CITATION MINING:

Citing Population Profiling using Bibliometrics and Text Mining

By

Dr. Ronald N. Kostoff
Office of Naval Research
Arlington, VA, USA

Dr. J. Antonio del Río
Centro de Investigación en Energía
Universidad Nacional Autónoma de México
Temixco, Mor. México

Mr. James A. Humenik
NOESIS, Inc.
Rockville, MD, USA

Lic. Esther Ofilia García
Centro de Investigación en Energía
Universidad Nacional Autónoma de México
Temixco, Mor. México

Lic. Ana María Ramírez
Centro de Investigación en Energía
Universidad Nacional Autónoma de México
Temixco, Mor. México

I. ABSTRACT

Objective: Proof-of-principle demonstration that breadth of research impact on the R&D community can be documented objectively and efficiently, and that documented research

user characteristics can be obtained efficiently to help guide future directions of sponsors' R&D.

Approach: The novel concept of Citation Mining, a combination of citation bibliometrics and text mining, is used for the demonstration. Citation Mining starts with a group of core papers whose impact is to be examined, retrieves the papers that cite these core papers, then analyzes the bibliometrics characteristics of the citing papers as well as their linguistic and thematic characteristics. The Science Citation Index is used as the source database for the core and citing papers, since its citation-based structure enables the capability to perform citation studies easily.

Results: This paper presents illustrative examples in photovoltaics (applied research) and sandpile dynamics (basic research) to show the types of output products possible. Bibliometric profiling is performed over a number of the citing papers' record fields to offer different perspectives on the citing (user) community. Text mining is performed on the aggregate citing papers, to identify aggregate citing community themes, and to identify extra-discipline and applications themes.

The photovoltaics applied research papers had on the order of hundreds of citations in aggregate. All of the citing papers ranged from applied research to applications, and their main themes were fully aligned with those of the aggregate cited papers. This is typically the case with applied research.

One of the sandpile dynamics basic research papers examined had on the order of 300 citations. Most of the citing papers were basic research whose main themes were aligned with those of the cited paper. This is typically the case with basic research. However, about twenty percent of the citing papers were research or development in other disciplines, or development within the same discipline.

Technical phrases in all ~300 citing paper Abstracts were extracted using text mining software, and correlated with citing paper Abstract themes obtained by reading all ~300 Abstracts. All of the citing paper Abstract themes that were not basic research and/ or were not aligned with the main themes of the cited paper had been previously identified using the text mining alone.

Conclusions: The combination of citation bibliometrics and text mining provides a synergy unavailable with each approach taken independently. Furthermore, text mining is a REQUIREMENT for a feasible comprehensive research impact determination. The integrated multi-generation citation analysis required for broad research impact determination of highly cited papers will produce thousands or tens or hundreds of thousands of citing paper Abstracts. Text mining allows the impacts of research on advanced development categories and/ or extra-discipline categories to be obtained without having to read all these citing paper Abstracts. The multi-field bibliometrics provide multiple documented perspectives on the users of the research, and indicate whether the documented audience reached is the desired target audience.

II. INTRODUCTION

The high technology component of the global economy is experiencing rapid growth. As a result, the qualitative and quantitative evaluation of research and development (R&D) investments is becoming increasingly important for different governments.

For example, in 1995, the total United States expenditure for research and development (R&D) was estimated to be 171 billion dollars ($171B), in current dollars [NSF, 1996]. The Federally-sponsored component was about $61B, and the industrially-sponsored component was about $102B. The total U. S. R&D expenditure was less than the

aggregate six next largest foreign R&D sponsors. Of the U. S. total R&D expenditures, about $30B was categorized as basic research, $40B as applied research, and $101B as development. Thus, global R&D expenditures are in the neighborhood of $450B annually, and growing.

Concomitant with these expenditure levels is the increasing motivation of the R&D sponsors to improve accountability. In the U. S., one example of this increased accountability was the passage of the Government Performance and Results Act (GPRA) of 1993. GPRA requires the Federal agencies to submit strategic plans, annual performance plans, and metrics to gauge progress toward the goals of the plans [GPRA, 1993]. Selection of metrics, especially for research evaluation and impact assessment, is very difficult. Combination of research evaluation tools and techniques, both qualitative and quantitative, may be necessary to provide a balanced picture of the value and impact of research [Kostoff, 1997].

The advent of very large R&D narrative databases, and more importantly the computer hardware and software that allow these database contents to be accessed and analyzed readily, offer the potential for comprehensive and credible quantitative analysis of research impact. In particular, the enhanced coverage of the research literature by the Web version of the Science Citation Index (SCI-~5300 leading research journals) allows a broad variety of bibliometric analyses of R&D units (papers, researchers, journals, institutions, countries, technical areas) to be performed.

However, only a small fraction of the potential myriad capabilities offered by the SCI has been, and is being, exploited. The main use of the SCI, as evidenced by many articles from past issues of Scientometrics, The Scientist, Science Watch, Current Contents, some of the information technology journals, and many of the academic medical and physical and chemical journals, has been to

compute numbers of citations received by the different R&D units mentioned above.

Use of the other SCI fields for bibliometric analysis has been published on a very sporadic basis, and typically only for one or two fields. In addition, when these few studies have been performed, the focus has been on relating citations or citation rates to these field variables. There do not appear to have been any citation studies performed for the specific purpose of user population profiling, where many of the available fields are examined in an integrated manner.

For example, a recent study [Steele, 2000] investigated the relationship between an article's citation rate and its degree of inter-disciplinarity (recognized-authorship, subject matter, and cited literature) in forestry. These limited other variables were examined in the context of their influence on citation rate, not for user profiling. In another study [Herring, 1999], a citation analysis was conducted to measure levels of interdisciplinary research and publishing in Internet search engine design and development.

A third study [Davidse, 1997] presented the results of an exploratory bibliometric study aiming at an analysis of basic high energy physics (HEP) research impact on fields other than physics, and particularly on application-oriented R&D. The paper focused on the 'knowledge flow' from physics to non-physics, and more specifically the flow from basic physics research to the 'applied world'. Journal, as well as research field, characteristics were reported, and the most frequently citing R&D groups were identified.

While the aggregation of citation number counts characteristic of almost all published citation studies identifies R&D units that have had (and have not had) impact on the user community, its potential impact on

decision-making is limited. Exploitation of the other types of information contained in the SCI and associated with the citation process offers the potential of providing R&D sponsors information that can help guide future directions of their R&D. In addition, the complete citation mining process described in the present paper has the potential to objectively document the impact of basic research on the R&D community. Use of the total process could help answer questions such as:

***What types of people and organizations are citing the research outputs; is this the desired target audience?**
***What development categories are citing the research outputs?**
***What technical disciplines are citing the research outputs?**
***What are the relationships between the citing technical disciplines and the cited technical disciplines?**

Examination of these other types of information available from the citation process is the focus of the present paper. The aim of the study is to demonstrate the power and capability of this new approach to citer profiling. It is necessary to stress that the aim is not to assess the productivity and magnitude of impact of any individual researcher, research group, laboratory, institution, or country. To perform such an assessment, the authors would need a charter and statistically representative data based on the unit of assessment. After the value and benefits of the method have been shown to the readers by the present study, then more detailed studies by the appropriate individuals/ institutions can focus on specific individual or comparative assessments.

In addition, text mining of the cited and citing papers will be performed in the present study, to supplement the information derived from the semi-structured field bibliometric analyses. Text mining will illuminate the trans-citation thematic relationships, and provide

insights of knowledge diffusion to intra-discipline research, advanced intra-discipline development, and extra-discipline research and development. The addition of text mining to citation bibliometrics will make feasible the large-scale multi-generation citation studies that are necessary to display the full impacts of research.

In the remainder of this paper, the other types of possible citation information from the SCI (bibliometric profiling) are identified, the scope of analysis for the present study is defined, sample computations are presented to provide some indication of what is possible from bibliometric profiling and trans-citation text mining, and conclusions from the specific analyses performed and from the experience with the evaluation process are presented. There is no claim on completeness for this initial study. The study results will help identify both additional studies that could be undertaken and lessons learned for improving these future studies based on the experiences from the present effort. Recommendations for future efforts in this area will be presented.

III. ANALYSIS

This section identifies the types of data contained in the SCI (circa early 2000), and the types of analyses that will be performed on this information.

III-A. Types of Data in SCI Fields

Figure 1 shows a record from the SCI, without the field tags. The actual paper that it represents is referred to in the following description as the 'full paper'. Starting from the top, the individual fields are:

1) Title - the complete title of the actual paper (Individual single wall ...).

2) Authors - all the authors of the actual paper (Tans SJ, Devoret, MH, ...). The full paper tends to provide a complete first name, while this record field provides only the first initial. Also, some authors may provide middle initials in some papers but not in others, resulting in different characterizations in this field for the same author in different papers. The same author may also change names (after marriage, for example), again resulting in different characterizations in this field for the same author in different papers. Finally, this field does not distinguish among different authors with the same name, resulting in confusion when trying to track specific authors with common names.
3) Source - journal name (Nature).
4) Issue/ Page(s)/ Publication Date - (386 6624 47 ...).
5) Document Type - (Article, notes, review).
6) Language - the language of the full paper, not necessarily the Abstract language (English).
7) Cited References - the number and names of the references cited in the full paper (see Figure 2). Typically, only some of the references are contained as full papers in the SCI. For those references not contained as SCI records, very skimp information is provided, and their associated paper must be obtained to provide the details.
8) Times Cited - the number and names of the papers (whose records are contained in the SCI) that cited the full paper. Thus, the number provided by this field is a lower bound. More importantly, many of the broader audience journals, that may discuss applications of the research, will not be represented here. For example, consider a magazine such as Popular Science, or Popular Mechanics, that may have an article written by a science reporter on the use of nano-tubes for packaging drugs or bio-warfare agents or for microcircuit wires. This article may reference the full paper on nano-tubes, but it wouldn't be represented in the SCI, and would not be available for the type of user profiling discussed in the present paper.

9) Related Records - not shown. This field lists the number and names of the papers (whose records are contained in the SCI) that share one or more common references with the full paper. They are listed in descending order by number of shared references, and then by publication date. While this field is a good starting point for identifying records related to the full paper record, it is not a unique indicator of relatedness. The authors have seen related papers that share no common references, and have seen papers that share common references that are not related. Typically, if papers share a few references beyond some threshold fraction, they tend to be related.
10) Abstract – the complete Abstract from the full paper. Not all records contain Abstracts. Not all Abstracts are very informative.
11) Keywords Plus – keywords supplied by the indexer. Some records have an additional field named Author Keywords, which contains author-supplied keywords. Not all records contain keywords of any type. Great care must be used when interpreting keywords. They are at a different level of abstraction from the Abstracts, provide a different perspective, and are used for different (sometimes non-technical) purposes than the full paper or Abstract [Kostoff, 1999].
12) Addresses – organizational and street addresses of the authors. For multiple authors, this can be a difficult field to interpret accurately. Different authors from the same organizational unit may describe their organizational level differently. The same organizational unit may be abbreviated differently by different authors. In the example below (Figure 1), there are seven authors, but only five separate addresses. The correspondence of authors to addresses is unclear.
13) Publisher – the publisher of the journal in which the full paper appeared.
14) Other fields.

FIGURE 1 – SCI RECORD

Figure 1

"Individual single-wall carbon nanotubes as quantum wires"
Tans SJ, Devoret MH, Dai HJ, Thess A, Smalley RE, Geerligs LJ, Dekker C
NATURE
386: (6624) 474-477 APR 3 1997
Document type: Article
Language: English
Cited References: 16
Times Cited: 259

Abstract:
Carbon nanotubes have been regarded since their discovery(1) as potential molecular quantum wires. In the case of multi-wall nanotubes, where many tubes are arranged in a coaxial fashion, the electrical properties of individual tubes have been shown to vary strongly from tube to tube(2,3), and to be characterized by disorder and localization(4). Single-wall nanotubes(5,6) (SWNTs) have recently been obtained with high yields and structural uniformity(7). Particular varieties of these highly symmetric structures have been predicted to be metallic, with electrical conduction occurring through only two electronic modes(8-10). Because of the structural symmetry and stiffness of SWNTs, their molecular wavefunctions may extend over the entire tube. Here we report electrical transport measurements on individual single-wall nanotubes that confirm these theoretical predictions. We find that SWNTs indeed act as genuine quantum wires, Electrical conduction seems to occur through well separated, discrete electron states that are quantum-mechanically coherent over long distance, that is at least from contact to contact (140 nm), Data in a magnetic field indicate shifting of these,states due to the Zeeman effect.

KeyWords Plus: **MICROTUBULES, TUBULES** **Addresses:** **DELFT UNIV TECHNOL, DEPT APPL PHYS, NL-2628 CJ DELFT, NETHERLANDS.** **DELFT UNIV TECHNOL, DIMES, NL-2628 CJ DELFT, NETHERLANDS.** **RICE UNIV, RICE QUANTUM INST, CTR NANOSCALE SCI & TECHNOL, HOUSTON, TX 77251.** **RICE UNIV, DEPT CHEM, HOUSTON, TX 77251.** **RICE UNIV, DEPT PHYS, HOUSTON, TX 77251.** **Publisher:** **MACMILLAN MAGAZINES LTD, LONDON** **IDS Number:** **WR256** **ISSN:** **0028-0836**

Figure 2 shows some of the cited references contained in the record of Figure 1. Some of the references will themselves be contained in the SCI, whereas other references will not.

FIGURE 2 – CITED REFERENCES IN ABOVE RECORD

Cited References				
Individual single-wall carbon nanotubes as quantum wires				
Tans SJ, Devoret MH, Dai HJ, et al.				
NATURE				
386: (6624) 474-477 APR 3 1997)				
Cited Author	Cited Work	Volume	Page	Year
BETHUNE DS	NATURE	363	605	1993
DAI HJ	SCIENCE	272	523	1996

	SCI FULLERENES			1996
DRESSELHAU	CARBO	382	54	1996
S MS	NATURE			1993
EBBESEN TW	PHYSICS FEW			1992
GEERLIGS IJ	ELECT NA	68	1579	1992
GRABERT H	SINGLE CHARGE	363	603	1993
HAMADA N	TUNNEL			
IIJIMA S	PHYS REV LETT			
	NATURE			

III-B. Types of Analyses Possible.

For citing population profiling, the overall study objectives must be defined first. This will determine the population to be cited, the fields of the cited records to be analyzed, and the statistical requirements for the citing population. The population to be cited could be an individual paper, the papers from a single author, the papers from an organizational unit, the papers from a technical discipline, the papers from a country, and/ or various combinations of the above. The fields of the citing records to be analyzed, including the types of information and analyses of potential interest, could include the following.

1) Title - text mining of the Abstract field could be performed to relate the themes in the cited paper to those of the citing papers. For those citing papers that do not contain Abstracts, the Title field could be substituted.

2) Authors - multi-author distribution profiles could be computed (e.g., number of papers with one author, number with two authors, etc). Also, if the disciplines of each author are known, discipline distribution profiles could be computed.

3) Source - journal name distributions, theme distributions, and development level distributions could

be computed (e.g., Science contains three citing papers, Nature contains four citing papers, etc; three journals are chemistry oriented, four are physics oriented, five are materials oriented, etc; three journals are basic research, four journals are applied research, five journals are early technology development, six journals are advanced technology development, etc)

4) Issue/ Page(s)/ Publication Date – time distribution of citing papers could be computed (e.g., three citing papers published six months after full paper publication, two published nine months after, four published fifteen months after, etc)

5) Document Type – distribution of different document types could be computed (e.g., three articles, four conference proceedings, etc)

6) Language – distribution over languages could be computed (e.g., three papers in English, four papers in Spanish, five papers in French, etc)

7) Cited References – A historical analysis of the problem can be performed from this field.

8) Times Cited – if the citing papers are of sufficient vintage, then their multiplier effect would be of interest, and could be computed. The distribution profile of times cited of the citing paper would be generated (e.g., ten citing papers were themselves cited two times, eight were cited three times, six were cited four times, etc).

9) Related Records –

10) Abstract – the texts of the citing paper Abstracts could be mined to examine relations between cited paper themes and the citing paper themes. Phrase frequency and proximity analyses could be performed.

11) Keywords Plus -

12) Addresses - distribution of names and types of institutions, and countries, could be generated. Institution and country combinations would be of special interest, and could be correlated with author combination distributions (e.g., four papers came from institution x, three papers came from institution y, etc; six papers came from government labs, five papers from academia, four papers from industry, etc; six papers came from Mexico, four papers from Brazil, three papers from Korea, etc; Mexico/ US/ Canada were combined in two papers, Mexico/ Brazil/ Argentina were combined in three papers, China/ Taiwan/ Korea were combined in three papers, etc)

In addition, text mining could be performed on the text fields (mainly the Abstract, but including the Title and Keywords) to supplement the analysis on the semi-structured and structured fields. Once the citing literature has been defined, the standard text mining analyses can be performed [Kostoff, 1999, 2000]. In addition, text mining will be used as a guide to identifying citing papers that are intra- and extra-discipline relative to the cited paper(s) of interest. This unique feature of text mining could become very important when identifying unique linguistic features (e.g., extra-disciplinary applications) in the thousands, or tens or hundreds of thousands, of citing papers characteristic of a comprehensive multi-generation citation analysis required for a comprehensive research impact study.

III-C. Focus of Present Study

The vehicles selected to illustrate the principles and techniques of citation mining include a comparison of a cited research unit from a developing country with a cited research unit from a developed country, and a comparison of a cited unit from a basic research field with a cited unit from an applied research field.

Specifically, the technique is being demonstrated using selected papers from a Mexican semiconductor applied research group (MA), a United States semiconductor applied research group (UA), a British fundamental research group (BF), and a United States fundamental research group (UF). These papers were selected based on the authors' familiarity with the topical matter, and the desire to examine papers that are reasonably cited.

Figure 2A – Cited Papers Used for Study

GROUP	PAPERS
MA	Nair P.K. Sem. Sc. Tech. 3 (1988) 134-145 Nair P.K. J Phys D - Appl Phys, 22 (1989) 829-836 Nair M.T.S. Sem. Sc. and Tech. 4 (1989) 191- 199 Nair M. T. S. J Appl Phys, 75 (1994) 1557-1564
UF	Jaeger HM, 1992, Science, V255, P1523
BF	Metha A, 1989, Physica A, V157, P1091 Metha A, 1991, Phys Rev Lett, V67, P394 Barker GC, 1992, Phys Rev A, V45, P3435 Metha A, 1996, Phys Rev E, V53, P92
UA	Tuttle, Prog. Photovoltaic v3, 235 (1995) Gabor, Appl. Phys. Lett. v65, 198 (1994) Tuttle, J. Appl. Phys. v78, 269 (1995) Tuttle, J. Appl. Phys. v77, 153 (1995) Nelson, J. Appl. Phys. v74 5757 (1993)

In addition, selection and banding of variables are key aspects of the study. While specific variable values are of interest in some cases (e.g., names of specific citing institutions), there tends to be substantial value in meta-level groupings (e.g., institution class, such as government, industry, academia). Objectives of the study are to demonstrate important variables, types of meta-level groupings providing the most information and insight, and those conditions under which non-dimensionalization become useful.

The specific variables considered for this study are presented below. Time and resource constraints allowed examination of only a fraction of these variables. The variables underlined were selected for analysis. The items in parentheses are the grouping bands.

***<u>multi-author distribution profiles</u> (1, 2, 3, 4, 5, 6, author paper...)**
***journal development category (basic research, applied research, technology development)**
***<u>journal discipline</u> (physical science, engineering science, environmental science, life science, multi-discipline [e.g., Science, Nature])**
***<u>citing paper discipline</u> (physical science, engineering science, environmental science, life science, multi-discipline)**
***citing paper development category (basic research, applied research, technology development)**
***<u>time profiles of citing papers</u>**
***<u>citing document types</u> (journal articles, notes, reviews, conference proceedings)**
***<u>citing paper language</u> (English, Spanish, French, German, Russian, Chinese)**
***citing institution (academia, government, industry)**
***<u>multi-institution distribution profiles</u> (1, 2 institutions represented, 3, 4, 5, ...)**
***cross-category institution profiles (e.g., 1 gov't/ 2 industry, 2 academia/ 1 industry, etc)**

***citing country** (fully developed, partially developed, developing)
***multi-country distribution profiles** (1, 2 countries represented, 3, 4, 5, ...)
***cross-category country profiles** (2 developed/ 1 developing, 3 partially developed/ 1 developing, etc)
***citing paper authors**
***citing paper gross references distributions** (n1 papers contain 0-10 references, n2 papers contain 11-20 references, …)
***citing paper specific references distributions** (Smith paper cited n1 times, Fernandez paper cited n2 times, …)

IV. RESULTS

This section presents the bibliometric and text mining results, showing the advantages and broad perspectives offered by these techniques, both alone and combined. The results are presented in graph and tabular forms.

IV-A. Bibliometric Profiling

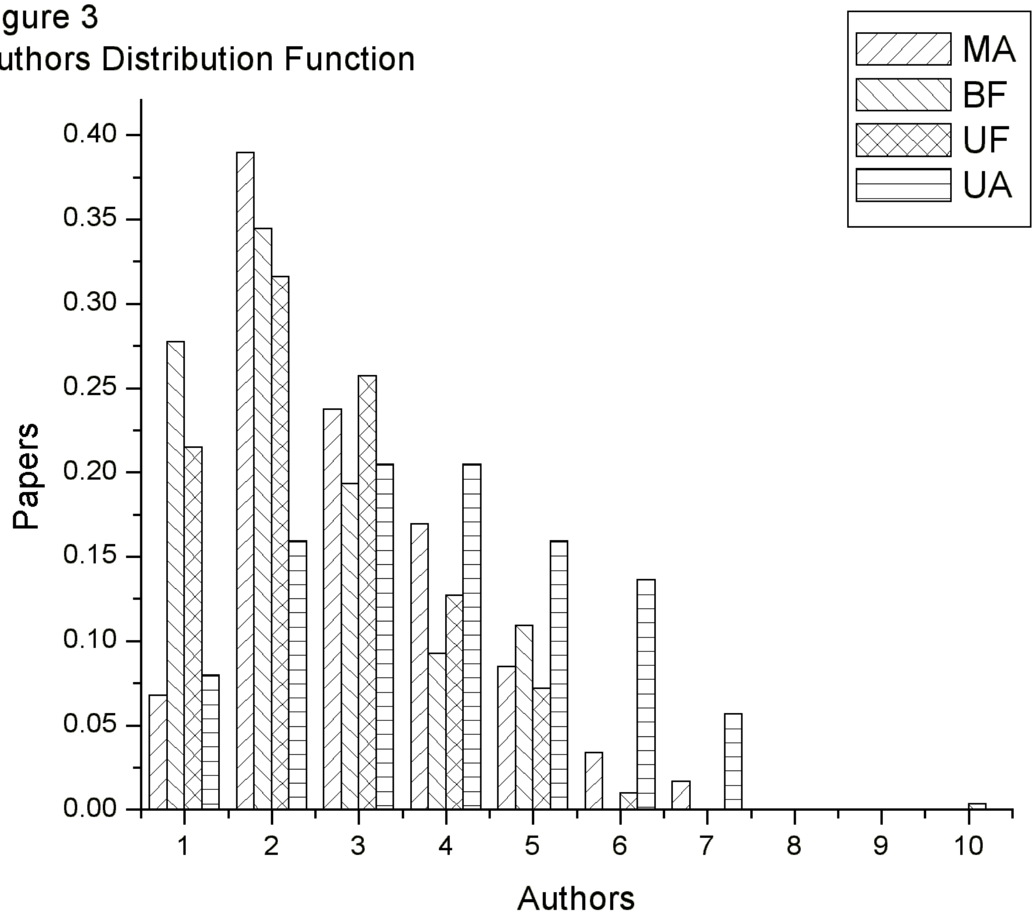
IV-A-1. Multi-author distribution profiles

Figure 3 contains a bar graph of multi-author distribution for the four sets analyzed. The most striking feature of this graph is the behavior at the wings. The papers citing basic research dominate the low end (single author), while the papers citing applied research dominate the high end (6-7 authors). The papers citing basic research (BF and UF) have a similar number of authors per paper, with a maximum in the frequency distribution at two authors per paper. The UA citing papers show gaussian-like authorship distribution with three and four authors per paper, while the MA group citing papers show a distribution similar to the groups citing fundamental research papers but with fewer single-author papers. These four sets show author distributions where 90% of the papers had less than six authors. These results confirm the diversity of collaborative group

compositions over different disciplines and levels of development.

Generally, as projects become more applied, they tend to become larger and more expensive, and require more resources. They also usually require the integration of multiple disciplines. Both these characteristics typically result in larger research groups, and hence in more contributors to a project and its resulting documents. Experimental work usually involves larger teams than theoretical work, while modeling and simulation activities tend to allow more individual efforts. The strong experimental emphasis of the two applied semiconductor groups, with little evidence of computer simulation shown, results in large teams on average. The more balanced theory/ experiment combination of the basic research group tends to suppress larger team efforts in favor of more individualized research. In addition, the intrinsic nature of sandpile vibration research, as opposed to elementary particle or fusion research, does not require large facilities and large research teams.

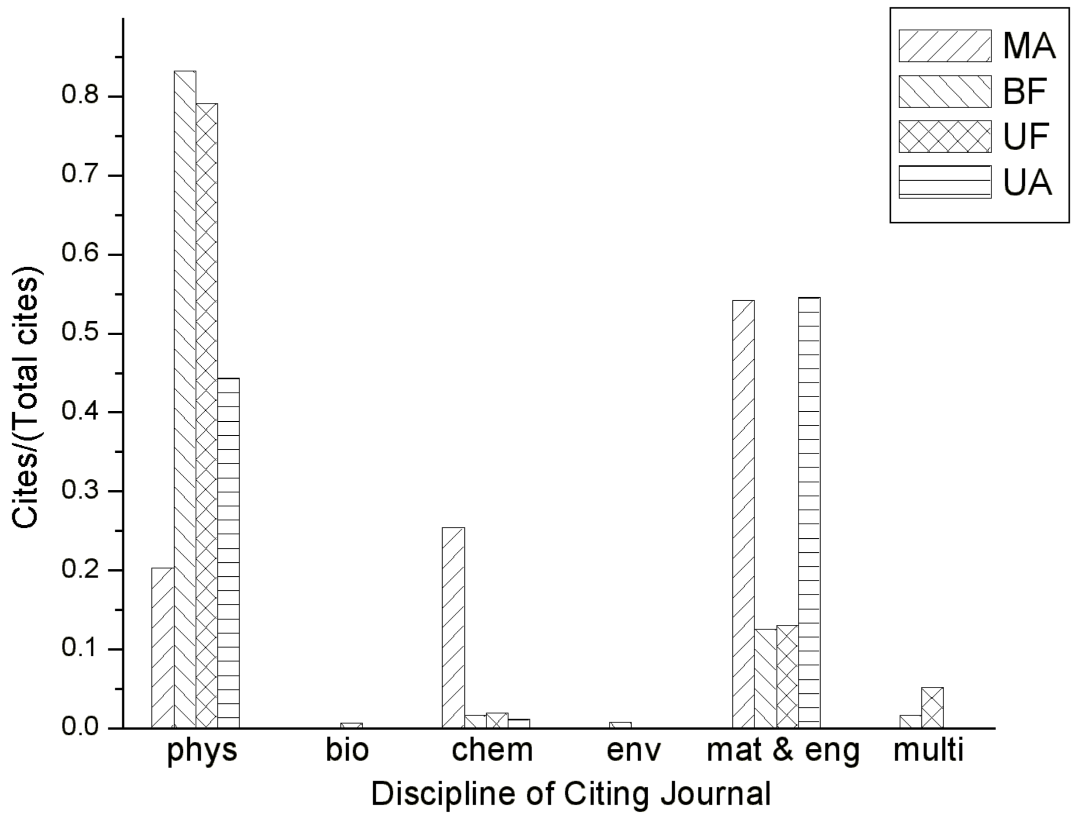
Figure 3
Authors Distribution Function



IV-A-2. Journal discipline

The citing paper discipline frequency is shown in Figure 4. Clearly, each paper set has defined its main discipline well. Also, there is a symmetry in the cross citing disciplines. UF and BF groups were cited more than 80% in fundamental journals and close to 10% in applied journals. Similarly, MA and UA groups were cited close to 50% in applied journals and 45% in fundamental journals. These journal discipline results suggest that the applications developed by the MA group have a strong impact on chemical papers, while the applications developed by the UA group strongly impact physics papers. A point to be stressed is that only the fundamental papers received cites in journals clearly outside of their disciplines.

Figure 4
Citing Journal Theme Distribution

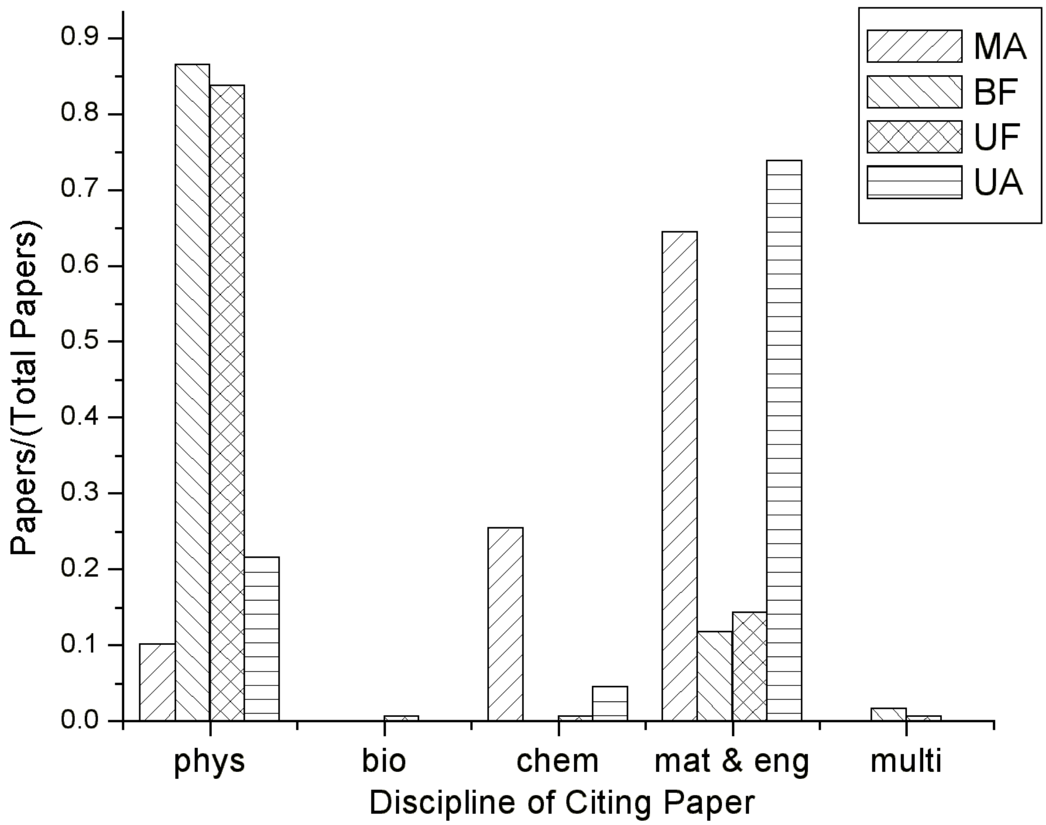


IV-A-3. Citing paper discipline

The discipline distribution of the citing papers, produced by analyzing the papers' Abstracts and Titles, is shown in Figure 5. It is slightly different from Figure 4. As concluded in the text mining section (IV-B), these free-text fields provide far more precise information than can be obtained from the journal discipline. Multi-disciplinary journals can publish uni-disciplinary papers from many different disciplines. Also, the journal categories are not a unique reflection of specific contents (e.g., an environmental journal can accept engineering papers, a materials journal can accept physics papers, etc). However, the chemical nature of

the papers/ journals impacted by the MA group is confirmed.

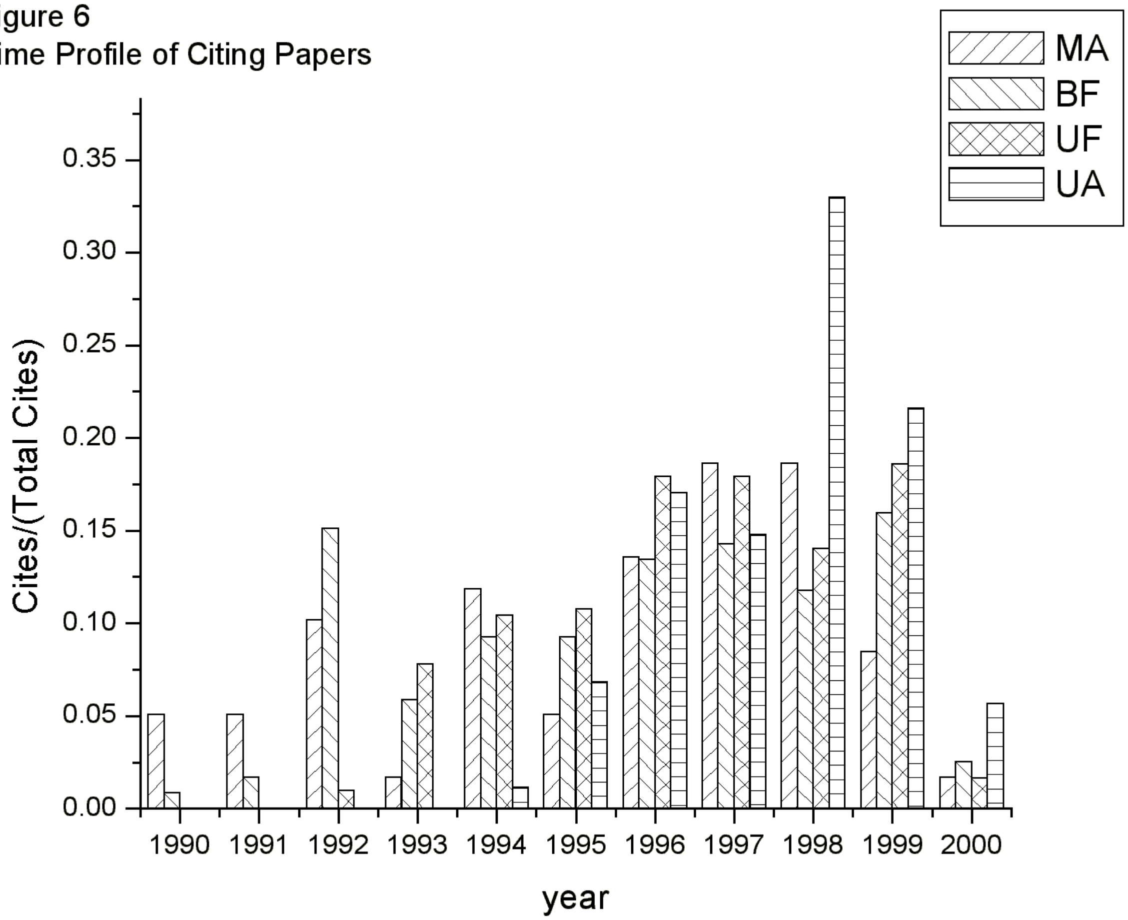
Figure 5
Citing Paper Theme Distribution



IV-A-4. Time profiles of citing papers

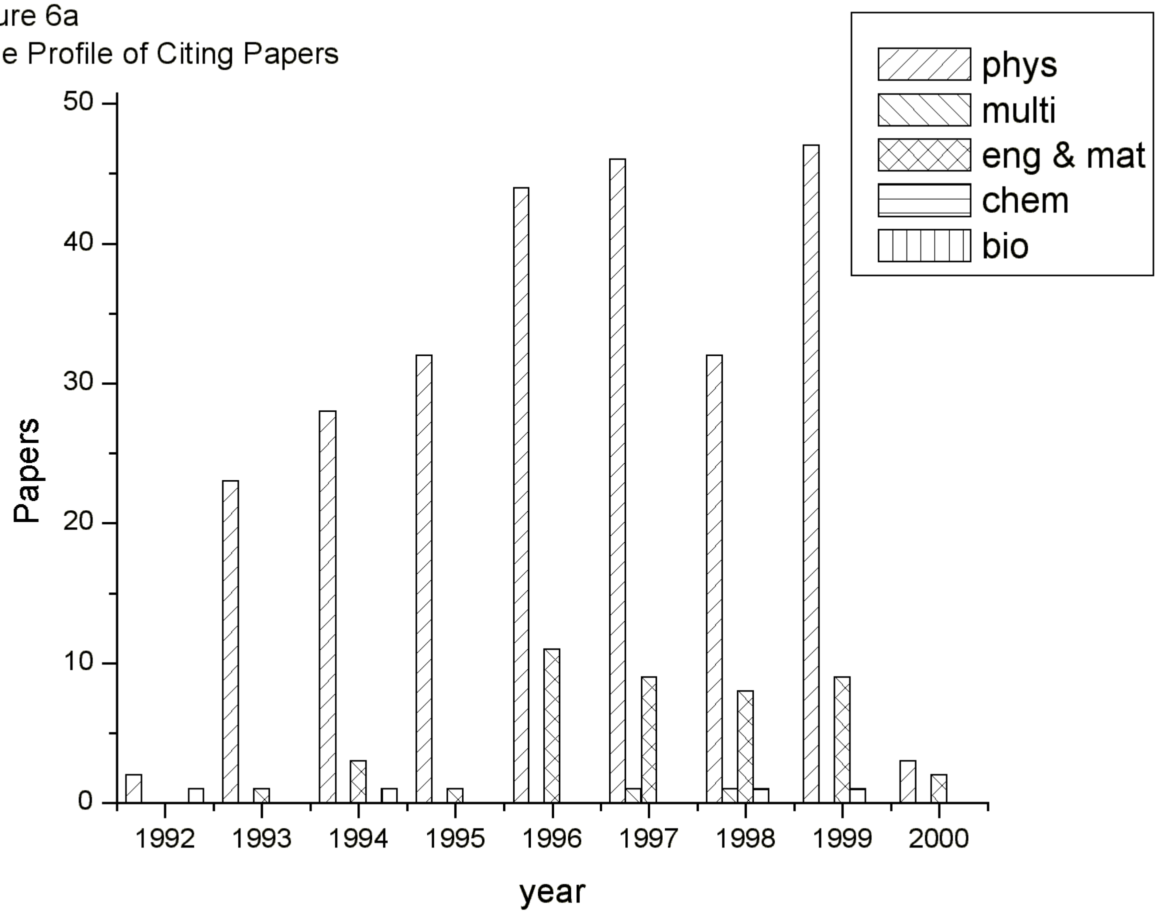
In three of the four sets analyzed, the component papers were published in different years. The MA set was published from 1989 to 1994, UA from 1994 to 1995, BF from 1989 to 1996, while UF includes only one paper published in 1992. Figure 6 shows a clear wave-like behavior of UA and BF, due partly to the different dates of paper publication. Also, most of the sets have between 10% and 20% of cites per year, while the UA set received 38% of the cites in 1998.

Figure 6
Time Profile of Citing Papers



The single highly-cited paper feature of the UF set allows additional analyses and perspectives. In Figure 6a, the UF citing paper disciplines are shown as a function of time. As time evolves, citing papers from disciplines other than that of the cited paper emerge. An important point is the four-year delay of the systematic appearance of engineering and materials science citing papers.

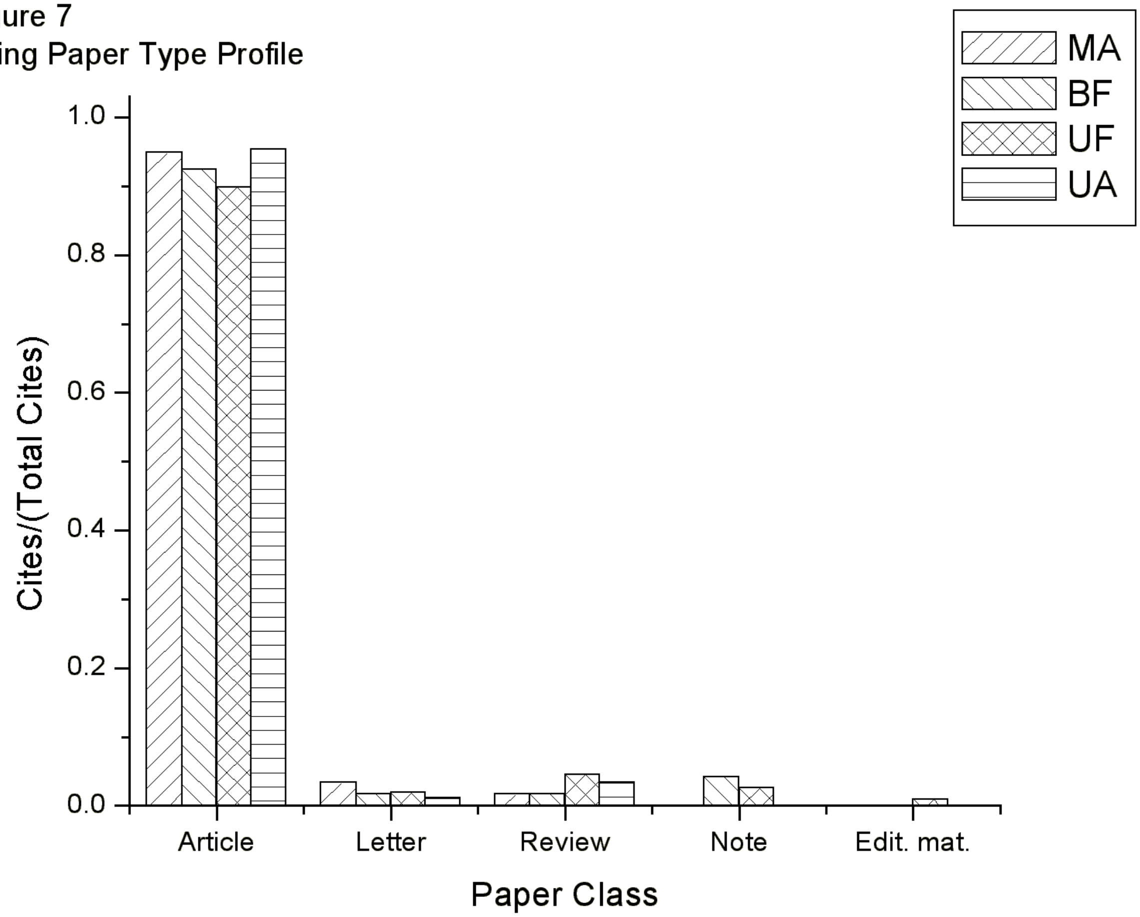
Figure 6a
Time Profile of Citing Papers



IV-A-5. Citing document types

Figure 7 shows that most cites appear in articles. The four analyzed sets are cited in review articles and letters. This indicates the relevance of the analyzed papers. One important point is that only the fundamental papers are cited in notes, and only the UF paper was cited in an editorial document.

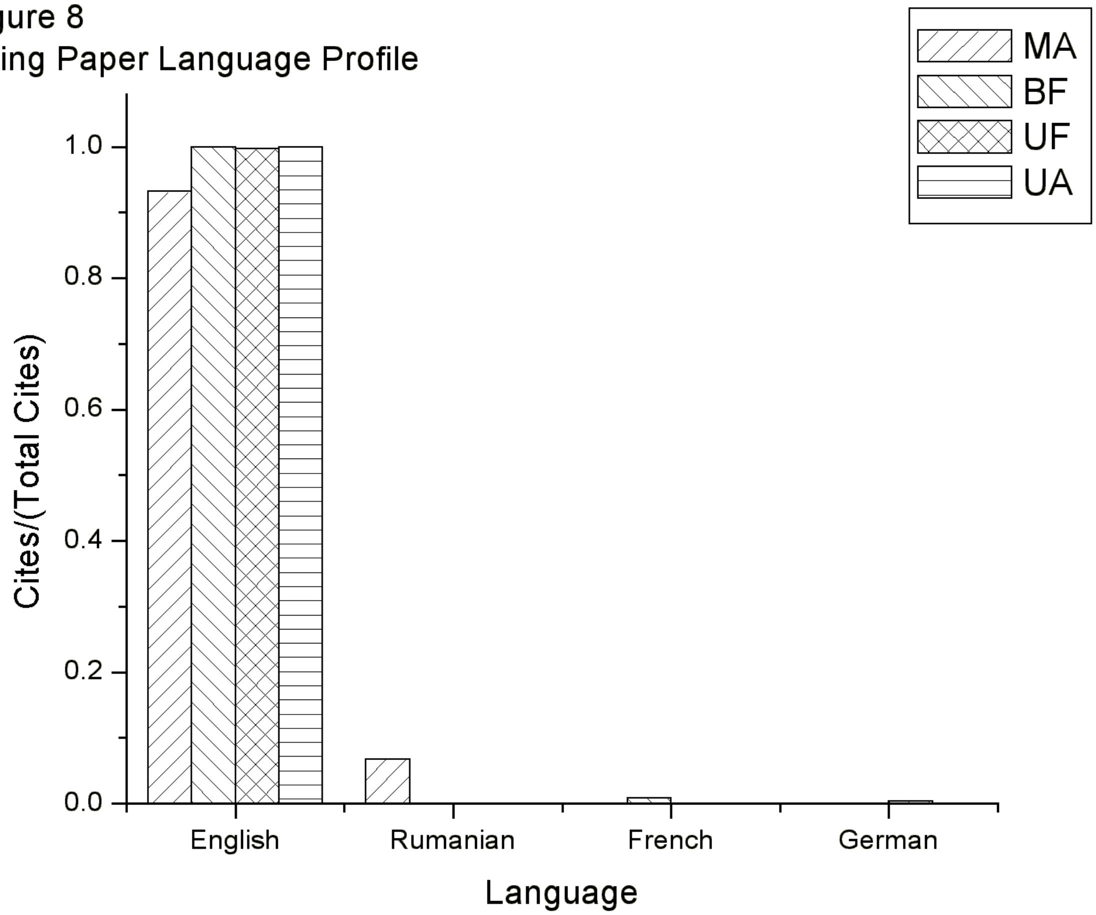
Figure 7
Citing Paper Type Profile



IV-A-6. Citing paper language

Figure 8 shows that English is the dominant language of all the paper sets analyzed. However, the surprising appearance of a significant number of citing papers written in Romanian for the MA set indicates that MA's work is important for at least one developing country.

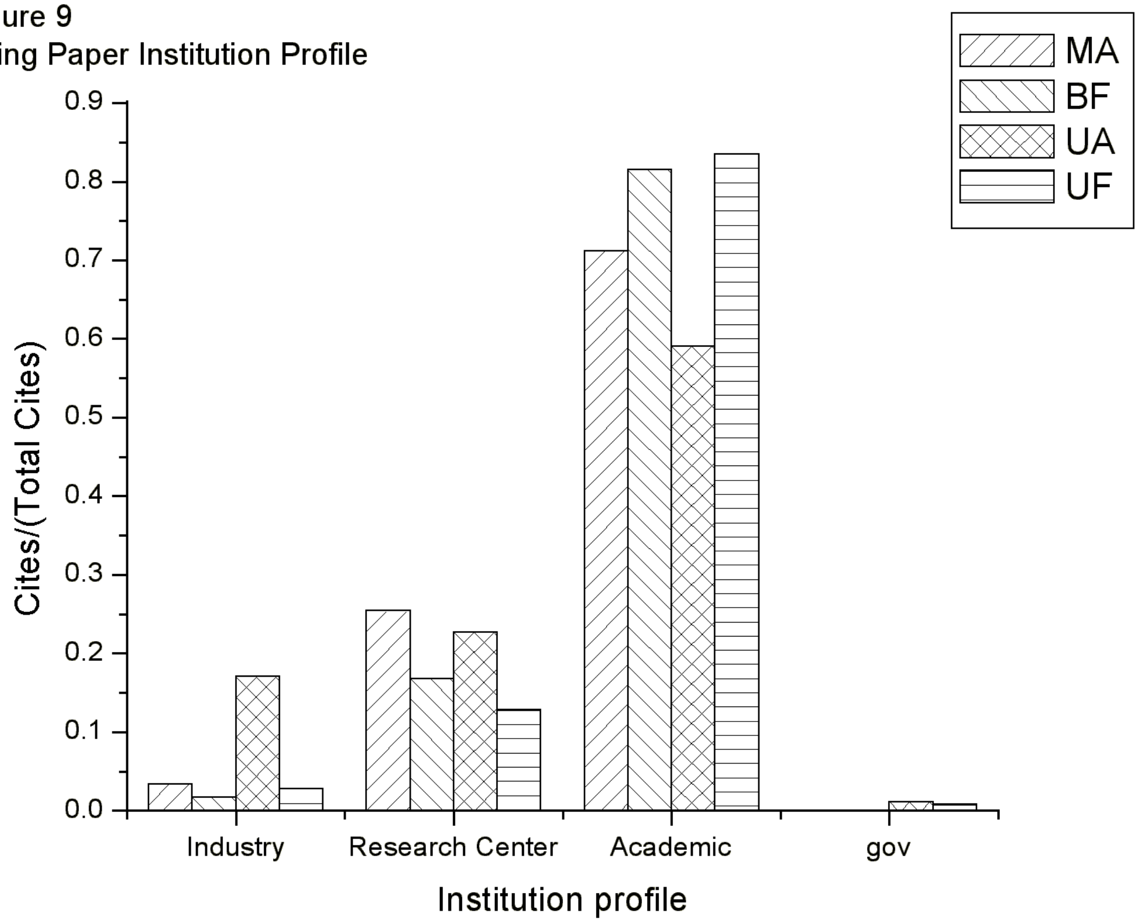
Figure 8
Citing Paper Language Profile



IV-A-7. Multi-institution distribution profiles

Figure 9 shows the profile of the citing institutions. Clearly, academia has the highest citing rates. Industry follows closely the advance of high-technological developments, but is not following directly the advances in fundamental research. Research Centers follow applied and fundamental research about equally. Direct government participation is not significant in the fields studied. Government/ national laboratories were classified under research centers.

Figure 9
Citing Paper Institution Profile

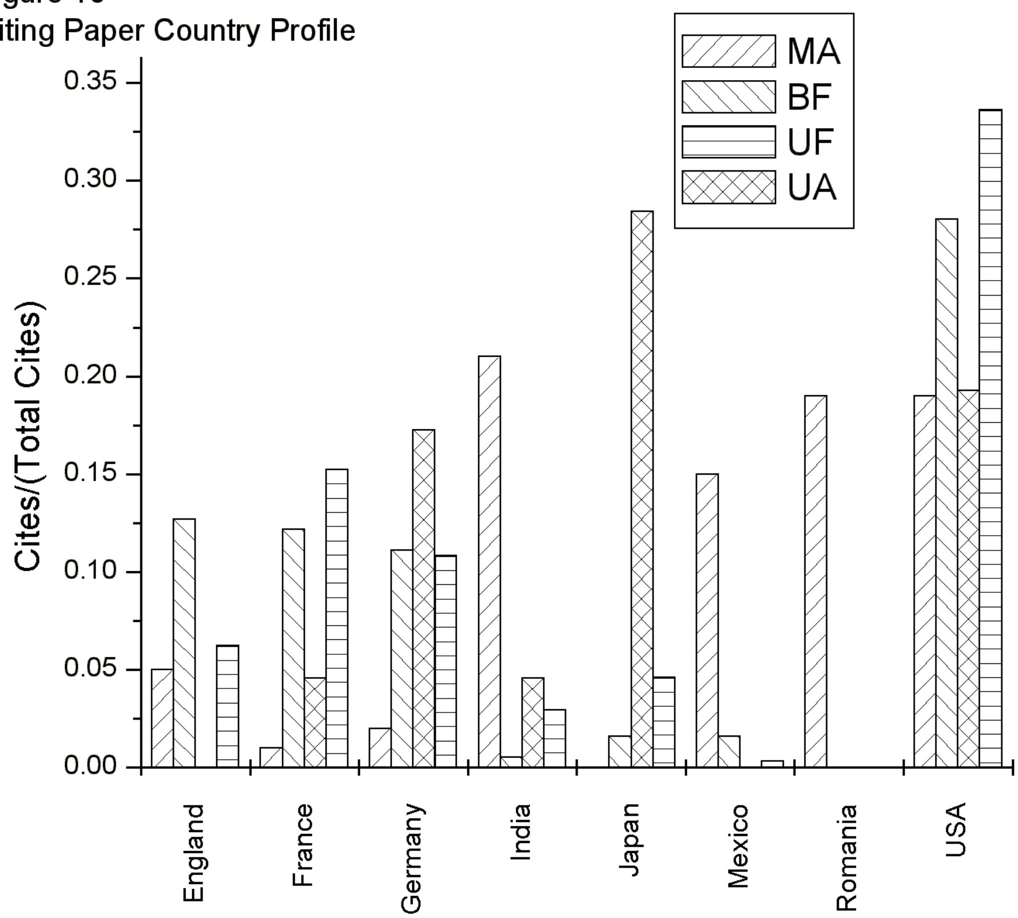


IV-A-8. Multi-country distribution profiles

There are 44 countries represented in the citing paper sets analyzed. Figure 10 shows only those countries with at least 10% of the citing countries for a set. USA has the most cites in aggregate. India has the largest cites of the MA set; Japan has the largest cites of the UA set. This fact is due to the different nature of the applied technology developed by MA and UA. The UA set contains work related to high technology, and the MA set is dedicated to explore low-cost technology. Therefore, this last set is cited by the less affluent countries of India, Romania and Mexico. India and Mexico also cite fundamental research, but not Romania. It is important to stress that if no low-cost technology papers were

considered, these latter countries would not appear in this graph, and only developed countries would appear. England does not cite UA works.

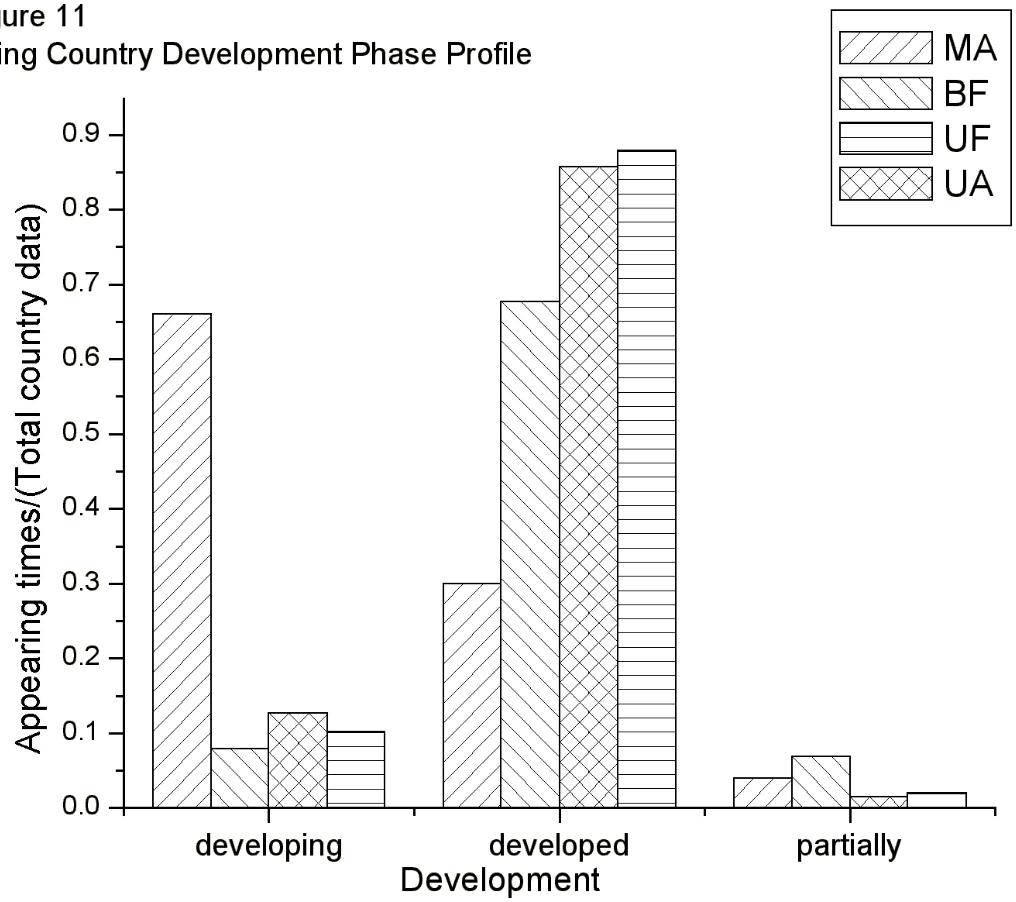
Figure 10
Citing Paper Country Profile



IV-A-9. Citing country profiles

Figure 11 shows clearly that the low-cost technology papers are cited by developing countries. Developed countries cite the mostly high-technology papers. The apparently smaller contribution of the partially developed countries in citing is due to the fact that this category characterizes relatively few countries, from the authors' perspectives.

Figure 11
Citing Country Development Phase Profile



IV-A-10. Citing author frequency

Tables 1 and 2 present the most citing authors for BF and UF sets. There is a common citing author who occupied the highest position in the frequency table in both sets (Hermann, HJ). Also, there is another common citing author (Nicodemi, M). This fact confirms that UF and BF are closely related research groups.

In Table 3, it is clear that the maximum citing author of the MA group is a Romanian researcher. However, the corresponding table for the UA set contains no relevant information for people outside the specific UA field of research (high efficient photovoltaic cells). Therefore, it is omitted.

TABLE 1 – BF CITING AUTHORS

BF Citing Authors		
Citing Author	Citing Times	Percentage
Herrmann, HJ	16	0.13
Jaeger, HM	11	0.09
Nagel, SR	11	0.09
Zhang, ZP	11	0.09
Nicodemi, M	10	0.08

TABLE 2 – UF CITING AUTHORS

UF Citing Authors		
Citing Author	Citing Times	Percentage
Herrmann, HJ	24	0.08
Nicodemi, M	14	0.05
Rahchenbach, J	11	0.04
Mehta, A	11	0.04
Makse, HA	11	0.04
Behringer, RP	11	0.04
Duran, J	10	0.03
Luding, S	9	0.03
Coniglio, A	8	0.02
Clement, E	8	0.02

TABLE 3 – MA CITING AUTHORS

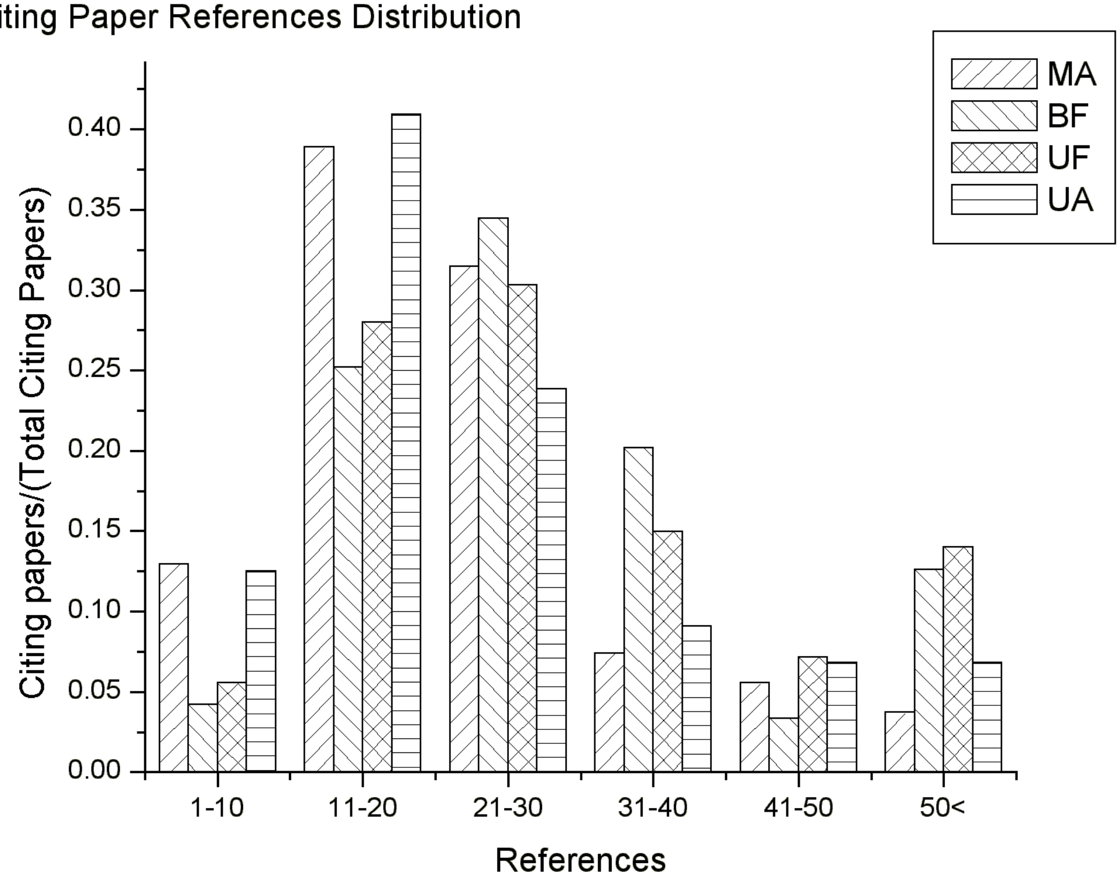
MA Citing Authors		
Citing Author	Citing Times	Percentage
Nascu C	7	0.12
Pop I	7	0.12
Bhushan S	6	0.10

Ionescu V	5	0.08

IV-A-11 References in citing papers

In Figure 12, it is clear that there are common features in the number of references in those papers that cite the core applied and fundamental papers, but there are also some differences. For instance, at the lower end of the spectrum (0-20), the applied papers' citing papers dominate. At the higher end of the spectrum (21-50+), the fundamental papers' citing papers dominate, with the exception of the BF anomaly at 41-50.

Figure 12
Citing Paper References Distribution



There are many possible reasons for these differences, and separating out the effects is complex. There are two different technical disciplines, and each has its citing culture and traditions. Also, each technical discipline

has a different level of research activity, and this could influence the magnitude of citations generated. Finally, there may be different citing practices in basic and applied research.

IV-A-12. Frequency of the references in citing papers
Frequency analysis of the references in the citing papers provides insight to co-cited papers, and allows a historical perspective to be obtained. The reference-frequency for the UF and BF citing papers is shown in Tables 4 and 5. In these two tables, Faraday's work (1831) appears within the twenty papers most cited in the UF and BF citing papers. This indicates the fundamental and seminal character of the experimental work performed by Faraday. Also, Reynolds work (1885) appears within the twenty most cited papers in the references of the BF set. These two references also indicate the longevity of the unsolved problems tackled by the UF and BF groups.

The highest frequency co-cited papers have three interesting characteristics. They are essentially all in the same general physics area, they are all published in fundamental science journals (mainly physics), and they are all relatively recent, indicating a dynamic research area with high turnover.

TABLE 4 - FREQUENCIES OF REFERENCES IN UF CITING PAPERS

Frequencies of References in UF Citing Papers		
Paper	Times	
JAEGER HM, 1992, SCIENCE, V255, P1523	307	100%
EVESQUE P, 1989, PHYS REV LETT, V62, P44	75	24.4%
GALLAS JAC, 1992, PHYS REV LETT, V69, P1371	72	23.4%
CHOO K, 1997, PHYS REV LETT, V79, P2975	68	22.1%
KNIGHT JB, 1993, PHYS REV LETT, V70, P3728	68	22.1%
ROSATO A, 1987, PHYS REV LETT, V58, P1038	64	20.8%
CAMPBELL CS, 1990, ANNU REV FLUID MECH, V22,	62	20.8%

P57		
TAGUCHI YH, 1992, PHYS REV LETT, V69, P1367	56	18.2%
JAEGER HM, 1989, PHYS REV LETT, V62, P40	52	16.9%
BAXTER GW, 1989, PHYS REV LETT, V62, P2825	52	16.9%
THOMPSON PA, 1991, PHYS REV LETT, V67, P1751	51	16.6%
BAK P, 1987, PHYS REV LETT, V59, P381	48	15.6%
CUNDALL PA, 1979, GEOTECHNIQUE, V29, P47	48	15.6%
CLEMENT E, 1992, PHYS REV LETT, V69, P1189	47	15.3%
JAEGER HM, 1996, REV MOD PHYS, V68, P1259	43	14.0%
DOUADY S, 1989, EUROPHYS LETT, V8, P621	43	14.0%
LAROCHE C, 1989, J PHYS-PARIS, V50, P669	42	13.7%
WILLIAMS JC, 1976, POWDER TECHNOL, V15, P 245	41	13.4%
HAFF PK, 1983, J FLUID MECH, V134, P401	38	12.4%
FARADAY M, 1831, PHIL T R SOC LONDON, V52, P299	37	12.5%
BAGNOLD RA, 1954, P ROY SOC LOND A MAT, V225, P49	37	12.5%

TABLE 5 - FREQUENCIES OF REFERENCES IN BF CITING PAPERS

Frequencies of References in BF Citing Papers		
Paper	Times	
MEHTA A, 1989, PHYSICA A, V157, P1091	63	52.9%
MEHTA A, 1991, PHYS REV LETT, V67, P394	42	35.3%
JAEGER HM, 1992, SCIENCE, V255, P1523	37	31.1%
EVESQUE P, 1989, PHYS REV LETT, V62, P44	33	27.7%
ROSATO A, 1987, PHYS REV LETT, V58, P1038	33	27.7%
BARKER GC, 1992, PHYS REV A, V45, P3435	32	26.9%
JAEGER HM, 1989, PHYS REV LETT, V62, P40	32	26.9%
EDWARDS SF, 1989, PHYSICA A, V157, P1080	28	23.5%
LAROCHE C, 1989, J PHYS-PARIS, V50, P699	23	19.3%
MEHTA A, 1996, PHYS REV E, V53, P92	23	19.3%
KNIGHT JB, 1995, PHYS REV E, V51, P3957	22	18.5%
CAMPBELL CS, 1990, ANNU REV FLUID MECH, V22, P57	21	17.6%
EDWARDS SF, 1991, J STAT PHYS, V62, P889	21	17.6%
REYNOLDS O, 1885, PHILOS MAG 5, V20, P469	20	16.8%

BAXTER GW, 1989, PHYS REV LETT, V62, P2825	19	16.0%
THOMPSON PA, 1991, PHYS REV LETT, V67, P1751	19	16.0%
CLEMENT E, 1991, EUROPHYS LETT, V16, P133	18	15.1%
FARADAY M, 1831, PHIL T R SOC LONDON, V52, P299	18	15.1%
KNIGHT JB, 1993, PHYS REV LETT, V70, P3728	18	15.1%
MEHTA A, 1994, GRANULAR MATTER	18	15.1%
BARKER GC, 1993, PHYS REV E, V47, P184	17	14.3%
GALLAS JAC, 1992, PHYS REV LETT, V69, P1371	17	14.3%
JAEGER HM, 1996, REV MOD PHYS, V68, P1259	17	14.3%

TABLE 6 – FREQUENCIES OF REFERENCES IN UA CITING PAPERS

Frequencies of References in UA Citing Papers		
Paper	Times	
GABOR AM, 1994, APPL PHYS LETT, V65, P198	35	39.8%
HEDSTROM J, 1993, P 23 IEEE PHOT SPEC, P364	26	29.5%
TUTTLE JR, 1995, PROG PHOTOVOLTAICS, V3, P383	26	29.5%
TUTTLE JR, 1995, J APPL PHYS, V77, P153	25	28.4%
SCHMID D, 1993, J APPL PHYS, V73, P2902	20	22.7%
ROCKETT A, 1991, J APPL PHYS, V70, PR81	17	19.3%
STOLT L, 1993, APPL PHYS LETT, V62, P597	14	15.9%
SHAY JL, 1975, TERNARY CHALCOPYRITE	12	13.6%
KLENK R, 1993, ADV MATER, V5, P144	10	11.4%
NELSON AJ, 1995, J APPL PHYS, V78, P269	10	11.4%
BOEHNKE UC, 1987, J MATER SCI, V22, P1635	9	10.2%
CONTRERAS MA, 1994, PROG PHOTOVOLTAICS R, V2, P287	9	10.2%
FEARHEILEY ML, 1986, SOL CELLS, V16, P91	9	10.2%
JAFFE JE, 1984, PHYS REV B, V29, P1882	8	9.1%
NELSON AJ, 1993, J APPL PHYS, V74, P5757	8	9.1%
TUTTLE JR, 1996, MATER RES SOC SYMP P, V426, P143	8	9.1%

TABLE 7 – FREQUENCIES OF REFERENCES IN MA CITING PAPERS

Frequencies of References in MA Citing Papers		
Paper	Times	
NAIR PK, 1989, J PHYS D APPL PHYS, V22, P829	23	25.84%
NAIR PK, 1988, SEMICOND SCI TECH, V3, P134	20	22.47%
NAIR MTS, 1994, J APPL PHYS, V75, P1557	15	16.85%
KAUR I, 1980, J ELECTROCHEM SOC, V127, P943	10	11.24%
MONDAL A, 1983, SOL ENERG MATER, V7, P431	10	11.24%
CHOPRA KL, 1983, THIN FILM SOLAR CELL	9	10.11%
BUBE RH, 1960, PHOTOCONDUCTIVITY SO	8	8.99%
NAIR MTS, 1989, SEMICOND SCI TECH, V4, P191	8	8.99%

This ends the bibliometric analysis. The following section illustrates the usefulness of text mining analysis.

IV-B. Text Mining

The purpose of the text mining is to perform trans-citation linguistic pattern analyses, and make trans-citation comparisons. A central hypothesis that this paper will attempt to prove is that the core themes of each cited paper will be reflected as core themes of its citing papers. In addition, there may be secondary themes contained in the linguistics of the citing papers, reflecting the diffusion of the information from the cited paper into other technical disciplines, and from research into technology development. It is hypothesized that the more fundamental the cited paper, the greater is the chance that there will be diffusion into other disciplines, and the text mining should be able to identify these additional themes.

Three types of studies were performed, and their results will be described in this section. Sub-section IV-B-1 describes the linguistic relationships between selected individual cited papers and their citing papers. In the four aggregate groups studied in this paper, there were a total of fourteen individual papers. Because of space limitations, three representative papers (Abstracts) are selected, and sample results are shown. While comparisons of many types of linguistic patterns are possible, again because of space limitations, only multi-word technical phrases and their frequencies are presented.

Sub-section IV-B-2 describes the linguistic relationships between the aggregated cited papers in each of the four groups and their aggregated citing papers. Again, only multi-word technical phrases and their frequencies are presented. Additionally, the major themes for each grouping of citing papers are categorized, and discussed in context.

IV-B-1. Relation between selected individual cited papers and their citing papers

IV-B-1-a. Applied Photo-voltaic Research Paper (U. S.)

The cited Abstract is reproduced in Figure 13. Its highest frequency technical phrases are shown in Figure 14, along with their respective frequencies. The highest frequency single, adjacent double, and adjacent triple word phrases from its citing papers aligned with the themes of the cited paper are shown in Figure 15, as well as the multi-word phrases that represent technical disciplines different from those of the cited Abstract. Asterisks (*) represent singular and plural.

The central themes and specific phrases used in the cited paper are replicated in the citing papers. There were no phrases in the citing papers that represented themes or disciplines significantly different from those in the

cited paper, above a frequency of unity. While there could possibly be phrases representative of different themes with a unity frequency, some minimal theme coherence was desired. The citing readership appears to be strongly concentrated in the thematic areas of the cited paper. One suspects that the audience obtained is the target audience for this paper, at least in terms of thematic interest, but the authors of this cited paper or any of the cited papers in this study were not contacted to confirm this conclusion.

FIGURE 13 – CITED PAPER ABSTRACT

In, Ga, and Se were coevaporated to form precursor films of (Inx,Ga1-x)2Se3. The precursors were then converted to CuInxGa1-xSe2 by exposure to a flux of Cu and Se. The final films were smooth, with tightly packed grains, and had a graded Ga content as a function of film depth. Photovoltaic devices made from these films showed good tolerance in device efficiency to variations in film composition. A device made from these films resulted in the highest total-area efficiency measured for any non-single-crystal, thin-film solar cell, at 15.9%.

FIGURE 14 – CITED PAPER THEMES

6 FILM*; 2 DEVICE; 2 EFFICIENCY; 2 GA; 2 SE

FIGURE 15 – CITING PAPER THEMES

15A – Highest Frequency Phrases Aligned with Cited Paper Theme

77 FILM*; 38 CU; 32 GA; 28 CIS; 27 CUINSE2; 26 CELLS; 26 DEPOSITION; 25 SOLAR; 22 DEFECT; 21 THIN; 20 CHALCOPYRITE; 20 PHASE; 19 CDS; 18 BAND; 18 SE-2; 18 DEVICE*; 16 CIGS; 10 EFFICIENCY; 23 SOLAR CELL*; 14 GA SE-2; 19 THIN FILM*; 9 ELECTRON MICROSCOPY; 7 BAND GAP; 7 INDUCED PHOTOPLEOCHROISM; 6 X-RAY DIFFRACTION; 5 CHEMICAL BATH; 5 CIGS FILMS; 5 FILMS DEPOSITED; 5 FORMATION

ENERGIES; 5 GLASS SUBSTRATES; 5 GRAIN SIZE; 5 SUBSTRATE TEMPERATURE; 18 CU IN GA; 5 TRANSMISSION ELECTRON MICROSCOPY; 4 GA SE-2 CIGS; 4 PHYSICAL VAPOR DEPOSITION; 3 CHEMICAL BATH DEPOSITION; 3 CU2SE AND CIS; 3 CUPT TYPE CIS; 3 ELECTRON MICROSCOPY TEM; 3 LINEARLY POLARIZED RADIATION; 3 SCANNING ELECTRON MICROSCOPY; 3 X-RAY PHOTOEMISSION SPECTROSCOPY

15B – Phrases Reflecting Themes other than Cited Paper

IV-B-1-b. Applied Photo-voltaic Research Paper (Mexico)

The cited Abstract is reproduced in Figure 16. Its highest frequency technical phrases are shown in Figure 17. The highest frequency single, adjacent double, and adjacent triple word phrases from its citing papers aligned with the central themes of the cited paper are shown in Figure 18, as well as the multi-word phrases that represent technical disciplines different from those of the cited Abstract.

The central themes and specific phrases used in the cited paper are replicated in the citing papers. There were two phrases in the citing papers that represented themes or disciplines other than those in the cited paper, above a frequency of unity. These additional themes reflected use of the solar coatings for automobile windows, in addition to the core architectural (building) applications. This is a very small extrapolation. Again, the citing readership appears to be strongly concentrated in the thematic areas of the cited paper.

FIGURE 16 – CITED PAPER ABSTRACT

Solar control coatings, required for architectural glazing applications in warm climates, must provide controlled optical transmission (~ 10-50%) of the solar radiation in the visible region and should reflect efficiently in the infrared (>0.7um) region to create a

cool interior in the buildings. Thin films of PbS and CuS on glass substrates, deposited from chemical baths, are shown to possess excellent solar control characteristics–superior or comparable to the metallic solar control coatings. For example, for an acceptable range of integrated optical transmittance (~10–20%) in the visible region, the integrated infrared reflectance for AM2 solar spectrum for the different glazing are: PbS coated glass, 50%: CuS goated glass, 14% stainless steel/Cu coated glass, 25% and tinted glass, 4%. The CuS and Pbs coatings also have the advantage of giving pleasant reflected colours (golden, purple, blue, etc), which improves the cosmetic appearance. This paper presents the basic requirements of solar control coatings and provides a comparison of the characteristics of PbS and CuS coatings against commercially available coatings.

FIGURE 17 – CITED PAPER THEMES

6 COATINGS; 6 SOLAR; 5 GLASS; 4 CONTROL; 4 CUS; 4 PBS; 3 REGION; 2 COATED; 2 GLAZING; 2 INFRARED; 2 INTEGRATED; 2 OPTICAL; 2 VISIBLE; 4 SOLAR CONTROL; 3 CONTROL COATINGS; 2 COATED GLASS; 2 VISIBLE REGION; 3 SOLAR CONTROL COATINGS; 2 PBS AND CUS

FIGURE 18 – CITING PAPERS THEMES

18A – Highest Frequency Phrases Aligned with Cited Paper Theme

152 FILM*; 64 SOLAR; 46 DEPOSITION; 33 CONTROL; 33 GLASS; 33 THIN; 32 COATINGS; 32 DEPOSITED; 27 OPTICAL; 26 PBS; 23 TEMPERATURE; 21 DEGREES; 20 CHEMICAL; 20 TRANSMITTANCE; 18 PROPERTIES; 35 SUBSTRATE*; 16 CUS; 33 SOLAR CONTROL; 30 THIN FILM*; 12 GLASS SUBSTRATES; 11 SOLAR RADIATION; 10 CHEMICAL DEPOSITION; 10 CONTROL COATINGS; 8 CHEMICALLY DEPOSITED; 8 FILMS DEPOSITED; 8 OPTICAL PROPERTIES; 8 OPTICAL TRANSMITTANCE; 7 AIR ANNEALING; 7 X-RAY DIFFRACTION; 6 CONTROL CHARACTERISTICS; 6 FILM THICKNESS; 15 SOLAR CONTROL

COATING*; 6 DEPOSITED ON GLASS; 4 OPTICAL TRANSMITTANCE SPECTRA

18B - Phrases Reflecting Themes other than Cited Paper

2 AUTOMOBILE; 2 ARCHITECTURAL AND AUTOMOBILE

IV-B-1-c. Basic Research Paper (U. K.)

The cited Abstract is reproduced in Figure 19. Its highest frequency technical phrases are shown in Figure 20. The highest frequency single, adjacent double, and adjacent triple word phrases from its citing papers aligned with the central themes of the cited paper are shown in Figure 21, as well as the multi-word phrases that represent technical disciplines different from those of the cited Abstract.

While the bulk of the citing paper themes are aligned with the cited paper central themes, there are a number of themes different and derived from those of the cited paper. This reflects one of the characteristics of a truly fundamental science paper, namely, the diffusion of basic scientific findings and principles to a variety of applications areas. Thus, the lessons learned from vibrating powders are extrapolated to the study of metals and alloys, traffic, aquifers, ice, food, nuclear and magnetic phenomena

FIGURE 19 - CITED PAPER ABSTRACT

The structure and dynamics of powders subjected to vibration are investigated by a nonsequential and cooperative computer-simulation approach in three dimensions. Starting from a microscopic model of the physics, we are able to probe independent and collective effects in the dynamics of vibrated powders, as well as in the resulting structures. In particular, we analyze the role of cooperative structures such as bridges, which are always present in reality and which cannot be formed

by purely sequential processes. We look in depth at the behavior of the volume fraction and coordination number as a function of the intensity of vibration, as well as at correlation functions describing contacts between neighboring grains, also as a function of intensity. Satisfying agreement with the qualitative predictions of earlier analytic work is obtained, and a framework is laid for future investigations.

FIGURE 20 – CITED PAPER THEMES

2 COOPERATIVE; 2 DYNAMICS; 2 FUNCTION; 2 INTENSITY; 2 POWDERS; 2 STRUCTURES; 2 VIBRATION

FIGURE 21 – CITING PAPERS' THEMES

21A – Highest Frequency Phrases Aligned with Cited Paper Theme

38 MODEL; 26 DENSITY; 25 GRANULAR; 24 PACKING; 22 RELAXATION; 21 DISTRIBUTION; 20 PARTICLES; 18 PORE; 17 SIZE 15 PACKED; 15 VIBRATION*; 14 DYNAMICS; 13 POWDER*; 13 DATA; 12 EXPERIMENTAL; GRAINS; 10 STRUCTURE*; 7 PACKING DENSITY; 7 PORE SIZE*; 6 EXPERIMENTAL DATA; 6 RANDOMLY PACKED; 6 SIZE DISTRIBUTION; 5 ENERGY DISSIPATION; 4 DISTRIBUTION FUNCTION; 4 EXPERIMENTAL RESULTS; 4 GRANULAR MATERIAL*; 4 PACKED BEDS; 4 RELAXATION TIME DISTRIBUTIONS; 3 COUPLED NONLINEAR STOCHASTIC; 3 PORE SIZE DISTRIBUTION; 3 RADIAL DISTRIBUTION FUNCTION; 2 ABSOLUTE STANDARD DEVIATIONS; 2 BEADS WERE CONFINED; 2 BEDS OF MONODISPERSE

21B – Phrases Reflecting Themes other than Cited Paper

4 ALLOY; 4 TRAFFIC; 3 CARS; 3 MAGNETIC; 2 AQUIFER; 2 CAR; 2 EXPRESSWAY; 2 FOOD; 2 FOODSTUFFS; 2 ICE; 2 NUCLEAR; 2 TUNGSTEN; 4 WATER PROTON; 3 PROTON TRANSVERSE; 3 TRAFFIC FLOW; 2 WATER PROTON TRANSVERSE

IV-B-2. Relation between each cited paper group and the group's citing papers

IV-B-2-a. Applied Photo-voltaic Research Group (U. S.)

The cited group's Abstracts are reproduced in Figure 22. The highest frequency single, adjacent double, and adjacent triple word phrases from their aggregate citing papers aligned with the central themes of the aggregate cited papers can be represented by the following taxonomy: Materials/ coatings; Structure/ properties; Experiments/ measurements/ instruments/ variables; Processes; Geometries/ regions; Phenomena; Outputs/ devices/ performance/ applications.

Representative phrases (CAPITALIZED) aligned with the central themes of the aggregate citing papers offer the following intra-discipline portrait of the citing aggregate. These papers reflect a mainly

experimental (SCANNING ELECTRON MICROSCOPY, TRANSMISSION ELECTRON MICROSCOPY, MEASUREMENTS WERE ACQUIRED, X-RAY PHOTOEMISSION SPECTROSCOPY, AUGER ELECTRON SPECTROSCOPY, ELECTRON MICROSCOPY, PHOTOEMISSION MEASUREMENTS, ROOM TEMPERATURE) examination of solar cell-related

applications (THIN FILM SOLAR CELLS, CIGS SOLAR CELLS, SOLAR CELL*, PHOTOVOLTAIC DEVICE*) with a central objective of

performance enhancement (HIGH EFFICIENCY, CONVERSION EFFICIENCY, CONFIRMED EFFICIENCY, DEVICE EFFICIENCY, TOTAL-AREA EFFICIENCY).

In aggregate, these studies focus on

materials and coatings (CU IN GA, GA SE-2 CIGS, CUINSE2 THIN FILMS, CU AND SE, THIN FILMS, CIGS FILM*, CDS, CIS, FILM*) through examination of their

structures and properties (VALENCE BAND ELECTRONIC STRUCTURE, GRAIN BOUNDARIES, CHALCOPYRITE STRUCTURE, HETEROJUNCTION, PROPERTIES),

growth processes (PHYSICAL VAPOR DEPOSITION, FILMS WERE DEPOSITED, LAYERS WERE DEPOSITED, CHEMICAL BATH DEPOSITION, GLASS SUBSTRATE*, SUBSTRATE TEMPERATURE, GROWTH),

physical phenomena (CORE LINES, OPTICAL ABSORPTION, INDUCED PHOTOPLEOCHROISM, SOLAR, OPTICAL, ENERGY), in

surface-related geometries and regions (NEAR SURFACE REGION, THIN, SURFACE).

In aggregate, the citing papers reflect a high tech approach to developing high performance solar conversion devices for electricity production.

There were no phrases aligned with extra-discipline themes.

FIGURE 22 – CITED GROUP'S ABSTRACTS

In, Ga, and Se were coevaporated to form precursor films of (Inx,Ga1-x)2Se3. The precursors were then converted to CuInxGa1-xSe2 by exposure to a flux of Cu and Se. The final films were smooth, with tightly packed grains, and had a graded Ga content as a function of film depth. Photovoltaic devices made from these films showed good tolerance in device efficiency to variations in film composition. A device made from these films resulted in the highest total-area efficiency measured for any non-single-crystal, thin-film solar cell, at 15.9%.

The surface versus bulk composition and electronic structure of polycrystalline CuInSe2 thin-film interfaces were studied by synchrotron radiation soft-x-ray photoemission spectroscopy. An n-type In2Se3/CuIn3Se5 surface layer forms on enhanced-grain polycrystalline

thin-film p-type CuInSe2 during fabrication. Enhanced-grain CuInSe2 films were sputter etched (500 V Ar) and analyzed in situ to determine core-level binding energies and Fermi-level positions for the n-type surface and the p-type CuInSe2 bulk within +/- 0.1 eV. The transition between the n-type surface and the p-type bulk was experimentally observed by noting the change in the position of the valence-band maximum relative to the Fermi level E(F). From these measurements, the valence-band offset DELTAE(v) between the layers was determined to be 0.50 eV. Measurement of the work functions phi was also completed and reveals phi = 4.75 eV for the In2Se3 (CuIn3Se5) surface layer and phi = 4.04 eV for the bulk CuInSe2. Combining these results allows construction of a surface band diagram for this device configuration as well as determination of the relationship between composition, electronic structure, and device performance.

We report a world-record total-area efficiency of 17.1% for a polycrystalline thin-film Cu(In,Ga)Se-2-based photovoltaic solar cell. The incorporation of Ga to raise the absorber bandgap has been accomplished successfully and in such a manner that an open-circuit voltage of 654 mV and a fill factor of greater than 77% have been achieved. We describe briefly the deposition process, the device structure, and the device performance characteristics.

X-ray photoemission spectroscopy (XPS) and positron annihilation spectroscopy (PAS) have been used to characterize the surface versus bulk composition, electronic, and physical structure of polycrystalline Cu(In,Ga)Se/sub 2/ thin-film interfaces. Angle-resolved high-resolution photoemission measurements on the valence-band electronic structure and Cu 2p, In 3d, Ga 2p, and Se 3d core lines were used to evaluate the surface and near surface chemistry of CuInSe/sub 2/ and Cu(In,Ga)Se2 device grade thin films. XPS compositional depth profiles were also acquired from the near surface

region. PAS was used as a nondestructive, depth-sensitive probe for open-volume-type defects. Results of these measurements are related to device efficiencies to show the
effects of compositional variations and defect concentrations in the near surface region on device performance. (20

The formation chemistry and growth dynamics of thin-film CuInSe2 grown by physical vapor deposition have been considered along the reaction path leading from the CuxSe:CuInSe2 two-phase region to single-phase CuInSe2. The (Cu2Se)_beta (CuInSe2)_(1-beta) (0 <beta<= 1) mixed-phase precursor is created in a manner consistent with a liquid-phase assisted growth process. At substrate temperatures above 500 C and in the presence of excess Se, the film structure is columnar through the film thickness with column diameters in the range of 2.05.0 m. Films deposited on glass are described as highly oriented with nearly exclusive (112) crystalline orientation. CuInSe2:CuxSe phase separation is identified and occurs primarily normal to the substrate plane at free surfaces. Single-phase CuInSe2 is created by the conversion of the CuxSe into CuInSe2 upon exposure to In and Se activity. Noninterrupted columnar growth continues at substrate temperatures above 500 C. The addition of In in excess of that required for conversion produces an In-rich near-surface region with a CuIn3Se5 surface chemistry. A model is developed that describes the growth process. The model provides a vision for the production of thin-film CuInSe2 in industrial scale systems. Photovoltaic devices incorporating Ga with total-area efficiencies of 14.4%16.4% have been produced by this process and variations on this process.

IV-B-2-b. Applied Photo-voltaic Research Group (Mexico)

The cited group’s Abstracts are shown in Figure 23. The highest frequency single, adjacent double, and adjacent triple word phrases from their aggregate citing papers

aligned with the central themes of the aggregate cited can be represented by the following taxonomy: Materials/ coatings; Structure/ properties; Experiments/ measurements/ instruments/ variables; Processes; Geometries/ regions; Outputs/ devices/ applications.

Representative phrases (CAPITALIZED) aligned with the central themes of the aggregate citing papers offer the following intra-discipline portrait of the citing aggregate. These papers reflect a modest

experimental (SCANNING ELECTRON MICROSCOPY, X-RAY PHOTOELECTRON SPECTROSCOPY, X-RAY DIFFRACTION, ROOM TEMPERATURE) examination of

growth and deposition processes (CHEMICAL BATH DEPOSITION, DEPOSITED ON GLASS, FILMS WERE DEPOSITED, DEPOSITED THIN FILMS, ANNEALING THE FILMS, AIR ANNEALING, DEPOSITION TECHNIQUE) for

solar absorption and transmission control-related applications (SOLAR CONTROL COATING*, SOLAR ABSORBER COATING, ARCHITECTURAL GLAZING APPLICATIONS, SOLAR CELL*). These studies in aggregate focus on

materials and coatings (FILM*, CDS, COATING*, CUS, CU, PBS, THIN FILMS, BISMUTH SULFIDE, CUXS FILMS, CDS THIN FILM*, CUS THIN FILM*, FILMS ON GLASS) through examination of their

structures and properties (OPTICAL AND ELECTRICAL, OPTICAL TRANSMITTANCE SPECTRA, OPTICAL BAND GAP, OPTICAL ENERGY BAND GAP, GLASS SUBSTRATES, OPTICAL PROPERTIES, SHEET RESISTANCE, ELECTRICAL PROPERTIES, FILM THICKNESS, ELECTRICAL CONDUCTIVITY, VISIBLE TRANSMITTANCE, ABSORPTION MEASUREMENTS, SOLAR ABSORPTANCE, THERMAL EMISSION, INTEGRATED TRANSMITTANCE), in

surface-related geometries and regions (THIN, SURFACE, FILM*, COATING*).

Relative to the previous group of citing papers, the present group of citing papers appears to have more of a low technology focus, concentrating on integral film properties and film growth techniques for solar absorption control. Physical phenomena within the material are not investigated to nearly the same degree of detail as in the previous citing papers.

While the vast majority of phrases were aligned with the central themes of the group's cited papers, whose applications focused on solar control for buildings, a very few phrases reflected a slightly different application focus of solar control for automobiles.

(AUTOMOBILE WINDOWS, AUTOMOBILE GLAZING APPLICATIONS).

FIGURE 23 - CITED GROUP'S ABSTRACTS

Cadium sulphide thin films showing photo-to-dark conductivity ratio up to 10° photoconductivity up to 3* 1 cm -1 for white illumination ~ 300 W m -2 can be prepared from chemical baths containing thiourea and triethanolamine complex of cadmium ions. The photocurrent decay time depends on the bath temperature and the duration of storage and it ranges from a few seconds to 10 4 s per decade. The optical transmission of the film also varies significantly: from about 10% to 70-80% (above the band-gap absorption), depending on the deposition conditions. The high activation energy ~ e V for dark conductivity as well as the high photosensitivity suggest the nearly stoichlometric nature of the films. The implications of these characteristics in various opto-electronic applications are discussed.

Solar control coatings, required for architectural glazing applications in warm climates, must provide controlled optical transmission (~ 10-50%) of the solar radiation in the visible region and should reflect efficiently in the infrared (>0.7um) region to create a

cool interior in the buildings. Thin films of PbS and CuS on glass substrates, deposited from chemical baths, are shown to possess excellent solar control characteristics-superior or comparable to the metallic solar control coatings. For example, for an acceptable range of integrated optical transmittance (~10-20%) in the visible region, the integrated infrared reflectance for AM2 solar spectrum for the different glazing are: PbS coated glass, 50%: CuS goated glass, 14% stainless steel/Cu coated glass, 25% and tinted glass, 4%. The CuS and Pbs coatings also have the advantage of giving pleasant reflected colours (golden, purple, blue, etc), which improves the cosmetic appearance. This paper presents the basic requirements of solar control coatings and provides a comparison of the characteristics of PbS and CuS coatings against commercially available coatings.

CuS thin films with a wide range of sheet resistances (r) and optical transmittance (T%) indicating different composition x have been obtained from chemical baths constituted from copper (11) chloride, triethanolamine and thiourea at appropriate pH (10-12). Depending on the deposition parameters, a range of combination of r = 30to 1 M and T%(500 nm) = 1 to 65 and a range of colour of reflected daylight (golden yellow, purple, blue, green, etc) can be obtained. The films have been found to be stable with respect to electrical and optical properties of storage under ambient. Various possible large area applications such as in architectural glazing, photothermal and photovoltaic conversions are discussed.

A method is presented for the deposition of CdS thin films of 0.05-0.7 mu m thickness from solutions at 50-70 degrees C containing citratocadmium(II) complex ions and thiourea. The films show an optical band gap E(g)>2.6 eV. Optical transmittance is about 80% for photon energy <E(g). The dark conductivity of the films is of the order of 10(-8) Omega(-1) cm(-1). The photosensitivity of these films is high, 10(6)-10(7) under illumination with tungsten halogen light of 1 kW m(-2). Annealing in air at

400-500 degrees C for 1 h converts the films to n type. It is possible to obtain sheet resistances of about 150 Omega for a 0.2 mu m film (i.e., conductivity of 300 Omega(-1) cm(-1)) by this process. Conversion of the films to n type is possible also by immersing the film in a 0.01 M HgCl2 solution for 15 min followed by air annealing for 1 h at 200 degrees C. The films show n-type dark conductivity of similar or equal to 0.05 Omega(-1) cm(-1) and photoconductivity of similar or equal to 1 Omega(-1) cm(-1). X-ray diffraction and x-ray photoelectron spectroscopic depth profile studies on the films show that the modification of the electrical characteristics is brought about through changes in composition of the surface layers in the films.

IV-B-2-c. Basic Research Group (U. K.)

The cited group's Abstracts are shown in Figure 24. The highest frequency single, adjacent double, and adjacent triple word phrases from their aggregate citing papers aligned with the central themes of the aggregate cited papers can be represented by the following taxonomy: Theory/ modeling; Experiments/ measurements/ instruments/ variables; Structure/ Properties; Phenomena.

Representative phrases (CAPITALIZED) aligned with the central themes of the aggregate citing papers offer the following intra-discipline portrait of the citing aggregate. These papers reflect an integrated

theoretical (COUPLED NONLINEAR STOCHASTIC EQUATIONS, COUPLED-MAP LATTICE MODEL, MOLECULAR DYNAMICS SIMULATIONS, MONTE CARLO SIMULATIONS, PERTURBATION THEORY MODE-COUPLING, NON-STANDARD THERMODYNAMIC PARAMETERS, GRANULAR TEMPERATURE, STATISTICAL MECHANICS, THERMODYNAMIC EQUILIBRIUM, THERMODYNAMIC THEORIES, THERMODYNAMIC CHARACTERISTICS, KINETIC THEORY) and

experimental (BEADS WERE CONFINED, EVENT SIZE DISTRIBUTION, RELAXATION TIME DISTRIBUTIONS, NUCLEAR

MAGNETIC RESONANCE, VIBRATION INTENSITY, ELECTROMAGNETIC VIBRATION EXCITER, IMAGE PROCESSING, VIBRATED GRANULAR, EXPERIMENT*, VIBRATION) effort targeted at studying the

motions of aggregated granular particles (MECHANICS OF POWDERS). The papers focus on examination of the

structure(s) and properties of vibrating sandpiles (ANGLE OF REPOSE, AXIS OF ROTATION, RADIAL DISTRIBUTION FUNCTION, CLUSTERS AND MOBILE GRAINS, MIXTURE OF GRAINS, GRANULAR MATERIAL*, GRANULAR MEDIA, MIXTURES OF PARTICLES, MONODISPERSE SPHERICAL BEADS, PROLATE ELLIPSOIDS, PORE SIZE*, PACKING DENSITY, BINARY MIXTURES, GRANULAR PILE, VOLUME FRACTION, RANDOMLY PACKED), and the

intrinsic phenomena of these collective systems (DILATANCY AND HYSTERESIS, EFFECT OF VIBRATION, ENERGY DISSIPATION, DENSITY FLUCTUATIONS), with emphasis on

segregation (AXIAL SEGREGATION, SEGREGATED STATE, SEGREGATED STATE, SEGREGATION PROCESSES),

relaxation (RELAXATION TIME DISTRIBUTIONS, INTERNAL RELAXATION DYNAMICS, RELAXATION OF TRANSIENTS, RELAXATIONAL BEHAVIOR, DENSITY RELAXATION, NMR RELAXATION, FASTER RELAXATION, SLOW RELAXATION, LONGITUDINAL RELAXATION, RELAXATION PROPERTIES),

fluidization (CONCEPT OF FLUIDIZATION, CONVECTION AND FLUIDIZATION, ONSET OF FLUIDIZATION, FLUIDIZED STATE, FLOWING PHASE, FLUIDIZATION TRANSITION, STREAMWISE DIRECTION),

avalanching (NONLINEAR INSTABILITY MECHANISM*, SANDPILE AVALANCHES, LARGE AVALANCHES, AVALANCH PROPAGATION, , AVALANCH*), and

collective behaviors (INDEPENDENT-PARTICLE AND COLLECTIVE, COLLECTIVE EXCITATIONS, COLLECTIVE RELAXATIONS, COLLECTIVE REORGANIZATION).

While the citing paper phrases reflected emphasis on studies of granular piles, the phenomenological results and insights on segregation, relaxation, fluidization, avalanching, and collective behavior were extrapolated to extra-discipline applications. These include

granular mixing (TYPES OF RICE, DEMONSTRATE THAT FOODSTUFFS, FOOD CARBOHYDRATES LIPIDS, LIPIDS AND PROTEINS, RESPONSES OF FOODSTUFFS, CHARACTERIZE FOODS),

traffic flow (TRAFFIC FLOW, CARS, EXPRESSWAY),

phase transition (GLASS TRANSITION, FILM BOILING, ICE),

materials (ALLOY, TUNGSTEN, RHEOLOGY),

multi-component systems (CHARACTERISTICS OF GELS, EMULSIONS AND POWDERS, GELS FOAMS EMULSIONS), and

micro systems (CLASSES OF MOLECULES, WATER PROTON TRANSVERSE).

FIGURE 24 – CITED GROUP'S ABSTRACTS

The authors use a new formation of the statistic mechanics of powders to develop a theory for mixture of grains of two different sizes. They map this problem onto the spin formulation of the eight-vertex model and reproduce the features of the phase separation diagram of the powder mixture that they would intuitively be led to expect. Finally, they discuss the insight afforded by this solution of the 'thermodynmaic' quantities of interest in the powder mixture.

We present a microscopic model of a granular pile submitted to vibration, emphasizing the competing roles of collective and single-particle excitations which lead to specific characteristics of the resulting state. The effect of vibration on a powder is next investigated by a

novel computer-simulation approach in three dimensions, which enables us to probe cooperative effects in the bulk; satisfying agreement with the predictions of the model is obtained.

The structure and dynamics of powders subjected to vibration are investigated by a nonsequential and cooperative computer-simulation approach in three dimensions. Starting from a microscopic model of the physics, we are able to probe independent and collective effects in the dynamics of vibrated powders, as well as in the resulting structures. In particular, we analyze the role of cooperative structures such as bridges, which are always present in reality and which cannot be formed by purely sequential processes. We look in depth at the behavior of the volume fraction and coordination number as a function of the intensity of vibration, as well as at correlation functions describing contacts between neighboring grains, also as a function of intensity. Satisfying agreement with the qualitative predictions of earlier analytic work is obtained, and a framework is laid for future investigations.

We present a set of coupled nonlinear stochastic equations in one space dimension, designed to model the surface of an evolving sandpile. These include nonlinear couplings to represent the constant transfer between relatively immobile clusters and mobile grains, incorporate the presence of tilt, and contain representations of inertia and evolving configurational disorder. The critical behavior of these henomenological equations is investigated numerically. It is found to be diverse, in the sense that different combinations of noise as well as different symmetries lead to nontrivial exponents. In the cases most directly comparable with previous studies, we find that our equations lead to a surface with a roughness exponent alpha(tilt)approximate to 0.40, to

be compared with the Edwards-Wilkinson and Kardar-Parisi-Zhang values, namely alpha(EW) = 1/4 and alpha(KPZ) = 1/3, respectively. This is, in our view, directly due to the effect of the tilt term. Finally we discuss our results, as well as possible modifications to our equations.

IV-B-2-d. Basic Research Group (U. S.)

The Abstract of the group's single highly cited paper is shown in Figure 25.

FIGURE 25 - CITED PAPER ABSTRACT

**Granular materials display a variety of behaviors that are in many ways
different from those of other substances. They cannot be easily classified as either solids or liquids. This has prompted the generation of analogies between the physics found in a simple sandpile and that found in complicated microscopic systems, such as flux motion in superconductors or spin glasses. Recently, the unusual behavior of granular systems has led to a number of new theories and to a new era of experimentation on granular systems.**

The highest frequency single, adjacent double, and adjacent triple word phrases from the aggregate citing papers aligned with the central themes of the cited paper can be represented by the following generic taxonomy: Theory/ modeling; Experiments/ measurements/ instruments/ variables; Structure/ Properties; Phenomena.

Representative phrases (CAPITALIZED) aligned with the central themes of the aggregate citing papers offer the following intra-discipline portrait of the citing aggregate. These papers reflect a balanced

theoretical/ modeling effort (AGREEMENT WITH EXPERIMENT*, MOLECULAR DYNAMICS SIMULATIONS, EQUATIONS OF MOTION, DIFFUSING VOID MODEL, MOLECULAR DYNAMICS DISSIPATION, NUMERICAL SIMULATIONS, MONTE CARLO SIMULATIONS, PARTICLE DYNAMICS SIMULATIONS, POWER LAW, KINETIC THEORY, STATISTICAL MECHANICS, THERMODYNAMIC LIMIT, GRANULAR TEMPERATURE, CELLULAR AUTOMAT*, PROBABILITY DISTRIBUTION) and

experimental effort (MAGNETIC RESONANCE IMAGING, VERTICAL GLASS PIPE, VELOCITY DISTRIBUTION FUNCTION*, EXPERIMENTAL RESULTS, EXPERIMENTAL DATA, EXPERIMENTAL OBSERVATIONS, VELOCITY PROFILES, FORCE BALANCE, IMAGE PROCESSING, VERTICAL SHAKER, CONVEYOR BELTS, POWER SPECTR*, DIGITAL HIGH SPEED PHOTOGRAPHY, CHARGE COUPLED DEVICE CAMERA) targeted at studying the motions of granular particles. The papers focus on examination of the

structure(s) and properties of vibrating sandpiles (ANGLE OF REPOSE, COEFFICIENT OF FRICTION, COEFFICIENT OF RESTITUTION, NUMBER OF BEADS, ROLLING AND STATIC, GRANULAR MATERIAL*, GRANULAR MEDIA, LARGE GRAINS, SMALL GRAINS, FREE SURFACE, GRANULAR MEDIUM, GRANULAR SYSTEMS, ROLLING GRAINS, GRANULAR FLOW, STATIC FRICTION, GRANULAR, GRAINS, PARTICLES), and the

intrinsic phenomena of these collective systems (DISSIPATION OF ENERGY, AGGLOMERATION AND DEAGGLOMERATION, COLLISIONS BETWEEN PARTICLES, ENERGY AND SHEAR, FRACTIONAL BROWNIAN MOTION, DENSITY WAVES, STEADY STATE, INELASTIC COLLISIONS, PARTICLE DYNAMICS, VIBRATIONAL ACCELERATION, DENSITY FLUCTUATIONS, WAVE AND ARCHING), with emphasis on

segregation (SEGREGATION PHENOMENA, SIZE SEGREGATION, SPONTANEOUS STRATIFICATION, AXIAL SEGREGATION, RADIAL SEGREGATION, SEGREGATE DOWNHILL, SEGREGATE UPHILL, PHASE SEGREGATION, DYNAMIC SEGREGATION, FLOW-INDUCED

SEGREGATION, FREE-SURFACE SEGREGATION, SEGREGATION AND STRATIFICATION, SEGREGATION OF GRANULAR),

relaxation (RELAXATION DYNAMICS, RELAXATION TIME TAU, RELAXATION),

avalanching (NONLINEAR INSTABILITY MECHANISM, AVALANCH*, DISTRIBUTION OF AVALANCH*, AVALANCH FLOW, AVALANCH DURATIONS, AVALANCH SIZE, DISCRETE AVALANCHES, HYSTERETIC TRANSITION),

fluidization (ONSET OF FLUIDIZATION, FLOW OF GRANULAR, FLOW OF SAND, FORMATION OF CONVESTION, FLOW IS INTERMITTENT, SURFACE FLOW, CONVECTION CELLS, CONVECTIVE MOTION, FLOW RATE, FLOW REGIME, FLOW PATTERNS), and

collective behaviors (COLLECTIVE PARTICLE MOTION, SELF-ORGANIZED CRITICAL DENSITY, SELF-ORGANIZED CRITICALITY, COOPERATIVE REARRANGEMENTS, LONG-RANGE CORRELATIONS, COLLECTIVE EFFECTS, LONG-RANGE FORCES).

While the citing paper phrases mainly reflected emphasis on studies of granular piles, the phenomenological results and insights on segregation, relaxation, fluidization, avalanching, and collective behavior were extrapolated to some extra-discipline applications. These include (sample category record Titles follow the phrases):

geological formations and processes (ROCK*, GEOLOGICAL, AEOLIAN, EARTHQUAKE*, PYRENEAN, SEISMIC, SEDIMENT*, LATE PALAOCENE, LOWER EOCENE, ROCK AVALANCHES, ROCK FRAGMENTATION, TECTONIC FORCING, CARBONATE TURBUDITE DEPOSITION, SAND DUNE*),
****Scaling of the critical slip distance for seismic faulting with shear strain in fault zones***
****Sedimentary evolution of the early Paleogene deep-water Gulf of Biscay***
****A fragmentation-spreading model for long-runout rock avalanches***

Distinguishing climatic and tectonic forcing of turbidite
The physics of debris flows

soil mechanics (SOIL*, SOIL MECHANICS, DEPTH OF PENETRATION, DECELERATION AND DEPTH, HILLSLOPES, HILLSLOPE GRADIENT),
Evidence for nonlinear, diffusive sediment transport on hillslopes and implications for landscape morphology
Analysis of vertical projectile penetration in granular soils

industrial applications (SCREW FEEDER*, INDUSTRIAL),
Precision dosing of powders by vibratory and screw feeders

interacting object dynamics (TRAFFIC, AIRSPACE, TRAFFIC CONGESTION, AIR TRAFFIC, TRAFFIC EQUATIONS, TRAFFIC FLOW, TRAFFIC MODEL, WAR GAME*).
Study on the dynamical behavior of a real-time distributed simulation system
Study on crowded two-dimensional airspace - Self-organized criticality
Derivation and empirical validation of a refined traffic flow
Kink soliton characterizing traffic congestion

materials (RHEOLOG*, METAL, UNTWINNED SINGLE CRYSTALS, CRYSTAL, MAGNETIZATION, MAGNETIC HYSTERESIS, MAGNETOMETER, FLUX JUMPS, CHEMICAL SHIFT TENSORS),
The controlled formation of ordered, sinusoidal structures by plasma oxidation of an elastomeric polymer
Vortex avalanches at one thousandth the superconducting transition temperature
Vanishing magnetization relaxation in the high-field quantum limit in yba2cu3o7 delta
Mesoscale self-assembly of hexagonal plates using lateral capillary forces: Synthesis using the "capillary bond"

Collective effects during crystal growth in the presence of mobile nonreactive impurities: experiments and simulations
Determination of the orientation of Si-29 chemical shift tensors using rotorsynchronized MAS NMR of single crystals: forsterite (Mg2SiO4)

films (POLYMER*, PDMS, FILM*, MONOLAYER*, POLYDIMETHYLSILOXANE, CONFINED FILMS, LIQUID FILMS, MOLECULAR FLUIDS, MOLECULAR LAYERS, ADSORBED POLYMER LAYERS, FILM THICKNESS, STICK-SLIP MOTION),
A model for the static friction behaviour of nanolubricated contacts
Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer
Stick to slip transition and adhesion of lubricated surfaces in moving contact

multi-phase systems (FOAM, PELLET*, SLUGGING, FLOW IMMUNOSENSORS, INTERCELLULAR, FLUIDIZED BED*, CELL PELLET*, FOAM PATTERN*, WATER TRANSPORT, DEBRIS-FLOW*, SUSPENSION*, SLURRIES, COLLOIDAL),
Advances in flow displacement immunoassay design
Foam pattern drift in a vibrating flask
Studies of cell pellets .1. electrical-properties and porosity
Rheophysical classification of concentrated suspensions and granular pastes
From bubbles to clusters in fluidized beds
Pseudo-solitons in fluidized-beds
Maximum-entropy analysis of disordered droplet

gas dynamics (GAS FLOW, GAS*, SHOCK WAVES, SHOCK FRONT, GRANULAR GAS*),
Statistics of shock waves in a two-dimensional granular flow
Scale invariant correlations in a driven dissipative gas
Steady uniform shear flow in a low density granular gas

micro particles (POLLEN*, POLLEN EXINE SCULPTURING, MOLECULAR, ATOMICALLY, SPINULE, GENETIC) and
The effects of genotype and ploidy level on pollen surface sculpturing in maize

microscale cooperative effects (TOKAMAK, PLASMA*, LATTICE GAS).
Sandpiles, silos and tokamak phenomenology: a brief review
Magnetic flux instabilities in superconducting niobium rings:
Logarithmic relaxations in a random-field lattice gas subject to gravity

It should be remembered that the phrases selected above had a threshold frequency of two. For the specific example of the U. S. basic research group, the impact of neglecting phrases with unity frequency was examined. The analysis was re-run using a unity frequency threshold, and the increase in volume of total phrases that had to be examined was enormous. As an alternative to identifying applications and extra-discipline impacts by searching for anomolous phrases, each of the ~300 citing paper Abstracts was read, and those Abstracts reflecting applications and extra-discipline impacts were identified.

All of the applications and extra-discipline papers identified from reading the Abstracts could be identified/ retrieved from examination of the anomolous text mining-derived phrases. The applications taxonomy of the previous section was validated as a reasonable classification of the applications and extra-discipline impacts. Identification of the applications and extra-discipline impacts most unrelated to the main themes of the cited paper was easiest because of the highly anomolous nature of their representative phrases. Identification of the intra-discipline applications was the most difficult, since the phraseology used was similar to that of the cited paper themes.

***The importance of this result should be emphasized*. A complete citation impact study will typically involve multiple generations of citations. For a citation impact study that involves large numbers of first-generation citing papers and/ or large numbers of succeeding generation citing papers, reading each citing paper Abstract to identify applications paper characteristics becomes infeasible. For example, the ~300 citing papers of the sandpile paper were themselves cited by ~3600 papers.**

Use of text mining capabilities, such as computational linguistics, allows only those applications and extra-discipline papers of interest to be identified, and the requisite information can then be obtained from reading the Abstracts. In addition, the computational linguistics provides a structure and categorization of these myriad applications, allowing the larger context of application themes to be displayed and understood. While phrase frequency algorithms were used for the present study, and proved adequate, specifically-tailored co-occurrence and clustering algorithms are being developed to improve the efficiency of the application papers identification and retrieval process.

The citing papers representing different categories of development and different disciplines from those of the cited paper are shown in the matrix of Figure 26, and portrayed graphically in Figure 27.

In Figure 26, the abscissa represents time. The ordinate, in the second column from the left, is a two character tensor quantity. The first number represents the level of development characterized by the citing paper (1=basic research; 2=applied research; 3=advanced development/ applications), and the second number represents the degree of alignment between the main themes of the citing and cited papers (1=strong alignment; 2=partial alignment; 3=little alignment).

Each matrix element represents the number of citing papers in each of the nine categories.

FIGURE 26 – DEVELOPMENT CATEGORY AND CITED PAPER THEME ALIGNMENT OF CITING PAPERS

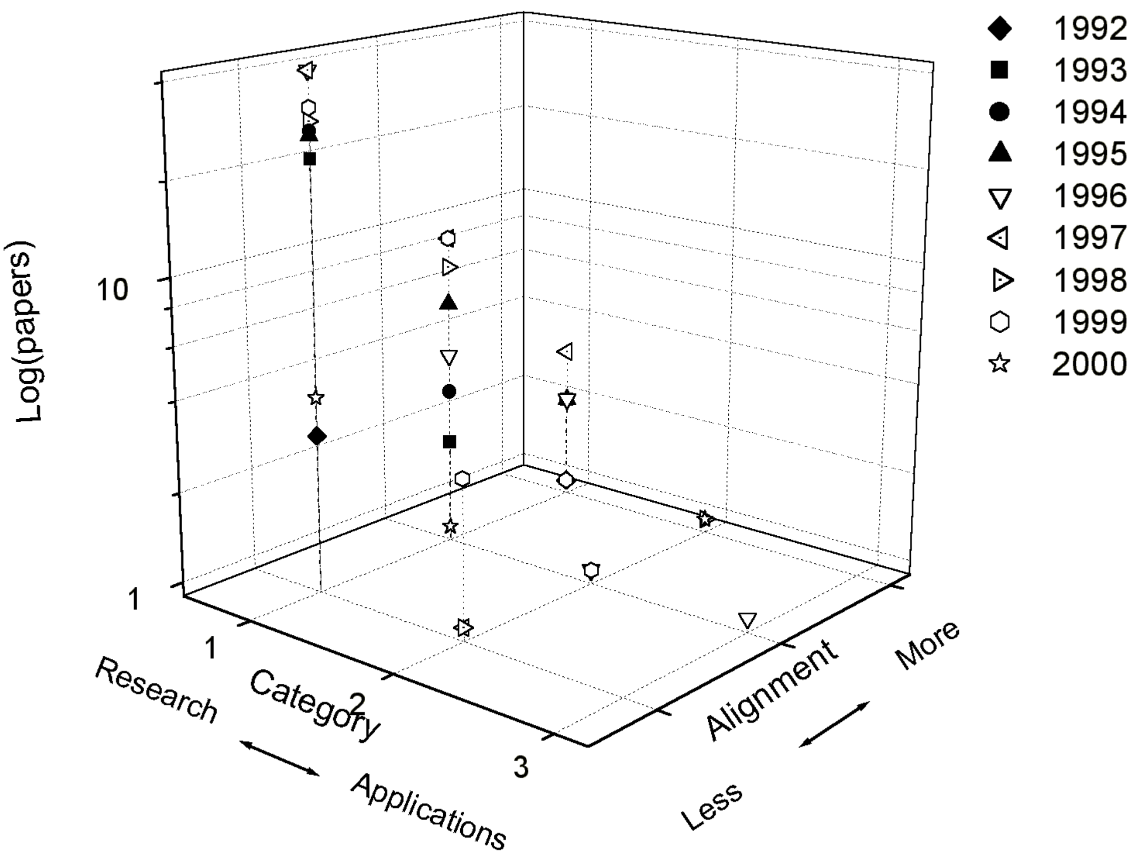
TECH DEV	33									
TECH DEV	32					1				
TECH DEV	31									
APPL RES	23							1	1	1
APPL RES	22					1			3	
APPL RES	21						1	1		
BAS RES	13	1	2	2	2	2	3		1	
BAS RES	12		2	3	6	4	10	8	10	1
BAS RES	11	3	23	28	27	43	43	30	33	4
		1992	1993	1994	1995	1996	1997	1998	1999	2000

TIME

CODE: MATRIX ELEMENT IS NUMBER OF PAPERS

In Figure 27, the axes are Category, Alignment and Papers

FIGURE 27 – CITING PAPER DISCIPLINES AND CATEGORY LEVELS



There are three interesting features on Figures 26/ 27. First, the tail of total annual citation counts is very long, and shows little sign of abating. This is one characteristic feature of a seminal paper.

Second, the fraction of extra-discipline basic research citing papers to total citing papers ranges from about 20-40% annually, with no latency period evident. This instant extra-disciplinary diffusion may have been due to the combination of intrinsic broad-based applicability of the subject matter and publication of the paper in a high-circulation science journal with very broad-based readership. The impact of journal type on impact diversity will be examined in a future study.

Third, there was a four-year latency period before the higher development category citing papers began to emerge. This correlates with the results from the bibliometrics component. From the present study, it is not possible to differentiate the reasons for this important result. The latency could have been due to the inability of the technology community to recognize the potential applications of the science immediately. Or, it could have been due to the information remaining in the basic research journals, and not reaching the applications community. Or the time that an application needs to be developed is of the order of four years. Thus, the basic science publication feature that may have contributed heavily to extra-discipline citations may also have limited higher development category citations for the latency period. This hypothesis will also be tested in a future study.

Finally, another variable was examined during the reading of the ~300 citing paper Abstracts. The alignment of the citing journal theme to the main theme of the cited paper was estimated for all citing papers. In essentially all cases, the citing paper theme could be subsumed within the citing journal theme. However, given the breadth of most journal themes, this result had minimal information content (e.g., citing paper X was published in a Physics journal vs. a Materials journal).

In Davidse's study of Physics papers citations (Davidse, 1997), a key metric used in cross-disciplinary citations/ impacts is the distinction between Physics and non-Physics papers. It is implicitly assumed that the flow from Physics to non-Physics papers is analogous to the flow from basic to applied. While it may be true for some cases, Figures 26 and 27 show that most extra-discipline flows are from basic physics research to basic research in the other disciplines.

Davidse used journal themes (based on the SCI journal classification taxonomy) as a proxy for citing paper

themes, with the level of resolution being at the gross discipline description, at best. There are many multi-discipline journals (e.g., Science, Nature, etc) that render a thematic distinction impossible. VanRaan's approach required such a computerized proxy representation, since tens of thousands of citing papers were analyzed.

In contrast, the present paper's approach of identifying impact themes through text mining allows a much more detailed and informative picture of the impact of research to be obtained. It represents the difference between stating that a "Physics paper impacted Geology research" and a "paper focused on sand-pile avalanches for surface smoothing impacted analyses of steep hillslope landslides".

V. INTEGRATED SUMMARY AND CONCLUSIONS

Exploitation of the other types of information contained in the SCI and associated with the citation process offers the potential for providing R&D sponsors information that can help guide future directions of their R&D. In addition, the complete citation mining process described in the present paper has the potential to objectively document the breadth of impact of basic research on the R&D community. The addition of text mining to citation bibliometrics will make feasible the large-scale multi-generation citation studies that are necessary to display the full impacts of research.

Highlights of the bibliometrics results include:

The journal discipline results suggest that the applications developed by the Mexican Applied photovoltaics group have a strong impact on chemical papers, while the applications developed by the U. S. Applied photovoltaics group strongly impact physics papers. A point to be stressed is that only the

fundamental papers received cites in journals clearly outside of their disciplines.

For the Fundamental cited papers, citing papers from disciplines other than that of the cited paper emerge with the passage of time. An important point is the four-year delay of the systematic appearance of engineering and materials science citing papers. Only the fundamental papers are cited in notes and only the U. S. Fundamental paper was cited in an editorial document.

Academia has the highest citing rates. Industry follows closely the advance of high-technological developments, but is not following directly the advances in fundamental research. Research Centers follow applied and fundamental research about equally. Direct government participation is not significant in the fields studied.

USA has the most cites. Based on the limited samples examined, mostly developing countries (e.g., India) cite the Mexican Applied photovoltaics set; while developed countries (e.g., Japan) cite the U. S. Applied photovoltaics set. This fact is due to the different nature of the applied technology developed by these two groups. The U. S. Applied photovoltaics set contains work related to high technology and the Mexican Applied photovoltaics set is dedicated to explore low-cost technology. Therefore, this last set is cited by the less affluent countries of India, Romania and Mexico. India and Mexico also cite fundamental research, but not Romania. Thus, in the limited sample, the low-cost technology papers are cited by developing countries, whereas the high-technology papers are cited mainly by the developed countries.

In terms of average references per paper, the applied papers dominate at the lower end of the spectrum (0-20). The fundamental papers dominate at the higher end of the spectrum (20-50+), with the exception of a British fundamental anomaly at 41-50.

The highest frequency co-cited papers have three interesting characteristics. They are essentially all in the same general physics area, they are all published in fundamental science journals (mainly physics), and they are all relatively recent, indicating a dynamic research area with high turnover.

Highlights of text mining results include:

The central themes and specific phrases used in the cited papers are replicated in the citing papers. In the applied cited papers, there were no phrases in the citing papers that represented themes or disciplines significantly different from those in the cited paper. One suspects that the audience obtained is the target audience for these cited papers, at least in terms of thematic interest, but the authors of these cited papers were not contacted to confirm this conclusion.

In the fundamental cited papers, about twenty percent of the citing papers contained phrases that represented themes or disciplines different from those in the cited paper. This reflects one of the characteristics of a truly fundamental science paper, namely, the diffusion of basic scientific findings and principles to a variety of applications areas. Thus, the lessons learned from vibrating powders are extrapolated to the study of metals and alloys, traffic, aquifers, ice, food, nuclear and magnetic phenomena .

The U. S. Applied photovoltaics citing papers can be represented by the following taxonomy: Materials/ coatings; Structure/ properties; Experiments/ measurements/ instruments/ variables; Processes; Geometries/ regions; Phenomena; Outputs/ devices/ performance/ applications. These papers reflect a mainly experimental examination of solar cell-related applications with a central objective of performance enhancement. In aggregate, these studies focus on

materials and coatings through examination of their structures and properties, growth processes, physical phenomena, in surface-related geometries and regions. These citing papers reflect a high tech approach to developing high performance solar conversion devices for electricity production.

The Mexican Applied photovoltaics papers can be represented by the following taxonomy: Materials/ coatings; Structure/ properties; Experiments/ measurements/ instruments/ variables; Processes; Geometries/ regions; Outputs/ devices/ applications. These papers reflect a modest experimental examination of growth and deposition processes for solar absorption and transmission control-related applications. These studies in aggregate focus on
materials and coatings through examination of their structures and properties, in surface-related geometries and regions.

Relative to the U. S. Applied photovoltaics group of citing papers, the Mexican Applied photovoltaics group of citing papers appears to have more of a low technology focus, concentrating on integral film properties and film growth techniques for solar absorption control. Physical phenomena within the material are not investigated to nearly the same degree of detail as in the U. S. group's citing papers.

The British Fundamental sandpile papers can be represented by the following taxonomy: Theory/ modeling; Experiments/ measurements/ instruments/ variables; Structure/ Properties; Phenomena. These papers reflect an integrated theoretical and experimental effort targeted at studying the motions of aggregated granular particles. The papers focus on examination of the structure(s) and properties of vibrating sandpiles, and the intrinsic phenomena of these collective systems, with emphasis on segregation, relaxation, fluidization, avalanching , and collective behaviors.

While the citing paper phrases reflected emphasis on studies of granular piles, the phenomenological results and insights on segregation, relaxation, fluidization, avalanching, and collective behavior were extrapolated to some extra-discipline applications. These include granular mixing, traffic flow, phase transition, multi-component systems, and micro systems.

The U. S. Fundamental sandpile papers can be represented by the following generic taxonomy: Theory/ modeling; Experiments/ measurements/ instruments/ variables; Structure/ Properties; Phenomena. These papers reflect a balanced theoretical/ modeling effort and experimental effort targeted at studying the motions of granular particles. The papers focus on examination of the structure(s) and properties of vibrating sandpiles, and the intrinsic phenomena of these collective systems, with emphasis on segregation, relaxation, avalanching, fluidization, and collective behaviors.

While the citing paper phrases reflected emphasis on studies of granular piles, the phenomenological results and insights on segregation, relaxation, fluidization, avalanching, and collective behavior were extrapolated to some extra-discipline applications. These include: geological formations and processes, soil mechanics, industrial applications, interacting object dynamics, materials, films, multi-phase systems, gas dynamics, micro particles and microscale cooperative effects.

As an alternative to identifying applications by searching for anomolous phrases, each of the ~300 citing paper Abstracts was read, and those Abstracts reflecting extra-discipline and technology applications were identified. All of the applications papers identified from reading the Abstracts could be identified/ retrieved from examination of the anomolous text mining-derived phrases. The applications taxonomy of the previous section was validated as a reasonable classification of

the applications. Identification of the applications most unrelated to the main themes of the cited paper was easiest because of the highly anomolous nature of their representative phrases. Identification of the intra-discipline applications was the most difficult, since the phraseology used was similar to that of the cited paper themes.

***The importance of this result should be emphasized*. A complete citation impact study will typically involve multiple generations of citations. For a citation impact study that involves large numbers of first-generation citing papers and/ or large numbers of succeeding generation citing papers, reading each citing paper Abstract to identify applications paper characteristics becomes infeasible. Use of text mining capabilities, such as computational linguistics, allows only those applications papers of interest to be identified, and the requisite information can then be obtained from reading the Abstracts. In addition, the computational linguistics provides a structure and categorization of these myriad applications, allowing the larger context of application themes to be displayed and understood. While phrase frequency algorithms were used for the present study, and proved adequate, specifically-tailored co-occurrence and clustering algorithms are being developed to improve the efficiency of the application papers identification and retrieval process.**

For the U. S. fundamental cited sandpile paper, there are three interesting features from detailed analysis of the citing paper Abstracts. First, the tail of total annual citation counts is very long, and shows little sign of abating. This is one characteristic feature of a seminal paper.

Second, the fraction of extra-discipline basic research citing papers to total citing papers ranges from about 20-40% annually, with no latency period evident. This instant extra-disciplinary diffusion may have been due to

the combination of intrinsic broad-based applicability of the subject matter and publication of the paper in a high-circulation science journal with very broad-based readership. The impact of journal type on impact diversity will be examined in a future study.

Third, there was a four-year latency period before the higher development category citing papers began to emerge. From the present study, it is not possible to differentiate the reasons for this important result. The latency could have been due to the inability of the technology community to recognize the potential applications of the science immediately. Or, it could have been due to the information remaining in the basic research journals, and not reaching the applications community. Thus, the basic science publication feature that may have contributed heavily to extra-discipline citations may also have limited higher development category citations for the latency period. This hypothesis will also be tested in a future study.

Finally, another variable was examined during the reading of the ~300 citing paper Abstracts. The alignment of the citing journal theme to the main theme of the cited paper was estimated for all citing papers. In essentially all cases, the citing paper theme could be subsumed within the citing journal theme. However, given the breadth of most journal themes, this result had minimal information content (e.g., citing paper X was published in a Physics journal vs. a Materials journal).

In contrast, the present paper's approach of identifying impact themes through text mining allows a much more detailed and informative picture of the impact of research to be obtained. It represents the difference between stating that a "Physics paper impacted Geology research" and a "paper focused on sand-pile avalanches for surface smoothing impacted analyses of steep hillslope landslides".

In summary, text mining is a requirement for making the total citation mining possible. Without text mining, either an overly general automated technique, such as journal classification, must be used to identify application areas, or tens or hundreds of thousands of Abstracts must be read. Text mining can locate small numbers of extra-discipline phrases (small signals) from large numbers of intra-discipline phrases (large clutter), and allow only those Abstracts of specific interest to be selected and read.

In addition, there is a very important message that emerges from the results of the present study relative to the sponsorship of basic research. Over the past decade, the trend in industry and government has been toward requirements-driven research (e.g., the term 'strategic research' is becoming used more widely in government agencies, and corporately-funded industrial research has strongly evolved into profit-center sponsored research). While this may be beneficial to the sponsoring organization from a short-term tactical perspective, the long-term strategical perspective may suffer. Would fundamental sandpile research receive funding from Tokomak, air traffic control, or materials programs, even though sandpile research could impact these or many other types of applications, as shown in this paper? It is necessary to stress that sponsorship of some unfettered research must be protected, for the strategic long-term benefits on global technology and applications!

VI. LESSONS LEARNED

There were three major lessons learned. First, in order to develop generalized conclusions about customer profiles and impacts of research, a substantial statistically representative number of cited papers must be examined. The amount of labor required should not be underestimated.

Second, in order to obtain a comprehensive understanding about the citing population and the broader impacts of research, multiple generations of citing papers must be examined. The method of allocating impact of the cited paper across generations of citing papers will be very important in estimating total impact of a cited paper population.

Third, citation mining requires expertise in the intra-disciplines of the citing papers, some expertise in the extra-disciplines of the citing papers, and expertise in understanding the information technology techniques required for the analyses. Intra-discipline expertise is required in order to be able to differentiate phrases and themes aligned with the cited paper(s) from those non-aligned, and to place the main pervasive themes and their relationships in the citing papers in larger context. Extra-discipline expertise is required in order to link the anomolous phrases to applications, and understand the sub-themes that are transcending disciplines and the insights that are extrapolating to different application areas. Thus, the citation mining process must intrinsically be flexible and recursive, since all the applications expertise areas will not be known beforehand, but will become evident as the study proceeds. Finally, intra- and extra-discipline expertise are necessary, but not sufficient conditions, for performing a complete and credible citation mining analysis. Selection of the appropriate citation mining tools and techniques will depend on the study objectives and the data required/ available, and will require an in-depth understanding of bibliometrics, citation analysis, and text mining. Incorporation of information technology expertise in citation mining is critical to success.

VII. FUTURE DIRECTIONS

The present study is a first step in understanding the types of results and insights available from multi-field

analysis of citing papers and trans-citation text mining. Future studies should add the following capabilities.

VII-A. Multi-Field Text Mining

A recent text mining study of Aircraft S&T demonstrated that analysis of different SCI database fields (Abstracts, Keywords) provided different perspectives on the contents of the aggregated papers. This result derives from the different purposes assigned by authors to each field. The present study text mined Abstracts only. Future studies should text mine other fields as well, and examine the effect of these different fields on trans-citation similarities. Fields recommended for study include Keywords, Titles, Abstracts, and Full-Text.

VII-B. Selection of Comparison Base

More base cited papers than were used in the present study are required to generalize the results. A strategy for cited paper selection must be developed for generalization based on scientific grounds.

Cited papers from different fundamental disciplines should be compared, to understand the effect of discipline. For example, physics, chemistry, mathematics, biology, etc., papers should be compared.

Papers from different levels of development in the same discipline should be compared, to understand the effect of level of development. For example, cited basic research, applied research, and technology development papers from one discipline should be compared.

Cited papers from different discipline sub-areas should be compared, to understand the effect of sub-area. For example, in the present study, the highly cited sandpile paper could be decomposed into a number of discipline sub-areas, such as segregation, relaxation, avalanching,

fluidization, and collective behaviors. Cited papers from each of these sub-disciplines should be compared.

Cited papers from different types of journals in the same discipline and level of development should be compared, to understand the effect of journal type. In the present study, the one highly cited sandpile paper was published in a general science journal of wide circulation. This paper appeared to have relatively high extra-disciplinary impact. Was this type of impact due to the intrinsic thematic nature of the paper, or was it due to the multi-disciplinary nature of the journal's readership (or both). Similar papers from as many different journal types as possible should be examined to understand this important finding.

VII-C. Text Mining and Bibliometric Profiling across Generations of Citations

The present study examined text mining and bibliometric profiling across one generation of citations (although there could have been some additional generations of citations within the large citing paper groups). Future studies should explicitly examine the impact diffusion through multiple generations of citations using bibliometric profiling and text mining. Since the numbers of documents can expand rapidly when generations of papers resulting from a highly cited paper are examined, methods for automating bibliometric profiling should be developed and investigated.

VII-D. Individual and Organizational Aggregations

There are a number of levels of individual aggregation that can be studied for both the text mining and the bibliometric profiling. The present study examined only the lowest levels (individual and small research group). Future studies should examine rigorously the effect of individual cited papers, research group, organization,

groups of organizations, countries, and global disciplines.

VII-E. Ordering of Citations in Text

Most, if not all, citation studies do not consider the order of citations in the text as part of the analysis. Yet, the citation sequencing may influence the natural logic flow within a paper, and could impact the association metrics for different papers. The citations in the full text may be viewed as a string, and may be decomposed for purposes of text similarity matching using standard methods of string matching (e.g., suffix tree decomposition). With the increasing availability of full text documents on-line, the study of citation ordering for text may have payoff, if successful.

VII-F. Statistical Representation

The present study selected sufficient citing papers to insure adequate representation within the different citing paper variable bands. Future studies need to pay attention to statistical representation for comparing across the different groups suggested in the previous sections.

VII-G. Importance of References on Document as Function of Linguistic Similarities

In citation studies, all references are treated as equal in their influence on the citing document. Obviously, some references will have far more influences on the final citing document than others. It might be instructive to ask selected document authors to rank the influence of different references on their published paper, then text mine the references and text to ascertain whether text similarity between reference and citing paper correlates with influence of the reference and citing paper. Full paper text mining might be more useful here.

VII-H. Integration of Citation Links with other Types of Document Links

The overall objective of S&T influence studies is to understand the larger puzzle, in which there are many pieces. The pieces need to be documented, but their connecting links need to be understood. Trans-citation analyses are one component for understanding links, but there are many other types of links, and these must be integrated with the trans-citation analyses for the larger picture. Future studies should identify the multiplicity of link types, and attempt to integrate these other types of link analyses with text mining and bibliometric profiling.

VII-I. Different Variable and Band Groupings

The present study treated each variable separately, and examined one band structure within each variable. Future studies should examine variable groupings, analogous to dimensionless quantities in engineering, to ascertain whether additional insight is gained from such analyses. Different band structures and definitions could also be examined for additional insights.

VII-J. Research Impact Quantification

The present paper demonstrated the identification of intra- and extra-discipline applications through Citation Mining. It did not address the quantification of research impact. However, an integrated Citation Mining study across many cited paper disciplines would provide a data base of discipline-discipline or discipline-applications relations. These relations could be transformed into a research impact network (where the nodes are research/ technology/ systems and the links are their relations), showing the impacts of a research discipline on other research, technology, and systems. The network links could be quantified from the citation

data, and the impacts of a research discipline or sub-discipline on other research, technology, or systems areas could be estimated from appropriate summations of link values along the desired network paths [Kostoff, 1994].

VI. REFERENCES

Davidse, R. J., and VanRaan, A. F. J., (1997). "Out of Particles: Impact of CERN, DESY, and SLAC Research to Fields other than Physics", Scientometrics, 40:2.

GPRA (1993). Government Performance and Results Act of 1993: Public Law (103-62).

Herring, S. D., (1999). "The Value of Inter-disciplinarity: A Study Based on the Design of Internet Search Engines", JASIS, 1 April.

Kostoff, R. N., (1994). "Research Impact Quantification". R&D Management. 24:3, July.

Kostoff, R. N., (1997). "Peer Review: The Appropriate GPRA Metric for Research", Science, Volume 277, 1 August.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1999), "Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography". Journal of the American Society for Information Science. 15 April.

Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. (2000). "Fullerene Roadmaps Using Bibliometrics and Database Tomography". Journal of Chemical Information and Computer Science. 40:1. Jan-Feb.

NSF (1996). Science and Engineering Indicators 1996, National Science Board, Government Printing Office, Washington, DC.

Steele, T. T., (2000). "The Impact of Interdisciplinary Research in the Environmental Science: A Forestry Case Study", JASIS, 15 March