Economics of Innovation and New Technology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713641545

Using patent technology codes to study technological change

Deborah Strumsky[a]; José Lobo[b]; Sander van der Leeuw[b]
[a] Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA [b] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

First published on: 13 June 2011

Economics of Innovation and New Technology
2011, 1–20, iFirst

Using patent technology codes to study technological change

Deborah Strumsky[a]*, José Lobo[b], and Sander van der Leeuw[b]

[a] *Department of Geography and Earth Sciences, University of North Carolina at Charlotte, McEniry 429, 9201 University City Boulevard, Charlotte, NC 28223, USA;* [b] *School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA*

(*Received 2 January 2011; final version received 4 April 2011*)

Much work on technological change agrees that the recombination of new and existing technological capabilities is one of the principal sources of technological novelty. Patented inventions can be seen as bundles of distinct technologies brought together to accomplish a specific outcome – and this is how the US Patent Office defines inventions. The technologies constituting inventions are identified by the US Patent Office through an elaborate system of technology codes. A combinatorial perspective on invention, emblematic of approaches to technological change informed by evolutionary economics and complexity science, is inherent in the use of technology codes to summarize what is technologically novel about a patented invention. The technology codes represent a set of consistent definitions of technologies and their components spanning 220 years of inventive activity, and are an underutilized data resource for identifying distinct technological capabilities, defining technology spaces, marking the arrival of technological novelty, measuring technological complexity, and empirically grounding the study of technological change. The present discussion provides an introduction to the use of patent technology codes as well as some basic empirics. Our results highlight the highly discriminating nature of the codes and their usefulness in characterizing the type of processes by which technological capabilities generate novelty.

Keywords: patents; technology codes; technological change; technological novelty

JEL Classification: O30; C81; N7

1. Introduction

Much work on technological change agrees that the recombination of new and existing technological capabilities is one of the principal sources of technological novelty (Schumpeter 1939; Kroeber 1948; Usher 1954; Allen 1977; Redman 1979; Nelson and Winter 1982; Basalla 1988; Von Hippel 1988; Henderson and Clark 1990; Kogut and Zander 1992; Tushman and Rosenkopf 1992; Christensen 1997; Hargadon and Sutton 1997; Levinthal 1997; Weitzman 1998; Auerswald et al. 2000; Kauffman, Lobo, and Macready 2000; Fleming 2001; Rivkin 2001; Fleming and Sorenson 2004; Frenken 2006; Arthur 2009). But the combinatorial view of technology, however compelling its metaphors,

*Corresponding author. Email: dstrumsky@uncc.edu

ISSN 1043-8599 print/ISSN 1476-8364 online

DOI: 10.1080/10438599.2011.578709
http://www.informaworld.com

analogies, and models are, runs into very serious empirical difficulties when it tries to identify and discretize technologies. This empirical challenge is encountered by just about any analysis of technology change: by just about any analysis of technology change: it is vastly easier to theorize about technology than it is to measure it (a difficulty on a par with measuring 'knowledge spillovers' (Krugman 1991; Rosenthal and Strange 2004)).

Technological change is often manifested in inventions – new artifacts, devices, processes, or materials (Gilfilan 1935; Higgs 1971; Rosenberg 1982; Hughes 1989; Mokyr 1990; Braudel 1992; Landes 1998; Jorgenson 2001; Khan 2005; Boot 2006). Some inventions, specifically patented inventions, leave behind a detailed evidentiary trail, and consequently patenting activity has become a widely used framework for studying the 'knowledge economy' (Acs and Audretsch 1989; Griliches 1990; Archibugi 1992; Jaffe, Trajtenberg, and Henderson 1993; Acs, Anselin, and Varga 2002; Jaffe and Trajtenberg 2002; Jaffe, Hall, and Trajtenberg 2005; Bettencourt, Lobo, and Strumsky 2007). We are aware of the many concerns about using patents as indicators of generic inventive activity, principally that not all inventions are patented and that many important types of inventions (e.g. organizational forms) cannot even be patented. We nevertheless agree with Lerner and Kortum (2000) and Jaffe, Hall, and Trajtenberg (2005) who see patents as a useful proxy for inventive activity.

The research literature on patenting has focused on using the patented inventions as instantiations of technology (Jaffe, Hall, and Trajtenberg 2005; Youtie, Iacopetta, and Graham 2008). Alternatively, inventions can be seen as distinct technological *capabilities* combined to accomplish a specific outcome (Rosenberg 1982, 1994). The laser presupposes the ability to construct highly reflective optical cavities, create light intensification mediums of sufficient purity, and supply light of specific wavelengths; the polymerase chain reaction results from (among various other capabilities) the ability to finely control thermal cycling (which requires the use of computers), and isolate short DNA fragments and a DNA polymerase (both of which require techniques from chemical engineering). The US Patent and Trademark Office (USPTO) in effect follows this approach and defines inventions as bundles of technological capabilities. The Patent Office identifies the technologies constituting inventions through an elaborate system of technology codes which categorize the technological features of a patented invention. A combinatorial perspective on invention, emblematic of approaches to technological change informed by evolutionary economics and complexity science, is inherent in the USPTO's use of technology codes to summarize what is technologically novel about a patented invention.

What is novel about an invention is described by its inventors, in technical and precise detail, in the patent's *claims*. The USPTO's patent examiner then categorizes the invention by encoding the information contained in the claims, using a system of numerical codes (technology codes). At any given time the existing set of technology codes available to a patent examiner is essentially a description of the current set of technological capabilities. A patented invention, by definition, represents technological novelty; thus with each new patent an examiner must decide which codes to use to capture the technological components in a patent, and whether or not new technology codes are required to capture the invention's novelty. The introduction of a new code sets in motion a retroactive reclassification of all previous patents that may have embodied that technology. The Patent Office's technology codes thus constitute a set of consistent definitions of technological capabilities spanning 220 years of inventive activity.

Researchers have primarily used the patent technology codes to select sets of patents from the patent record and group them into particular taxonomies (Hall, Jaffe, and Trajtenberg 2001; Moser and Nicholas 2004; Marx, Strumsky, and Fleming 2009). We argue here that the technology codes are an underutilized data resource for identifying distinct technological capabilities, defining technology spaces, marking the arrival of technological novelty, measuring technological complexity, and empirically grounding the study of technological change. The present discussion is meant to provide an introduction as to how patent technology codes are assigned and what information they convey robustly and in sufficient detail so that researchers can exploit the empirical and modeling opportunities afforded by the codes. We provide some basic empirics for the codes – highlighting their discriminating nature – and use them to characterize the nature and rate of technological change since the mid-nineteenth century.

We see the technology codes as complementary to that other patent metric that has been put to widespread use tracing relationships among inventions, assessing their usefulness and quantifying their novelty, namely, patent citations (Trajtenberg 1990; Trajtenberg, Henderson, and Jaffe 1997; Hall, Jaffe, and Trajtenberg 2001). An important analytic difference between using citations or technology codes is that citations treat *inventions* as the fundamental unit of analysis for studying technology while the codes refer to the *technological capabilities* constituting the inventions as the fundamental particles. As a marker of quality and usefulness – once the first quality hurdle is cleared by having an application for a patent granted – being cited by other inventions is surely a good sign. And a hallmark of 'General Purpose Technologies', which really refer to inventions with transformative reach such as the steam engine, electricity or the personal computer, is surely that other inventions utilize them (Helpman 1998). But it is important to remember that patent citations first and foremost serve a legal function: delimiting the legal scope of the property rights awarded by a patent.

If the research focus is on the origin of technological novelty embedded in inventions – refinement, recombination or creation de novo? – a finer-grained measure than citations is needed. The absence of references to previous patents does not imply that an invention does not use previously existing technologies. Consider the case of the most cited invention: the polymerase chain reaction (US patent # 4,683,202) which did not cite any 'prior art' and which has been cited 2135 times. ('Prior art' constitutes all information that has been made available to the public in any form before the submission of a patent application that might be relevant to a patent's claims of originality.) While this patent did not utilize prior art, it surely combined existing technological capabilities (Pisano 2006). Given how sparse the citation record is, relying on citations to study the origins of technological novelty is bound to result in a somewhat distorted picture. The distribution of citations is extremely skewed with a mean of 5.4 and a median of 1. Of the 8,493,636 patents successfully applied for between 1790 and September of 2010, almost 40% have no citations.

Furthermore, while measures of a patent's originality and generality inform us about the inventions that engendered them and the inventions they engendered, they provide little information regarding the capabilities that had to be originated or combined to create a new device, process or material. Consider the case of the second most cited patent, that for the bubble jet printer (US patent # 4,723,129): it cites 14 patents and has been cited 1962 times. The inventions cited by the bubble jet printer involve the formation of fine droplets, circuitry design, thermal actuators, and photocopying. As innovative as the bubble jet printer was, all of its constituent technologies (described by the patent's 10 technology codes) existed prior to its invention. Ultimately, patent citations and patent technology codes measure different things and thus make it possible to answer different questions. We expect that much can be learned about technological change by comparing, contrasting and integrating insights gained by using the two types of data.

Our discussion is organized as follows. The next section discusses in some detail how what is novel in a patent is described in the patent's claims, while Section 3 describes how the information contained in the claims is classified by assigning the patent technology codes. Basic empirics for the codes are provided in Section 4 and their use in quantifying technological change is the topic of Section 5. Section 6 concludes with a discussion of research possibilities using the codes.

2. Patent claims

A patent is an intellectual property right granted by the Government of the United States of America to an inventor; it excludes others from making, using, offering for sale, or selling the invention throughout the USA, or importing the invention into the USA, for a limited time (generally 20 years from the filing date) in exchange for public disclosure of the invention when the patent is granted. (This right was established over 200 years ago in Article 1, Section 8 of the United States Constitution: 'To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries'.) Section 101 of the US Patent Law specifies four categories of inventions or discoveries that are eligible for the protection of a patent: processes, machines, manufactures and compositions of matter. The USPTO grants three types of patents: utility patents – also referred to as 'a patents for invention' – are issued for the invention of 'new and useful' processes, machines, artifacts, or compositions of matter; design patents, which are granted for the ornamental design of a functional item; and plant patents which are conferred for new varieties of plants or seeds. Among these patents, 92% granted by the USPTO are utility patents. Though it is the case that most patents have been granted to inventions involving machines or the transformation of one physical article into another, the US Supreme Court has reaffirmed the patentability of business methods, computer programs and mathematical algorithms (561 US Supreme Court, 2010).

By definition, a patent is granted by the USPTO to an invention that is both novel – not obvious to others skilled in the same field – and useful (a recent US Supreme Court Decision, 550 US 398 (2007), considerably lifted the bar on the non-obviousness requirement). A patent is intended to be limited to only one invention consisting of several closely related and indivisible (i.e. integrated) technologies that, acting together, accomplish a specified task; in patent law this is know as the 'unity of invention'. This is a subtle point that often leads to unwarranted criticism of inventors, firms and the USPTO. In plain terms, a jet engine cannot be patented but the numerous components of the jet engine can. In the case of inventions which accomplish multiple and separable tasks, the inventors are required to file separate patent applications for each distinct task. The narrow and circumscribed nature of patented inventions has implications for how patents can be used to study the origins of new technologies. Consider the case of a firm that develops a new automobile that incorporates 10 distinct and novel technologies. The firm cannot seek a patent for the entire new car but must file 10 patents enumerating the 10 indivisible inventions each with its own set of claims of technological novelty.

What is new in a patentable invention is specified by the patent's authors in the numbered *claims* which also serve to define the scope of the legal protection granted by the patent. The claims state, in technical and precise terms, the subject matter of the invention (or discovery), as well as its purpose, principal properties, how it operates and the methods it employs. Claims can read, for example, as 'A hydrocarbon transfer system comprising a processing vessel and a tanker vessel . . . ,' or as 'A method of bandwidth reservations through a switch by employing counters associated with input-output queues . . . .' For patent examiners, the claims are the principal factor in determining whether a proposed invention infringes on an existing patent, and whether it meets the criteria of novelty and non-obviousness. Claims have been a necessary part of US patent applications since the enactment of the Patent Act of 1836. Inventors are obviously motivated to seek the broadest set of monopoly rights, and therefore to be expansive with respect to how many claims they make in their applications.

It would at first seem that the number of claims in a patent corresponds to a higher degree of technological novelty but alas this is not the case, the reason being that there are two types of claims. A claim may be exemplary or dependent: exemplary claims are *independent* claims of novelty or discovery and must meet the criteria of patentability without reference to another claim in the same application, whereas dependent claims refer to or limit another claim in the same patent application. By law, a patent application must contain at least one independent claim. (As of September 2010, 98.6% of patents have a single exemplary claim.) A large number of total claims on a patent almost invariably results from the patent having a large number of dependent claims, and this in turn implies a narrowing of the legal scope of protection for the patent, not a high degree of novelty.

We can illustrate this with two examples. The patent with the highest number of claims is a computer search algorithm for insurance and financial products owned by Ryan Evalulife Systems, Inc. (patent # 6,684,189); although this patent lists 887 dependent claims, it has a single exemplary claim. (Patent # 6,684,189 has received 21 citations to date, which is not at all unusually high.) The majority of the dependent claims qualify the exemplary claim by enumerating each specific type of insurance or financial product the algorithm can search on (documents with hyperlinks, typed or handwritten documents, etc.) and the criteria than can be applied to the search (such as gender, health condition etc.). Or consider patent # 6,609,967 (an apparatus for air recirculation in a ventilation system) with a total of 100 claims and a single exemplary claim. This patent's 99 dependent claims are specifications regarding air flow direction and measurement as well as the number of sensor devices necessary to conduct such measurements. The dependent claims do not touch on or expand the invention's technological novelty.

Dependent claims are important not because they expand upon the novelty of the invention's technology but because they specify the legally protected characteristics of an invention. Even if a court finds the primary claim was wrongly granted, dependent claims may still provide valid grounds for granting a patent. A patent with a large number of claims therefore has significant qualifications concerning the scope of its monopoly rights. Such detailed specifications are required so as not to infringe on other patents (such as other ventilation systems utilizing fewer sensors or a search algorithm that searched on a different set of financial products). Nor is a high number of claims a predictor of the usefulness of an invention; for utility patents granted since 1975 (about 3.4 million patents), the correlation between the number of claims in a patent and the number of citations received within 8 years of the patent being granted is only 0.11.

In technology fields that are 'crowded' – that is to say, areas of technology that already have many patents and technical publications – the examiners must be more careful to avoid conferring duplicate rights of discovery. Patent applications in crowded areas are more likely to have restrictive-dependent claims, and are more likely to be challenged in a declaratory judgment by another party or in an infringement action brought by the patentee themselves. Accordingly, older technologies and mature industries are more likely to see a higher average number of claims. As existing technological capabilities become more fully exploited over time, the crowding effect will manifest itself as a general increase in the number of total claims over all (Tong and Frame 1994; van Zeebroeck, van Pottelsberghe, and Guellec 2006).

Figure 1 shows the mean number of claims per utility patent granted by the USPTO for 5-year windows from 1950 to 2009 (the plot also shows, for purposes of comparison, the

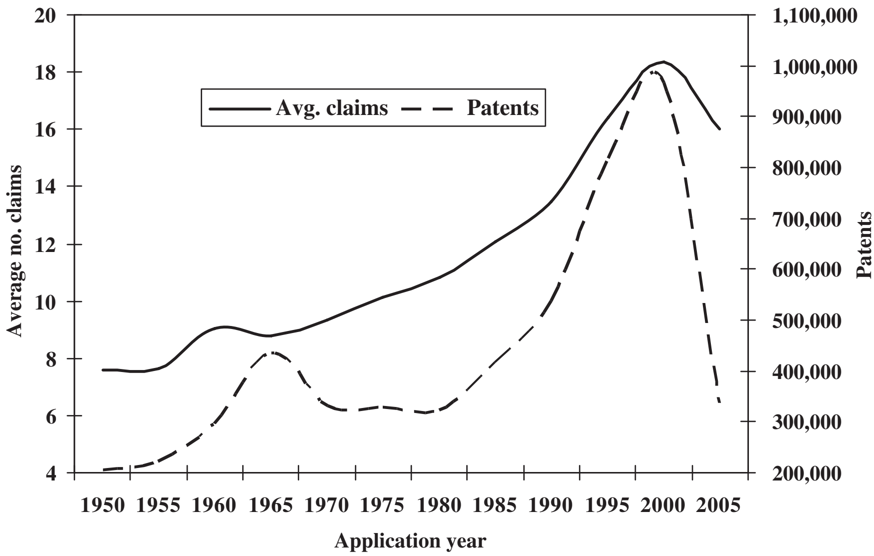

Figure 1. Number of claims per granted patent and number of utility patents successfully applied for, 1950–2009 (results shown are calculated for 5-year windows starting in the indicated year).

total number of patents successfully applied for during each of the 5-year windows). We count patents on the year they were successfully applied for (the application year) so as to count inventions as close as possible to the time they were invented. There is considerable variation across patents in their number of claims as indicated by the coefficient of variation which remained around 0.8 throughout the period. The increase in the number of total claims is quite evident, and two very different processes are driving this increase: there has been an increase in the number of technological capabilities being brought together by patents, and the legal scope of patents is narrowing (which increases the number of dependent claims containing limitations or restrictions). There is no evident relationship between the level of patenting and the increase in the average number of claims (the sharp drop-off in the number of patents after the 2000 window is an artifact of the length of time it takes for a patent application to be reviewed, which on average is about 6 years).

3. Technology codes

The USPTO uses a system of numerical *technology* (or *classification*) *codes* to categorize the technologies invoked by approved claims. The codes make it possible to group patents according to similarly claimed subject matter, thereby allowing for patents to be searchable. Classification codes are used by patent examiners when searching for relevant prior art during a patent application examination process. Multiple-dependent claims may fall under a single technology code, though identical redundant codes appear only once. The process of assigning the technology codes to a patent begins by determining the type of subject matter, that is, whether the invention concerns processes, machines, manufactures and/or compositions of matter. Once the general subject matter is determined, every claim, whether independent or dependent, must be considered separately for classification, and in some instances encoding a single claim may require more than one technology code. The passage from claims to codes is not the result of applying precisely defined and algorithmic-like procedures, but rather depends on the expertise and experience, that is on the craft, of the patent examiners. We emphasize that the reduction of the number of claims to a more limited number of codes is not informative or revealing about the nature of the technologies constituting an invention.

The present USPTO patent classification system is based on a classification scheme created in 1900; prior to 1900 the coding reflected economic activities of the period with somewhat informal subdivisions created to keep the collections to a manageable size. Since then, the Patent Office has specifically avoided classifying technologies based on industries or end-uses as such a scheme carries the risk of granting multiple patents to the same invention. In 1872, for example, the patent classification system distinguished between cooling devices based on the product being cooled (such as beer or milk). Today the technology used to cool beer and milk is treated as being the same, and should be identifiable in any search performed for cooling devices. The classification system – which is legally mandated to provide an exhaustive reference to all patentable subject matter – has been periodically revised since 1900.

A technology (or classification) code consists of two parts: a technology class and a technology subclass. A class is a major category of patentable technology, while a subclass is a refined subset within a class. Subclasses have very detailed definitions and some subclasses are nested within hierarchical relationships to other subclasses. There are currently 481 classes and about 150,000 subclasses used by the USPTO; the active technology codes number 214,091.

Once codes have been assigned to a patent, one of the classifications must be designated as the original or primary classification. If all the codes occur in the same class, then the most specific or detailed code is chosen. If codes are in different classes, the primary classification is generally determined by the most comprehensive claim. Just as the minimum criterion for a patent is a single claim of novelty, a patent must have at least one code, but there are no limits as to how many codes may be assigned to a patent. Other than the code selected as the original classification, the order in which technology codes are listed on a patent is arbitrary. We refer to the list of *n* technology codes categorizing a patent as an *n-tuple*. (For the results presented here, *n-tuples* are treated as combinations and not as permutations.)

Consider the example of the patent for an early global positioning system submitted in 1984 by NASA researchers (patent number 4,445,118): the patent has 24 claims and is described by 2 technology codes. The patent's primary claim can be paraphrased as a system for determining the position of a user, by transmitting data between the user and remote resources to determine a user's spatial coordinates. This claim was encoded using technology code 342/357.395; class 342 is defined as 'Communication: Directive Radio Wave Systems and Devices' (i.e. radar and radio navigation), while subclass 357.395 (which is nested within a larger subclass, 350) codes for devices used for sending or receiving radio signals that vary according to the relative direction or position of the sender or receiver and which are controlled from the ground. All the other claims on this patent enumerate how the signals are processed (e.g. from a relay tower or a satellite), and all the dependent claims are subsumed under technology code 342/356 which is described as the directed radio signal processing that includes the use of synchronous satellite transmission. Or take the example of a recent invention (patent # 7,785,861, granted on 31 August 2010) for a genetically altered bacterium that converts sunlight and carbon dioxide into ingredients of the diesel fuel. The patent for this potentially very consequential invention contains 16 claims which are summarized by just two technology codes: 435/252.3, which refers to a genetically modified micro-organism containing fused-bacteria for producing a desired compound, and code 435/471 which refers to the introduction of a polynucleotide molecule into a bacterium. (Appendix 1 presents three other examples illustrating how technology codes are used to categorize inventions.)

'Classification orders' are issued monthly by the USPTO's Technology Centers, with each order detailing changes to the classification system over the past month. (The Technology Centers lodge the Examining Groups which are in turn divided into Art Units. The Art Units house the patent examiners and are responsible for specifying the patentable technologies for their respective subject matter. There are currently nine Technology Centers within the USPTO: Biotechnology and Organic Chemistry; Chemical and Materials Engineering; Computer Architecture, Software and Information Security; Computer Networks, Multiplex Communication, Video Distribution and Security; Communications; Semiconductors, Electrical and Optical Systems and Components; Designs; Transportation, Construction, Electronic Commerce, Agriculture; National Security; and Mechanical Engineering, Manufacturing and Products.) Classification orders can indicate the introduction of new codes, changes in the definitions of existing codes or the elimination of codes. A 'Primary Patent Examiner' can suggest to his/her 'Supervisory Patent Examiner' the need for a new classification code for categorizing a patent submission; if after deliberation by the appropriate Technology Center, a new code is created, the US Patent Classification System is revised and retroactive reclassification of all previously granted patents takes place. *Technology codes therefore provide a consistent classification scheme making it possible to compare patent technologies across more than 200 years of inventive activity*. The process of reclassification of course takes time, and the introduction of new codes identifying new technologies occurs some time after the new capabilities exist in the economy. The classification manual is updated every two months and archived every June and December. When the set of technology codes is revised, the USPTO reviews all granted patents and reclassifies those meeting the criteria of the new codes. Nanotechnology provides an example of this process. In October 2004, the USPTO announced the creation of a new class (977) for patents in nanotechnology. The USPTO then reviewed all patents issued before class 977 was created and reclassified those meeting the 977 criteria. As a result, according to the USPTO, the first nanotechnology patent was granted in 1986.

Once the encoding process is completed, a patent's set of technology codes encapsulates all the aspects of novelty set forth in the claims and specifies the technological capabilities embodied in the invention. Therefore the larger the number of technology codes, the larger the number of distinct technologies, constituting the patent. Since the technologies denoted by the codes are assumed to tightly interact in order to accomplish a specific task, it seems plausible to also consider a larger number of codes as indicative of greater technological complexity (although not necessarily greater novelty). Two patents with the same set of technology codes embody the same technologies, albeit with different claims to novelty and scope of legal protection. A patented invention is in effect an *instantiated* combination of technologies from the defined set of all possible combinations of technologies.

Our emphasis on tracking the technological capabilities present in patented inventions distinguish our approach from previous efforts which have studied technological recombination using patent technology classes (as in Fleming and Sorenson 2001; Nesta 2008; Antonelli, Krafft, and Quatraro 2010; Quatraro 2010). As previously mentioned, technology classes identify broad subject areas much like the two or three-digit codes do in the North American Industry Classification System (NAICS) used by governmental statistical offices in the USA, Canada and Mexico to classify business establishments according to the type of economic activity. There are 24 two-digits NAICS codes (such as manufacturing, construction, finance and insurance) and 1174 six-digits NAICS codes (such as plastic bag manufacturing, residential remodeling and portfolio management). For some questions, the level of detail provided by two-digits NAICS codes is amply adequate, while other types of analyses require the fine-grained detail afforded by the six-digit codes. Similarly, with patent classes and patent technology codes: choosing to use one or the other as a way to identify technologies is a matter of what level of detail is adequate to the analytic task at hand. We also note that 15% of all patents granted by the USPTO since 1975 are classified using only one technology class and almost 40% of all granted patents are assigned no more than two classes. Given our emphasis on understanding technological change and recombination as involving *capabilities*, technology classes (which denote inventions' subject matter) are too coarse-grained.

4. Basic empirics

We now present some basic empirics for the patent technology codes (without loss of generality we use data for utility patents). In order to dampen noise the time-series results are shown for 5-year windows starting in 1835 and ending in 2009, with the starting date chosen due to the paucity of observations between 1790 and 1834. (Construction of the patent database is described in Appendix 2.) As previously discussed, one of the most attractive features of the system of technology codes is how the Patent Office reclassifies all patents when new codes are introduced, thereby creating a historically consistent dataset permitting 'apple to apples' comparisons. It is therefore possible to get a measure of accrued technological capabilities, up to a certain point in time, by recording how many codes have been utilized to classify patents up to that date: we refer to this measure as *total used technology codes*. The cumulative total is distinguished from the *number of active codes*, which is the number of codes actually used by patent examiners to categorize utility patents during a specific period. (The results shown here are based on the April 2010 *Classification Manual*.)

Figure 2 shows the total number of technology codes used (i.e. the number of codes instantiated in patents granted up to the end of a given time window) and the total number of active codes. Since 1900 about 80% of the total number of available codes are actually used by patent examiners, indicating that the set of codes is well matched to the technological features of contemporary inventions – that is, there is not much technological 'dead wood' in the set of technology codes. There was a pronounced increase in the total used codes after the year 2000 due to the creation of new codes, starting in the late 1990s, to accommodate software and search technologies.

The mean and median numbers of technology codes assigned to utility patents, calculated within each 5-year window, are shown on Figure 3. The mean number of codes categorizing patents has increased over the past 175 years, an indication of the steadily growing technological complexity of inventions. The mean number of codes increased from 3.2. in 1945 to 4.4 in 2005 – a significant increase considering that the process by which codes are assigned to patents aims at economy of description (the 75th percentile has consistently had a value of 5). The mean number of codes categorizing patents is distinctly smaller than the number of claims, however – a testimony to the descriptive efficiency of the technology codes. There has been significant, and increasing, variability, though, as indicated by the difference between mean and median values, and a coefficient of variation of about 0.75 for every time window (much of this variability is due to some inventions having upwards of 20 codes assigned to them).

Figure 4 shows the cumulative distribution of patents by *n-tuples* (for all patents granted between 1835 and 2009). Sixty-two percent of all patents have been categorized by no more than three technology codes and four codes have sufficed to describe almost 75% of

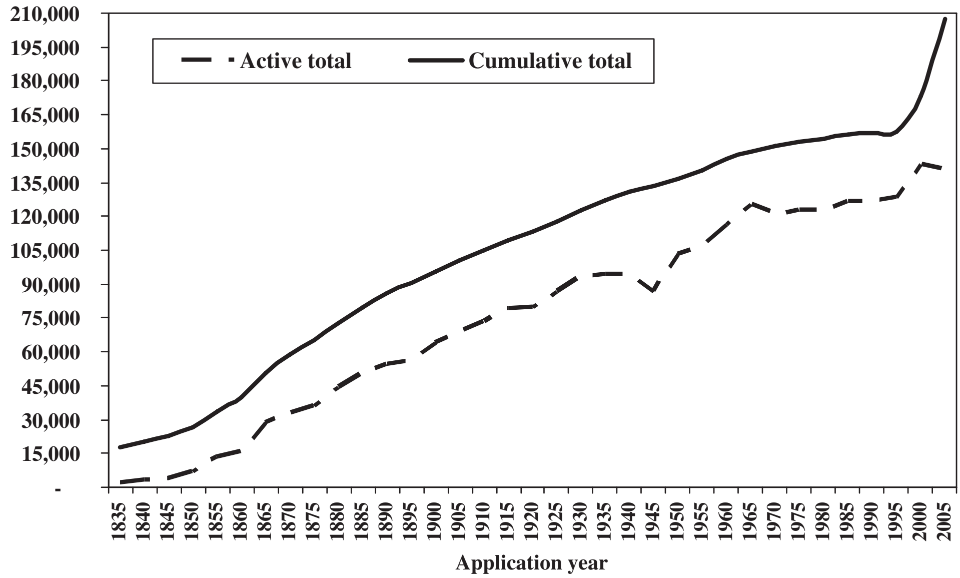

Figure 2. Cumulative total and active number of USPTO technology codes, 1835–2005 (results shown are yearly totals added over 5-year windows).

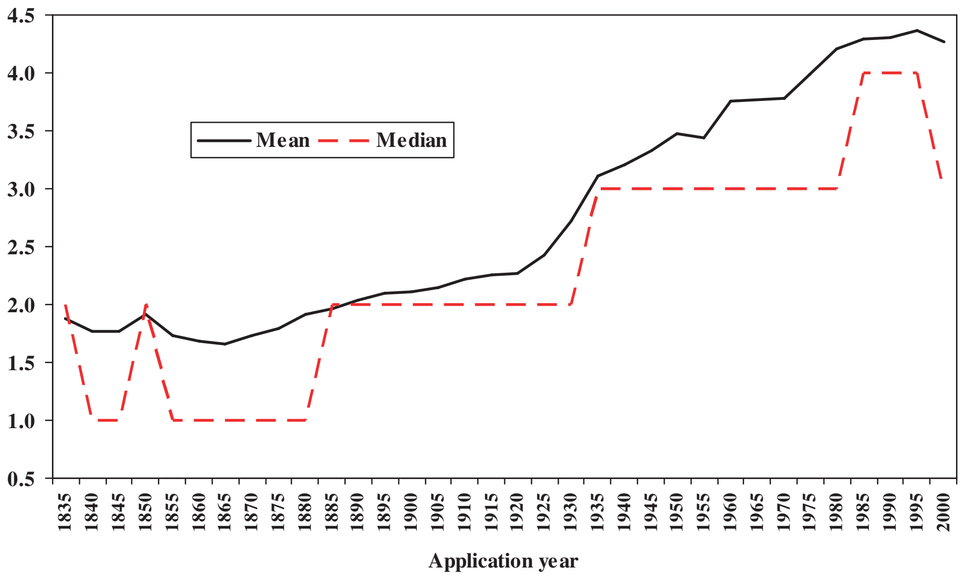

Figure 3. Mean and median number of technology codes assigned to USPTO granted patents, 1835–2005 (calculated for 5-year windows).

all patents granted over the 175 years covered by the data. Clearly, the codes provide the vocabulary for precise and parsimonious descriptions of inventions. Figure 5 depicts the change over time for the proportion of patents categorized by *n-tuples* of different sizes (the percentages are calculated over 5-year windows). Whereas one technology code is enough to categorize almost 60% of patents granted between 1835 and 1870, by 2009 only

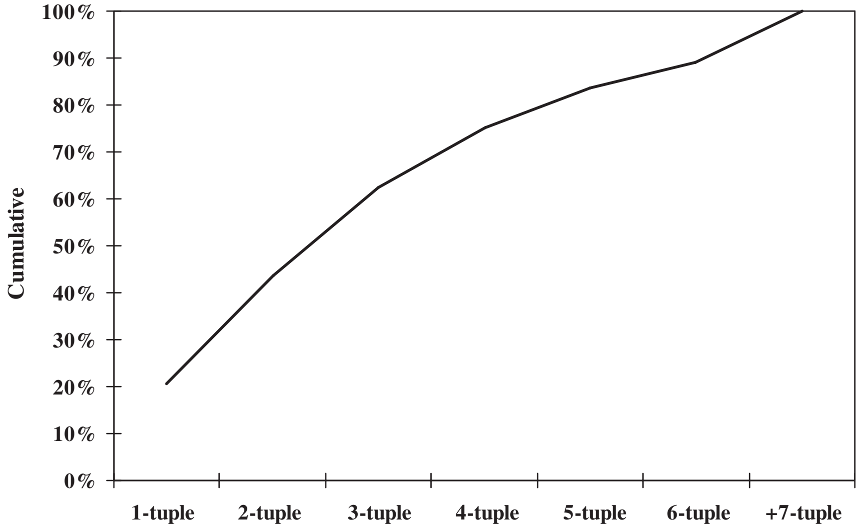

Figure 4. Distribution by *n*-tuples (the number of technology codes are used in describing a patent) of all patents granted between 1835 and 2009.

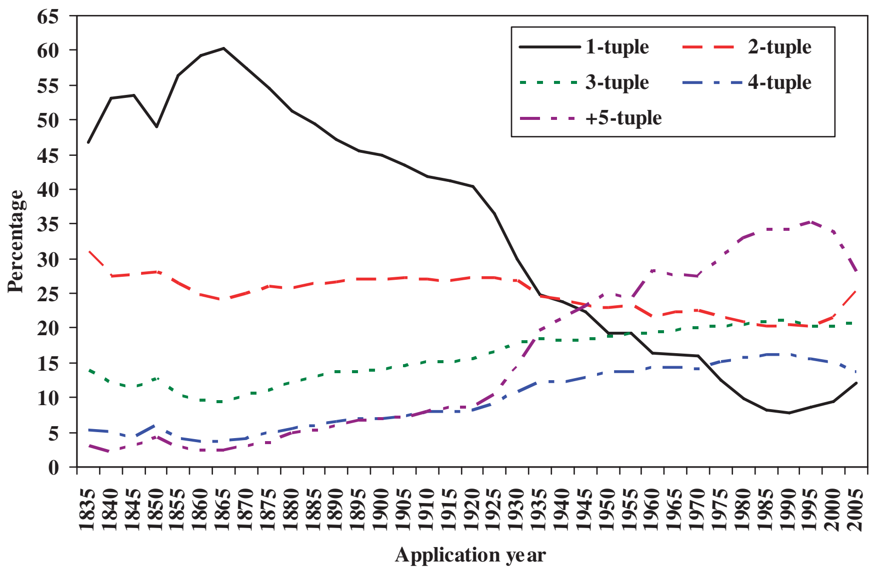

Figure 5. Proportion of patents categorized by 1, 2, 3, 4, and 5 and more codes by the 5-year window.

10% of patents could be described with only one code. The proportion categorized by two technology codes decreased by much less, from 31% to 25%. The proportions accounted by descriptions using 3 and 4 codes both increased (to 21% and 14%, respectively). Over the past 50 years, a third of all inventions have required 5 or more codes indicating that a significant portion of recent inventive activity consists of combining several distinct technologies (rather than the creation of truly novel technologies).

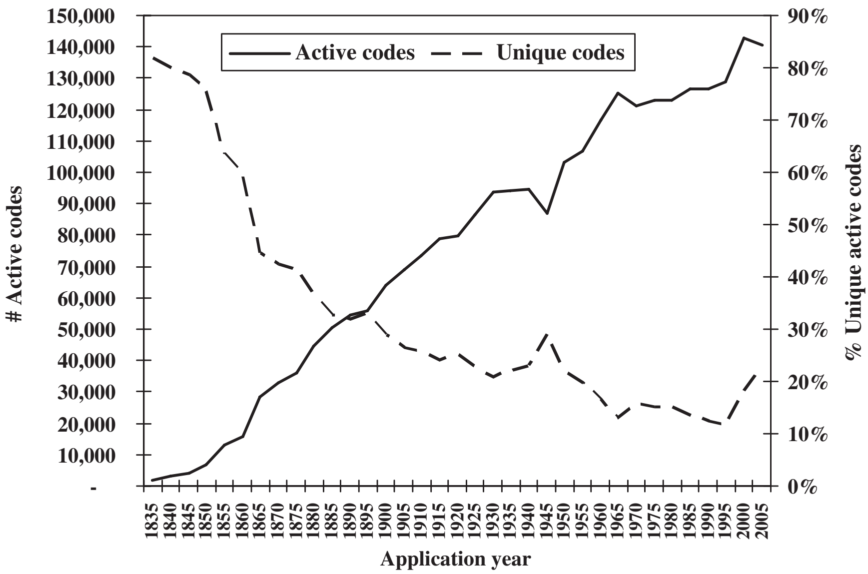

Figure 6. Number of active technology codes (codes used by patent examiners to categorize utility patents during the time window) and the percentage of active codes which were used only once per 5-year windows (1835–2005).

Since individual technology codes describe specific technological capabilities, multiple codes must be combined by patent examiners in order to generate comprehensive descriptions of inventions. If inventions have become more complex, one would then expect the proportion of active codes used only once to decrease, as the *n-tuples* built by examiners invoke more codes per patent. Figure 6 contrasts the increase in the number of active codes used with the decrease in the proportion of active codes used only once. By the end of

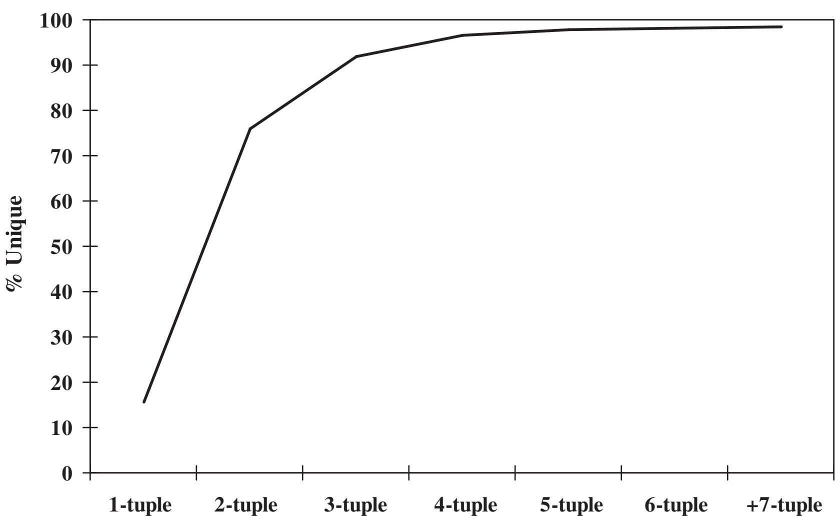

Figure 7. Percentage of differently sized *n*-tuples which are have been used only once to categorize a patent (1835–2009).

the 1990s, only 12% of active codes saw duty only once. There was a pronounced increase over the 2000–2009 period, though, in the percentage of codes used only once (from 12% to 24%). This increase is most likely due to the creation of codes, starting in 1996, for accommodating patents for software and search algorithms (if patents for software and algorithms are removed from the data, the downward trend visible over most of the period covered by the plot in Figure 6 is restored for the last decade).

The USPTO's classification system aims to describe the technological features of patents that make the inventions novel. The classification of patents must capture the characteristics of inventions as distinctly as possible, otherwise inventors and examiners would be unable to conduct efficient 'prior art' searches, and the classification system would be incapable of issuing patent rights on a timescale relevant to the inventors. We find that the *n-tuples* are indeed very discriminating. Figure 7 shows the percentage of differently sized *n-tuples* which are unique – have been used only once to categorize a patent – over the period from 1835 to 2009. Almost 80% of code combinations consisting of two codes (2-*tuples*) have only been used once; in the case of 3-*tuples*, the percentage of unique usages is an astounding 92%.

5. Tracking technological change

It is our principal contention that patent technology codes can be used to track and characterize technological change and in what follows, we illustrate some of the analytic possibilities facilitated by the codes. Technological change can be detected by noting the difference, between two measurement periods, in the number of technology codes instantiated by the patents granted during the two periods in questions. This difference measures the number of technological functionalities and capabilities used in the later period which were not available in the previous period. An obligatory comparison is that between the number of patents granted during a 5-year window and the total number of technology codes available during the same period for classifying inventions.

Figure 8 shows (a) the time-series for the total number of technology codes instantiated by patents granted up to the last year of the 5-year windows, and (b) the time-series for the number of patents successfully applied for within a 5-year window. (For ease of comparing the growth trajectories, the two series are expressed in natural logarithms.) Initially the two series track each other as the inventions generated during the 1840s, 1850s, and 1860s were mainly single-code inventions. (Judging by the growth rates of inventions and technological capabilities, the inventive golden age in the USA occurred in the first half of the nineteenth century.) Starting in 1870, the two series start to diverge as inventions began to combine technological capabilities more often than using single capabilities. The growth in the number of codes has been quite steady since the start of the twentieth century while patenting activity increased markedly in the post-WWII period, stalled during the 1970s and then experienced a sharp increase starting in 1980. The years since 1980 have seen a combinatorial explosion of invention with almost 40% of all patents granted by the USPTO from 1835 to 2009 occurring in the two decades between 1990 and 2009. No doubt some of this inventive surge was due to technological factors, but the effects of changes in patent law enacted during the 1980s – which combined to increase the value of holding a patent and creating more incentives for firms to patent their inventions – should not be ignored (Jaffe and Lerner 2004). The decrease in patenting observed after 2005 is nothing more than a consequence of the fact that many of the patents applied for after 2005 are still under review (the average length of the patent review process is a little bit over 6 years).

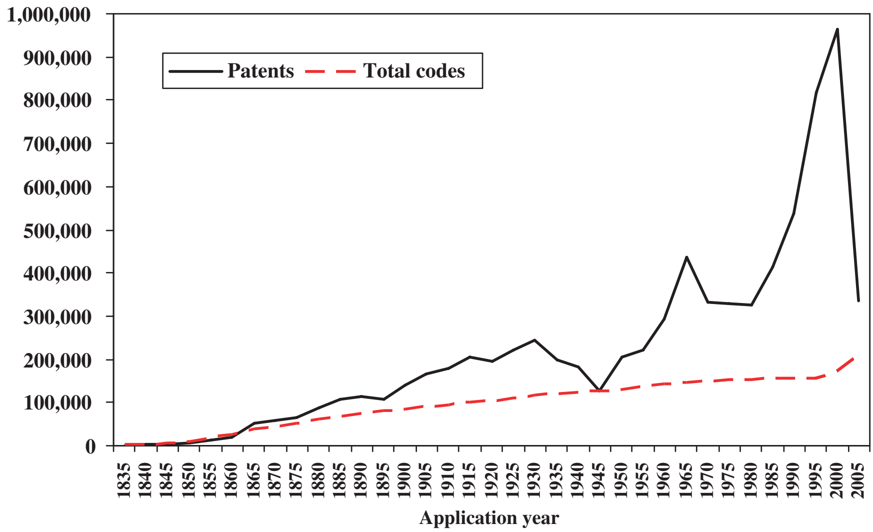

Figure 8. Total of used technology codes and patents generated by 5-year windows.

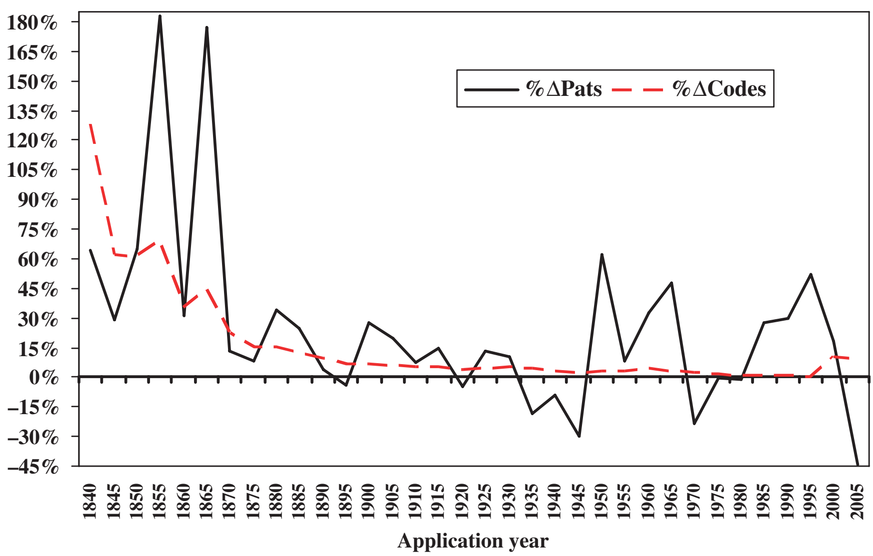

Figure 9. Percent change in patents and new technology codes per 5-year windows (1840–2005).

Figure 9 takes a closer look at the change in technological capabilities and patenting by showing the percentage change, between 5-year windows, in the number of total technology codes and in the number of patents granted. The rate of change for patenting has been quite volatile, but since the late nineteenth century the rate at which new technology codes have been introduced has been much steadier. The two very noticeable spikes in the percent change in patenting occurred during the 5 years before and the 5 years after the Civil War. And from 1945 to 1999, what many consider to be the coming of age of high technology, the rate at which new codes were introduced is a paltry 2% per period. The growth in new technologies has perhaps been experienced as faster because the sheer number of existing technological capabilities already in existence is so vast. (The acceleration in the growth rate of technology codes experienced from 1995 to 2005 was due mainly to the introduction of over 6200 distinct new technology codes for categorizing software-related inventions.)

The ratio of the number of cumulative codes to the number of patents in technology codes, shown in Figure 10, captures how the economy's inventive efforts have been directed. A value for the ratio greater than 1 indicates that inventive effort was channeled towards developing new technological capabilities; a value less than 1 signifies that inventive effort was directed toward combining existing capabilities. From 1870 onwards, the value of this ratio has been less than one (except for the WWII period), and declining, indicating that invention in the USA has been driven much more by the combination of existing technologies than by the generation of truly new technological capabilities. Another perspective is afforded by the ratio of the percent change in technology codes over the percent change in patents, a sort of elasticity, shown in Figure 11. A value for this ratio of less than 1 signifies that inventive effort was directed toward combining existing capabilities and a value close to 1 indicates that new technological capabilities have been added to the economy at the same rate that inventions have been created. The sharp declines in the ratio coincide closely with major economic shocks and periods of war, specifically the panics of 1893 and 1896, World War I and World War II, the oil price crisis of the early 1970s and the most recent recession starting in 2007. These are periods of high uncertainty in which the risks associated with seeking new technological capabilities are higher than the risks associated with improving existing capabilities, thus firms favor invention in existing technologies during hard economic times or periods of national emergencies. (The ratio becomes negative for those periods, where patenting output declined in relation to the previous period.) Both of the ratios shown in Figures 9 and 10 show an uptick during the 1995–2005 period, evidence again of the upsurge in patenting and the creation of new technological capabilities which occurred during that decade.

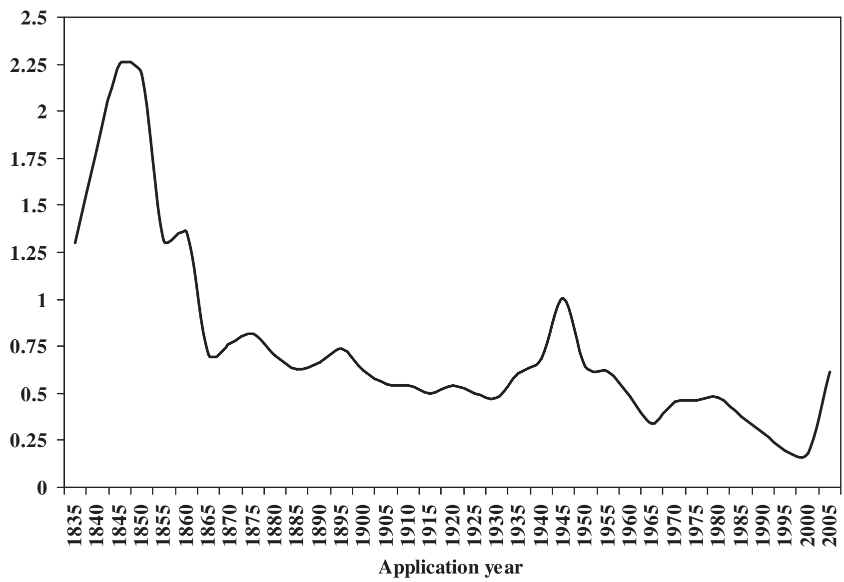

Figure 10. Ratio of total used technology codes to patents per 5-year windows (1835–2005).

Figure 11. Ratio of the change in total used technology codes to the change in patent output.

6. Research possibilities

To the somewhat crowded research domain which uses patent data, we want to add another data stream which we think can help illuminate the dynamics of technological change: the technology codes used by the US Patent Office to classify a patented invention's technological novelty. The technology codes provide a consistent classification scheme making it possible to compare the technological capabilities instantiated by patented inventions, and our results show that the codes are efficient and discriminating in classifying inventions. By focusing on the technologies constituting inventions, rather than on the inventions themselves, researchers can use the technology codes to track changes in the functionalities and capabilities available to the economy at any one time.

Our results suggest that invention mainly involves the combination and recombination of previously existing technological capabilities rather than the development of totally new capabilities – and we will pursue this question in greater depth in a subsequent research effort. We briefly sketch other research paths that seem promising from this perspective, hoping that other researchers can add to the list.

- The fine-grained perspective made possible by technology codes could help identify 'General Purpose Technological Capabilities' to complement the work done on General Purpose Technologies which has focused on ensembles of technologies.
- Using the patent technology codes, one could define a technology space with the codes as nodes in a graph and edges linking technologies which co-appear in an invention. The degree of connectivity in sub-regions of the space (which define technological domains) could be construed as an indicator of the complexity of that sub-space.
- While predicting the invention seems foolhardy, identifying possible pathways for technological change that can be put on a serious footing by asking whether past inventive activity makes some regions of technology space more or less likely to be explored and filled-in. Starting from a fundamental invention, one could examine how regions of technology space become populated. It may be possible to make

probabilistic assessments about which regions of the search space are most likely to develop the next recombinant cluster.

- Another approach for determining if predictability is inherent in the system of technology codes emerges from asking whether the codes define a *language*. It is a common feature of languages that some word sequences are more common than others, and that the appearance of some words, or sequences of words, in a sentence is highly predictive of other words or phrases. Preliminary results indicate that the system of patent technology codes is a highly discriminating and precise language. Some technology codes could then be predictive of the occurrence of other technologies.

Acknowledgements

We benefitted from comments by Luis Bettencourt, Fabrizio Lillo, John Miller, Bruce Kogut, William Kerr, and Tihamer von Ghyczy. The work presented in this paper was largely done during a 3-week research stay at the Santa Fe Institute during the summer of 2010 – we gratefully acknowledge the Institute's hospitality.

References

Acs, Z.J., L. Anselin, and A. Varga. 2002. Patents and innovation counts as measures of regional production of new knowledge. *Research Policy* 31, no. 7: 1069–85.

Acs, Z., and D. Audretsch. 1989. Patents as a measure of innovative activity. *Kylos* 42, no. 2: 171–80.

Allen, T. 1977. *Managing the flow of technology*. Cambridge, MA: MIT Press.

Antonelli, C., J. Krafft, and F. Quatraro. 2010. Recombinant knowledge and growth: The case of ICTs. *Structural Change and Economic Dynamics* 21, no. 1: 50–69.

Archibugi, D. 1992. Patenting as an indicator of technological innovation: A review. *Science and Public Policy* 16, no. 6: 357–68.

Arthur, B. 2009. *The Nature of Technology: What it is and How it Evolves.* New York: Free Press.

Auerswald, P., S. Kauffman, J. Lobo, and K. Shell. 2000. The production recipes approach to modeling technological innovation: An application to learning by doing. *Journal of Economic Dynamics and Control* 24, no. 3: 389–450.

Basalla, G. 1988. *The evolution of technology*. Cambridge, MA: MIT Press.

Bettencourt, L.M.A., J. Lobo, and D. Strumsky. 2007. Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research Policy* 36, no. 1: 107–20.

Boot, M. 2006. *War made new: Technology, warfare, and the course of history: 1500 to today.* New York: Gotham Books.

Braudel, F. 1992. *Civilization and capitalism 15th–18th century*. Vols I, II, III. Berkeley: University of California Press.

Christensen, C. 1997. *The innovator's dilemma: When new technologies cause great firms to fail.* Cambridge, MA: Harvard Business Press.

Fleming, L. 2001. Recombinant uncertainty in technological search. *Management Science* 47, no. 1: 1019–39.

Fleming, L., and O. Sorenson. 2001. Technology as a complex adaptive system: Evidence from patent data. *Research Policy* 30, no. 7: 117–32.

Fleming, L., and O. Sorenson. 2004. Science as a map in technological search. *Strategic Management Journal* 25, nos. 8–9: 909–28.

Frenken, K. 2006. *Innovation, evolution and complexity theory.* Northampton, MA: Edward Elgar Publishing, Inc.

Gilfilan. S. 1935. *Inventing the ship*. Chicago: Follett Publishing Co.

Griliches, Z. 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28, no. 4: 1661–707.

Hall, B.H., A. Jaffe, and M. Trajtenberg. 2001. The NBER patent citations data file: Lessons, insights and methodological tools. NBER working paper 8498, Cambridge, MA.

Hargadon, A., and R. Sutton. 1997. Technology brokering and innovation in a product development firm. *Administrative Science Quarterly* 42, no. 4: 716–49.

Helpman, E. 1998. *General purpose technologies and economic growth*. Cambridge, MA: MIT Press.

Henderson, R., and K. Clark. 1990. Architectural innovation: The reconfiguration of existing product technologies and failure of established firms. *Administrative Science Quarterly* 35, no. 1: 9–30.

Higgs, R. 1971. American inventiveness, 1870–1920. *Journal of Political Economy* 79, no. 3: 661–7.

Hughes, T. 1989. *American genesis: A century of invention and technological enthusiasm*. New York: Viking Press.

Jaffe, A., B.H. Hall, and M. Trajtenberg. 2005. Market value and patent citations: A first look. *RAND Journal of Economics* 36, no. 1: 16–38.

Jaffe, A., and J. Lerner. 2004. *Innovation and its discontents: How our patent system is endangering innovation and progress and what to do about it.* Princeton: Princeton University Press.

Jaffe, A.B., and M. Trajtenberg. 2002. *Patents, citations, and innovations: A window on the knowledge economy*. Cambridge, MA: MIT Press.

Jaffe, A.B., M. Trajtenberg, and R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108, no. 3: 577–98.

Jorgenson, D.W. 2001. Information technology and the US economy. *American Economic Review* 91, no. 1: 1–32.

Kauffman, S., J. Lobo, and W.G. Macready. 2000. Optimal search on a technology landscape. *Journal of Economic Behavior and Organization* 43, no. 2: 141–66.

Khan, Z. 2005. *The democratization of invention: Patents and copyrights in American economic development, 1790–1920.* Cambridge: Cambridge University Press.

Kogut, B., and U. Zander. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science* 3, no. 3: 383–97.

Kroeber, A. 1948. *Anthropology*. New York: Harcourt Brace Jovanovic, Inc.

Krugman, P. 1991. *Geography and trade.* Cambridge, MA: MIT Press.

Landes, D. 1998. *The wealth and poverty of nations: Why some are so rich and some so poor.* New York: W. W. Norton & Company.

Lerner, J., and S. Kortum. 2000. Assessing the contribution of venture capital to innovation. *RAND Journal of Economics* 31, no. 4: 674–92.

Levinthal, D. 1997. Adaptation on rugged landscapes. *Management Science* 43, no. 7: 934–50.

Marx, M., D. Strumsky, and L. Fleming. 2009. Mobility, skills, and the Michigan non-compete experiment. *Management Science* 55, no. 6: 875–89.

Mokyr, J. 1990. *The lever of riches: Technological creativity and economic progress*. New York: Oxford University Press.

Moser, P., and T. Nicholas. 2004. Was electricity a general purpose technology? Evidence from historical patent citations. *American Economic Review* 94, no. 2: 388–94.

Nelson, R.N., and S.G.Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Nesta, L. 2008. Knowledge and productivity in the world's largest manufacturing operations. *Journal of Economic Behavior and Organization* 67, nos. 3–4: 886–902.

Pisano, G.P. 2006. *Science business: The promise, the reality and the future of biotech*. Boston: Harvard Business School Press.

Quatraro, F. 2010. Knowledge coherence, variety and economic growth: Manufacturing evidence from Italian regions. *Research Policy* 39, no. 10: 1289–302.

Redman, C. 1979. *The rise of civilization: From early farmers to urban society in the ancient near east.* New York: W.H. Freeman and Company.

Rivkin, J.W. 2001. Reproducing knowledge: Replication without imitation at moderate complexity. *Organization Science* 12, no. 3: 274–93.

Rosenberg, N. 1982. *Inside the black box: Technology and economics*. Cambridge: Cambridge University Press.

Rosenberg, N. 1994. *Exploring the black box: Technology, economics, and history*. Cambridge: Cambridge University Press.

Rosenthal, S.S., and W.C. Strange. 2004. Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, vol. 4, ed. J.V. Henderson and J.F. Thisse, 2119–71. Amsterdam: Elsevier Press.

Schumpeter, J. 1939. *Business cycles*. New York: McGraw-Hill.

Tong, X.S., and J.D. Frame. 1994. Measuring national technological performance with patent claims data. *Research Policy* 23, no. 2: 133–41.

Trajtenberg, M. 1990. A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics* 21, no. 1: 172–87.

Trajtenberg, M., R. Henderson, and A. Jaffe. 1997. University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology* 5, no. 1: 19–50.
Tushman, M., and L. Rosenkopf. 1992. Organizational determinants of technological change: Toward a sociology of technological evolution. *Research in Organizational Behavior* 14: 311–47.
Usher, A.P. 1954. *A history of mechanical inventions.* New York: Dover Publication.
Von Hippel, E. 1988. *The sources of innovation.* New York: Oxford University Press.
Weitzman, M.L. 1998. Recombinant growth. *The Quarterly Journal of Economics* 113, no. 2: 331–60.
Youtie, J., M. Iacopetta, and S. Graham. 2008. Assessing the nature of nanotechnology: Can we uncover an emerging general purpose technology? *The Journal of Technology Transfer* 33, no. 3: 315–29.
van Zeebroeck, N., B. van Pottelsberghe, and D. Guellec. 2006. Claiming more: The increased voluminosity of patent applications and its determinants. Working Papers CEB 06-018.RS, Université Libre de Bruxelles, Brussels.

Appendix 1. Examples of using technology codes to categorize patents

We present three additional examples showing how the USPTO's technology codes categorize inventions. Patent number 6,823,725 (filed on 12 December 2000 and granted on 30 November 2004) is for a 'Linear distance sensor and the use thereof as actuator for motor vehicles'. This patent list 13 claims and is categorized by three codes – 73/114.01, 324/207.13 and 324/207.24 – which together specify the technological capabilities brought together by the invention. Code 73/114.01 indicates that the invention involves a process or an apparatus for performing a test on an engine in which the combustion of the fuel takes place within a cylinder; code 324/207.13 indicates that the invention involves measuring the change in a magnetic field; and code 324/207.24 specifies that the invention involves measuring motion in a straight line.

Patent 5,029,133 (filed on 30 August 1990 and granted on 2 July 1991) is for 'A VLSI chip having improved test access'. This patent lists two claims and is categorized by four codes: 365/189.02, 365/201, 714/703 and 714/731. Code 365/189.02 indicates that the apparatus involves the static storage and retrieval of information through the transmission of plural signals over a single signal path; code 365/201 categorizes the invention as involving the testing of electronic memory for defects or erroneous information; code 714/703 signifies that the invention involves a process or apparatus for detecting and correcting errors in electrical pulse or pulse coded data and in which the proper operation of the error detection/correction or fault detection/recovery apparatus itself is verified; and code 714/731 informs us that the invention utilizes a timing function or a clock-pulse generator for causing the various parts of the device to operate on a common time base.

As a third example, consider patent number 6,569,641 (filed on 5 June 1995 and granted on 27 May 2003) for a 'Process for obtaining DNA, RNA, peptides, polypeptides, or protein by recombinant DNA technique'. This invention lists 21 claims and is categorized by eight codes – 435/69.1, 435/440, 435/471; 435/476, 435/6, 435/71.1, 435/91.1, and 435/91.4 – all pertaining to class 435 ('Chemistry: Molecular Biology and Microbiology'), which classifies a process using a micro-organism or enzyme to synthesize a chemical product. For the sake of brevity, we will unpack only two of the patent's technology codes. Code 435/69.1 identifies the invention as involving a process involving the use of recombinant DNA techniques in a process of synthesis of a protein or polypeptide while code 435/476 tells us that the invention involves a process, wherein the nucleic acid is a plasmid or episome (and therefore capable of autonomous replication in the micro-organism cell).

Appendix 2. Patent data

The USPTO makes available data on the technology codes assigned to every patent it has granted in the optical disc format (the Cassis Patent Assignments File); the data cover the period from 1790 to the most recent month that data have been archived. The information provided in the Cassis file includes the date a patent was granted, but does not provide information about the date the patent was applied for. So as to count patents close to the time they were invented, we choose to count them in the year they were successfully applied. In order to count patents by application year, the patents in the Cassis file have been matched with the patents recorded in a database constructed by one of us (Strumsky) using data provided by the USPTO. This patent database, which covers slightly over 8 million patent records, includes information on filing data for every utility patent granted by the USPTO since 1963.

For utility patents granted prior to 1963, the application year has been obtained from citation records and direct searches. This still left about 13% of all patents, applied for between 1790 and 1960, without an application year. For these patents, the year the patent was granted was used to replace missing application years. Prior to the 1980s, the time between the filing of an application and the granting of a patent was significantly less than now (usually taking place within 24 months), and since we use 5-year windows to report our results, we do not believe that using the year granted instead of the filing year in those patents introduced any significant bias.