Scientometrics (2015) 102:1965–1985
DOI 10.1007/s11192-014-1493-2

Technology structural implications from the extension of a patent search method

Christopher L. Benson · Christopher L. Magee

Received: 14 July 2014 / Published online: 2 December 2014
© Akadémiai Kiadó, Budapest, Hungary 2014

Abstract Many areas of academic and industrial work make use of the notion of a 'technology'. This paper attempts to reduce the ambiguity around the definition of what constitutes a 'technology' by extension of a method described previously that finds highly relevant patent sets for specified technological fields. The method relies on a less ambiguous definition that includes both a functional component and a component consisting of the underlying knowledge in a technological field to form a two-component definition. These two components form a useful definition of a technology that allows for objective, repeatable and thus comparable analysis of specific technologies. 28 technological domains are investigated: the extension of an earlier technique is shown to be capable of finding highly relevant and complete patent sets for each of the technologies. Overall, about 500,000 patents from 1976 to 2012 are classified into these 28 domains. The patents in each of these sets are not only highly relevant to the domain of interest but there are relatively low numbers of patents classified into any two of these domains (total patents classified in two domains are 2.9 % of the total patents and the great majority of patent class pairs have zero overlap with a few of the 378 patent class pairs containing the bulk of the doubly listed patents). On the other hand, the patents within a given domain cite patents in other domains about 90 % of the time. These results suggest that technology can be usefully decomposed to distinct units but that the inventions in these relatively tightly contained units depend upon widely spread additional knowledge.

Electronic supplementary material The online version of this article (doi:10.1007/s11192-014-1493-2) contains supplementary material, which is available to authorized users.

C. L. Benson (✉)
Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: cbenson@mit.edu

C. L. Magee
Engineering Systems Division and Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: cmagee@mit.edu

Keywords Patent searching · Technology decomposition · Technological relatedness

Mathematical Subject Classification (2000) 68P20

JEL Classification O31 · O32 · O34

Introduction

There are many applications in academic and industrial settings where studying a specific 'technology' is useful (Pavitt 1984) and thus differentiating among specific descriptions of a "technology" is an important research agenda. A very important goal of some of this work is to more closely link technologies and industries in order to more clearly understand the economic impact of various technologies (Evenson and Putnam 1988; Verspagen et al. 1994). Indeed, an extensive program (Schmoch et al. 2003; Schmoch 2008) has developed a preliminary "concordance" between the industrial classes as economically viable and the International Patent Classes (IPC) as representative of technologies. The concordance has been empirically developed based upon patenting activity of firms in various industries. However, the concept of "a technology" and "an industry" are often if not usually conflated whereas technologies as technically understood cut across industries as shown in great detail by Rosenberg (1979). Cockburn and Griliches (1987) describe their attempts to categorize industries for their study on patent valuations in the following words

> An industry in this sense is quite clearly defined at the conceptual level, but (as usual) is difficult to define in practice

In any event, the issue in this paper is *not* linking technologies to specific industries. While recognizing the great potential value of a more direct linkage of technological and economic data, this work is aimed at understanding non-economic relationships among technologies; such relationships are also of some interest. For example, a potential long-term trend towards convergence of technologies (Luan et al. (2013) requires an objective means for defining distinct technologies that are seen to merge over time. Importantly, the emergence of new technologies also requires an objective understanding of interactions among technologies over time. As a second example, there has been much recent interest in how specific renewable energy technologies have been adopted in comparison to one another (Jacobsson and Johnson 2000; Neij 1997). Other studies have looked at variation of R&D spending across several technologies (Levin 1988). Others are concerned about how the rate of technical improvement changes for a specific technology (Benson and Magee 2014; Benson 2014). Business leaders are often searching for specific 'technologies' for investment or how they relate to a competitive analysis (Utterback and Acee 2005; Bower and Christensen 1995). We refer to these and other types of analysis of technologies as the field of 'technological research.' At this point in time, the patent classification systems in use (the US Patent Classification System, UPC, the European Patent Classification Systems, ECLA and the International Patent Classification System, IPC) do not adequately allow such comparisons.[1] Choi and Hwang (2013) describe the

[1] The Cooperative Patent Classification (CPC) system currently under the joint development of the European and US Patent offices (see http://www.cooperativepatentclassification.org/obj.html) will be interesting to understand and test when fully available and used by both patent offices.

need for an unambiguous and less time-consuming method for selecting of a set of inventions that describe a particular technological field.

> Regarding the limitations of this research, all the patents within the fields of interest could not be collected due to the ambiguous boundaries between technical fields. Also, most technical fields, not only those of LED and wireless broadband, have a vast amount of patents, taking a great deal of time and manpower to extract and refine processes of patent data. In this research, the target field was therefore narrowed down for analysis thanks to consultation with experts. (Choi and Hwang 2013)

These examples use the term ‘technology’ in many different ways with varying levels of specificity. Technological cross-analysis requires an objective and consistent definition of what constitutes a technology. Nonetheless, flexibility in the definition is necessary because (1) some “technologies” can be sub-categories of other “technologies” and (2) the wide range of purposes for studies where the unit of analysis is a ‘technology’. Thus, an ideal taxonomic structure for technology should place emphasis on specificity, repeatability and flexibility across many different types of analyses. One auspicious starting point for such a structure is to utilize a two-part definition for specific technologies.

Many widely used taxonomic structures include definitions that consist of two components: form and function in a product, form and structure in a piece of literature or a society (from an ethnographers point of view), form and content in art, or prescriptive and descriptive grammar. In all of these definitions, one component takes a top-down ‘functional’ view of the subject, and the other component consists of a bottom-up ‘compositional’ approach. The following section will describe a number of previous attempts at defining technology from both top-down and bottom-up approaches.

Background

Functional definition of a technology (top-down)

One of the least repeatable and generalizable aspects of technological research is the selection of the unit of analysis, a problem that was explored in Magee et al. (2014). Many different units of analysis are used in technological research and are shown on a continuum in Fig. 1. Some studies have examined specific inventions at specific times, such as Nelson’s (1962) or Riordan and Hoddeson’s (1997) study of the invention of the transistor. Similarly, Tushman and Anderson’s (1986) list of technological discontinuities or Girifalco’s (1991) list of innovations since the 18th century attempt to focus on singular

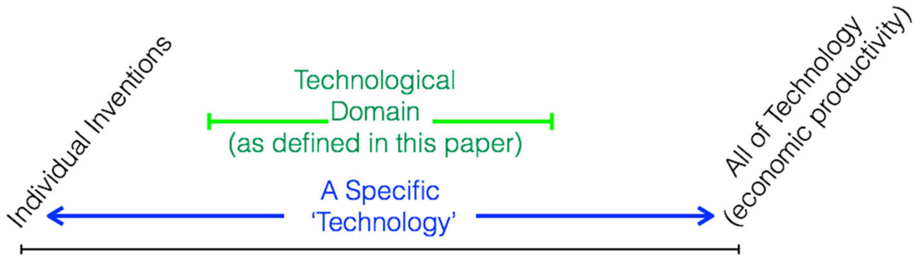


Fig. 1 Range of technological unit of analysis in technological change research and a technological domain as used in this paper

inventions. Others, such as Solow (1956), have studied technology as a single integrated unit, in an attempt to explain economic growth that is not caused by additional labor or capital. More commonly, researchers attempt to study specific technological fields (called technological research in the introduction). Studying technologies at this intermediary level mitigates the subjectivity and lack of breadth inherent in selecting individual inventions, while allowing for greater specificity and deeper analysis than when studying all of technology at once.

There is still much ambiguity in the intermediate unit of analysis "technological field." Arthur (2007) posited that any technology has two main elements. The first element is that any technology is 'a means to fulfill a human purpose.' Examples of purposes he notes include 'to power an aircraft', or 'to sequence a DNA sample,' or to 'generate electricity.' Arthur's second element of technology is that it must take advantage of a particular effect or phenomenon. This effect could be something like the conversion of light to electrons through the photoelectric effect, or the mathematical principles that govern radio waves; the effects do not necessarily need to be physical, they can be scientifically, mathematically, or even socially based. Thus Arthur's definition of a technology is:

> a technology is a means to fulfill a purpose, and it does this by exploiting some effect. (Arthur 2007)

Earlier, Dosi (1982) presented a similar definition that incorporates the different embodiments of knowledge that are represented by a technology.

> Let us define technology as a set of pieces of knowledge both directly 'practical' (related to concrete problems and devices) and 'theoretical' (but practically applicable although no necessarily already applied) know-how, methods, procedures, experience of successes and failures and also of course physical devices and equipment. (Dosi 1982).

Dosi's definition of technology includes the practical knowledge that is related to the domain which is often embodied as patents, theoretical knowledge that is associated, but not necessarily used yet which can be things such as scientific articles and finally the specific artifacts that represent the technology which are often the end products or enabling tools used to make the products.

In Magee et al. (2014), many of the underlying concepts behind Dosi and Arthur's definitions are maintained, while the definition of technology is further modified to move closer to the goal of a specific, repeatable and flexible denotation. First, due to the significant and different uses of the term 'technology', the term used in both that and this paper is *Technological Domain (TD)*, which provides clear differentiation from the other uses of the term 'technology'.

> A *technological domain* can be defined as: The set of artifacts that fulfill a specific generic function utilizing a particular, recognizable body of knowledge. (Magee et al. 2014)

This definition is more specific in terms of the set of artifacts (which includes systems, processes and algorithms as well as devices) than Arthur's use of the term 'means.' Additionally, the term purpose is less ambiguous when it is described as a specific generic function. The precision in this term provides more clarity about the relationship between a domain and their performance characteristics and links the technological domain to its economic purpose. Finally, the term 'some effect' has been replaced by 'a particular, recognizable body of knowledge,' in an attempt to more closely link the technological

Table 1 List of technological discontinuities for three fields—adapted from Tushman and Anderson (1986)

Industry	Year	Event	Importance	Type of discontinuity	Locus of innovation		Probability
					New firms	Existing firms	
Cement	1872	First production of portland cement in the United states	Discovery of proper raw materials and importation of knowledge opens new industry	Niche opening	10 of 10	1 of 10	
	1896	Patent for process burning powdered coal as fuel	Permits economical use of efficient rotary kilns	Competence-destroying	4 of 5	1 of 5	0.333
	1909	Edison patents long kiln (150 ft.)	Higher output with less cost	Competence-enhancing	1 of 6	5 of 6	0.001·
	1966	Dundee cement installs huge kiln, far larger than any previous	Use of process control permits operation of very efficient kilns	Competence-enhancing	1 of 8	7 of 8	0.000·
Airlines	1924	First airline	Mail contracts make transport feasible	Niche opening	9 o 10	1 of 10	
	1936	DC3 airplane	First large and fast enough to carry passengers economically	Competence-enhancing	0 of 4	4 of 4	0.005·
	1959	First jet airplane in commercial use	Speed changed economics of flying	Competence-enhancing	0 of 4	4 of 4	0.005·
	1969	Widebody jets debut	Much greater capacity and efficiency	Competence-enhancing	0 of 4	4 of 4	0.005·
Minicomputer manufacture	1956	Burroughs E-101	First computer under $50,000	Niche opening	1 of 8	7 of 8	
	1965	Digital equipment corp. PDP-8	First integrated-circuit minicomputer	Competence-destroying	3 of 6	3 of 6	0.019·
	1971	Data general supernova SC	Semiconductor memory much faster than core	Competence-enhancing	0 of 7	7 of 7	0.533

· The dot denotes $p < 0.01$

domain with the underlying knowledge that it is based upon and reduced uncertainty about unknown effects that are not yet considered 'knowledge' that may crosscut several technological domains.

It is also important to note the areas in which this definition is intentionally non-specific. The two terms to take notice of are the 'set of artifacts…' and '…a recognizable body of knowledge.' These two terms allow for a technological domain to be as broad as 'semiconductors' or as narrow as 'industrial stereolithography 3D printers'. The fact that this definition does not require a certain level of specificity makes it more flexible and able to represent a large set of potential technological domains. Another benefit of this flexibility is that it is likely impossible to create a specific set of technological domains that uniquely map the entire space of technology, and technological change is strongly dynamic so that one time-invariant best structure is not a practical or worthwhile goal. This flexible definition of a technological domain allows for the scale and scope of a domain to be adapted to the goals of the specified research. The range of the technological domain as defined in this paper is shown in schematically Fig. 1.

Composition (bottom-up) definition of technology: locating a set of patents that represents a technological domain

Difficulties in creating accurate and complete lists of inventions

One of the main strategies used by many technological change researchers is to explain differences in technologies by analyzing the underlying inventions that make up each 'technology'. Excellent examples that do such analyses as a function of time include Gilfillan (1935), Hunter (1949), Passer (1953), Fogel (1964) and Rosenberg (1982). All of these studies note that their listing is incomplete because of small incremental improvements too numerous to fully determine. On the other hand, some researchers attempt to categorize only the important specific technical improvements in a technological field, for example Tushman and Anderson's (1986) paper on technological discontinuities. They claim to demonstrate that 'technology evolves through periods of incremental change punctuated by technological breakthroughs.' In this and many other cases of invention categorization, there is both a lesser and a greater classification relating to the 'breakthroughs' and the 'incremental' inventions—with most examples attempting to focus heavily on the greater classification (i.e. only listing breakthrough inventions within a field).

While the definitions of the greater or lower classification are often given, they are also almost always subjective and open to interpretation. This means that often times the decision of whether an invention is upper or lower class can be different based upon the researcher, which reduces the repeatability of the theories derived from these subjective determinations. For example, in their review of breakthrough inventions, Tushman and Anderson described the process of selecting their innovations as easy, but have very little detail regarding their selection process beyond that.

> Technological discontinuities were relatively easy to identify because a few innovations so markedly advanced the state of the art that they clearly stand out from less dramatic improvements (Tushman and Anderson 1986)

The result of their simple search is Table 1 below that lists the technological discontinuities for three technological fields.

Analyzing Table 1, one perceives a wide variety of artifacts and inventions that are classified as breakthrough, including the first production of commercial cement and the introduction of a longer (150 ft) kiln for producing cement. It is possible that these “breakthroughs” received a significant amount of attention; however, it is certain (Rosenberg 1982) that they were enabled by other inventions that are less well known. This is a significant issue because for every Watt steam engine that gets the majority of the credit, there are a number of Wilkinson boring machine that enabled the engine to have precise and concentric cylinders; for every transistor there is a point rectifier for a radio that demonstrated the initial principle first. The purpose of these examples is to show that while we may remember one specific invention (or even a specific artifact such as the DC-3) as being the most important, it is often one of many inventions (or a combination of many) that together were able to create a new and successful product or product class. Thus, a quantitative and repeatable methodology of relating inventions to a specific technological domain is required for an adequate compositional approach- assessing publicity is not adequate.

Patents as a proxy for inventions

Patent data has been widely used for categorizing inventions into specific technological areas in recent years. Patents are an attractive choice for analyzing technological change because they are: generalizable, objective, quantitative and yet contain extensive qualitative information. Patents include a strong majority of technical fields over a long period of time, and thus allow for easier generalization of the research. Moreover, there are *specific criteria* for an invention to be patented and professional experienced evaluators creating an objective standard as to what counts as an invention. Each patent is well tracked and includes a wealth of quantitative meta-data and qualitatively detailed text allowing for many types of analyses. Of course, it is certain that patents do not contain all developments in a domain due to secrecy, lack of interest in protecting the innovation and other reasons detailed in Mones et al. (2014) and elsewhere. Not all inventions are patented and inventions are not equivalent to all aspects of a technological domain. Our use of patents in this work makes the assumption that the technical characteristics of the field are captured in the patent data so that interactions of fields found in patents would be largely unchanged if one had access to all information about the technological domain. At present, there is no method known for testing this assumption since the non-patented technical changes are largely not documented so it remains a potential limitation of the work. However, even obtaining the relevant patents is challenging (Stefanov and Tait 2011; Alberts et al. 2011).

In selecting a set of patents that represent a technological domain, it is important that the set be complete and contains a high percentage of patents that are relevant to the field of interest. Completeness is the number of relevant patents in that set divided by the total number of relevant patents in the entire United States patent database (a number that can never be known for sure). Similarly, the relevance of a patent set resulting from a search is defined as the number of relevant patents in that set divided by the total number of patents in the same set. A large number of patent-searching techniques were explored and their completeness and relevance evaluated by Benson and Magee (2013). Benson and Magee (2013) also developed a robust, repeatable method initially called the Hybrid-Keyword Classification method for selecting a set of highly relevant and complete patent sets that represent a particular technological domain.

Classification overlap method (COM)

The method developed in Benson and Magee (2013) involves searching for keywords that are selected as potentially important in the technological domain of interest and analyzing the patents in each of the sets retrieved with the keyword analysis by quantitative metrics assessing the *patent classes* of the sets. The patents that are in both the most likely UPC patent class as well as the most likely IPC patent class are then taken as the patents in the domain. The basic intuition behind this classification overlap method is that the US patent examiners (who make all the classification assignments in the US patent system that we utilize) differentially utilize—at least implicitly—the two systems in ways that go beyond the sub-classifications in each system. Thus, additional confirming evidence of the nature of the technology in a patent is obtained by requiring that the patent be in both the top IPC and UPC classes. The fact that each patent is classified in several IPC and UPC classes allows this dual classification to not be over-restrictive thus resulting in reasonably good completeness as well as relevancy. Each possible set is assessed by reading of patents in the potential set by two different technically-knowledgable people who independently judge the relevancy of the patents to the technological domain of interest.[2] 300 Patents are read for each set which results in an overall relevancy assessment for the patent set that is ± 5.7 %.[3] Although COM was empirically supported in the previous work, applying it to 28 cases in the present work (rather than only five as in the initial work) has uncovered some worthwhile enhancements. The enhancements were sometimes necessary because of the reduced scope of some domains in the present work but also were developed semi-empirically and iteratively (using patent relevance testing as feedback) to a greater extent than in the previous work.

With these enhancements, a more general method—renamed the Classification Overlap Method (COM)—has been developed and was elaborated in Benson (2014). The COM is repeatable and can be used by many different types of users, including those who are not well versed in the complexity of the patent system. Figure 2 shows an overview of the COM method with the components that are different from the previous method highlighted. The most direct uses of the COM are identical to that of the previous method and are discussed in depth in Benson and Magee (2013). This paper will emphasize more advanced emendations to the direct COM method and will refrain from repeating the cases that were previously described. The most important of these changes will now be discussed in turn.

Multiple combinations of UPCs and IPCs

The major difference between the previous method and the COM is the increased focus on the overlap of more than two patent classes to select final patent sets. The previous method largely relied upon the overlap between one IPC and one UPC, with an occasional inclusion of another IPC or another UPC. The COM places more emphasis on evaluating the combinations of overlaps between three *or more* classification codes in an attempt to find the most complete and relevant patent sets as is shown in Fig. 3 using two

[2] In the two cases where the two raters differed by more than 7 % in the relevancy rating, a third rater was used and in both cases, a different overlap was used. Thus, in all cases, the relevancy rating given is the average of the two (closely agreeing) raters.

[3] This percentage follows from a standard sampling test for very large data sets that states that the uncertainty range at 95 % confidence is determine by $1/(N)^{1/2}$.

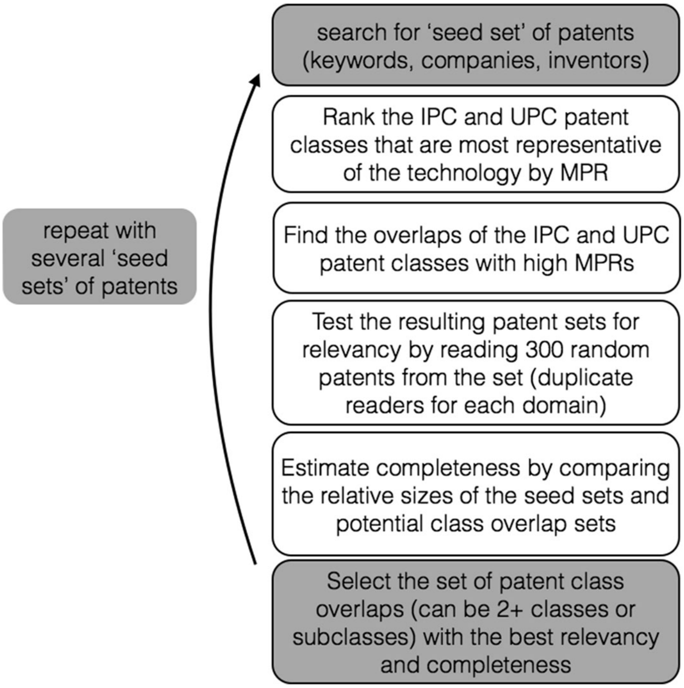


Fig. 2 Process flow of the COM with the largest differences from the previous method (Benson and Magee 2013) highlighted

representative IPC classes and two representative UPC classes. In all cases, analysis of mean precision and recall (MPR), tests of relevancy and analysis of completeness guide selection of the Classes chosen for overlap in this method.

The standard one UPC and one IPC overlap is represented by combining sectors A&B for IPC_1/UPC_1, combining B&C&D for IPC_1/UPC_2 and by combining D&E for IPC_2/UPC_2. When more than one of the IPC or UPCs has a fairly high MPR there can be a 2:1 overlap such as combining B&C&D&E for $IPC_1/IPC_2/UPC_2$ or A&B&C&D for $IPC_1/UPC_1/UPC_2$ although this latter grouping is unlikely to be the final patent set as the addition of UPC1 to the set only adds overlap A, which is relatively small and therefore may not significantly add completeness to the UPC_2/IPC_1 patent set. In other cases, there are technological domains that are best represented by 2 completely separate patent class overlaps, such as IPC_1/UPC_1 and IPC_2/UPC_2: these combinations are represented by A&B and D&E. Finally, there are some situations where relevance testing indicates that an IPC or UPC class is NOT related to the particular TD, in which case it is possible to exclude patent sets in the same way that one would include them in an overlap. For example, one could create a patent set such as IPC_1/UPC_2 NOT UPC_1, which would result in just sectors C&D (rather than A&B&C&D). This Boolean set selection adds further flexibility and specificity to the COM. It is important to note that the overlap of only UPCs or IPCs (i.e. UPC1/UPC2 or IPC1/IPC2) has not shown to result in useful patent sets, as it is the information contained in the two separate classification systems that provides the essence of COM effectiveness (Benson and Magee 2013).

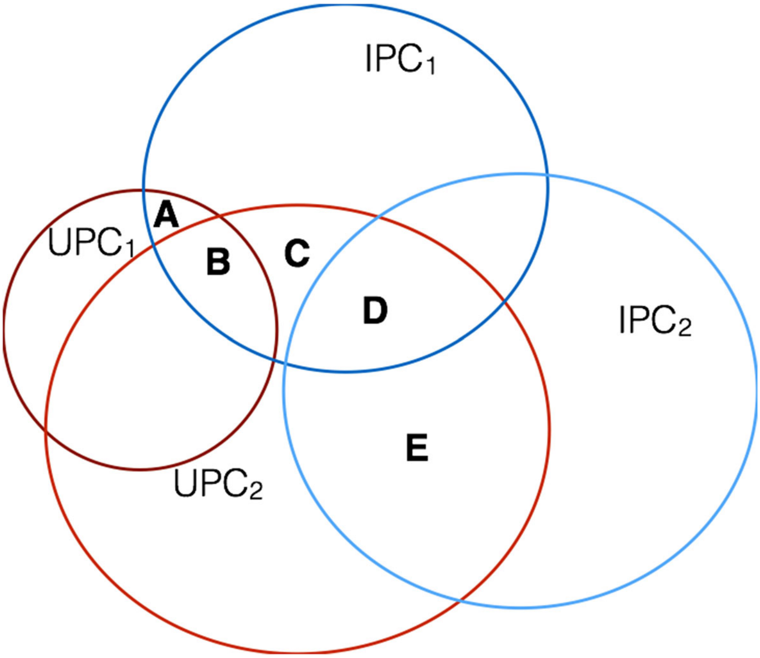


Fig. 3 Different types of overlap types between multiple IPCs and UPCs using the COM with specific sectors labeled

Lower level hierarchy classifications

The method described in Benson and Magee (2013) was designed to work at the primary level of the UPC (the number before the '/') and the 4 digit level of the IPC (ex: H01L). The COM allows for the selection of highly relevant patent sets by overlaps of IPC and UPC classes at lower level hierarchy classifications in each of the patent classification systems. An example of this is 3D printing, where the primary UPC located is 264 (Plastic and nonmetallic article shaping or treating: processes), however the more appropriate patent class for SLA 3D Printing is 264/401 (Stereolithographic Shaping From Liquid Precursor). This same approach can be applied to the IPCs in SLA 3D printing with the primary IPC being B29C (Shaping Or Joining Of Plastics; Shaping Of Substances In A Plastic State, In General; After-Treatment Of The Shaped Products, E.G. Repairing) and the appropriate IPC being B29C35/08 (Heating, cooling or curing, e.g. crosslinking, vulcanising; Apparatus therefor… by wave energy or particle radiation). These lower level hierarchy classifications are overlapped and tested for relevancy and completeness in the same way as described above for the higher level hierarchy classifications to find the appropriate patent sets.

Pre-searching using known company names or inventors

Another enhancement found effective in the current expanded effort is that the COM makes use of 'seed sets' of patents that can be found using more than just a keyword search. Locating a seed set of patents using keywords works very well for most technological domains, however, in some cases searching for the patents that are assigned to companies or particular inventors that are known to operate in a particular technological domain can act as a useful supplement to the initial keyword search. This technique was used in selecting the patents for the Genome Sequencing technological domain, as there were a few well-known organizations that worked on Genome Sequencing (e.g. Affymetrix, Oxford Nanopore Sciences, Sequenom, Illumina, Knome, Broad Institute) and thus helped located the final patent classification codes.

Results

To demonstrate the applicability of the two-component definition of technology and in particular the effectiveness of the COM, 28 technological domains are analyzed in this paper.

Functional definition of 28 domains

Magee et al. (2014) defined the 28 domains within their functional performance categories as is shown in Table 2. The first row of the table is the operand on which the domain acts, and the first column of the table shows the operation that the technological domain performs.

The previous work done on this functional technology classification system shows that the 9 types of classifications represent a relatively complete overview of all possible technologies. The 28 domains analyzed in this paper fall into 8 out the 9 (with matter storage being the exception) possible operand-operation classifications and thus represent a very wide range of technological functions. Magee et al. (2014) describe in further detail the other components of the top-down functional definitions.

Compositional definition of 28 domains

Using the COM it was possible to locate a relatively complete and relevant set of patents for 28 technological domains, which demonstrates the COM to be applicable across a wide variety of different technical areas and hierarchy levels. These new results confirm that US patent examiners are using the two classification systems in distinctly different ways and thus the information of using both UPC and IPC codes (all that are coded on each patent) returns more highly relevant patent sets to the technological domains than is possible by simple use of one or the other classification systems. The present results add substantially to the strength of this conclusion which was first made in Benson and Magee (2013).

Patent sets were found for one half of the 28 domains by the direct COM using the overlap of the UPC and IPC classes with the highest MPR. Patent sets for another 8 domains were located with the COM using the overlap between multiple UPC and IPC classification codes as illustrated in Fig. 3. Finally, 6 of the domains used the COM with lower level patent class hierarchies or keyword modification.

Using the direct COM to define 14 technological domains

Patent sets were found for one half of the 28 domains by the direct COM using the overlap of the UPC and IPC classes with the highest MPR. This result shows the ease of which highly relevant and complete data sets can be located using the COM. All of the patent sets except electrical information transmission that were located using the direct COM had empirical relevancy assessments higher that 80 %. Table 3 shows a summary of the patent sets selected for the 14 TD using the direct COM method.

Although each of these 14 TDs results from the overlap of one IPC and one UPC, the size of the resulting patent sets varies from 1,744 patents (camera sensitivity) to 149,491 patents (integrated circuit processors). Many different seed sets were evaluated for each of the TDs in order to find the most relevant and complete final set. The highly- automated nature of the COM makes it possible to test a large number of seed sets to help ensure that

Table 2 The 28 domains studied in the paper classified by functional technological classifications with operands and operations, adapted from Magee et al. (2014)

	Information	Energy	Matter
Storage	Integrated circuit memory Magnetic memory Optical memory	Batteries Capacitors Flywheel	
Transfer	Coaxial telecom Optical telecom Wireless telecom	Electrical power transmission	Aircraft transport
Transformation	Integrated circuit processors Electronic computation Camera sensitivity MRI CT scan Genome sequencing	Combustion engines Electrical motors Solar PV Wind turbines Fuel cell Incandescent lighting LED lighting	Milling machine 3D printing Photolithography Superconductivity

Table 3 Patent sets for the 14 domains that were found using the direct COM including the UPC and IPC classes used in the overlap

TD	Size	Relevancy (%)	Patent class overlap
Camera sensitivity	1,744	86	257 AND H04N
Capacitor energy storage	5,944	84	361 AND H01G
Electric motors	17,869	86	310 AND H02K
Electrical energy transmission	10,375	86	363 AND H02M
Electrical information transmission	44,910	67	439 AND H01R
Electronic computation	13,204	97	712 AND G06F
Integrated circuit information storage	49,018	81	365 AND G11C
Integrated circuit processors	149,491	81	257 AND H01L
LED artificial illumination	3,792	85	313 AND H01L
Magnetic information storage	33,576	93	360 AND G11B
Milling machines	2,315	93	409 AND B23C
Optical information storage	23,543	82	369 AND G11B
Solar photovoltaic energy generation	5,203	85	136 AND H01L
Superconductivity	1,776	85	505 AND H01L

minimal relevant patents are missed. For example, when searching for patents in the 'Electric Motor' TD, 20 different keywords were used to populate seed patent sets as shown in Table 4.

After the IPCs and UPCs from each seed set are located, several of the overlaps are tested based upon the MPR variables. In Table 4, classes H02K appears to have the dominant MPR independent of search seeds but UPC classes 290, 318 and 310 all appear potentially viable. Thus crosses of each of these UPC classes with IPC class H02K are tested with the relevancy results shown in Table 5.

Table 4 Seed patent sets used to located final patent set for 'Electric Motors' technological domain along with the number of patents in the seed set and the corresponding UPC and IPC with the highest MPR

Search term	Size of seed patent set	IPC	MPR for IPC	UPC	MPR for UPC
Electric motor	37,459	H02K[a]	0.15	310[b]	0.12
Stator	20,019	H02K[2]	0.37	310[3]	0.322
Rotor	44,367	H02K[2]	0.26	310[3]	0.2
Electric machine	14,098	B23H[c]	0.2	310[3]	0.14
Generator	591,838	G06F[d]	0.17	365[e]	0.1
Electric generator	62,238	H02P[f]	0.075	290 [g]	0.16
Winding currents	10,188	H02P[7]	0.14	318 [h]	0.13
Brushless motor	2,137	H02K[2]	0.244	318[9]	0.294
Electromagnetic coil	7,087	H01F[i]	0.07	335[j]	0.12
Electric primary mover	25	H02P[7]	0.16	290[8]	0.123
Motor	152,382	H02P[7]	0.296	318[9]	0.28
Rotary motor	8,163	H02K[2]	0.06	310[3]	0.06
Electric windings	10,795	H02K[2]	0.178	310[3]	0.153
Mechanical commutator	319	H02K[2]	0.196	310[3]	0.189
Electric commutator	1,677	H02K[2]	0.25	310[3]	0.26
Squirrel cage motor	238	H02K[2]	0.23	310[3]	0.236
Wound rotor	1605	H02K[2]	0.3456	310[3]	0.34
Permanent magnet motor	3,688	H02K[2]	0.333	310[3]	0.312
Brushless AC	115	H02P[7]	0.236	318[9]	0.222
Induction motor	3,126	H02P[7]	0.232	318[9]	0.272

[a] (Dynamo-Electric Machines)

[b] (Electrical generator or motor structure

[c] (Working of metal by the action of a high concentration of electric current on a workpiece using an electrode which takes the place of a tool; such working combined with other forms of working of metal)

[d] (Electric digital data processing)

[e] (Static information storage and retrieval)

[f] (Control or regulation of electric motors, generators, or dynamo-electric converters; controlling transformers, reactors or choke coils)

[g] (Prime-mover dynamo plants

[h] (Electricity: motive power systems

[i] (Magnets; inductances; transformers; selection of materials for their magnetic properties)

[j] (Electricity: magnetically operated switches, magnets, and electromagnets

Table 5 IPC and UPC overlaps along with patent set size and relevancy ratios

Patent class overlap set	Number of patents	Relevancy (%)
290^8 AND H02K^2	768	16
318^9 AND H02K^2	2,754	55
310^3 AND H02K^2	18,575	85

Table 6 Resulting patent set overlaps for the 'Combustion Engine' technological domain demonstrating the use of three patent classification codes in the overlap

Patent class overlap set	Number of patents	Relevancy (%)
123[a] AND F02B[b]	13,431	95
123 AND F01L[c]	6,719	98
123 AND (F01L OR F02B)	19,640	96

[a] (Internal Combustion Engines)

[b] (Internal-combustion piston engines; combustion engines in general)

[c] (Cyclically operating valves for machines or engines)

Table 5 shows that the 310/H02K overlap is the most preferable patent set because it has the largest number of patents and a much higher empirical relevancy ratio. This process was repeated for each of the 28 TDs with details shown in Benson (2014).

Multiple UPC or IPC classes used in the COM overlap for 8 technological domains

As was illustrated in Fig. 3, the COM can be adapted to use the overlap of more than two patent classifications as long as there is at least one UPC and one IPC (i.e. the overlap between 3 UPCs obviates the essential power of the COM)). For example, after analyzing 21 seed patent sets for the combustion engine TD, the two IPC/UPC overlaps in the first two lines of Table 6 (123/F02B and 123/F01L) were both found to have very high relevancy and a relatively large number of patents. The third line of Table 6 shows that when combined they make an even larger patent set still with a very high relevancy ratio. Additionally there is very little overlap between the two patent sets, as is shown by the small discrepancy between the combined set ($n = 19{,}640$) and the addition of each of the sizes of the individual sets ($13{,}431 + 6{,}719 = 20{,}150$). The large total patent set size and the high relevancy indicates that the combined patent set 123/F02B/F01L is the most representative patent set for the combustion engine TD.

Out of the 28 TDs, 8 of the patent sets were located by using the overlaps of 3 or more classifications. The patent sets found using 3 or more classification and the COM are given in Table 7. Note that the first five of these are relatively simple combinations of three classes but the last three are more complex with illustration in Fig. 3 and details in Benson (2014).

Further COM modifications

While many of the TDs were relatively easy to find using the COM, there were a few that required deeper searching and more sophisticated applications of the COM.

Table 7 Summary of patent sets for the 8 patent sets that were found using the COM with overlap of 3 + patent classifications including the classifications used in the overlap

TD	Size	Relevancy (%)	Patent class overlap
Combustion engines	19,094	96	123 AND (F01L OR F02B)
Computed tomography (CT)	6,817	88	378 AND (A61B OR G01 N)
Photolithography	14,975	87	(430 OR 355) AND G03F
Wind turbine energy generation	2,498	94	(416 OR 290) AND F03D
Wireless Information Transmission	39,675	94	455 AND (H01L OR H04B)
Incandescent Artificial Illumination	642	89	(313 AND H01K) AND NOT (H01J1 OR F21 V)
Magnet Resonance Imaging (MRI)	1,778	86	(324 AND A61B) OR (600 AND G01R)
Optical Information Transmission	36,494	82	(398 AND H04B) OR (385 AND G02B)

An example of this is the search for the 'Genome Sequencing' TD. The result of the seed set analysis showed that clearly the US patent class 435[4] was the most related UPC, and that the IPC could be a number of options including C12 N,[5] G01 N.[6] All of the IPCs were tested for relevancy and none of the direct COM overlaps resulted in a highly relevant set. The next step was to look closer into the lower level hierarchy patent classification codes by searching for patents from companies that were known to be working in this space:

(AN:(Affymetrix) OR AN:(Oxford Nanopore Sciences) OR AN:(Sequenom) OR AN:(454 Life Sciences) OR AN:(Illumina) OR AN:(Knome) OR AN:(Complete Genomics) OR AN:(Broad Institute)) AND (abst:(sequencing) OR ttl:(sequencing))

This search results revealed lower level UPCs such as 435/6.11 *(Nucleic acid based assay involving a hybridization step with a nucleic acid probe, involving a single nucleotide polymorphism (SNP), involving pharmacogenetics, involving genotyping, involving haplotyping, or involving detection of DNA methylation gene expression)* or 435/6.12 *(With significant amplification step (e.g., polymerase chain reaction (PCR),* etc.*))*. These more specific UPCs were combined with the international patent class C12Q for the final data set.

((CCL:(435/6.11) OR CCL:(435/6.12)) AND ICL:(C12Q)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

Which resulted in a patent set with 4,861 patents with a 0.74 relevancy ratio. The summary of the 6 TDs in which the COM modifications were used is shown in Table 8.

[4] (Chemistry: molecular biology and microbiology).

[5] (Micro-organisms or enzymes; compositions thereof).

[6] (Investigating or analyzing materials by determining their chemical or physical properties).

Implications for understanding structure of technology

Overlap of the patent sets

The COM is a technology-patent search engine; therefore, the patent sets that are located for each technology are not required to be exclusive of other technological domains (i.e. solar PV patents can also be integrated circuit patents). One of the results of locating these patent sets is the ability to analyze the overlap between the patents. Because each patent can be multiply listed in a number of different UPCs and IPCs, some patents will be present in multiple patent sets in the patents selected to represent the 28 TDs examined for this research. The question we examine is how large the overlaps are.

In order to quantify the overlap between the patents, each patent set was compared with each of the other 27 domains in order to find the overlap ratio between the two patent sets. This ratio is shown in Eq. 1, with Pi and P_j representing all the patents in domains i and j.

$$\frac{\mathrm{count}(P_i \cap P_j)}{\min(P_i, P_j)} \tag{1}$$

Note that Eq. 1 gives an overlap ratio of zero when there are no patents that are present in both sets and an overlap ratio of 1 when all of the patents in the smaller set are contained in the larger set. The ratio is also defined so that the overlap is identical for any two domains; thus, there are 378 possible overlap ratios [$(28^2 - 28)/2$] in our 28 domains. Since patents in our sets on average are classified into 4.61 UPC's and 2.4 IPC's each, it is possible that we similarly have large overlaps and this is what is tested here.

The first result is that 225 of these 378 possible overlaps have zero patents in both sets. Moreover, another 135 have very small overlap ratios (< 0.001)—see Fig. 4. Thus, there is either zero or quite low overlap among the great majority of our cases and some of the apparent overlap may be due to our non-perfect relevancy of classification. For example, there are three patents that are present in both the Electrochemical Battery Energy Storage TD and the Aircraft Transport TD, and there are 16,122 patents in the Batteries TD and 8,629 patents in the Aircraft TD, therefore the overlap of Aircraft with Batteries is 0.0003 as is shown below.

$$\frac{P_{\mathrm{batteries}} \cap P_{\mathrm{aircraft}}}{\min(P_{\mathrm{batteries}}, P_{\mathrm{aircraft}})} = \frac{3}{\min(16122, 8629)} = 0.0003 \tag{2}$$

The three patents are shown in Table S1 of the SI and clearly show that the three patents in question are related to fuel cells but not batteries so this is another case of zero overlap of relevant patents.

Thus, it is clear that the vast majority of possible patent overlaps between different domains is zero or near zero; however there are 7 overlaps between domains that share more than 10 % (> 0.1 in Fig. 4) of their patents between the two domains. Table S2 in the SI shows all seven of these cases and their overlap ratio. In all of these cases, the two domains share a common patent classification code whether it be IPC or UPC. For example, Solar PV (136 and H01L) and Integrated Circuits (257 and H01L) both share the large international patent classification H01L (Semiconductor Devices; Electric Solid State Devices not Otherwise Provided for) therefore the 2,221 patents that are in both technological domains need only to be listed in *three* patent classes: 257,136 and H01L (remembering that the average patent is list in about seven classes).

Table 8 Patent sets for the 6 patent sets that were found using the COM with modifications

TD	Size	Relevancy (%)	Patent class overlap
3D-Printing (industrial stereolithography)	251	93	264/401 AND B29C35/08
Aircraft transport	8,629	79	244 AND (B64D OR B64C) AND NOT ('canopy' OR 'parachute' or 'helicopter')
Electrochemical battery Energy storage	16,122	83	(429 AND H01M) AND NOT 'fuel cell'
Flywheel energy storage	154	70	74/572 AND (F16F15 OR H02K7)
Fuel cell energy production	7,368	97	(429 AND H01M) AND 'fuel cell'
Genome sequencing	3,990	74	(435/6.11 OR 435/6.12) AND C12Q

The classes are also given now, but usually deeper in the patent classification hierarchy

Moreover, the overlaps in listing do in these cases (except perhaps for batteries and fuel cells) represent close technology relatedness. Our highest number of overlapped patents (3,189) is between the magnetic and the optical information storage domains (see all overlap numbers in SI Table S4). These domains (see Table 3) share IPC G11B with magnetic storage patents found from the cross with UPC 360 and optical storage crossed with UPC 369. Table S3 in the SI shows the title and abstracts for a few of these doubly listed patents. These five patents (and others we have examined) show that optical and magnetic storage are not quite mutually exclusive domains and that some of the inventions are clearly opto-magneto storage inventions. In this case, even our "clean top 100" patents[7] share 10 patents (see Table 5 in the SI). Table S5 in the SI shows that only 20 of the clean top 100 are listed in two such lists with 1/2 of them in the clearly converging optical and magnetic memory information storage.

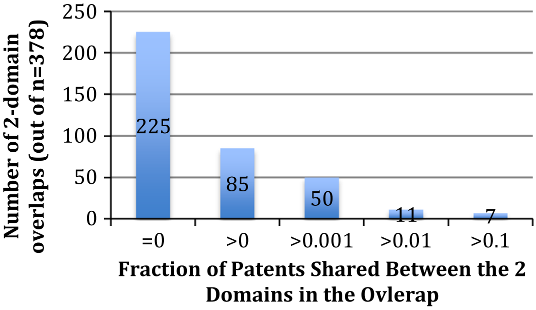
Thus, the patent overlaps among our domains show some reality for convergence between technologies. However, the extensive mutual exclusivity of the patents in these 28 domains indicates that the COM patent searching method can be effectively used as the compositional definition of technology.

Coverage of the US patent database by our patents and their citations

In this paper, 28 domains were identified using the COM, there are certainly many more domains that could be classified using this methodology. The total number of patents (counting the duplicates only once) in all of the TDs studied in this paper is 496,733 and the number of cited patents analyzed was 2,619,355, which can be compared to the 4,666,574 patents that were issued between 1976 and 2013 (uspto.gov 2014). This means that just over 10 % of the total patents have been categorized into TDs, and that the cited patents represent nearly 56 % of the total patents issued. Realistically, the number of technological domains that would comprise nearly all of the patents could be in the range of 300–1,000 TDs based upon the number of patents and domains that were analyzed in this study. Of course, the number of domains would depend upon the scope of choice for

[7] When doing the readings for the relevancy ratings, we read up to 150 of the most cited patents and eliminated those we found (consensus of two readers) were not relevant until we had the relevant, most cited 100 patents in the domain.

Fig. 4 Number of potential domains domain overlaps (out of a possible 378) that share the indicated fraction of patents between two domains



the chosen technological definition (function and knowledge base). The fact that the citations by the patents in our domains comes from a much larger percentage of the total patents (56 %) than these domains ($\sim$ 10 %) is not surprising since we find that the typical set of citations of patents within its own domain is $\sim$ 2–20 % with the average internal citations within a domain being $\sim$ 10 %. Thus, typically patents use knowledge from a much wider part of the technological landscape than their own closely related domain.[8] The very wide knowledge base tapped by these largely mutually exclusive sets of patents is the second aspect of technological structure implied by the results reported here.

Conclusions

The use of a top-down two-component definition for technologies enables a very effective bottom-up compositional definition of a set of 28 technological domains. The two components of the top-down definition are generic functions and particular recognizable bodies of knowledge.

The method used for executing the patent search in the compositional definition is an extended version of a method previously described (Benson and Magee 2013). The extension involves more emphasis upon multiple (more than the basic 2) IPC and UPC class listings to be utilized in the gathering of the final patent set. As in the earlier work, the effectiveness of this method indicates that the US Patent examiners are using the two classification systems differently enough to make the joint groups of patents more aligned with (relevant to) the technological domains defined by our top-down two component approach than patent sets using a singular classification system. It is possible that the new more detailed IPC classification scheme being jointly developed by the US and European Patent office-the Cooperative Patent Classification (CPC) system will have a single structure that works as well but this is unknown at this time (EPO and USPTO 2014). Since classification overlap is the essential element, the extended method is named the Classification Overlap Method (COM). Over a wide range of technological domains, the COM is shown in this paper to yield highly relevant sets of patents where relevance is empirically assessed by reading of patents. The COM is also shown here to give a fairly complete set of patents as assessed by use of multiple seed patent sets and analysis of all of the resulting possible overlaps.

Although the relevance and completeness of the 28 patent sets is a key aspect of evaluating the effectiveness of COM in patent search, technological structural implications

[8] As an example, none of the patents in the camera sensitivity domain are doubly listed in the wireless telecommunication domain but nonetheless, there are 79 citations from patents listed in camera sensitivity to patents in the wireless domain.

arise from further analysis of the patents in the 28 domains. In particular, we find remarkably low overlap among patents in the various domains. We find that more than 80 % of the pairs of potential overlaps in fact have zero (or very near zero) overlap. In the seven domain pairs (out of 378 total possible pairs) where more than 0.01 of the patents are listed in both domains, there is clear evidence of "technology convergence". In these cases, we note the importance of the very large Integrated Circuits domain and find further support for the idea that this technological domain is a "general purpose technology".

While the first structural implication is mutual exclusivity of the patent sets derived from use of the two component top-down definition, the citation distribution is much more widespread with only 10 % of citations by patents in a domain being to other patents in that domain. Assuming that citations represent use of knowledge in the domain, the structure of technology appears to be well-defined domains that nonetheless widely use knowledge from throughout the technological landscape.

Limitations of the current study and further useful work includes continued improvement of the COM and continued use of the method to further explore overall technological structure. Although our method for assessing relevancy (dual readers of all patents with resolution by three participants when rare discrepancies appear) is effective, it is time-consuming and the most "non-automated" and potentially subjective part of the COM. Thus, research to assess relevancy by natural language processing (NLP) as demonstrated by Park et al. (2013) and Moeller and Moehrle (2014) is a very worthwhile avenue to pursue. Such work might not only be able to make further improvements in the COM but also might lead to further technological structural findings.

Our first structural implication is extensive mutual exclusivity of the patent compositional execution for the 28 domains studied here. We also find a few cases where technology convergence in the sense discussed in Luan et al. (2013) is clearly occurring between separately defined domains. The major limitation of the current conclusions is that while extensive, 28 domains are only about 10 % of the total patent set and thus the existence of mutual exclusivity mixed with some convergence cannot be described quantitatively with reliability. The solution to this limitation is much more ($\sim \times 10$) extensive domain definition using the two part top-down approach described here followed by use of the COM to arrive at the compositional definition in terms of patent sets. Analysis of the overlap structure of this wider array of patent sets would do much to clarify current technological structure and could be done as a function of time to explore changes in convergence between domains and divergence or the appearance of new domains over time. Our second structural implication (very broad tapping of knowledge even in mutually-exclusive domains) appears more reliable. Nonetheless, examination of the citation network among a more complete set of domains as defined here would yield much additional knowledge about technological structure.

Acknowledgments The authors would like to thank Subarna Basent and Guillaume Baldo for participating in reading thousands of patents to help determine the relevancy rankings for each of the domains. The support of the SUTD-MIT International Design center for this work is also gratefully acknowledged.

References

Alberts, D., Yang, C. B., Fobare-deponio, D., Koubek, K., Robins, S., Rodgers, M., et al. (2011). In M. Lupu, K. Mayer, J. Tait, & A. J. Trippe (Eds.), *Current challenges in patent information retrieval* (Vol. 29). Berlin: Springer. doi:10.1007/978-3-642-19231-9

Arthur, W. B. (2007). The structure of invention. *Research Policy, 36*(2), 274–287. doi:10.1016/j.respol.2006.11.005.

Benson, C. L. (2014). *Cross-domain comparison of quantitative technology improvement using patent derived characteristics*. New York: MIT.
Benson, C. L., & Magee, C. L. (2013). A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. *Scientometrics, 1*(1), 1–12. doi:10.5539/emr.v1n1p2.
Benson, C., & Magee, C. L. (2014). On improvement rates for renewable energy technologies: Solar PV, wind turbines, capacitors, and batteries. *Renewable energy*. http://www.sciencedirect.com/science/article/pii/S0960148114001451.
Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: Catching the wave. *Harvard Business Review 1*, 43–53.
Choi, J., & Hwang, Y. (2013). Patent keyword network analysis for improving technology development efficiency. *Technological Forecasting and Social Change, 83*, 170–182.
Cockburn, I., & Griliches, Z. (1987). Industry effects and appropriability measures in the stock market's valuation of R&D and patents. *Worker paper no. 2468*. Cambridge, MA: NBER.
Dosi, G. (1982). Technological paradigms and technological trajectories. *Research Policy, 11*(3), 147–162. doi:10.1016/0048-7333(82)90016-6.
EPO & USPTO. (2014). *Cooperative patent classification—Objectives*. http://www.cooperativepatentclassification.org/obj.html.
Evenson, R., & Putnam, J. (1988). *The Yale-Canada patent flow concordance*. Yale University, Economic Growth Centre Working Paper.
Fogel, R. W. (1964). *Railroads ad American economic growth: Essays in econometric history*. Baltimore, MD: John Hopkins Press.
Gilfillan, S. C. (1935). *Inventing the ship*. Chicago: Follett Publishing Company.
Girifalco, L. A. (1991). *Dynamics of technological change*. New York: Van Nostrand Reinhold.
Hunter, L. (1949). *Steamships on the Western Rivers: An economic and technological history*. Cambridge, MA: Harvard University Press.
Jacobsson, S., & Johnson, A. (2000). The diffusion of renewable energy technology: An analytical framework and key issues for research. *Energy policy*, *28*(9), 625–640.
Levin, R. (1988). Appropriability, R&D spending, and technological performance. *American Economic Review, 78*(2), 424–428.
Luan, C., Liu, Z., & Wang, X. (2013). Divergence and convergence: technology-relatedness evolution in solar energy industry. *Scientometrics, 97*(2), 461–475. doi:10.1007/s11192-013-1057-x.
Magee, C. L., Basnet, S., Funk, J. L., & Benson, C. L. (2014). *Quantitative empirical trends in technical performance*. MIT, ESD working paper ESD-WP-2014-22. http://esd.mit.edu/WPS/2014/esd-wp-2014-22.pdf.
Moeller, A., & Moehrle, M. G. (2014). *Scientometrics,*. doi:10.1007/s11192-014-1446-9.
Mones, E., Pollner, P., & Vicsek, T. (2014). Universal hierarchical behavior of citation networks. *Journal of Statistical Mechanics: Theory and Experiment, 2014*(5), 1–17.
Neij, L. (1997). Use of experience curves to analyse the prospects for diffusion and adoption of renewable energy technology. *Energy Policy, 23*(97), 1099–1107.
Nelson, R. (1962). *The link between science and invention: The case of the transistor* (pp. 549–584). The rate and direction of inventive activity: Economic and Social Factors.
Park, H., Kim, K., Choi, S., & Yoon, J. (2013). A patent intelligence system for strategic technology planning. *Expert Systems with Applications, 40*, 2373–2390.
Passer, H. C. (1953). *Electrical manufacturers: 1875–1900: A study in competition, entrepreneurhsip, and economic growth*. Cambridge, MA: Harvard University Press.
Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy, 13*(1984), 343–373.
Riordan, M., & Hoddeson, L. (1997). *Crystal fire: The birth of the information age*. New York: W. W. Norton & Company.
Rosenberg, N. (1979). Technological interdependence in the American economy. *Technology and Culture, 1*, 25–50.
Rosenberg, N. (1982). *Inside the black box: Technology and economics*. Cambridge, UK: Cambridge University Press.
Schmoch, U. (2008). *Concept of a technology classification for country comparison*. Final report to the World Intellectual Property Organization (WIPO). http://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_41/ipc_ce_41_5-annex1.pdf.
Schmoch, U., Laville, F. Patel, P., & Freitsch, R. (2003). *Linking technology areas to industrial sectors*. Final report to the European commission, DG research. ftp://cordis.europa.eu/pub/indicators/docs/ind_report_isi_ost_spru.pdf.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*,*70*(1), 65–94. doi:10.2307/1884513

Stefanov, V., & Tait, J. I. (2011). An introduction to contemporary search technology. In *Current challenges in patent information retrieval* (pp. 45–65). Berlin: Springer.

Tushman, M., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly, 31*(3), 439–465.

USPTO. (2014). *Decription of patent types*. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/patdesc.htm.

Utterback, J. M., & Acee, H. J. (2005). Disruptive technologies: an expanded view. *International Journal of Innovation Management, 9*(1), 1–17. doi:10.1142/S1363919605001162.

Verspagen, B., Morgastel, T. V., & Slabbers, M. (1994). *MERIT concordance table: IPC—ISIC (rev. 2)*. Maastricht: MERIT Research Memorandum 2/94-004.