSEAU, R., (2002), International visibility of Chinese scientific journals, *Scientometrics*, 53 (3) : 389–405.

SCHUBERT, A., BRAUN, T. (1986) Relative indicators and relational charts for comparative assessment of publication output and citation impact, *Scientometrics*, 9 : 281–291.

SEN, B. K. (1992), Normalized Impact Factor, *J. Documentation,* 48 : 318–329.

VAN LEEUWEN, TH. N., MOED, H. F., REEDIJK, J. (1999), Critical comments on Institute for Scientific Information impact factors: a sample of inorganic molecular chemistry journals, *Journal of Information Science*, 25 (6) : 489–498.

ZHANG, H. Q., ZHANG, Y. H. (1997), Scientometric study on research performance in China, *Information Processing & Management*, 33 (1) : 81–89.