DISCERN: ONLINE DATA APPENDIX / A. ARORA, S. BELENZON & L. SHEER[1]

This Appendix describes the methodology used to construct our database of publicly listed U.S. headquartered firms matched to assignees of patents from the United States Patent and Trademark Office (USPTO) and scientific publications from the Web of Science for the period 1980-2015. Data users should cite the related forthcoming AER paper: *"Knowledge spillovers and corporate investment in scientific research"* (Arora, Belenzon, Sheer).

We introduce a major data extension and improvement to the historical NBER patent dataset (Hall, Jaffe, and Trajtenberg and others, 2001; Bessen, 2006), which should be valuable for all researchers working with patent and publication data. In updating the data to match between Compustat and patents to 2015, we address two major challenges: name changes and ownership changes. These challenges are central to how patents are assigned to firms over time. To be consistent over the sample period, we reconstruct the complete historical data covered in the NBER data files. About 30% of the Compustat firms in our sample change their name at least once. Accounting for name changes improves the accuracy and scope of matches to patents (and other assets), ownership structure, and dynamic reassignments of GVKEY codes to companies. Dynamic reassignment means that, for instance, if a sample firm merges with another firm, the patents of the merged firm are included in the stock of patents linked to the Compustat record from that point onward, but not before. For ownership and subsidiary data, we rely on a wide range of M&A data, including SDC, historical snapshots of ORBIS files for 2002-2015, 10-K SEC filings, and NBER2006 as well as perform extensive manual checks that help us uncover firms' structure and ownership changes before proceeding to the patent match. Thus, we have extended and improved the NBER patent data. In this Appendix, we document our data construction work, present several examples ("case studies"), and outline the improvements we made to existing NBER historical patent data.

We combine data from six main sources: (i) company and accounting information from U.S. Compustat 2018 (Standard & Poor's (S&P), 2018b), (ii) scientific publications from Web of Science (Clarivate Analytics, 2016), (iii) USPTO patents and their non-patent literature (NPL) citations from PatStat (European Patent Office, 2016); (iv) subsidiary data from historical snapshots of ORBIS files for 2002-2015 (Bureau van Dijk, 2018); (v) mergers and acquisition data from SDC Platinum (Securities Data Company (SDC) Platinum, 2018), and (vi) company name changes from WRDS's "CRSP Monthly Stock" (Center for Research in Security Prices (CRSP), 2018ab).

We match (i) corporate subsidiaries to Compustat ultimate owner (UO) firms; (ii) acquisition data to Compustat companies and their related subsidiaries; (iii) patent data to Compustat companies and their related subsidiaries; (iv) scientific publications to Compustat companies and their related subsidiaries; and (v) patent citations to scientific articles. We discuss the details of our methodology below.

A. ACCOUNTING DATA PANEL

Our methodology builds and improves on the NBER patent database (Hall et al., 2001; Bessen, 2006), by extending the time period by a decade (now from 1980 to 2015) and implementing several methodological improvements for the complete sample period.

We start with all North American Compustat records obtained through WRDS in August 2018 and select companies with active records and positive R&D expenses for at least one year during our sample period, 1980-2015[2]. We exclude firms that are not headquartered in the United States based on their current headquarter location. After matching the remaining firms to patent assignees from the USPTO, we further restrict our sample to ultimate-owner[3] (UO) Compustat firms with at

[1] Arora: Duke University, Fuqua School of Business and NBER, ashish.arora@duke.edu; Belenzon: Duke University, Fuqua School of Business and NBER, sharon.belenzon@duke.edu; Sheer: Duke University, Fuqua School of Business, lia.sheer@duke.edu;
We thank Jim Bessen, Nick Bloom, Wesley Cohen, Alfonso Gambardella, Bronwyn Hall, David Hounshell, Adam Jaffe, Brian Lucking, David Mowery, Mark Schankerman, Scott Stern, Manuel Trajtenberg, John Van Reenen, and seminar participants at NBER summer institute and NBER Innovation Information Initiative for helpful comments. We thank Bernardo Dionisi, Honggi Lee, Dror Shvadron, and JK Suh for excellent research assistance. All remaining errors are ours.

[2] We define an active record as a year with positive common shares traded (CSHTR_F). We do this to avoid including years with data based on prospectus submitted by the focal company as part of the filing process before the firm became publicly traded.

[3] Compustat database does not link parent companies to subsidiaries, however we supplement the data with subsidiary level data. Following NBER 2006, we aggregate the data to the parent company level which we call ultimate owner (UO).

least one patent during our sample period. A UO firm enters the sample once it is publicly traded and has at least one patent in stock and remains in our data until the end of the sample period unless it is acquired, dissolved, or taken private. All UO firms in our final sample have at least 3 consecutive years of active records in Compustat. Our final estimation sample consists of an unbalanced panel of 4,520 UO firms and 60,885 firm-year observations.[4] The process of defining a UO firm and its related subsidiaries is explained below.

We face several challenges when working with Compustat data, as following.

1) **Unique company identifier over time.** Compustat uses GVKEY to track companies over time[5]. However, a single company may correspond to multiple GVKEYs within the Compustat database due to changes in ownership and other accounting changes over the sample period (e.g., the pet food company Ralston Purina is listed under two different GVKEYs: (i) 1980-1993 under "RALSTON PURINA-CONSOLIDATED" (GVKEY 008935) and (ii) 1993-2000 under "RALSTON PURINA CO" (GVKEY 028701)). The Compustat database does not link related company identifiers, making it difficult to track companies over time only based on GVKEY.

2) **Name changes.** While scientific publications and patent records contain the owner's name at the time of their publication, companies appear in the Compustat file under their most current name with no records of previous names. Company names may change over the course of our sample period due to general name changes[6] and M&As[7], including reverse takeovers[8]. About 30% of the Compustat firms in our sample change their name at least once. A company with a name change (which might have been accompanied by an ownership change) without a corresponding change in its GVKEY in Compustat may lead us to assign the record incorrectly to its most recent owner for the complete sample period. Without historical information on the record's ownership, we cannot correctly link patents and scientific publications to their relevant financial records.

3) **Ownership structure.** A parent company and a majority-owned subsidiary may have different identification numbers and records within Compustat. While innovative activities typically take place inside numerous subsidiaries, we aggregate the data to the UO level. Since the Compustat database does not link parent companies and majority-owned publicly traded subsidiaries, comprehensive manual checks and investigations are required.[9] We further link non-publicly traded subsidiaries to their UO firm based on historical snapshots of ORBIS files.

4) **Changes in ownership.** Ownership of a firm can change throughout the sample period due to mergers, acquisitions, and spinoffs[10]. While firms typically stop being traded independently after an M&A, their existing stock of publications and patents must be reassigned to the new owner. Moreover, in many cases, the acquiring entities continue to file patents and produce scientific publications post-acquisition. Compustat data do not provide information on ownership changes. Thus, we rely on SDC Platinum's M&A data and ORBIS to track ownership changes at the UO level as well as at the subsidiary level. Using historical snapshots of ORBIS files for 2002-2015, we are able not only to identify ownership changes at the subsidiary level but also new subsidiaries and changes in subsidiary names.

---

[4] See "panel_do.do" file for exact details on the construction of the final panel file.
[5] GVKEY code remains the same, regardless of changes in TICKER, CUSIP, and firm names and thus is preferred on the later as a firm identifier for Compustat records. Compustat database only provides the most recent TICKER, CUSIP and name for each security with no historical info available.
[6] e.g., name abbreviations (for example, "MINNESOTA MINING AND MANUFACTURING" changed its name in 2002 to "3M"),
[7] e.g., "WESTINGHOUSE ELECTRIC CORP" (GVKEY 011436) purchased "CBS INC" in 1995 and changed its own name to "CBS CORPORATION" in 1997 keeping the same GVKEY Compustat firm identifier.
[8] e.g., in 1993 the private company Dentsply International Inc acquired the public company GENDEX CORPORATION (GVKEY 013700) in a reverse takeover and became publicly traded under the "DENTSPLY INTERNATIONAL INC" name and the original GVKEY.
[9] e.g., Thermo Electron's publicly traded majority-owned spun-out subsidiaries (all of which returned to be privately owned after 1999) need to be accounted under the parent company THERMO ELECTRON CORP (GVKEY 010530) for the complete period.
[10] e.g., "AT&T CORP" (GVKEY 001581) stopped being traded independently in 2005 after it was acquired by "SBC COMMUNICATIONS INC" (GVKEY 009899) which in turn changed its own name to "AT&T INC".

We implement the following procedures to manage these challenges.

I. NAME CHANGES

One of our key contributions is identifying name changes of Compustat firms over the sample years 1980-2015. To the best of our knowledge, this has not been done consistently for a broad range of companies across many industries over a third of a century. Past research mainly considers the name that appears for each record in the most recent Compustat file (CONM variable) as the relevant name for the complete period the security was traded. The variable CONM, however, is the current name of the Compustat record as of the date the file was downloaded with no historical name information provided by Compustat. As shown above, company name changes may not be accompanied by changes in the original GVKEY firm identifier on Compustat, leading to assigning a record to its most recent holder for the complete sample period. Matching the original assignee name to a current Compustat file can result in misallocation of patents and publications. As companies change names, we wish to carry forward past patents and publications assigned to the original name as well as make sure that new patents and publications are assigned to the correct UO firm. Instead of building on the most recent Compustat name, we link our Compustat records to WRDS's "CRSP Monthly Stock" file, which records historical names for each month the security was traded and perform extensive manual checks using SEC filings to validate all related names for our sample period. We find that in our sample, 30 percent of Compustat records have more than one related name[11]. Accounting for all historical names significantly improves the accuracy and scope of the matches we perform across various databases as well as the linkage to relevant financial data. We elaborate on our name change methodology below, using several examples.

Example 1: SEALED POWER and GENERAL SIGNAL

The following example underscores the mismatching consequences of not accounting properly for name and ownership changes and how it affects the existing NBER patent data.

Up to the year 1998, SEALED POWER and GENERAL SIGNAL are two distinct entities. Historical Compustat records include the following records for these companies up to 1998:

1) GVKEY 9556, related names:

   i. SEALED POWER CORP (1962-1988) – original name
   ii. SPX CORP (1988-1997) -name changes retroactively in Compustat

2) GVKEY 5087, related name: GENERAL SIGNAL CORP (1950-1997)

In 1998, SPX Corp acquired General Signal Corp in a reverse merger transaction, and General's GVKEY (5087) became the new security of SPX traded retroactively under the new name "SPX CORP". At the same time, the original SPX records are renamed retroactively in Compustat as "SPX CORP-OLD" and stopped being traded. Current Compustat records include the following records for these companies for the complete period they are traded:

1) GVKEY 9556, related name: SPX CORP-OLD

2) GVKEY 5087, related name: SPX CORP

Our approach is to treat these GVKEYs as two separate companies up to 1997 accounting for all relevant names (SEALED POWER CORP, SPX CORP for GVKEY 9556 and GENERAL SIGNAL CORP for GVKEY 5087) in our matches and to connect the SPX CORP name to General's original GVKEY (5087) only from 1998.

When we examine the NBER 2006 patent dataset, we find that the two companies are collapsed under the same company (same PDPCO id) and that for the purpose of Compustat accounting information General's original GVKEY (5087) is used for the complete period while the original SPX GVKEY (9556) is disregarded:

[11] This is comparable to the findings of Wu (2010), who finds that during 1925-2000 over 30% of CRSP-listed firms changed their names at some point after going public. For name changes occurring between 1980-2000 the paper finds that the top 3 reason for name changes are: (i) M&As & restructure activity (36%); (ii) change in focus of operation (17%); (iii) brand or subsidiary name adoption (12%)

Table 1. Data for SPX Corp in NBER 2006

current name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
SPX CORP	5087	1950	2006	5087	1	1950	2006
SPX CORP-OLD	9556	1962	1997	5087	-1		

Note: PDPCO is NBER's Patent Data Project (PDP) unique company id. FIRSTYR is the first year GVKEY company has data. LASTYR is the last year a GVKEY company has data. PDPSEQ is the GVKEY sequence within PDPCO. If PDPSEQ=-1, the related GVKEY is disregarded. BEGYR is the beginning year for GVKEY within PDPCO. ENDYR is the last year for GVKEY within PDPCO. All patents related to SPX CORP will be accounted under GVKEY 5087 from 1950 to 2006, while the original SPX GVKEY (9556) is disregarded.

Practically, this means that all the patents of SPX CORP are matched to General's financial data up to 1998. To verify, we tracked the NBER files and confirmed that indeed SPX patents pre-1998 are matched to General's GVKEY. Moreover, patents related to "GENERAL SIGNAL CORP" (757 patents without considering related subsidiaries) as well as "SEALED POWER CORP" (36 patents without considering related subsidiaries) are located in the 2006 NBER raw patent match but are not assigned to any Compustat record.

The NBER patent data file does not track ownership and name changes of GVKEYs over time. However, as shown in this example, using the current Compustat name can be misleading. The availability of data on historical name changes enables us to have a better understanding of the firms included in our sample and their origin. We are able to improve the accuracy of their match to the different databases (by using the complete history of firm names) and their linkage to relevant financial data. To be consistent over the sample period, we reconstruct the complete historical data covered in the NBER data files.

Compiling historical names

To locate historical names, we use the WRDS's "CRSP Monthly Stock" file, which includes historical monthly information on names for each security alongside its historical CUSIP code and a unique permanent security identification number assigned by CRSP, the PERMNO code, which is kept constant throughout the trading period regardless of changes in name or capital structure.[12] We compute for each name the starting and end years based on their trading dates in the "CRSP Monthly Stock" file.

Using WRDS "CRSP/Compustat Merged Database - Linking Table", we link each PERMNO to Compustat GVKEY code. The crosswalk between CRSP and Compustat is not obvious as it first seems. As shown above, a PERMNO can have multiple GVKEYs related to it- in such case, we apply a dynamic match between a PERMNO and Compustat accounting data. However, CRSP also includes cases where under the same GVKEY there are several PERMNO codes. This is mainly due to significant M&As, including reverse acquisition, that occurred during the years when the firm was not listed. For example, in some cases, the merge between CRSP to Compustat results in a firm name related to more than one GVKEY identifier. For those cases, we manually checked using 10K-SEC fillings the years that the name was relevant for each GVKEY. Also, there is a difference in coverage between CRSP and Compustat for the early sample years[13] – we added missing information from Compustat and manually checked for historical names wherever possible.

[12] For example, while SPHERIX INC is related to 2 different GVKEYs (002237 for 1980-2013 and 018738 for 2013-current) it has a unique PERMNO code for the entire period (18148). Similarly, Google Inc PERMNO code is 90319 and it remains the same after the company reorganized as ALPHABET INC in 2015.

[13] There are differences between CRSP and Compustat coverage- for example, CRSP only includes firms listed in USA major exchanges and specifically excludes regional exchanges, while Compustat includes all 10-K filer firms in North America. Moreover, CRSP coverage for major exchanges has expanded gradually over the years (e.g., ARCA was only added from 2006).

Our main firm identifier PERMNO_ADJ builds on the original CRSP PERMNO id with several adjustments[14]. (i) In cases where under the same GVKEY, we find several PERMNO codes we replace it with one main PERMNO code[15] – for example, OWENS Corning GVKEY (008214) was split to two PERMNO codes 24811 and 91531 due to it being unlisted between 2003-2005. However, we keep PERMNO_ADJ the same for the complete period (24811). (ii) We manually add a PERMNO_ADJ code for firms in our Compustat sample that did not appear in the "CRSP Monthly Stock" file due to coverage differences.
We further perform extensive manual checks on the name list, including identifying and distinguishing companies with similar names[16]. Finally, we cleaned and standardized firm names as CRSP tends to abbreviate long words in the company name that it provides. We located those cases and manually corrected them to avoid mismatches.[17]

Standardizing firm names

Prior to matching, we standardize firm names to reconcile company names that may be spelled differently across databases. We compose a standardization code used on both the source and the target names to increase the number of exact matches.

Each company name was first standardized by converting all strings to uppercase characters and cleaning all non-alphabetic characters as well as Compustat related indicators (e.g., -OLD, -NEW, -CL A) and other common words (e.g., THE).

Additionally, an important step in standardizing the company names is standardizing abbreviations. We formed a list that includes over 80 abbreviated words matched to their various original words. For example, LABORATORIES, LABORATORY, LABS, LABO, LABORATORIE, LABORATARI, LABORATARIO, LABORATARIA, LABORATORIET, LABORATORYS, and LABORATORIUM were all abbreviated to "LAB". The list was compiled from the most frequently abbreviated words in WOS affiliation field (accordingly, the list is targeted to our sample). This list is presented in Table 2.

Table 2. Most frequent abbreviated words

ADV	AEROSP	AGR	AMER	ANAL	ANALYT	ANIM	APPL	APPLICAT
ASSOC	AUTOMAT	BIOL	BIOMED	BIOPHARM	BIOSCI	BIOSURG	BIOSYS	BIOTECH
BIOTHERAPEUT	CHEM	CLIN	COMMUN	COMP	CORP	CTR	DEV	DIAGNOST
DYNAM	EDUC	ELECTR	ENGN	ENVIRONM	FAVORS	GEN	GENET	GRAPH
GRP	HLDG	HLTHCR	HOSP	INC	IND	INFO	INNOVAT	INST
INSTR	INTERACT	INTL	INVEST	LAB	LTD	MAT	MED	MFG
MICROELECTR	MICROSYS	MOLEC	NATL	NAVIGAT	NEUROSCI	NUTR	ONCOL	ORTHOPAED
PHARM	PHOTON	PHYS	PROD	RES	SCI	SECUR	SEMICOND	SERV
SFTWR	SOLUT	SURG	SYS	TECH	TEL	TELECOM	THERAPEUT	TRANSPORTAT

For each standardized name, we create a cleaner, fully-standardized name by omitting the legal entity endings and other general words (e.g., INC, CORP, LTD, PLC, LAB, PHARMACEUTICAL), where possible, to maximize match rates (e.g., "XEROX CORP" was standardized to "XEROX", "ABBOTT LABORATORIES" to "ABBOTT"). However, in cases where the company name is too short, generic or can match to other strings within the affiliation field, we

[14] It is consistent with NBER2006's PDPCO firm id.
[15] In the final accounting data panel, we split firms based on big jumps in sales, patents or publications. For example, we split PERMNO_ADJ 66093 to the period before and after SBC Communications Inc acquired AT&T Corp and became AT&T Inc. PERMNO_ADJ_LONG is the final UO identifier in the accounting data panel after the split.
[16] For instance, RACKABLE SYSTEMS INC (GVKEY 162907) changed its name to SILICON GRAPHICS INTL CORP after it acquired the public company SILICON GRAPHICS INC (GVKEY 012679) in 2009 – we need to make sure that we count SILICON GRAPHICS related publications and patents under RACKABLE's GVKEY only from 2009. Similarly, we need to distinguish between the original BIOGEN INC (GVKEY 002226) and the new BIOGEN INC (GVKEY 024468) that was formed only after the merger with IDEC PHARMACEUTICALS CORP in 2003.
[17] It is also worth mentioning that the "CRSP Monthly Stock" file reports acronym firm names with extra space between the initial letters (e.g., E G & G INC and not EG&G INC). This has to be taken into consideration when performing matches to other databases that do not use this format.

preserved the original standardized name to avoid mismatches and extensive manual checks on the match results. For example, omitting the legal entity from "QUANTUM CORP" would result in a potential mismatch between "QUANTUM" and "TEXAS STATE UNIV CTR APPL QUANTUM ELECTR DEPT".

The last step in name standardization is to locate abbreviations that are commonly used by companies instead of their official names. For example, "INTERNATIONAL BUSINESS MACHINES CORP", will also appear under its common abbreviation "IBM" and ''GENERAL ELECTRIC CO" under "GE". We also add the names of prominent R&D laboratories affiliated with companies, such as the T.J. Watson Research Center (IBM) and Bell Labs (initially AT&T and later under Lucent technologies), as authors often omit the name of the company when the address of the laboratory is stated as the publication address.

Constructing the name list

All our matching is done at the firm name level. We assign each firm name a unique identifier ID_NAME and indicate the first, and last year the name is relevant for a PERMNO_ADJ. We then perform dynamic matching of names to PERMNO_ADJ based on SDC's M&A data. M&A reassignment includes up to five reassignments per name over the sample period (explained in further details below). PERMNO_ADJs are then dynamically linked to GVKEYs[18]. We further link non-publicly traded subsidiaries to their UO firm. Related subsidiary names are reassigned accordingly up to five times to UO firms. For further details on the ownership methodology, see Section B below.

Our UO and subsidiary historical standardized name lists ("DISCERN_UO_name_list.dta" and "DISCERN_SUB_name_list.dta", respectively), including the dynamic reassignment, are publicly available for researches to match to their database of interest. Main variables of the name list file are described below:

Variable name	Description
NAME_STD	Historical standardized UO firm names (1980-2015) for firms that are included in our initial Compustat sample[19] and their related subsidiaries.
ID_NAME	Name ID unique at name_std-permno_adj1
PERMNO_ADJ$_{0\text{-}5}$	UO firm id: up to 5 owners + "0" is usually the pre-IPO owner if applicable.
NAME_ACQ$_{0\text{-}5}$	Related UO name
FYEAR$_{0\text{-}5}$	First-year for ID_NAME within PERMNO_ADJ.
NYEAR$_{0\text{-}5}$	Last-year for ID_NAME within PERMNO_ADJ

I. DYNAMIC REASSIGNMENT

We build on the strategy used by NBER patent match (2006) to perform a dynamic reassignment for our subset of UO Compustat firms (see Figure 1). The dynamic reassignment accounts for: (i) changes in Compustat identification numbers (challenge 1 above) - dynamically matching Compustat accounting information for firms that are related to more than one GVKEY record, and (ii) M&A reassignment based on SDC data and construction of a complete name

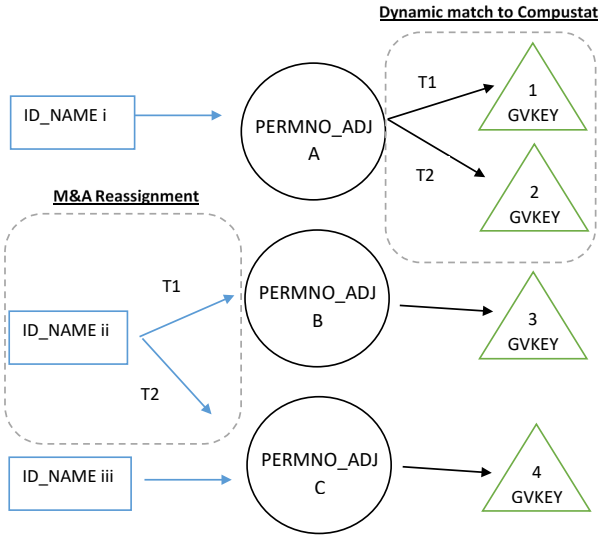
[18] For the link between PERMNO_ADJ and GVKEYs see "permno_gvkey.dta" file. In the final panel file, we further split UO firms based on big jumps in sales, patents, or publications and our unique UO firm identifier in the accounting data panel is labeled as PERMNO_ADJ_LONG.

[19] The UO list, "DISCERN_UO_name_list.dta", includes only names of UO parent firms included in our initial Compustat sample. Exceptional are names of top laboratories and names of majority-owned publicly traded subsidiaries that appeared in our initial Compustat sample and were collapsed under the UO parent firm. The subsidiary name list, "DISCERN_SUB_name_list.dta", includes all related subsidiaries as explained in Section B below. The standardization code that was used to standardize the names is available under NAME_STD.do file. Standardized names include legal entity and other common words - in cases where users want to match to a cleaner version of the name, they should apply their own script to clean the names further. When matching the name list to other databases, users should include extensive manual inspection to matched results. Special care should be given to companies with similar names and to generic company names. While the name lists include all names related to the initial sample, the panel file includes only the final estimation sample firms.

history for the period 1980-2015 (Challenge 4 above). For M&A reassignment, we include up to five ownership reassignments for each firm name that appears in our initial Compustat subsample and acquired by another firm in our sample. Unless a name is reassigned to another PERMNO_ADJ, it stays with the focal firm until the end of the sample (or the firm's trading period). We dynamically reassign related patents and scientific publications of the acquired UO firm and its related subsidiaries to acquirer firms accordingly (will be discussed in more detail below).

Each PERMNO_ADJ is then linked to Compustat GVKEYs. For cases where there are changes in Compustat identification numbers over the sample period, we dynamically match PERMNO_ADJ to GVKEYs. In the final accounting data panel, we further split firms based on big jumps in sales, patents, or publications. PERMNO_ADJ_LONG is the final UO identifier in the accounting data panel after the split.

Figure 1. Description of dynamic changes



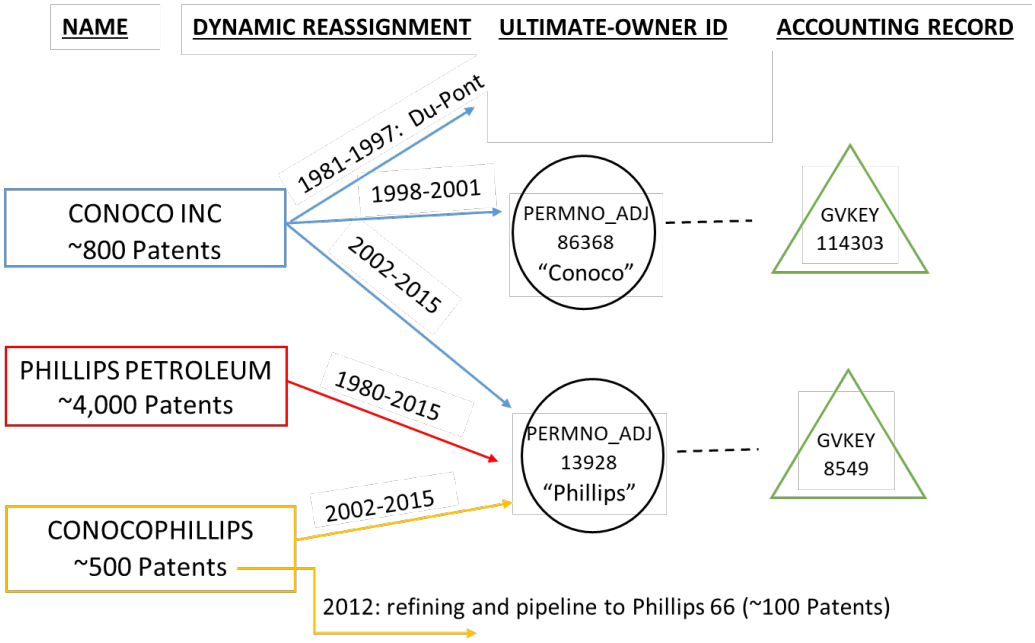
Note: This figure illustrates the dynamic structure of the data. The dynamic reassignment accounts for: (i) changes in Compustat identification numbers (GVKEY), and (ii) M&A reassignment. Throughout the sample periods (T1 and T2), each name (ID_NAME) can be assigned to more than one firm (PERMNO_ADJ) and each firm can be linked to more than one Compustat record (GVKEY). For example, for the GENERAL SIGNAL – SPX CORP case study: SPX CORP as two related GVKEYs: 9556 (1980-1997) and 5087 (1998-2015). GENERAL SIGNAL has two related PERMNO_ADJ: 12095 (1980-1997) and 55212 (1998-

Example 2: CONOCO and PHILLIPS PETROLEUM

This example illustrates the importance of the dynamic structure of our data.

In 1981, Conoco was acquired by Dupont, which has later spun it off as a publicly traded company, which was eventually acquired by the publicly traded company, Phillips Petroleum, in 2002. The merged entity was renamed ConocoPhillips. When we examine current Compustat records, we would only locate the name ConocoPhillips with no record of Philips Petroleum. Compustat does not provide any info on the owner of the record prior to the merger. We use the CRSP monthly stock file to locate all historical names of related securities.

Figure 2. Conoco-Phillips historical names and related patents, 1980-2015

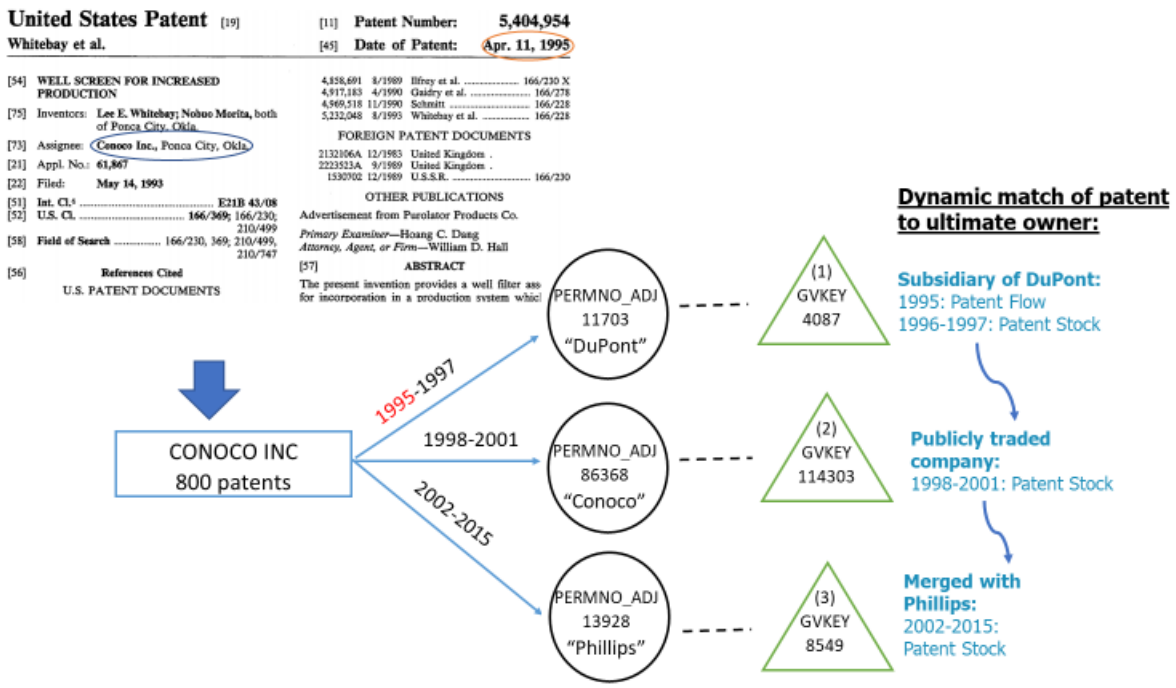


Notes: The figure illustrates the historical names and the dynamic structure of the data related to Conoco-Philips. Historical names allow us to account for patents assigned to firms in our sample that historically operated under a different name. Here, we locate the majority of granted patents- 4,000 patents - under Phillips Petroleum, the original owner pre-merger with Conono. Furthermore, each name can be assigned throughout the sample period to more than one firm (PERMO_ADJ). For ownership changes, we match the merged firm's patents only after the M&A and not before. In this example, Conoco Inc is matched to Phillips only after the merger in 2002.

Historical names are important for matching patents (and other assets) for the following reasons. (i) They allow us to account for patents assigned to firms in our sample that earlier data missed because the focal firm operated under a different name: under Phillips's name, we locate the majority of granted patents. Four thousand patents that were issued to Phillips Petroleum that were not matched previously without the historical name info. (ii) For ownership changes- we match the merged firm's patents only after the M&A and not before. In this case, Conoco is matched to ConocoPhillips only after the merger in 2002. (iii) Historical names also help match subsidiary data as UO names appear in ORBIS files as of the year the file was recorded (e.g., Chevron-Phillips JV formed in 2000 that we match at the subsidiary level).

In addition to locating historical firm names, we do extensive work on ownership, which enables us to match firm names dynamically to more than one UO-Firm.

Figure 3. Conoco-Phillips dynamic match



Note: patents can be assigned throughout the sample period to more than one UO firm (PERMO_ADJ). The figure illustrates the dynamic match of patent num. 5404954 to its related UO firm in each period. The patent would be included in Dupont's patent flow for 1995. It will also be counted under Dupont's patent-stock for 1996-1997. However, from 1998 this patent would be transferred dynamically from Dupont to Conoco's patent stock. Similarly, in 2002 the patent would transfer to ConocoPhilips patent stock.

Figure 3 illustrates the process of dynamic matching. Patent "5404954" was granted to Conoco Inc in 1995. At that time, Conoco was a subsidiary of Dupont. In our data, this patent would be included in Dupont's patent flow for 1995. It will also be counted under Dupont's patent-stock for 1996-1997. However, from 1998- when Conoco is spun-off as an independent publicly traded company, this patent would be transferred dynamically from Dupont to Conoco's patent stock. Similarly, in 2002 the patent would move on to ConocoPhilips patent stock.

A different patent, which is issued to Phillips Petroleum in 1999, for instance, would be part of the patent flow assigned to Phillips in 1999 and be counted under the patent stock for Phillips Petroleum till 2002, and then would move on to become part of ConocoPhilips patent stock. The dynamic reassignments are based on our dynamic name list, as shown in Figure 4. We put much effort into tracking these ownership changes. We will elaborate on our ownership methodology below.

Figure 4. Example of dynamic name list for Conoco-Phillips:

ID Name	Name std	fyear 0	nyear 0	Permno Adj_0	Name ACQ_0	Fyear 1	Nyear 1	Permno Adj_1	Name ACQ_1	fyear 2	nyear 2	Permno Adj_2	Name ACQ_2
2384	CONOCO INC	1981	1997	11703	DU PONT E I DE NEMOURS & CO	1998	2001	86368	CONOCO INC	2002	2015	13928	PHILLIPS PETR CO
7325	PHILLIPS PETR CO					1980	2002	13928	PHILLIPS PETR CO	2003	2015	13928	CONOCO PHILLIPS
2385	CONOCO PHILLIPS					2002	2015	13928	CONOCO PHILLIPS				
7324	PHILLIPS 66	1980	2011	13928	CONOCO PHILLIPS	2012	2015	13356	PHILLIPS 66				

Note: This table presents the dynamic reassignment name list related to Conoco-Philips. It illustrates the historical names and the dynamic structure of the data. ID_NAME is the unique standardized name id. NAME_STD is the standardized firm name. PERMNO_ADJ(0-5) is the UO firm id. A name can be matched dynamically up to 5 times (1-5) and to an additional pre-IPO owner if applicable (0). NAME_ACQ(0-5) is the related UO name. FYEAR(0-5) is the first-year for ID_NAME within PERMNO_ADJ. NYEAR(0-5) is the last-year for ID_NAME within PERMNO_ADJ.

Example 3: TIME-WARNER and AMERICAN ONLINE

This example illustrates how properly accounting for name and ownership changes improve the accuracy of patent flow as well as the dynamic reassignment of patents.

Warner Communication and its subsidiaries were independent and publicly traded companies until their merger with Time Inc in 1989 when Time-Warner Inc was formed. In the second half of 2000, Time-Warner was merged with American Online to form AOL Time Warner. In 2003 the company dropped the "AOL" from its name and was renamed Time-Warner Inc. AOL remained a subsidiary until it was spun-out in 2009.

The NBER 2006 patent match reveals:

1) Warner Communication and its related subsidiary patents are correctly matched to WARNER COMMUNICATIONS INC (GVKEY 11284) up to the merger with Time Inc. However, they are not dynamically assigned after 1988 to Time Warner or any other company, implying that the patent stock and patent flow of Time-Warner (and later AOL Time-Warner) from patents related to Warner communication and its subsidiaries (e.g., Warner Bros, WEA Manufacturing (before it was acquired) – above 60 patents up to 2006) are below the true value after the acquisition in1989.

2) TIME-WARNER related patents from 1991 to 2000 (before the merger with American-Online Inc in late 2000) are matched incorrectly to GVKEY 25056, which during those years was solely AMERICAN-ONLINE INC original Compustat financial records. The current name of GVKEY 25056, TIME WARNER INC, which is likely to have misled NBER to link the Time Warner patents to it, was only adopted retroactively in 2003 when the "AOL" was dropped from the official name. Moreover, AMERICAN ONLINE INC and AOL related patents (152 patents up to 2006 based on NBER raw patent match) are not linked to any Compustat record. AOL-TIME WARNER related patents, on the other hand, are matched to a "Pro-Form" Compustat record that is active for only two years 1999-2000: AOL TIME WARNER INC-PRO FORM (GVKEY 142022). All of this implies that AOL Time Warner's flow of patents is below the true level throughout the period.

Having a complete history of names enables us to correctly identify each Compustat record and its origin and dynamically match each firm name in our sample to the correct financial records accordingly: (i) AMER ONLINE INC (and later AOL) is matched from 1980 until its spinout in 2009 to GVKEY 25056 and after to AOL INC (GVKEY 183920). (ii) Warner Communication is matched up to the merger with Time Inc to WARNER COMMUNICATIONS INC (GVKEY 11284) and later dynamically transferred ending up in AOL -Time Warner GVKEY (25056) starting 2001. (iii) AOL -Time Warner is matched to AOL -TIME WARNER (GVKEY 25056) starting 2001 after the merger was approved. (iv) As a side note- Time Inc is not included as an UO in our sample as it did not have R&D expenses, but it is included as a subsidiary name under the Time-Warner UO company.

Example 4: PHARMACIA & UPJOHN and MONSANTO

This example demonstrates that having a complete history of names enables us to correctly identify each Compustat record's historical ownership and dynamically match each firm name in our sample to its relevant financial records in each period. For instance, linking each patent to its correct financial record can be a concern for papers that link patents to market value, specifically those distinguishing different types (e.g., high vs. low cited patents), which rely on the specific patent that was matched and not only the quantity.[20]

In 1995 original Pharmacia merged with Upjohn to form Pharmacia & Upjohn. In 2000, original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). Between 2000-2002 the new Pharmacia gradually spun off its agricultural operations to a newly created subsidiary, Monsanto Company (New Monsanto). In 2003 the new Pharmacia was acquired by Pfizer and is now a wholly-owned subsidiary of Pfizer. Table 3 illustrates how our methodology allows us to compute patent stock and flow for each GVKEY record correctly.

[20] The following are additional examples: (I) Patents of Honeywell before the merger with Allied Signal (3,112 patents) are incorrectly linked to Allied Signal's GVKEY (001300) up to 1999, while the financial records of the original Honeywell Inc are disregarded (GVKEY 5693). (II) Patents of TELEDYNE INC (GVKEY 10405) pre-merger with the publicly traded ALLEGHENY LUDLUM CORP in 1996 (to form ALLEGHENY TELEDYNE INC, which in 1999 was renamed ALLEGHENY TECHNOLOGIES INC after TELEDYNE was spun-off as free-standing public company) are not linked GVKEY 10405 (634 patents up to 1999, of which 597 patents are pre-1996 merger). In addition, ALLEGHENY LUDLUM CORP's (GVKEY 13708) patents (254 patents, of which 240 patents pre-1996 merger) were not dynamically moved to TELEDYNE INC post-merger. This means that in 1996 (post-merger) the patent stock of GVKEY 10405 is missing at least 789 patents (not including related subsidiary patents). (III) For the new Biogen Inc (GVKEY 24468) NBER does not include patents of IDEC pharmaceuticals, who was the owner of the security before Biogen and IDEC merged in 2003 (40 patents).

Table 3. PHARMACIA & UPJOHN and MONSANTO dynamic match

Period	related GVKEY	Relevant Compustat name for period	Most recent Compustat name	Comments	Patent flow per period per our strategy (based on NBER raw patent match, w/o subsidiaries)	Original NBER match
1950-1994	11040	UPJOHN CO	PHARMACIA & UPJOHN INC	Original Upjohn before merger with Pharmacia	**2,091** Upjohn related patents	N/A
1995-1999	11040	PHARMACIA & UPJOHN INC	PHARMACIA & UPJOHN INC	**1995**: Upjohn merged with original Pharmacia to form Pharmacia & Upjohn	**479** Pharmacia &/ Upjohn related patents	N/A
1950-1999	7536	MONSANTO CO	PHARMACIA CORP	Original Monsanto before merger with Pharmacia & Upjohn	**3,228** Monsanto related patents	**2,733** Pharmacia &/ Upjohn related patents (including patents of Pharmacia before it merged with Upjohn). **While Monsanto's 3,228 patents are not linked.**
2000-2002	7536	PHARMACIA CORP ("new Pharmacia")	PHARMACIA CORP	**2000:** original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). All of PHARMACIA, UPJOHN and PHARMACIA & UPJOHN patents are transferred here from 2000. Monsanto's patents are redirected to the new Monsanto spin-off company.	**304** Pharmacia &/ Upjohn related patents	**304** Pharmacia &/ Upjohn related patents
2000-2015	140760	MONSANTO CO ("new Monsanto")	MONSANTO CO	**2000-2002:** Pharmacia Corporation (New Pharmacia)gradually spun-off its agricultural operations to a new publicly traded company, Monsanto Co (New Monsanto). All Monsanto related patents are transferred here from 2000.	**553** Monsanto related patents (2000-2006)	**553** Monsanto related patents (2000-2006). *NBER links Monsanto's patents to GVKEY 140760 from 1997 - while records for 1997-1999 are available on Compustat, they are based on prospective filings when Monsanto was still traded under GVKEY 140760.
2003-2015	8530	PFIZER INC	PFIZER INC	**2003:** Pharmacia Corporation (New Pharmacia) was acquired by Pfizer and is now a wholly owned subsidiary of Pfizer. All of PHARMACIA, UPJOHN and PHARMACIA & UPJOHN patents are transferred here from 2003.	**472** Pharmacia &/ Upjohn related patents(up to 2006)	**472** Pharmacia &/ Upjohn related patents(up to 2006)

Note: his table presents the comparison between NBER 2006 and our data for dynamic patent reassignment for Phamacia-Monsanto related patents at the GVKEY-Period level. Most recent Compustat name is based on Compustat 2018 file. Relevant Compustat name for the period is the historical firm name based on CRSP Monthly Stock file. Patent flow per our strategy is based on NBER raw patent match data for the relevant Compustat name excluding subsidiaries. Patent flow per original NBER match is based on NBER 2006 data.

II. AGGREGATING DATA TO THE UO FIRM LEVEL

To merge parent Compustat companies and their independent majority-owned publicly traded Compustat subsidiaries (Challenge 3 above), we locate related firms in our initial Compustat subsample based on name similarity as well as by matching the firm names to ORBIS subsidiary data. Where needed, we perform manual checks to confirm majority ownership using SEC 10-K filings. We aggregate the data to the UO parent-company level, accordingly.[21] We further link private subsidiaries to their UO firm based on ORBIS data (will be explained separately below). Accordingly, if a firm's subsidiary publishes scientific articles while the parent company is the assignee registered on the firm's patents, we record both at the UO level and a citation from a patent to a publication would be considered as an internal citation.

B. OWNERSHIP STRUCTURE

Dealing with ownership changes has been a major effort of this project, especially in regard to reconstructing and improving the NBER patent database. We unpack firms' ownership structure by constructing firm-level data before proceeding to patent match. Ownership may change over the years of our sample due to changes at the UO Compustat firm level as well as at the subsidiary level. We rely on two main sources to construct ownership data: (i) SDC Platinum and (ii) historical snapshots of ORBIS files.

I. SDC M&A MATCH

Ownership changes of the UO Compustat firms in our sample are tracked through the SDC Platinum database with each firm name dynamically matched to up to five PERMNO_ADJ between the years 1980 and 2015. Based on M&A deals available in SDC Platinum from 1980 to 2015, we downloaded detailed information on the acquirer and target firm names, acquirer and target firm CUSIPs, types of deals, execution dates, and percentage of shares owned after each transaction. We exclude deals that we identify as asset or business unit acquisitions.

We restrict the sample to deals involving a change in ownership that resulted in majority ownership (more than 50% of shares) for the acquirer. Execution dates are used to define the years a target firm begins or ends (in case of several acquisitions during the sample period) being owned by an acquirer. We then standardized both target and acquirer names similar to the standardization done for Compustat firm names. We match each deal's target and acquirer firm to our list of Compustat firms using both CUSIP numbers and all standardized historical names. It is important to use historical data as the information is recorded on SDC at the time of acquisition. We retain deals where both acquirer and target firms are matched to a Compustat firm in our sample. We track up to five ownership changes for each target firm name after it enters Compustat and one additional reassignment before it became publicly traded if relevant (i.e., if it was a subsidiary of another Compustat firm in our sample prior to its IPO)[22].

We perform extensive manual checks, including identifying and distinguishing companies with similar names (e.g., old vs. new Pharmacia). We Assume that if a firm is acquired, all its patents and publications are transferred to the acquirer firm.

[21] For example, GENZYME CORP (GVKEY 12233) - after verifying ownership on SEC filings: GENZYME MOLECULAR ONCOLOGY (GVKEY 117298), GENZYME TISSUE REPAIR (GVKEY 118653), GENZYME SURGICAL PRODUCTS (GVKEY 121742) and GENZYME BIOSURGERY (GVKEY 143176) are all accounted under their parent company GENZYME CORP (GVKEY 12233). While, GENZYME TRANSGENICS CORP (a.k.a. GTC BIOTHERAPEUTICS, GVKEY 028563) is a standalone alone company in our data as it was not majority-owned by GENZYME CORP after it spun-off.

[22] For example, Vysis Inc first enters our sample as a subsidiary of Amoco (1991-1997) and is then spun-off and becomes an UO firm in our sample as an independent publicly traded company in 1998 and eventually acquired and becomes a subsidiary of Abbott in 2001.

Example 5: NABISCO

This example illustrates how we account for ownership changes in our data. During our sample period, Nabisco has changed ownership four times. In 1981 Nabisco merged with the publicly traded company Standard Brands to form Nabisco Brands. Then, in 1985 R.J. Reynolds merged with Nabisco Brands to create RJR Nabisco, which eventually became Nabisco Group holding after the tobacco business was spun out in 1999. In 2000, Nabisco was acquired by Phillip Morris, which combined Nabisco with its Kraft brand. Finally, in 2001 Kraft (together with Nabisco) was spun out as a publicly traded company that later on became Mondelez International Inc. In our dataset all Nabisco related patents and publications are dynamically transferred between Compustat records and UO firms based on its ownership throughout the years:

Table 4. Nabisco dynamic match

Years	related GVKEY	Original owner	Current Compustat name	Comments
1981-1985	7674	STANDARD BRANDS INC	NABISCO BRANDS INC	1981: Standard Brands company merged with Nabisco Inc to form Nabisco Brands Inc.
1986-1999	9113	R J REYNOLDS IND INC	NABISCO GROUP HOLDINGS CORP	1985: R.J. Reynolds Industries merged with Nabisco Brands to form R J R Nabisco Inc
2000	8543	PHILIP MORRIS COS INC	ALTRIA GROUP INC	2000: Nabisco was acquired by Phillip Morris
2001-2015	142953	KRAFT FOODS INC	MONDELEZ INTERNATIONAL INC	2001: Kraft together with Nabisco split from Phillip Morris

Note: This table presents the dynamic reassignment for Nabisco related patents at the GVKEY-Period level. Current Compustat name is based on Compustat 2018 file. Original Compustat owner for the period is the historical firm name based on CRSP Monthly Stock file.

Examining NBER 2006, we find that for the purpose of Compustat accounting information, all Nabisco related patents are linked to GVKEY 9113 from 1950 to 1999. Though the current name related to GVKEY 9113 is "Nabisco Group Holding Corp", based on the historical name information, we know that up to the merger of R.J. Reynolds with Nabisco it belonged solely to R.J. Reynolds. Reynold's patents, on the other hand (Over 419 patents for the period before it spun-out of RJR Nabisco and not including patents of acquired companies such as Heublein Inc), are not assigned by NBER to GVKEY 9113 and they are only being linked to Compustat records after the tobacco business spun-out of RJR Nabisco and became independently traded again under GVKEY 120877 (eventually merging with U.S. operations of British American Tobacco to form Reynolds American Inc). As a result, in 1998, the patent stock in NBER for GVKEY 9113 ("Nabisco Group Holding Corp") is 495 (consisting solely of Nabisco matched patents), whereas it should be 914 if it included R.J. Reynolds related patents. Furthermore, NBER does not dynamically move Nabisco's patent-stock or account for its patent flow after 1999 when it was bought by Philip Morris and eventually became part of Kraft (a total of 529 Nabisco related patents up to 2006).

Table 5. Data for Nabisco in NBER 2006

Current compustat record name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
NABISCO GROUP HOLDINGS CORP	**9113**	**1950**	**1999**	**9113**	**1**	**1950**	**1999**
NABISCO INC	7675	1950	1980	9113	-1		
NABISCO BRANDS INC	7674	1950	1984	9113	-1		
NABISCO HLDGS CORP -CL A	31427	1993	1999	9113	-1		

Note: PDPCO is NBER's Patent Data Project (PDP) unique com y id. FIRSTYR is the first year GVKEY company has data. LASTYR is the last year a GVKEY company has data. PDPSEQ is the GVKEY sequence within PDPCO. If PDPSEQ=-1, the related GVKEY is disregarded. BEGYR is the beginning year for GVKEY within PDPCO. ENDYR is the last year for GVKEY within PDPCO. All patents related to Nabisco will be accounted under GVKEY 9113 from 1950 to 1999, while all other related GVKEYs are disregarded.

Example 6: CHEMTURA CORPORATION

An example that illustrates how having historical names helps account for ownership changes in our data and accurately compute the patent stock. Chemtura Corporation traces back to the chemical corporation Crompton & Knowles that was founded in the 19th century. In 1996, Uniroyal Chemical Corporation merged with Crompton & Knowles. In 1999, Crompton & Knowles merged with the publicly traded company Witco to form Crompton Corporation. In 2005, Crompton acquired the publicly traded company Great Lakes Chemical Company, Inc., to form Chemtura Corporation, while Great Lakes Chemical Corporation continued to exist as a subsidiary company of Chemtura.

Based on our strategy, we consider all historical names of the current Chemtura Corporation (PERMNO_ADJ 38420) including:

1) CROMPTON & KNOWLES CORP starting 1980
2) CK WITCO CORP starting 1999
3) CROMPTON CORP starting 2000
4) CHEMTURA CORP starting 2005

Most importantly, because we consider the complete set of historical names, we are able to locate all the relevant M&As throughout the years of the publicly traded firms that exist as an independently traded company in our data prior to an acquisition. Accordingly, we dynamically transfer them post-acquisition to PERMNO_ADJ 38420:

1) Uniroyal Chemical Corporation (acquired 1996)
2) Witco Corp (acquired 1999)
3) Great Lakes Chemical (acquired 2005)

When we examine NBER 2006 patent dataset, we find that the only name that was matched to CHEMTURA CORP (GVKEY 3607) is "CHEMTURA CORP" (PDPASS 13245038). As the Chemtura name was adopted in 2005, only one patent was matched for that name. In addition, none of the acquired publicly traded companies were dynamically transferred to CHEMTURA CORP post-acquisition. It is likely that a lack of information on historical names led NBER to rely on post-acquisition name (Chemtura) and thus prevented it from accounting for the M&A activities.

By considering all previous names (without their subsidiaries and the acquired companies) related to GVKEY 3607: (i) Crompton & Knowles Corp; (ii) CK Witco Corp and (iii) Crompton Corp - based on the NBER raw patent match, we locate 220 additional patents up to 2006 that were not linked to any Compustat record that should be assigned to *Crompton & Knowles* (77 patents), *CK Witco* (26 patents)), and *Crompton* (117 patents). In addition, the acquired *Uniroyal Chemical Corp* has a patent stock of 379 patents in 2006 (out of which 185 patents are post-acquisition), and the acquired Witco company has a patent stock of 405 in 2006 (out of which 62 patents are from post-acquisition period), and *Great Lake Chemicals* has a patent stock of 183 in 2006 (out of which three patents are in 2006, the year after the company was acquired).

Overall, applying our strategy to the raw NBER patent match, we find a patent stock of **1,187** patents in 2006 for GVKEY 3607 as opposed to **1** patent in NBER.

II. ORBIS SUBSIDIARY MATCH

Due to the complexity of measuring large firms' innovative activities, which typically take place inside numerous subsidiaries, we aggregate the data to the ultimate-owner-parent-company level based on majority ownership. There are several challenges in keeping track of subsidiaries owned by UO Compustat firms, which may publish and patent in their own name. First, many of these subsidiaries are private, and manual checks are sometimes required to verify which of the several similarly named companies was acquired by the firm. Furthermore, subsidiary ownership may change over the years. Companies may spin out their subsidiaries, some of which might go public or sold to other firms, where

they are maintained as stand-alone subsidiaries and continue to patent or publish. Tracking subsidiary ownership is the main challenge we deal with and is explained below.

For firms with at least 50 patents[23] over the sample years at the PERMNO_ADJ UO level, we collect all related domestic and international subsidiary names using ORBIS and SEC filings, as explained below.

We obtained historical ORBIS files for years 2002 to 2015, which provide us with snapshots of ownership structures for each of the years. Using historical snapshots of ORBIS files, we are able not only to identify ownership changes at the subsidiary level but also new established subsidiaries.[24]

We start by standardizing the names of "Global Ultimate Owner" (GUO) firms and match the names to standardized historical Compustat names of firms with more than 50 patents at the UO level. Once again, it is important to use historical names for this match as the names in each of our ORBIS files appear as of the year the file was recorded.

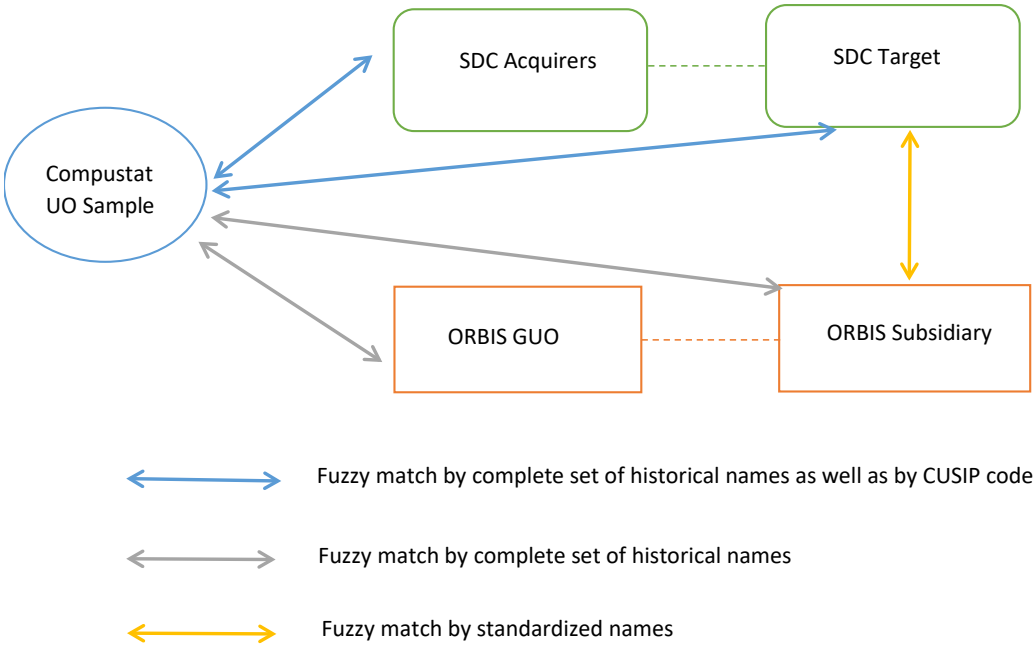
Next, we link the subsidiaries of the successfully matched ORBIS owners to the PERMENO_ADJ of the corresponding parent firms. We restrict our sample to subsidiaries that are majority-owned by the parent firm. After standardizing each subsidiary name similar to the standardization done for Compustat names, we obtain the first and last year it appears under a PERMENO_ADJ during 2002-2015. To avoid duplicated matching efforts, in many cases, we drop subsidiaries that have the same organic name as the parent UO firm as they were already matched at the UO Compustat level. Some subsidiary names appear under more than one PERMNO_ADJ due to acquisitions throughout the years. Because we use yearly snapshots of ownership structure from ORBIS, we are able to account for name changes of subsidiaries over the period.

For firms that exit Compustat before 2002, we manually collect subsidiary names based on their latest available 10-K SEC filing[25] as well as rely on the NBER patent database for pre-2002 ownership data.

Since our sample starts from 1980 and the ORBIS files are only from 2002, we try our best to account for ownership changes of the subsidiaries for the years preceding 2002 using SDC and Compustat databases. We elaborate on our approach below.

[23] At the UO level we match for all subsample firms related subsidiaries with organic names.
[24] One caveat is that the coverage of subsidiaries in the first few years of data files is incomplete.
[25] We do so for top 100 firms based on R&D spending.

Figure 5. Subsidiary matching Description



1) Fuzzy match between standardized subsidiary names and standardized SDC target name. For the matched result, we locate:

a) Cases where the acquirer firm is a UO Compustat firm in our sample, which include:

(i) Cases where the acquirer firm has the same PERMNO_ADJ as the parent firm of the subsidiaries. These cases confirm the direct acquisition of the subsidiary by the parent firm and provide us with the start date of the subsidiary (the year of acquisition) under the parent firm.

(ii) Cases where the acquirer firm is a UO Compustat firm in our sample that was acquired by the parent firm of the subsidiary (i.e., the PERMNO_ADJ of the acquirer and the PERMNO_ADJ of the parent of the subsidiary are related through acquisition). These cases confirm an indirect acquisition of the subsidiary by the parent firm and provide us with the start year of the subsidiary under the parent firm – i.e., year the ORBIS parent firm acquired the Compustat acquirer firm or the year of acquisition of the subsidiary (the latest).

b) cases where the acquirer firm is not a UO Compustat firm in our sample:

(i) if the CUSIP code of the UO parent firm related to the target firm (as indicated in SDC file) is the same as a CUSIP code related to the PERMNO_ADJ of the ORBIS parent of the subsidiary, it indicates that the subsidiary was acquired from the parent firm by the acquirer and provides us with the end date for the subsidiary under the parent firm – the year of acquisition.

(ii) For each acquirer firm's direct CUSIP code, we search the complete SDC file for a deal where it was acquired by a firm with a CUSIP code related to the PERMNO_ADJ of the ORBIS parent of the subsidiary. These cases indicate indirect acquisitions, in which the subsidiary was acquired by a non-Compustat sample firm that was itself acquired by the subsidiary's ORBIS parent firm. Such cases provide us the start year of the subsidiary under the parent firm –i.e., the year the ORBIS parent firm acquired the non-Compustat acquirer firm or the date of acquisition of the subsidiary (the latest).

2) Fuzzy match of cleaned subsidiary names

As the subsidiary name list includes closely related firm names with different legal entities, we use a clean version of the names that omits legal entity and other common words and we fuzzy match it to both clean Compustat names and the list of clean subsidiary names we found relevant acquisitions for in (1) above.

The fuzzy match to Compustat enables us to link each matched subsidiary name to the dynamic year sequence we constructed for UO Compustat firms. For the fuzzy match to the list of acquired subsidiaries, we adopt the relevant start &/end year we located in (1) above to all related subsidiaries.

3) As an additional check, we manually go over subsidiaries that did not match under 1) or 2) above and appear under more than one parent firm in our ORBIS sample or have more than 100 matched publications or patents.[26] For these cases, we check online sources and manually adjust their start and end date. Finally, for subsidiaries, we were not able to identify the start or end year- we assume that they belong to the UO firm from its start date until the end date. However, if the UO firm appeared in ORBIS files for more than three years before the subsidiary was first linked to it, we adopt the first year the subsidiary is connected to the parent ORBIS firm as the start date of the subsidiary, under the assumption that it was acquired during that year by the parent firm.

All subsidiaries are assumed to move with their parent firm in cases where the parent firm is acquired unless a subsidiary has a different end date from its parent firm, or it is related to the Compustat dynamic year sequence. Moreover, we do not account for reassignment of patents that are not part of the ownership changes that we document.

C. MATCHING

We perform several matches to construct our data, including (1) matching patent data to Compustat companies and their related subsidiaries; (2) matching scientific publications to Compustat companies and their related subsidiaries; (3) mapping patent citations to publications. We discuss each of these procedures below.

I. MATCHING PATENT DATA TO COMPUSTAT COMPANIES AND THEIR SUBSIDIARIES

After obtaining our initial subsample of firms and the various firm names, we proceed to match our firm sample to assignees of the patents granted by USPTO[27] using PatStat, which includes approximately 5.3 million patents for years 1980 through 2015.

We first remove published patent applications (i.e., publication numbers longer than 7 characters), non-utility patents, including Design, Reissue, Plant and T documents, and reexamination certificates. Next, we remove patents assigned to individuals or government entities (for example, an assignee that includes the string "DECEASED" or "U.S. DEPARTMENT"). We are then left with 4.97 million granted utility patents.

To compare assignee names to the standardized firm names in our sample, we standardize assignee names similar to the firm name standardization explained above. Assignee name standardization includes converting names to upper case, removing excess spaces, cleaning non-alphanumeric characters, and replacing legal entity endings, including commonly abbreviated terms (for example, "CORPORATION" is replaced with "CORP"; "LABORATORIES" and "LABS" with "LAB"). At the end of this process, we are left with 897K unique standardized assignee names.

The matching strategy includes several distinct steps. We begin by matching firm names to assignees using an exact match. We then perform several fuzzy matching techniques to account for names that are slightly different but are in fact, the same entities. Extensive manual checks at the assignee name and patent level were performed to ensure the quality of the matches.

[26] When matching the subsidiary name list to other databases users should include extensive manual inspection to matched results, including manually verifying the start and end year for top matched result that differ from the top 100 matches that we manually verified. Special care should be given to companies with similar names and to generic company names.

[27] We limited our data sources to USPTO data to make the project manageable in terms of matching. Since our firm sample is limited to U.S. HQ firms, we believe it is reasonable to focus only on USPTO data.

UO Level Matching

(1) Exact Matching

Exact matching was conducted by comparing assignee names to firm names. The matching was carried out twice, both for standardized and for original names. An additional match was conducted after dropping legal entities. The latter step was performed to account for firms whose names differ only by the legal entity. Extensive manual checks are performed to verify the matches. Special care was taken in cases where firm or assignee names are generic, when several different firms share a common portion of a name, or when firm names contain a common given or family name. To resolve ambiguities, we performed web searches and examined the actual patent documents.

(2) Fuzzy Matching

For the remaining assignee names not matched during the exact matching process, fuzzy matching was performed to find each of the assignee names from the firm names to catch cases where assignee and firm names do not match exactly but are, in fact, the same firm. Some names are misspelled or include additional letters that prevent an exact match. In other cases, patent assignee names include a specific division title ("ROCKWELL BODY AND CHASSIS SYSTEMS", "ROCKWELL SOFTWARE"), a licensing unit ("MICROSOFT TECHNOLOGY LICENSING LTD", "RCA LICENSING"), or a geographic branch or firm location ("BIOSENSE WEBSTER ISRAEL LTD").

Fuzzy matching was performed using the FuzzyWuzzy library in Python (i.e., Token Set function), and using term frequency-inverse document frequency (TF-IDF).

FuzzyWuzzy uses a slightly modified Levenshtein distance to calculate similarities between two strings. More specifically, a vector is created for each assignee name using the words contained in it and then compared to the entire list of firm names (that are also vectorized) to find potential matches. When comparing two vectors, the same elements (i.e., words) contained in both vectors are marked as “matched”, and the similarity between the remaining, different elements are calculated using the Levenshtein distance algorithm after sorting the elements alphabetically. The similarity score between the two strings is higher when the elements that match exactly make up a larger portion of the strings and when the remaining (unmatched) part has a small distance based on the Levenshtein distance. To account for multiple scores that indicate a strong match, the top ten potential matches with the highest scores are examined manually to identify the most appropriate match.

An additional fuzzy match was done by converting the assignee and firm names into a term frequency-inverse document frequency (TF-IDF) matrix and calculating a cosine similarity score for each pair of assignee and firm name. This method is widely used to take care of typos and variations of spelling in textual string matching. By increasing weights of unique words and reducing the weights of common words in the corpus, the TF-IDF algorithm improves the relevancy of cosine similarity measures that are calculated between each pair of names.

An additional search of the top 300 patenting firm names was conducted to find matching assignee names that were not matched through the initial fuzzy match process. In this step, we search for assignee names with at least five related patents that contain any of the fully standardized firm names after the removal of legal entities. Through this process, we include subsidiaries that have the same organic name as the parent UO firm (For example, "EMERSON" firm name matched with "EMERSON CLIMATE TECH", a division within the firm). The search was conducted through a script that receives the list of assignee names and fully standardized firm names and automatically produces all matching pairs. In each search result pair, a firm name is contained within the assignee name string. Following the search, a complete manual check was conducted among all search results to mark the legitimate matches.

As a final check, we employed RAs to verify that the assignees with more than 100 patents were correctly matched by the fuzzy matching algorithm. The RAs went through the fuzzy matched names to confirm that they are in fact, the right match. Existing matches were invalidated when they were not the right match, and new matches were added when more appropriate matches were found.

Subsidiary Level Matching

(1) Exact Matching

Exact matching was conducted in a similar fashion to the UO level matching process. Original and standardized versions of the assignee names were compared to the list of standardized subsidiary names, and manual checks were performed in cases where the name was generic.

(2) Fuzzy Matching

The fuzzy match for subsidiaries was done by converting the assignee and firm names into a term frequency-inverse document frequency (TF-IDF) matrix and calculating a cosine similarity score for each pair of assignee and standardized firm name. To reduce the size of the task, results were limited to assignees with at least 30 patents, and identification of matches was conducted by manually comparing the top-scoring assignee-firm pairs for each assignee.

Overall, this process yields 1.35 million patents mapped to 4,520 U.S. headquartered Compustat firms and their subsidiaries via patent number and NAME_ID. These patents account for about 50% of all utility patent grants from U.S. Origin. When a patent has several sample firm assignees, we match the patent to multiple firms and assign fractional patent ownership to each assignee (i.e., 1/number of sample assignees). Patents enter our sample once the related UO firm is publicly traded and not before[28]. Any patent that enters the data remains until the end of the sample period unless the related firm is acquired by an out of sample firm, dissolved, or taken private. In the case of ownership change within the sample, patents are dynamically matched to up to five UO firms. Moreover, we do not account for reassignment of patents that are not part of the ownership changes that we document.[29]

II. MATCHING SCIENTIFIC PUBLICATIONS TO COMPUSTAT FIRMS AND THEIR SUBSIDIARIES

We procced by matching our firm names to publication data to capture their investment in science. We obtain publications data from the Web of Science database (previously known as ISI Web of Knowledge). We include articles from journals covered in the "Science Citation Index" and "Conference Proceedings Citation Index - Science," while excluding social sciences, arts, and humanities articles.

Each publication record contains detailed information including the title of the publication, authors, journal, and our primary variable of interest, an affiliation field with name and address of the publishing institute or company in case of a corporate publication. This field can include more than one listing in case of a collaborative publication, for example, "*TEXAS INSTRUMENTS INC*, DEPT DATAPATH VLSI PROD SEMICOND GRP 8330 LBJ FREEWAY, POB 655303, DALLAS, TX 75265 USA | *SUN MICROSYST INC*, MT VIEW, CA USA".

We apply a many-to-many fuzzy matching algorithm between each standardized name and the affiliation field for each publication (approximately 47 million publications, 8 million conference proceedings and 60 thousand names) while allowing for more than one firm to be matched to each publication (to allow for collaborative publications).

We first standardize the affiliation string of each Web of Science publication similar to the name standardization process explained above. The standardization removes special characters such as ampersands and words that indicate legal entities such as "INC" or "CORP". It also ensures that common words such as "technology" and "chemicals" that frequently appear in company names are abbreviated in the same manner[30].

Second, we perform exact matching on company names and publication affiliation string using regular expressions. In addition, we calculate Levenshtein edit distances between company name-publication affiliation pairs. This step is

[28] Furthermore, we do not account for patents in gap years that the related UO firm is not publicly traded. To adjust for potential previous stock of patents, we apply a 15% inflation to the patent stock at the year a name enters the sample.

[29] Specific details on construction of patent flow and patent stock variables are provided under "patent_do.do" and "panel_do.do" files. The main patent output file is "DISCERN_patent_database_1980_2015_final1.dta" – it presents ownership at grant year.

[30] For instance, the word "technology" in a company name can be plural ("technologies") or abbreviated ("technol", "tech"). These special cases are abbreviated to "TECH" in our standardization code.

necessary because misspellings are common (e.g., BRISTOL MYERS SQUIBB misspelled as BRISTOL MEYERS SQUIBB). Since the company name in a publication affiliation is typically embedded in a longer string, which includes buildings, street names, cities, zip-codes, and country names, even correct matches will incur large distances. Therefore, we use a "partial" Levenshtein distance, which calculates the edit distance between the shortest common segment between two strings. That is our "partial" edit distance for the company name "IBM" and affiliation "IBM Corp, SSD, San Jose, CA 951953 USA" will be zero, whereas a raw Levenshtein distance would be 35.

Third, we conduct manual checks on fuzzy-matched company name-publication affiliation pairs. In particular, we exclude matches from company names to eponymous buildings (e.g., Gillette Hall), schools (e.g., Heinz College), hospitals (e.g., Du Pont Children's Hospital), charitable foundations, and endowed chairs. We also conduct manual checks on company-publication pairs with zero edit distances (exact matches) if the company names overlap with a common last name (e.g., ABBOTT), a geographic/historical location (e.g., BABYLON, BRISTOL), or branch of science & engineering such as "APPLIED MATERIALS" or "SEMICONDUCTOR", as these are especially prone to being false positive matches. We also ensure that similar but distinct company names do not match to the same affiliation field (e.g., NORTHROP and GRUMMAN before their merger in 1994 are treated as separate companies and will not match to NORTHROP GRUMANN). In cases where company names are the same, we verify matches by comparing the address listed within Compustat to the address in the publication data. For example, to distinguish between "THERATECH INC / UTAH" and "THERATECH INC", we verify that the address of the firm under the affiliation field is in Salt Lake City.

At the end of this procedure, we obtain a match between a WOS record ID and our NAME_ID. We find approximately 800 thousand unique articles from more than 10 thousand different journals that were published from 1980 through 2015, with at least one author employed by our sample of Compustat firms and their subsidiaries. When a publication has several sample firm affiliates, we match the publication to multiple firms and assign fractional publication ownership to each firm (i.e., 1/number of sample affiliates). For the sample of patenting firms, publications enter our sample once the related UO firm is publicly traded and not before[31]. Any publication that enters the data remains until the end of the sample period unless the related firm it is acquired by an out of sample firm, dissolved, or taken private. In the case of ownership change within the sample, publications are dynamically matched to up to five UO firms.[32]

III. MATCHING NPL PATENT CITATIONS TO WEB OF SCIENCE ARTICLES

Patent citations to science are obtained from the Non-Patent Literature (NPL) citations section located at the front page of patents taken from the PatStat database. An example of a front-page patent citation to non-patent literature is provided in Figure 7. We obtain all NPLs related to patents granted in the period 1980-2015 (including corporate sample firm patents and non-corporate patents). We first remove NPL citations that we identify as non-publication references (e.g., reference that includes the string "PATENT ABSTRACT", "U.S. APPLICATION NO.", "US COURT", "PRODUCT INFORMATION", "DATA SHEET", "WHITEPAPER"). We then proceed to match NPLs to corporate publications from Web of Science (approximately 10M citations and 800K corporate publications). This step presents a significant challenge due to differences in structure between NPL and publication string text- NPL patent citations to publications are highly non-standardized (see Table 7 for examples). We begin with a many-to-many match, allowing more than one publication to be matched to each NPL. For each possible records pair, we construct a score that captures the degree of textual overlap between the title, journal, authors, and publication year. To exclude mismatches, we use a more detailed matching algorithm that is based on different sources of publication information: standardized authors' names, number of authors, article title, journal name, and year of publication. The matching algorithm accounts for misspelling, unstructured text, incomplete references, and other issues that may cause mismatches.

[31] Furthermore, we do not account for publications in gap years that the related UO firm is not publicly traded. To adjust for potential previous stock of publications, we apply a 15% inflation to the publication stock at the year a name enters the sample.

[32] Specific details on construction of publication flow and publication stock variables are provided under "pub_do.do" and "panel_do.do" files. The main publication output file is "DISCERN_pub_database_1980_2015_final1.dta" – it presents ownership at publication year.

We will use the example below to illustrate the complication of the match and the algorithm we applied to detect a match.[33]

The first step is to match the publication's "Title" field and the title that is located within the citation string. There are two main problems: (i) the position of the title within the citation is not fixed and (ii) there may be a small variation in the title (e.g., "GIVE" vs. "GIVES") and thus an exact match may not perform well. To overcome these problems, we implement a fuzzy matching algorithm. After we standardize and clean the different strings, we measure the length-difference between the citation string and the publication title string. Then, using STATA's "STRDIST" command, we calculated the distance between the two strings. We use the difference between the length difference and distance as a measure of proximity of the titles. We supplement this measure with an exact match of the first part of the title. In some cases, the title is missing from the citation string. In such cases, we rely more on other available features to determine the final match.

Second, we match between the publication's "Authors" field and the authors listed within the citation string. As with the title, we cannot identify the exact location where the authors are contained within the citation string since the location varies from one citation to another. In addition, there are several differences in how names are written: (i) Last name only vs. full names; (ii) name vs. initials (e.g., LIN KS vs. LIN KUN SHAN); (iii) listing of all authors vs. one author followed (or not) by "et al."; (iv) order of last and first names within the string. To verify a match by authors, we first count the number of authors listed in the publication record. We then check whether the citation string contains "et al.". To mitigate the name variation problem, we implement an algorithm that matches different variations of the authors' name to the citation (including the transformation of last and/or first and/or middle name to initials and changes in the order listed). In cases where several authors are listed under the publication and "et al." does not appear within the citation, we perform a one-to-many match between the citation and each author and impose that at least 80% of the authors must be matched to the citation to determine a match. For cases where several authors are listed in the publication and only one is matched within the citation while "et al." is omitted, we rely more on match results in other features to determine the final match.

Next, we match journal information including standardized journal's name, publication year, page numbers and volume, while accounting for typos, abbreviations (e.g., "INTERNATIONAL ELECTRONICS" vs. "ELECTRONICS") and differences in the format of the string between the datasets (e.g., "VOL. 53, NO. 3" vs. "53(3)").

Finally, we use different combinations of the match results for the various features (title, authors, and journal information) according to their relative importance to determine a final match[34]. We perform extensive manual checks to confirm matches[35]. At the end of this procedure, we obtain unique identification numbers for the citation, the citing patent, and the cited publication.

We then focus on citations made by corporate sample patents. We further differentiate between internal citations (patent citation by the focal firm's patent to its own publication) and external corporate citations (patent citation to the focal firm's publication by other corporate patents). The Dynamic match of patents and publications allows us to classify an internal or external citation based on the owners of the citing patent and the cited paper at the time the paper is published. For the purpose of classifying internal or external citation, we rely on the original UO firm the publication was affiliated with at its publication year[36]. For external citations from the corporate sample firms, we further construct segment

[33] The following example (first line in Table 7) illustrates the matching challenge. NPL citation: LIN, KUN SHAN, ET AL., SOFTWARE RULES GIVES PERSONAL COMPUTER REAL WORD POWER, INTERNATIONAL ELECTRONICS, VOL. 53, NO. 3, FEB. 10, 1981, PP. 122 125.
Matched Publication: Title: SOFTWARE RULES GIVE PERSONAL-COMPUTER REAL WORD POWER, Authors: LIN KS, FRANTZ GA, GOUDIE K, Journal information: ELECTRONICS 54 (3): 122-125 1981.

[34] A sample algorithm is provided under "NPL_cleaning_exp.do" file

[35] There are several cases where the NPL reference is a citation to a working paper and we are able to match it to the final published paper that appears on WOS database – we consider those as matches.

[36] i.e., if Company B acquires Company A (let's assume A is a Compustat firm in our sample pre-acquisition): Citations by B's patents post-acquisition to A's publications that were published pre-acquisition are classified as external citations. However, citation from B's patents to A's publications published post-acquisition are classified as internal citations. Moreover, as opposed to publication and patent stock variables, citations do not move dynamically between firms in case of acquisition.

proximity (Standard & Poor's (S&P), 2018a; Bloom, 2013) measures between the cited and the citing firms, as explained in the main text.

Following the above procedures, we obtain 71 thousand unique corporate cited publications (9 percent of corporate publications), by 143 thousand unique corporate citing patents. Of the cited publications, 61 percent receive only external corporate citations, and the remaining receive at least one internal citation[37]. The temporal structure of citations and publications are illustrated in Figure 8.

D. COMPARISON OF OUR DATA TO NBER PATENT DATA, FOR 1980-2006

We match 780 thousand patents for 1980-2006 (Figure 6). We compare our sample for 1980-2006 to NBER 2006 patent data for U.S. headquarter firms and their related subsidiaries looking at a specific patent assigned to a GVKEY at a grant year (Table 6).

Figure 6. Patents assigned to U.S. HQ public corporations and their related subsidiaries

Table 6 presents the comparison results. For this period, we match about 80% of the patent-GVKEY matches as in NBER. We find an additional 18% patents due to: (i) improved dynamic linkage of patents to GVKEYs (e.g., Pharmacia), and (ii) linkage of additional patents based on historical name information, wider M&A coverage, and improved matching techniques (e.g., Phillips). In 1% of the cases, we find the same assignment as NBER, but these matches are irrelevant for our sample (e.g., Rhone-Poulenc). Lastly, in about 1% of the cases, we are unable to include the NBER matches for a variety of reasons, including possible mistakes on our end.

[37] Specific details on construction of NPL citation variables are provided under "npl_do.do" file. The main NPL output file is "DISCERN_corp_NPL_output_80_15_final.dta".

Table 6. Comparison to with NBER for 1980-2006: Patent-GVKEY Assignments, U.S. HQ Firms

Comparison 1980-2006	% Patents	Examples
Agreement	80	
Matched to different GVKEY	4	Improved dynamic matching to Compustat records using historical name >>> Patents of the merged company included under the GVKEY from acquisition, but not before. Example: PHARMACIA: we matched to PHARMACIA & UPJOHN's GVKEY pre-2000 instead to MONSANTO.
Only our Sample	14	Newly matched patents due to (i) availability of historical names; (ii) better M&A data; and (iii) Improved matching. e.g., PHILLIPS PETROLEUM CO: 4000+ patents pre-merger with Conoco Inc in 2002; MONSANTO: 2000+ patents pre-merger with Pharmacia;
Only NBER- we matched but irrelevant gvkey-year	1	(i) NBER match (incorrectly) based on 2006 Compstat name: e.g., ~1000 patents of RHÔNE-POULENC patents matched to RORER's GVKEY pre-merger in 1990; (ii) Improved subsidiary coverage: e.g., ~450 patents of HUGHES AIRCRAFT are incorrectly linked to GM's GVKEY pre-1985 acquisition;
Only NBER	1	(i) Withdrawn patents: ~600 patents (ii) Misc. couldn't verify connection, typos, and possible mistakes by us

REFERENCES

Bessen, James. 2009. "NBER PDP Project user documentation: Matching patent data to Compustat firms." Unpublished documentation. http://users.nber.org/~jbessen/matchdoc.pdf. Data available at: https://sites.google.com/site/patentdataproject/Home/downloads?authuser=0 (Accessed: April, 2016).

Bloom, N., M. Schankerman, and J. Van Reenen. 2013. "Identifying technology spillovers and product market rivalry." Econometrica, 81(4): 1347–1393.

Bureau van Dijk. 2018. ORBIS Ownership files, 2002-2015. Bureau van Dijk, Chicago, IL. Provided via a Duke University subscription service.

Center for Research in Security Prices (CRSP). 2018a. CRSP Compustat Merged (Monthly), 1980-2015. Available to Duke University through Wharton Research Data Services (WRDS). https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/ccm_m/linktable/index.cfm?navId=137 (Accessed: August 2018).

Center for Research in Security Prices (CRSP). 2018b. CRSP Stock (Monthly), 1980-2015. Available to Duke University through Wharton Research Data Services (WRDS). https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/stock_m/msf.cfm?navId=145 (Accessed: August 2018).

Clarivate Analytics. 2016. Web of Science (WoS) Core Collection XML, 1900-2016. Clarivate Analytics, Philadelphia, PA. Obtained from Clarivate Analytics by license in 2017.

European Patent Office. 2016. The EPO Worldwide Patent Statistical Database, 1980-2015. PATSTAT Global single edition 2016. Obtained from EPO by license in 2016.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER patent citation data file: Lessons, insights and methodological tools." NBER Working Paper 8498, National Bureau of Economic Research. https://www.nber.org/papers/w8498. Data available at: http://data.nber.org/patents/ (Accessed: April 2016).

Securities Data Company (SDC) Platinum. 2018. Mergers & Acquisitions module, 1980-2015. Refinitiv. Provided via a Duke University subscription service.

Standard & Poor's (S&P). 2018a. Compustat Segments dataset, 1980-2015. Available to Duke University through Wharton Research Data Services (WRDS). https://wrds-web.wharton.upenn.edu/wrds/ds/comp/seghistd/index.cfm?navId=87 (Accessed: August 2018).

Standard & Poor's (S&P). 2018b. North America Annual Compustat, 1980-2015. Available to Duke University through Wharton Research Data Services (WRDS). https://wrds-web.wharton.upenn.edu/wrds/ds/comp/funda/index.cfm?navId=80 (accessed August 2018).

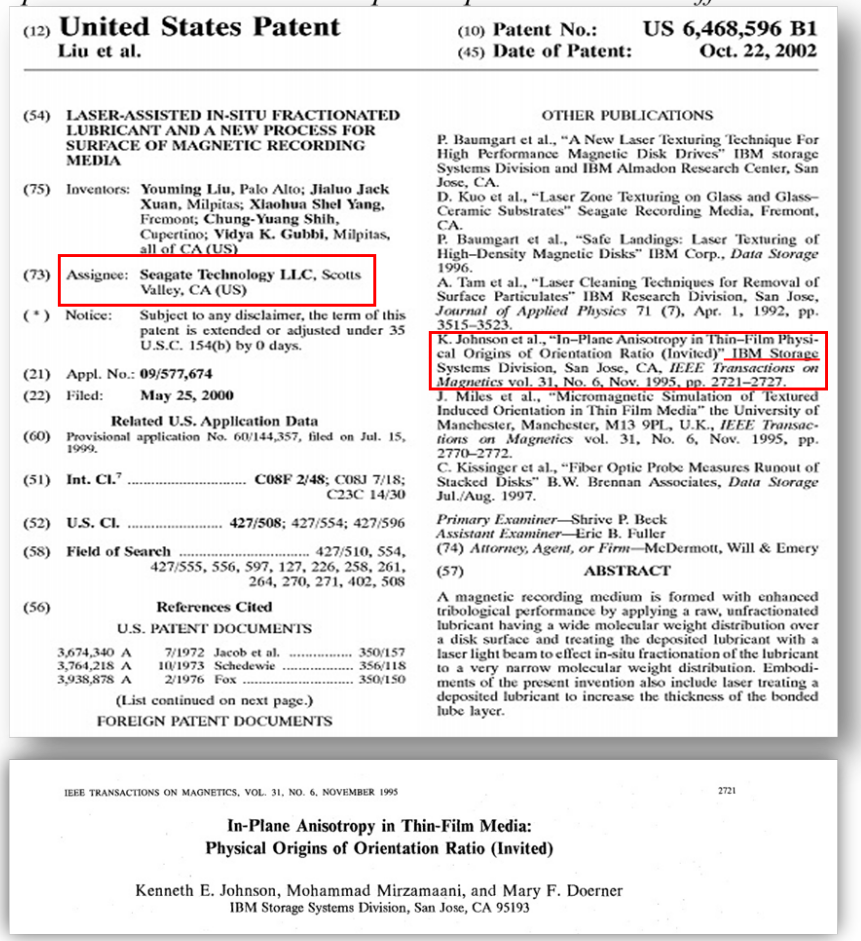
Wu, Y., 2010. What's in a name? What leads a firm to change its name and what the new name foreshadows. Journal of Banking & Finance, 34(6), pp.1344-1359.

Table 7. Matching Citations to Scientific Publications - Examples

Citation	Publication info			Comment
	Title	Authors	Journal information	
LIN, KUN SHAN, ET AL., SOFTWARE RULES ***GIVES*** PERSONAL COMPUTER REAL WORD POWER , **INTERNATIONAL ELECTRONICS**, VOL. **53,** NO. 3, FEB. 10, **1981**, PP. **122 125**.	"SOFTWARE RULES ***GIVE*** PERSONAL-COMPUTER REAL WORD POWER"	LIN KS, FRANTZ GA, GOUDIE K	**ELECTRONICS 54** (3): 122-125 1981	Typo in title and journal Vol.; initials vs. full name
U. WACHSMANN, R. F. H. FISCHER AND J.B. HUBER, MULTILEVEL CODES: THEORETICAL CONCEPTS AND PRACTICAL DESIGN RULES, **IEEE TRANS INFORM. THEORY**, VOL. 45, NO. 5, PP. **1361-1391**, JUL. **1999**.	"MULTILEVEL CODES: THEORETICAL CONCEPTS AND PRACTICAL DESIGN RULES"	WACHSMANN U, FISCHER RFH, HUBER JB	**IEEE TRANSACTIONS ON INFORMATION THEORY** 45 (5): 1361-1391 JUL 1999	Several names listed; variation in journal name
DESIGN CHARACTERISTICS OF GAS JET GENERATORS, **BORISOV, 1979**, PP. 21 25.	"DESIGN CHARACTERISTICS OF GAS-JET GENERATORS"	BORISOV YY	SOVIET PHYSICS ACOUSTICS-USSR 26 (1): 21-25 **1980**	Typo in year; diff in location of title within the citation
KERNS, SHERRA E., THE DESIGN OF RADIATION HARDENED ICS FOR SPACE: A COMPENDIUM OF APPROACHES, PROCEEDINGS OF THE IEEE, NOV. 1988, PP. 1470 1509.	"THE DESIGN OF RADIATION-HARDENED ICS FOR SPACE - A COMPENDIUM OF APPROACHES"	**KERNS SE**, SHAFER BD, ROCKETT LR, PRIDMORE JS, BERNDT DF, VANVONNO N, BARBER FE	PROCEEDINGS OF THE IEEE 76 (11): 1470-1509 NOV 1988	Several authors w/o "et al."
GENESTIER ET AL (BLOOD, 1997, VOL. 90, PP. 3629-3639).	"FAS-INDEPENDENT APOPTOSIS OF ACTIVATED T CELLS INDUCED BY ANTIBODIES TO THE HLA CLASS I ALPHA 1 DOMAIN"	GENESTIER L, PAILLOT R, BONNEFOYBERARD N, MEFFRE G, FLACHER M, FEVRE D, LIU YJ, LEBOUTEILLER P, WALDMANN H, ENGELHARD VH, BANCHEREAU J, REVILLARD JP	BLOOD 90 (9): 3629-3639 NOV 1 1997	No title within citation- however, perfect match in all other features
STEPHEN M. BEBGE, LYLE D. BIGHLEY AND DONALD C. MONKHOUSE PHARMACEUTICAL SALTS JOURNAL OF PHARMACEUTICAL SCIENCES, 1977, 66, 1-19.	"PHARMACEUTICAL SALTS"	**BERGE SM, BIGHLEY LD, MONKHOUSE DC**	JOURNAL OF PHARMACEUTICAL SCIENCES 66 (1): 1-19 1977	Several names listed; variation of names
L. YOUNG AND D. SHEENA, **METHODS & DESIGNS:** SURVEY OF EYE MOVEMENT RECORDING METHODS, BEHAV. RES. METHODS INSTRUM., VOL. 5, PP. 397-429, 1975.	"SURVEY OF EYE-MOVEMENT RECORDING METHODS"	YOUNG LR, SHEENA D	BEHAVIOR RESEARCH METHODS &INSTRUMENTATION 7 (5): 397-429 1975	diff in title
MICROWAVE JOURNAL, VOL. 22, NO. 2, FEB. 1979, DEDAHAM US PP. 51 52, **H. C. CHAPPELL.**	"DESIGNING IMPEDANCE MATCHED IN-PHASE POWER DIVIDERS"	CHAPPELL HC	MICROWAVE JOURNAL 22 (2): 51-52 1979	no title - however, perfect match in all other features; diff position of author's name within citation

Figure 7. External and Internal citation, matching process

(i) Example of an external citation to IBM's publication : the patent owner and cited corporate publication are different



(12) **United States Patent**
Liu et al.

(10) **Patent No.: US 6,468,596 B1**
(45) **Date of Patent: Oct. 22, 2002**

(54) **LASER-ASSISTED IN-SITU FRACTIONATED LUBRICANT AND A NEW PROCESS FOR SURFACE OF MAGNETIC RECORDING MEDIA**

(75) Inventors: **Youming Liu**, Palo Alto; **Jialuo Jack Xuan**, Milpitas; **Xiaohua Shel Yang**, Fremont; **Chung-Yuang Shih**, Cupertino; **Vidya K. Gubbi**, Milpitas, all of CA (US)

(73) Assignee: **Seagate Technology LLC**, Scotts Valley, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/577,674**

(22) Filed: **May 25, 2000**

Related U.S. Application Data

(60) Provisional application No. 60/144,357, filed on Jul. 15, 1999.

(51) **Int. Cl.**[7] **C08F 2/48**; C08J 7/18; C23C 14/30

(52) **U.S. Cl.** **427/508**; 427/554; 427/596

(58) **Field of Search** 427/510, 554, 427/555, 556, 597, 127, 226, 258, 261, 264, 270, 271, 402, 508

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,674,340 A	7/1972	Jacob et al.	350/157
3,764,218 A	10/1973	Schedewie	356/118
3,938,878 A	2/1976	Fox	350/150

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

OTHER PUBLICATIONS

P. Baumgart et al., "A New Laser Texturing Technique For High Performance Magnetic Disk Drives" IBM storage Systems Division and IBM Almadon Research Center, San Jose, CA.

D. Kuo et al., "Laser Zone Texturing on Glass and Glass–Ceramic Substrates" Seagate Recording Media, Fremont, CA.

P. Baumgart et al., "Safe Landings: Laser Texturing of High–Density Magnetic Disks" IBM Corp., *Data Storage* 1996.

A. Tam et al., "Laser Cleaning Techniques for Removal of Surface Particulates" IBM Research Division, San Jose, *Journal of Applied Physics* 71 (7), Apr. 1, 1992, pp. 3515–3523.

K. Johnson et al., "In–Plane Anisotropy in Thin–Film Physical Origins of Orientation Ratio (Invited)" IBM Storage Systems Division, San Jose, CA, *IEEE Transactions on Magnetics* vol. 31, No. 6, Nov. 1995, pp. 2721–2727.

J. Miles et al., "Micromagnetic Simulation of Textured Induced Orientation in Thin Film Media" the University of Manchester, Manchester, M13 9PL, U.K., *IEEE Transactions on Magnetics* vol. 31, No. 6, Nov. 1995, pp. 2770–2772.

C. Kissinger et al., "Fiber Optic Probe Measures Runout of Stacked Disks" B.W. Brennan Associates, *Data Storage* Jul./Aug. 1997.

Primary Examiner—Shrive P. Beck
Assistant Examiner—Eric B. Fuller
(74) *Attorney, Agent, or Firm*—McDermott, Will & Emery

(57) **ABSTRACT**

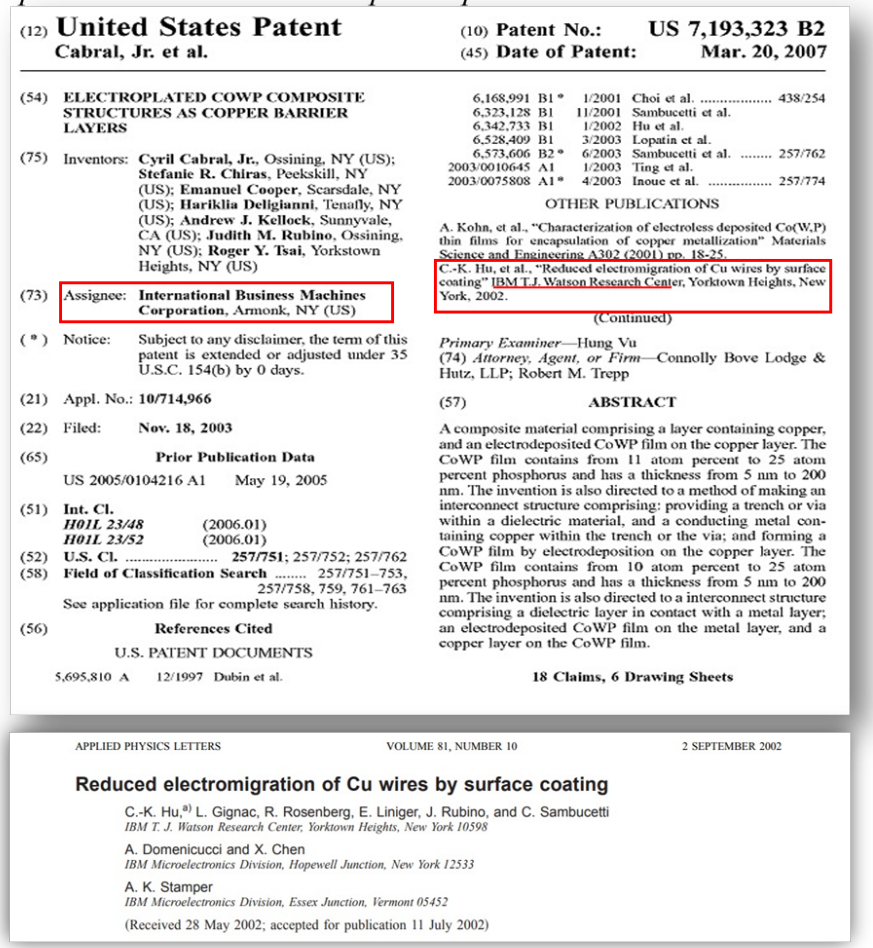
A magnetic recording medium is formed with enhanced tribological performance by applying a raw, unfractionated lubricant having a wide molecular weight distribution over a disk surface and treating the deposited lubricant with a laser light beam to effect in-situ fractionation of the lubricant to a very narrow molecular weight distribution. Embodiments of the present invention also include laser treating a deposited lubricant to increase the thickness of the bonded lube layer.

IEEE TRANSACTIONS ON MAGNETICS, VOL. 31, NO. 6, NOVEMBER 1995 2721

In-Plane Anisotropy in Thin-Film Media: Physical Origins of Orientation Ratio (Invited)

Kenneth E. Johnson, Mohammad Mirzamaani, and Mary F. Doerner
IBM Storage Systems Division, San Jose, CA 95193

(ii) Example of an internal citation to IBM's publication : the patent owner and cited corporate publication are the same



(12) **United States Patent**
Cabral, Jr. et al.

(10) **Patent No.: US 7,193,323 B2**
(45) **Date of Patent: Mar. 20, 2007**

(54) **ELECTROPLATED COWP COMPOSITE STRUCTURES AS COPPER BARRIER LAYERS**

(75) Inventors: **Cyril Cabral, Jr.**, Ossining, NY (US); **Stefanie R. Chiras**, Peekskill, NY (US); **Emanuel Cooper**, Scarsdale, NY (US); **Hariklia Deligianni**, Tenafly, NY (US); **Andrew J. Kellock**, Sunnyvale, CA (US); **Judith M. Rubino**, Ossining, NY (US); **Roger Y. Tsai**, Yorktown Heights, NY (US)

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/714,966**

(22) Filed: **Nov. 18, 2003**

(65) **Prior Publication Data**

US 2005/0104216 A1 May 19, 2005

(51) **Int. Cl.**
H01L 23/48 (2006.01)
H01L 23/52 (2006.01)

(52) **U.S. Cl.** **257/751**; 257/752; 257/762

(58) **Field of Classification Search** 257/751–753, 257/758, 759, 761–763
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,695,810 A	12/1997	Dubin et al.	
6,168,991 B1 *	1/2001	Choi et al.	438/254
6,323,128 B1	11/2001	Sambucetti et al.	
6,342,733 B1	1/2002	Hu et al.	
6,528,409 B1	3/2003	Lopatin et al.	
6,573,606 B2 *	6/2003	Sambucetti et al.	257/762
2003/0010645 A1	1/2003	Ting et al.	
2003/0075808 A1 *	4/2003	Inoue et al.	257/774

OTHER PUBLICATIONS

A. Kohn, et al., "Characterization of electroless deposited Co(W,P) thin films for encapsulation of copper metallization" Materials Science and Engineering A302 (2001) pp. 18-25.

C.-K. Hu, et al., "Reduced electromigration of Cu wires by surface coating" IBM T.J. Watson Research Center, Yorktown Heights, New York, 2002.

(Continued)

Primary Examiner—Hung Vu
(74) *Attorney, Agent, or Firm*—Connolly Bove Lodge & Hutz, LLP; Robert M. Trepp

(57) **ABSTRACT**

A composite material comprising a layer containing copper, and an electrodeposited CoWP film on the copper layer. The CoWP film contains from 11 atom percent to 25 atom percent phosphorus and has a thickness from 5 nm to 200 nm. The invention is also directed to a method of making an interconnect structure comprising: providing a trench or via within a dielectric material, and a conducting metal containing copper within the trench or the via; and forming a CoWP film by electrodeposition on the copper layer. The CoWP film contains from 10 atom percent to 25 atom percent phosphorus and has a thickness from 5 nm to 200 nm. The invention is also directed to a interconnect structure comprising a dielectric layer in contact with a metal layer; an electrodeposited CoWP film on the metal layer, and a copper layer on the CoWP film.

18 Claims, 6 Drawing Sheets

APPLIED PHYSICS LETTERS VOLUME 81, NUMBER 10 2 SEPTEMBER 2002

Reduced electromigration of Cu wires by surface coating

C.-K. Hu,[a)] L. Gignac, R. Rosenberg, E. Liniger, J. Rubino, and C. Sambucetti
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598

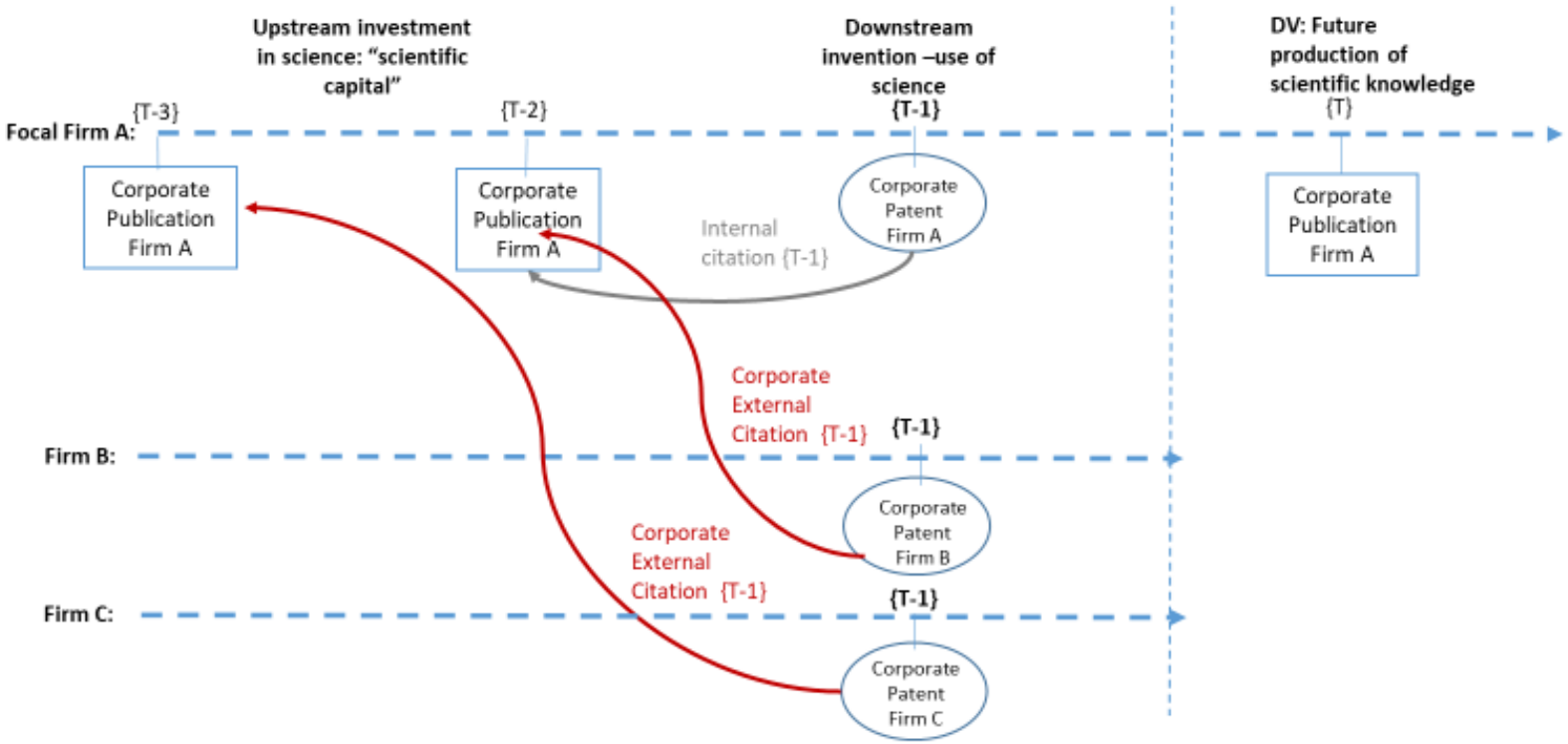
A. Domenicucci and X. Chen
IBM Microelectronics Division, Hopewell Junction, New York 12533

A. K. Stamper
IBM Microelectronics Division, Essex Junction, Vermont 05452

(Received 28 May 2002; accepted for publication 11 July 2002)

Note: this figure presents examples of front-page patent reference to non-patent literature. Below each patent reference is the related scientific publication that is being cited. Example (i) is an external patent citation to IBM's publication and example (ii) is an internal patent citation to IBM's publication.

Figure 8. Timeline- Production and Use of Research



Note: this figure illustrates the temporal structure of citations and publications. At time T-1 the focal firm (Firm A) has: (i) one Internal citation, and (ii) two corporate external citations from patents filed by sample Compustat firms (Firm B and Firm C).