NBER WORKING PAPER SERIES

UNCOVERING GPTS WITH PATENT DATA

Bronwyn H. Hall
Manuel Trajtenberg

Working Paper 10901
http://www.nber.org/papers/w10901

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2004

Uncovering GPTS with Patent Data
Bronwyn H. Hall and Manuel Trajtenberg
NBER Working Paper No. 10901
November 2004
JEL No. O330, O310, L860

ABSTRACT

This paper asks the question: Can we see evidence of General Purpose Technologies in patent data? Using data on three million US patents granted between 1967 and 1999, and their citations received between 1975 and 2002, we construct a number of measures of GPTs, including generality, number of citations, and patent class growth, for patents themselves and for the patents that cite the patents. A selection of the top twenty patents in the tails of the distribution of several of these measures yields a set of mostly ICT technologies, of which the most important are those underlying transactions on the internet and object-oriented software. We conclude with a brief discussion of the problems we encountered in developing our measures and suggestions for future work in this area.

Bronwyn H. Hall
Department of Economics
549 Evans Hall, #3880
University of California-Berkeley
Berkeley, CA 94720-3880
and NBER
bhhall@econ.berkeley.edu

Manuel Trajtenberg
Eitan Berglas School of Economics
Tel-Aviv University
Tel-Aviv 69978
ISRAEL
manuel@post.tau.ac.il

Uncovering GPTs with Patent Data[1]

Bronwyn H. Hall and Manuel Trajtenberg

1 *Introduction*

In 'The Computer and the Dynamo,' Paul David makes a persuasive case for considering the process by which the electric dynamo spread throughout the economy during the turn of the last century and the process by which the use of information technology (specifically, computing technology) is currently being spread throughout different industries as similar manifestations of the diffusion of 'General Purpose Technologies,' a term introduced into the economics literature by Bresnahan and Trajtenberg (1995). All these authors, as well as Helpman and Trajtenberg (1998a,b), emphasize the singular contribution to economic growth made by this type of technology, because of its ability to transform the means and methods of production in a wide variety of industries.

At the same time and using historical data, David (1989, 1990), Rosenberg (1976), and others have argued that the diffusion of these technologies throughout the economy may take decades rather than years because of coordination problems and the need for complementary investments (both tangible and intangible) in using industries. For this reason it may take some time for the benefits of the technologies to be manifest in economic growth. On the theoretical side, Bresnahan and Trajtenberg (1995) have

[1]A preliminary version of this paper was presented at the conference 'New Frontiers in the Economics of Innovation and New Technology,' held in honour of Paul A. David at the Accademia delle Scienze, Torino, 20-21 May, 2000. We are grateful to participants in that conference, especially Paul David, John Cantwell, Giovanni Dosi, Ove Granstrand, and Ed Steinmueller, for comments on the earlier draft. The first author thanks the Centre for Business Research, Judge Institute of Management, University of Cambridge for hospitality while this version was being written.

studied the non-optimality of innovation and diffusion when a decentralized market system is called upon to try to solve the coordination problem between technology-innovating and technology-using industries. However there has been relatively little empirical and econometric work that incorporates the insights of these various authors to analyze specific technologies.

Our modest goal in this paper is to see what might be learned about the existence and technological development of General Purpose Technologies (hereafter, GPTs) through the examination of patent data, including the citations made to other patents. Such measures would be useful both to help identify GPTs in their early stages of development and also as proxies for the various rates of technical change called for in a fully developed growth model such as that in Helpman and Trajtenberg (1998b). In doing this exploration we are also motivated by the observation that not all technologies, or indeed, R&D dollars are equal, but that economists too often ignore that fact, primarily because of data limitations. As has been pointed out by others before us, patenting measures have the potential to allow more detailed analysis of the 'direction' as well as the 'rate' of technical change.[2]

Although such an exploration might be made using data from a variety of countries, our focus here is on the use of US patent data, where the citations have a well-defined meaning and also where they have been computerized since 1977, enabling us to work with them relatively easily. Given the importance of the United States as a locus of technical change in the late twentieth century, we do not feel that this limitation to US patenting activity is a serious drawback for a preliminary investigation of this kind.

We begin with the definition (description) of GPTs offered by Helpman and Trajtenberg (1998a):

1. They are extremely pervasive and used in many sectors of the economy. Historical examples are the steam engine and the electric dynamo (the engine of electrification). Contemporary examples are the semiconductor and perhaps the internet.
2. Because they are pervasive and therefore important, they are subject to continuous technical advance after they are first introduced, with sustained performance improvements.
3. Effective use of these technologies requires complementary investment in the using sectors; at the same time, the GPT enhances the productivity of R&D in the downstream sector. It is these points that are emphasized by David.

Using this definition, the contribution of the effort described here is to define measures using patents and citations that quantify the insights of David and Trajtenberg and their co-authors.

Our study is subject to a variety of limitations, however: first, it is based on patent data, which provides imperfect coverage of innovative activity, as not all innovations are patented or patentable.[3] Second, it relies heavily on the US Patent Office classification system for technology, treating each three-digit patent class as roughly comparable in the 'size' of a technology. An examination of the classes suggests that this is unlikely to be strictly true (for example, the chemistry of inorganic compounds is a single class, whereas there are multiple optics classes). Making use of the subclasses to refine the class measures would be a formidable task, because subclasses are spawned within the three-digit class 'ad libidum' and may descend either from the main class or

---

[2] Griliches (1990); Pavitt (1988).
[3] In the United States environment, this statement is increasingly less true, although the converse, that not all patent subject matter is innovative, may be becoming more true.

from another subclass. Thus some subclasses are more 'important' than others, but this fact has to be uncovered by a tedious search of the text on the USPTO website. Rather than attempting to construct our own classification system in this way, we chose to look at measures based on the primary IPC class of the patent.[4] Finally, because patent citation data is only available in computerized form in 1975, and because severe truncation due to the application-grant lag and the citation lag sets in by around 1995 our period of study is necessarily fairly short and emphasizes the 1980s and 1990s. However, truncation in the later years means that we are unable (yet) to fully explore the implications of changes in information processing technology during the very recent past.

Because of time and resource constraints, we have focused the large part of our analysis on an extremely small subset of the nearly three million patents available to us, the 780 most highly cited patents that were granted between 1967 and 1999. There is considerable evidence that the value or importance distribution of patents is highly skewed, with most patents being unimportant and a few being highly valuable.[5] We expect that one reason for this finding is that true GPT patents are concentrated among the highly cited patents, so the current endeavour is centred on those patents, which represent the extreme tail of a very skew distribution.

In order to understand how patent data might help us identify GPTs and explore their development and diffusion, it is necessary first to understand something more about

---

[4] As a general rule, the USPTO does not classify patents individually into IPC classes, but relies on a map based on US classes and subclasses to determine them. This is not ideal, but does mean that they incorporate some subclass information.

[5] For recent evidence on this point, see Harhoff et al (1999) for the results of a survey of patent owners, and Hall, Jaffe, and Trajtenberg (2000) for results showing that the market value-citation relationship is highly nonlinear, with firms owning highly cited patents subject to very large premia, as well as a graph showing the frequency distribution of patent citations.

patent citations. This is the subject of the next section. We then discuss the GPT-related measures we have constructed from the patent data, and show how our sample of highly-cited patents differs in various dimensions from the population as a whole.

2 *Patent citations*[6]

A key data item in the patent document is 'References Cited – US Patent Documents' (hereafter we refer to these just as 'citations'). Patent citations serve an important legal function, since they delimit the scope of the property rights awarded by the patent. Thus, if patent B cites patent A, it implies that patent A represents a piece of previously existing knowledge upon which patent B builds, and over which B cannot have a claim. The applicant has a legal duty to disclose any knowledge of the 'prior art,' but the decision regarding which patents to cite ultimately rests with the patent examiner, who is supposed to be an expert in the area and hence to be able to identify relevant prior art that the applicant misses or conceals. The presumption is thus that citations are informative of links between patented innovations. First, citations made may constitute a 'paper trail' for spillovers, i.e. the fact that patent B cites patent A may be indicative of knowledge flowing from A to B; second, citations ***received*** may be telling of the 'importance' of the cited patent.[7] The following quote provides support for the latter presumption:

> '..the examiner searches the…patent file. His purpose is to identify any prior disclosures of technology… which might be similar to the claimed invention and limit the scope of patent protection...or which, generally, reveal

[6] This description of the meaning of patent citations is drawn from Hall, Jaffe, and Trajtenberg (2002).

[7] See Jaffe, Trajtenberg and Fogarty (2000) for evidence from a survey of inventors on the role of citations in both senses.

> the state of the technology to which the invention is directed. If such documents are found...they are "cited"... if a single document is cited in numerous patents, the technology revealed in that document is apparently involved in many developmental efforts. Thus, the number of times a patent document is cited may be a measure of its technological significance.' (OTAF, 1976, p. 167)

The aspect of citations that is important for the present effort is that they provide a record of the link between the present invention and previous inventions. Thus they can tell us both the extent to which a particular line of technology is being developed (if they are made to patents in the same technology area) and whether a particular invention is used in wide variety of applications (if they are made to patents in different technology areas). In principle, given that we know which firm owns the relevant patents, it is possible to ask these question both using the technology field, which is a classification made by the US Patent Office,[8] and using the industry in which the patent falls, as indicated by the firm to which it is assigned.

3 Measures of GPTs

The definition of GPTs paraphrased in the introduction suggests that the following characteristics apply to the patents associated with GPT innovations: 1) they will have many citations from outside their particular technology area or perhaps from industries outside the one in which the patented invention was made; 2) they will have many citations within their technology area, and the citations will indicate a pattern of cumulative innovation, or trace out a technology trajectory; 3) more speculatively, citing

[8] The USPTO has developed over the years a highly elaborate classification system for the technologies to which the patented inventions belong, consisting of about 400 main (three-digit) patent classes, and over 120,000 patent subclasses. This system is being updated continuously, reflecting the rapid changes in the technologies themselves. Trajtenberg, Jaffe, and Hall have developed a higher-level classification, by which the 400 classes are aggregated into 36 two-digit technological sub-categories, and these in turn are

technologies will be subject to a burst of innovative activity as complementary goods are developed; and 4) given the length of time it takes for a GPT to pervade the economy, citation lags for patents in this area may be longer than average. In this section of the paper we report on the construction of a number of proxies for these characteristics. We use these proxies to identify patents that are in the extreme tail of the distribution of patent characteristics, in an effort to identify some candidate GPTs. Nor surprisingly, we find that looking at a single characteristics may be misleading, so in the later sections of the paper we use a more multivariate approach to refine the analysis.

It is well known that the distribution of patent values and patent citations is very skew with almost half of all patents receiving zero or one cite and less than 0.1 per cent receiving more than 100 cites (see Hall et al 2005 for evidence on both points). Observations (1) and (2) above also suggest that GPT patents are likely to be highly cited. Therefore we began our investigation by focusing on highly cited patents. We identified these patents by requiring that the number of citations the patent received be greater than three times the number received by the patent in the 99th percentile of the distribution. The results of this selection process are shown in Table 1. It yielded 780 patents granted between 1967 and 1999 that were ultimately granted, together with the name and type of their assignee (owner), the three-digit patent classification, and similar information on ALL the patents issued between 1975 and 2002 that cited this patent, for a total of 159,822 citations. Table 1 also makes it clear how skew the citation distribution is: our sample of 780 patents is about one out of 3700 patents, whereas the 160 thousand citations are one out of 100 citations (there are sixteen million citations in

further aggregated into six main categories: Chemical, Computers and Communications (C&C), Drugs and Medical (D&M), Electrical and Electronics (E&E), Mechanical, and Others.

all). Thus our patents are 37 times more likely to be cited than predicted by the average probability.

Table 1 about here

3.1 Generality

Observation (1) above suggests the use of a measure that is similar to the Trajtenberg, Jaffe, and Henderson 'Generality' measure, which is defined in the following way:

$$Generality \equiv G_i = 1 - \sum_j^{n_i} s_{ij}^2$$

where s_{ij} denotes the percentage of citations received by patent i that belong to patent class j, out of n_i patent classes (note that the sum is the Herfindahl concentration index). Thus, if a patent is cited by subsequent patents that belong to a wide range of fields the measure will be high, whereas if most citations are concentrated in a few fields it will be low (close to zero).[9]

Point (2) suggests that even if Generality is relatively high, the absolute number of citations should also be high, implying that there may still be a large number of citations in the patent's own technology class. It also suggests that 'second-generation' citations be examined. We implement this using two variables, the average number of citations to the citing patents, and the average generality of the citing patent.

In actual measurement, the preceding two predictions interact in ways which make our task a bit more complex. Like patents, citation counts are a discrete random variable bounded below by zero. This means that fewer citations in total imply that fewer classes will be observed to have citations than should be observed were the total

[9] Note that Generality is not defined if a patent receives no citations, and is zero by construction when a patent receives only one. We have omitted such patents in the tables and graphs shown in this paper. They comprise about one quarter of all patents in our sample.

number of citations larger. That is, n_i is biased downward by the fact that fractional citations are not observed, and Generality will tend to be lower when there are fewer citations. This is quite visible in the graph of average Generality over time shown in Figure 1, where we show two different versions of Generality, one based on US patent classes, and another based on the International Patent Classification, as assigned to these patents by the USPTO. Note that the average of either Generality measures begins to decline fairly steeply in 1993-95, at the same time as our measure of average citations per patent turns down sharply due to the effects of lag truncation (see Figure 2). In this case, this is a spurious rather than real decline in Generality, due to the fact that our patent grant data ends in 2002, and therefore our application-dated data around 1999, so that patents in the years after about 1994 have had less chance to receive citations.[10]

Figures 1 and 2 about here

Using a simple binomial model of the probability of observing a citation in a given cell, Hall (2002) shows that an unbiased estimate of the generality of the *i*th patent can be computed using the following correction:

$$\tilde{G}_i = \frac{N_i}{N_i - 1} G_i$$

where N_i is the number of citations observed. Note that this measure is not defined when $N_i < 2$ and will be fairly noisy when N_i is small. We have used this bias correction for the first three generality measures described below.

The US patent classification system has grown over time in ways that make it not ideal for the purpose we have in mind here. Generality measures essentially assume that

[10] A typical citation lag distribution is shown in Figure 2. This curve was estimated from the observed data using the methodology of Trajtenberg, Jaffe, and Henderson (1997). See Appendix D of Hall, Jaffe, and Trajtenberg (2000) for details. From the graph it appears that over half the citations ever made are made in the first six years since the cited patent's application date.

all categories are equidistant from each other if they are to be compared, but this is not the case for the US patent class system. Therefore, we explore the use of generality measures based on five different classification systems:

1. US patent class (approximately 400 cells).
2. Hall-Jaffe-Trajtenberg technology subcategories (36 cells).
3. Main International Patent Class (approximately 1200 cells).
4. Industry classification based on Silverman's IPC-SIC concordance (Silverman 2002) for industry of manufacture, aggregated to the Hall-Vopel (1997) level (37 cells).
5. Industry classification based on Silverman's IPC-SIC concordance for industry of use, aggregated to Hall-Vopel level (37 cells).

The rationales for these various choices are the following: First, in addition to using the US patent classification system (measure 1), we also constructed generality based on the more equal groupings of technologies constructed by Hall, Jaffe, and Trajtenberg (2002) from the patent classes (measure 2) and from the International Patent Classification system main four-digit classes (measure 3), which is more detailed that the US patent classification system.

Second, it could be argued that a GPT is not likely to manifest itself as a series of citations by patents in different technology classes, but rather as citations by firms in different industries. To consider this possibility, we would like to base a measure on the shares of citations that come from firms in different industries at the roughly two and one half-digit level. That is, our fourth measure is a Herfindahl for patent citation dispersion across industries rather than across technologies. Based on the discussion of GPT diffusion to using industries in the introduction, an industry-based measure would seem to be intrinsically preferred for this exercise. There are basically two ways to construct such a measure: the first uses the industry of ownership of the patents based on identification of the patent assignees and the second determines the industry for each patent class/subclass from some type of industry-patent class concordance. The first approach is difficult to implement in practice, given the number of patents that are unassigned, and the number of assignees that are not identifiable, either because they are small firms or because they are foreign firms for which we do not yet have a match to other data sources.[11]

Therefore we used the SIC-technology class concordances of Silverman (2002) to assign these patents and their citations to industries of manufacture and of use. Then we collapsed the distribution of citations by SIC codes into a 37-element vector of industries using the SIC-industry correspondence given in Appendix Table 1, and used this vector to construct the generality measures 4 and 5. The computation of these measures was able to make use of all the patents rather than just those held by US industry.[12] One major drawback of using the Silverman concordance ought to be mentioned, especially in the light of our subsequent findings: it is based on assignments to industry of manufacture and use made by Canadian patent examiners between 1990 and 1993. This means that it will do a poor job on patents in technologies related to the growth of the internet and software, because there were unlikely to be many of these in the Canadian patent system prior to 1994.

Figures 3 and 4 show the distribution of two of the computed Generality indices for the highly cited patents, Figure 3 the US patent class index and Figure 4 the index

---

[11] Slightly fewer than half the patents granted between 1967 and 1999 are assigned to US corporations that we can identify (see Hall et al 2002). However, many of these are in multiple industries so the primary industry assignment may not be relevant for the particular patent or citation that we are using.

[12]The actual industry classification we use was developed by Hall and Vopel (1997) from an earlier classification used by Levin and Reiss (1984). It is based on four-digit SICs aggregated up to a level that

based on Silverman's industry of manufacture map. As the figures show, these indices range from zero to one and the measure based on the industry of manufacture has a somewhat different distribution from that based on the US classification system. In Appendix Table A3, we show the correlation matrix for all 5 generality measures for our highly cited sample. Although they are generally fairly highly correlated, the industry of use measure (5) is not very correlated with the US class-based measures (1 and 3), and the industry of manufacture (4) not very correlated with the US class measure (1).

Figures 3 and 4 about here

Table 2 shows the twenty highly cited patents which also have the highest generality, where generality is measured by each of the five measures. In general, the most general patents are those in chemicals, especially when we consider the industry of use. Looking at the industry of manufacture, those in other technologies seem to be the most general. However, looking by US class, we can see the drawback of this generality measure: there are a number of chemical classes that are all essentially the same large class (the series 532-570), whereas in the case of some of the physics-based classes, there is only a single class. This fact will tend bias the index toward generality in the chemicals case; however, the fact that the IPC classification produces a similar result is somewhat reassuring.

Table 2 about here

3.2 Patenting growth

Observation (3) suggested that we look at patent classes with rapid growth in patenting. Using the entire patent database aggregated to patent class, we constructed three sub-periods (1975-83, 1984-92, and 1993-99) and computed the average growth

is coarse enough to include most, but not all, whole firms in the US manufacturing sector. We augmented this industry list with ten industries for the non-manufacturing sector. See Appendix Table A1 for details.

within class for each of the periods.[13] The results are shown in Table 3. As might be expected, in all three periods, the patent classes with rapid growth are dominated by the information and data processing classes (395 and 7xx), with the addition of the new multicellular biotechnology class 800 in the latter two periods. Highly cited patents are slightly more common in rapidly growing classes, although only a few of these classes have significant numbers of highly cited patents and the difference may not be very significant. It does appear that the patent classes that are growing rapidly include technologies that have more of the character of what we think of as GPTs, but that although highly cited patents are two to three times more likely to be found in rapidly growing classes (as we might expect if citations tend to come from the same class), they do not seem to be disproportionate in these classes.

Table 3 about here

Another way of looking at the growth in patenting following the introduction of a GPT is to look at the growth of the patent classes that cite such a technology. The hypothesis is that innovations which build on a GPTlike innovation will themselves spawn many new innovations. Table 4 shows the patent classes for the top twenty patents in terms of the growth of their citing patent classes, both for the highly cited patents and for all patents, excluding those that are highly cited. The message is clear: using this measure, almost all the patent classes identified are in computing and communications technology, and most are in data processing technologies more narrowly defined.

Table 4 about here

[13] We have omitted patent classes with fewer than 10 patents at the end of each period.

3.3 Citation lags

Finally, observation (4) suggested that the average citation lag to GPT patents might be longer. Of course, citations to patent data for a fixed time period such as ours (1967-1999) are always subject to truncation. For this reason, we look at mean citation lags that are large relative to the average citation lag for patents applied for in the same year.

Table 5 shows that the twenty highly cited patents with long lags (greater than 70 per cent of the average citation lag) are typically in older technologies. It is noteworthy that there are none in the chemicals or electrical industries and only five in computing and drugs, most of which are to surgical innovations. The only highly cited computing patent with long citation lags is a patent on an aspect of computer architecture taken out by Siemens in 1976; this patent has a mean citation lag of 23 years and is noteworthy because it is has essentially no citations until after it expired in 1994. It now has over 200. In general, given the fact that long lags by themselves often simply identify older and slower-moving technologies such as packaging, we will want to use this indicator in combination with our other indicators when looking for GPTs.

Table 5 about here

3.4 Summary

The GPT measures we have identified (the five generality indexes, the generality of citing patents, within class growth in patenting, growth in citing patent classes, and the average citation lag) are promising, but clearly give contradictory messages when taken separately. The goal is to combine them in a reasonable way to give an indication of the types of evidence GPTs leave in the patent statistics. We explore solutions to this problem in section 5 of this paper, but first we summarize the relationship between them and the probability that a patent is highly cited.

4 *Highly cited patents*

Table 6 shows that the highly cited patents differ in almost all respects from the population of all patents, and also from a four per cent sample of patents with at least one cite that we will use later as a control sample. They take longer to be issued, they have about twice as many claims, they are more likely to have a US origin, and more likely to be assigned to a US corporation, more likely to have multiple assignees, and have higher citation lags on average.[14] They also have higher generality, no matter how generality is measured, and are in patent classes that are growing faster than average. Although the patents that cite them more likely to be cited themselves, they have only slightly higher generality than citing patents in general.

Table 6 about here

More than half of the highly cited patents are in two of our six main technology classes: computing hardware and software, and drugs and medical instruments. Of course, these are indeed the technology classes where we expect to find modern day GPTs. In Appendix Table A2, we broke this down, in order to identify the important technologies more precisely. Highly cited patents are more than twice as likely to be found in computer hardware and software, computer peripherals, surgery and medical instruments, genetic technologies, miscellaneous drugs, and semiconductors.

Table 7 shows a series of probit estimations for the probability that a patent with at least one cite will be highly cited, in order to provide a multivariate summary of the data in Table 6.[15] The table shows the derivative of the probability with respect to the

[14] Unlike the case of generality measures, the mean citation lag is linear in the citation counts and therefore not a biased estimate, conditional on the total number of citations. It is, however, truncated at the end of the period, but this truncation affects both highly cited and non-highly cited patents equally.

[15] The sample used is the 780 highly cited patents plus the four per cent random sample of patents with at least one cite shown in Table 2. The fact that we use a random sample rather than a population affects the

independent variable that are implied by the coefficient estimates. In the case of dummy variables, it shows the change in probability when the variable changes from zero to one. Because the probability of being one of the highly cited patents in the sample is very small (0.77 per cent), the values in the table are small. Taking the grant lag as an example, the interpretation is that an additional year between application and grant increases the probability of being highly cited by 0.06 per cent, or from 0.77 per cent to 0.83 per cent at the mean. Being a patent in the drugs and medical category increases the probability by 2.9 per cent, which is a very large change at the mean probability.

Table 7 about here

The table confirms the univariate differences between highly cited and all patents. In addition, this table shows that variations over time in the probability of high citation do not greatly affect the coefficients (compare column 4 to column 2). The only generality measure that enters significantly and positively in this regression is that based on the US class; the others were all insignificant (IPC, technology subcategory, and industry of use) or slightly negative (industry of manufacture). Also note that highly cited patents are far more likely to be cited by patents that are themselves cited by patents in many technology classes, once we control for the other differences between highly cited and other patents.

5 Identifying GPT patents

It is not obvious how to combine these measures to choose a sample of GPT patents. In this first investigation of the topic, we have chosen simply to look for patents that are outliers in several of the categories, on the grounds that such patents are likely to

constant term in this probit regression, so we do not report it. The other coefficient estimates will not be affected by this procedure, although the interpretation of the changes in probability will depend on the average probability in the sample used.

give us an idea of the technologies that have given birth to the most subsequent inventive activity in the largest number of technological areas. Accordingly, we began with the 780 highly cites patents and then we chose a set of patents that fell in the top twenty per cent of these patents according to generality, citing patent generality, and the subsequent five year growth of the patent's class. We performed this exercise for each of the five generality measures in turn.

Table 8 shows the result: twenty patents out of the 780 were selected, many by several of the different criteria. Selection by each of the five generality measures is indicated by the presence of the measure in the table. Of these patents, all but two were in technologies related to information and communication technology (ICT). The remaining two are the oldest (applied for in 1970) and cover a process that is useful in the making of paper, and in sewage and waste treatment, and absorbable sutures for surgery. All but one of the patents cover US inventions, five from California, three from New Jersey, and the remainder from a number of other states. The sole exception comes from a Toronto-based company. All but one of the patents were assigned to corporations at the time they were taken out; the exception was a patent for a method of compressing audio and video data for transmission.

Table 8 about here

The ICT–related patents cover a range of technologies: integrated circuit manufacturing, handheld computers, spread spectrum technology, and so forth. What is noteworthy is the number of patents that relate to internet transactions (e-commerce) and software development, especially object-oriented programming. Some of the e-commerce patents greatly precede the actual use of the technology. For example, the celebrated Freeny patent (US4528643, shown in Figure 5) was applied for in 1983 and issued in 1985, but has been successfully asserted against such corporations as Microsoft and Apple almost to the present day.[16] The fact that the original use for which this patent was contemplated is unlikely to have been internet-based e-commerce reminds us to be cautious in our interpretation of the results in the table: we do not argue that the patents we identify are necessarily the source of the GPT itself, but we do suggest that by identifying them via the subsequent growth and generality in their citing patents, we are observing the symptoms of the diffusion and development of a General Purpose Technology.

Figure 5 about here

Looking at the actual ICT patents in Table 8 (rather than at the classes in which they have been placed) yields the following summary: seven are related to object-oriented and windows-based software, four to internet commerce and communication, three to audio-video applications, two to handheld computing, and 1 each to telecommunications and semiconductor manufacturing. Thus the specific technologies identified as being both general and spawning rapid patenting growth are those related to the effective use of the computer, especially for interacting and transacting over distance. That is, they are not computing hardware patents per se, but patents on the technologies that allow a network of computers to operate together effectively, and to interact with the users of those computers. This seems to us to characterize the GPT of the 1980s and 1990s, and we would therefore declare our prospecting exercise a qualified success.[17]

---

[16] This patent is currently owned by E-Data Corporation and was agreesively asserted by that company in the US beginning in 1996. http://www.prpnet.net/7604.html

[17] Note that the industry of manufacture and industry of use measures do not identify the software and internet patents as GPTs, for reasons discussed earlier: they have been obtained using a concordance that did not really admit these as patentable areas.

6 Conclusions

Many empirical papers close with interpretive cautions and calls for further research. This paper is no exception, but the caution and the call are stronger than usual. For reasons of limited time and computing power we have not been able to explore the validity and use of the measures we have constructed as much as we would like and we encourage further work in this area. In particular, all of the generality measures suffer from the fact that they treat technologies that are closely related but not in the same class in the same way that they treat very distant technologies. This inevitably means that generality may be overestimated in some cases and underestimated in others. One suggestion for future research would be to construct a weighted generality measure, where the weights are inversely related to the overall probability that one class cites another class.

A second area of concern has to do with changes in the strategic uses of patents during the two decades we have studied. These changes are not unrelated to the growth in importance of ICT technologies but they may also have had an distorting impact on some of the measure we have used. In particular, as Hall and Ziedonis (2001) have shown, one reason for rapid growth in semiconductor patenting after the mid-1980s is a conscious decision on the part of some major firms to build up their patent portfolios in order to fend off litigation and negotiate cross-licensing agreements. This type of strategy has spread throughout the industry and the consequences for patenting by firms such as IBM, Lucent, and Hewlett-Packard has been confirmed by Bessen and Hunt (2004) and Hall (2005). The implication is that citations to earlier patents in the ICT sector may be growing rapidly partly because of a strategic shift as well as because the underlying technology is growing in importance and diffusing throughout the economy. Sorting this out from our data will require more attention to the time series behaviour of the indicators, improved generality measures, and more detailed investigation of the firms involved. In the interim, this paper has demonstrated the potential validity of patent-based measures of GPTs and we hope it will spur further investigation into the use of patent data in this way.

References

Moses Abramovitz and Paul A. David (1996), ‘Technological change, intangible investments and growth in the knowledge-based economy: The US historical perspective,’ in D. Foray and B. A. Lundvall, ***Employment and growth in the knowledge-based economy***, Paris: OECD.

Bessen, James and Robert M. Hunt, (2004), ‘An Empirical Look At Software Patents,’ Philadelphia: Federal Reserve Bank Working Paper No. 03-17/R

Bresnahan, Timothy J., and Manuel Trajtenberg (1995), ‘General Purpose Technologies: ‘Engines of Growth’?’ ***Journal of Econometrics*** 95, 83-108.

David, Paul A. (1990), ‘The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox,’ ***American Economic Review*** 80, 355-61.

David, Paul A (1991), Computer and Dynamo: The Modern Productivity Paradox in a Not-too-distant Mirror,’ in ***Technology and Productivity***, Paris: OECD. Also Stanford University: CEPR Discussion Paper No. 172 (1989).

David, Paul A., and Gavin Wright (2003), ‘General Purpose Technologies and Productivity Surges: Historical Reflections on the Future of the ICT Revolution,’ in P. A. David and M. Thomas (eds.), ***The Economic Future in Historical Perspective***, Oxford: Oxford University Press for the British Academy.

Griliches, Zvi (1990), 'Patent Statistics as Economic Indicators: A Survey,' ***Journal of Economic Literature*** 28, 1661-1707.

Hall, Bronwyn H. (2005), 'Exploring the Patent Explosion,' ***Journal of Technology Transfer*** 30, 35-48.

Hall, Bronwyn H (2002), 'A Note on the Bias in the Herfindahl Based on Count Data,' in Jaffe, A. and M. Trajtenberg (eds.), ***Patents, Citations, and Innovation***, Cambridge, MA: MIT Press.

Hall, Bronwyn H., and Katrin Vopel (1997), 'Market Value, Market Share, and Innovation,' NBER, the University of California at Berkeley, and the University of Mannheim. Industry classification available at http://emlab.berkeley.edu/users/bhhall/index.html

Hall, Bronwyn H. and Rosemarie Ham Ziedonis (2001), 'The Determinants of Patenting in the U.S. Semiconductor Industry, 1980-1994,' ***Rand Journal of Economics*** 32, 101-128.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg (2002), 'The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools,' in A. Jaffe and M. Trajtenberg (eds), ***Patents, Citations and Innovations***, Cambridge, MA: The MIT Press. Also NBER Working Paper No. 8498 (October 2001).

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg (2005), 'Market Value and Patent Citations' ***Rand Journal of Economics***, Spring. Also NBER Working Paper No. 7741 (June 2000).

Harhoff, Dietmar, Francis Narin, F.M. Scherer, and Katrin Vopel (1999), "Citation Frequency and the Value of Patented Inventions," ***The Review of Economics and Statistics*** 81(3), 511-515.

Helpman, Elhanan and Trajtenberg, Manuel (1998a), 'Diffusion of General Purpose Technologies,' in Helpman, E. (ed.), ***General Purpose Technologies and Economic Growth***, Cambridge: MIT Press, pp. 85-119.

Helpman, Elhanan and Trajtenberg, Manuel (1998b), 'A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies,' in Helpman, E. (ed.), ***General Purpose Technologies and Economic Growth***, Cambridge: MIT Press, pp. 85-119.

Jaffe, Adam., Manuel Trajtenberg, and Michael Fogarty (2000), 'Knowledge Spillovers and Patent Citations: Evidence from A Survey of Inventors,' ***American Economic Review***, Papers and Proceedings (May), 215-218.

Office of Technology Assessment and Forecast (OTAF) (1976), ***US Department of Commerce, Patent and Trademark Office, Sixth Report***, Washington, DC: US Government Printing Office.

Pavitt, Keith (1988), 'Uses and Abuses of Patent Statistics,' in A. F. J. van Raan *(ed.)*, ***Handbook of Quantitative Studies of Science and Technology***, Amsterdam: Elsevier Science Publishers, pp. 509-536.

Rosenberg, Nathan (1976), 'Technological Change in the Machine Tool Industry, 1840-1910,' in ***Perspectives in Technology***, Cambridge: Cambridge University Press, pp. 9-31.

Sakakibara, Mariko and Lee G. Branstetter (2001), 'Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms,' ***Rand Journal of Economics*** 32 (1), 77-100.

Silverman, Brian (2002), 'International Patent Classification – US SIC Concordance,' http://www.rotman.utoronto.ca/~silverman/ipcsic/ipcsicfiles.ZIP

Trajtenberg, Manuel, Adam Jaffe, and Rebecca Henderson (1997), 'University versus Corporate Patents: A Window on the Basicness of Invention,' ***Economics of Innovation and New Technology*** 5 (1), 19-50.

Figure 1
Average Measures of Generality and Originality

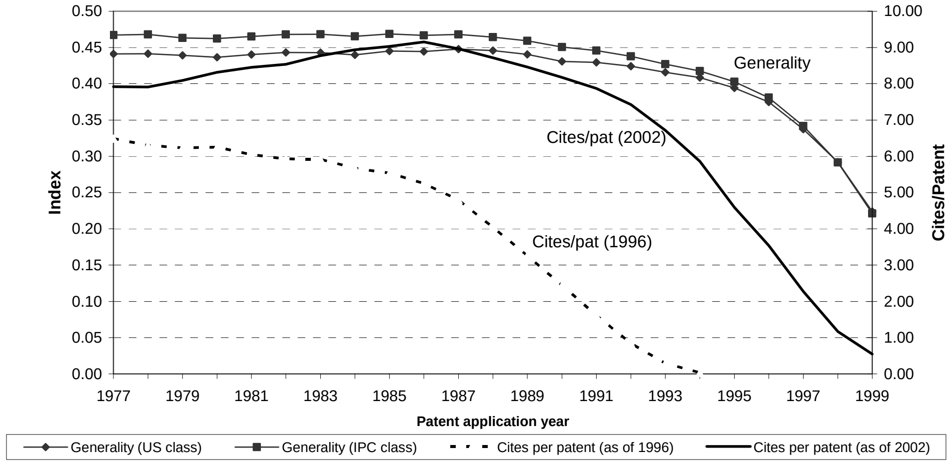
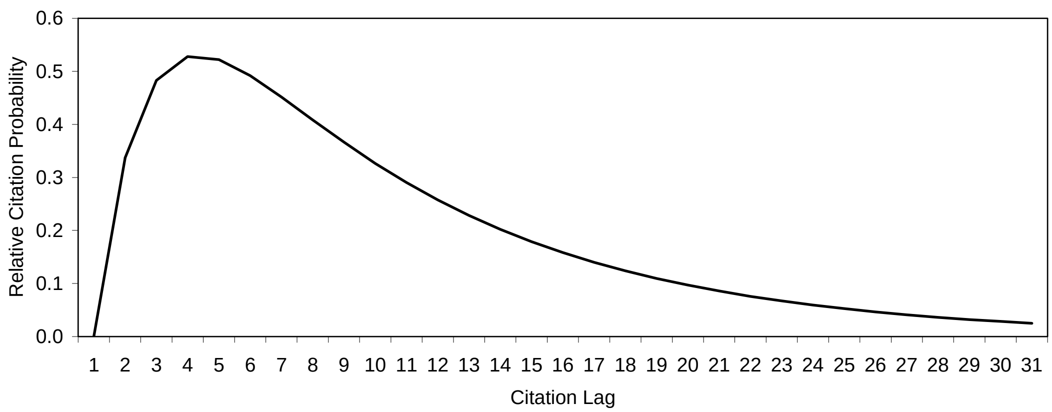


Figure 2
Citation Lag Distribution (1976-1994)
Trajtenberg, Jaffe, and Henderson Methodology



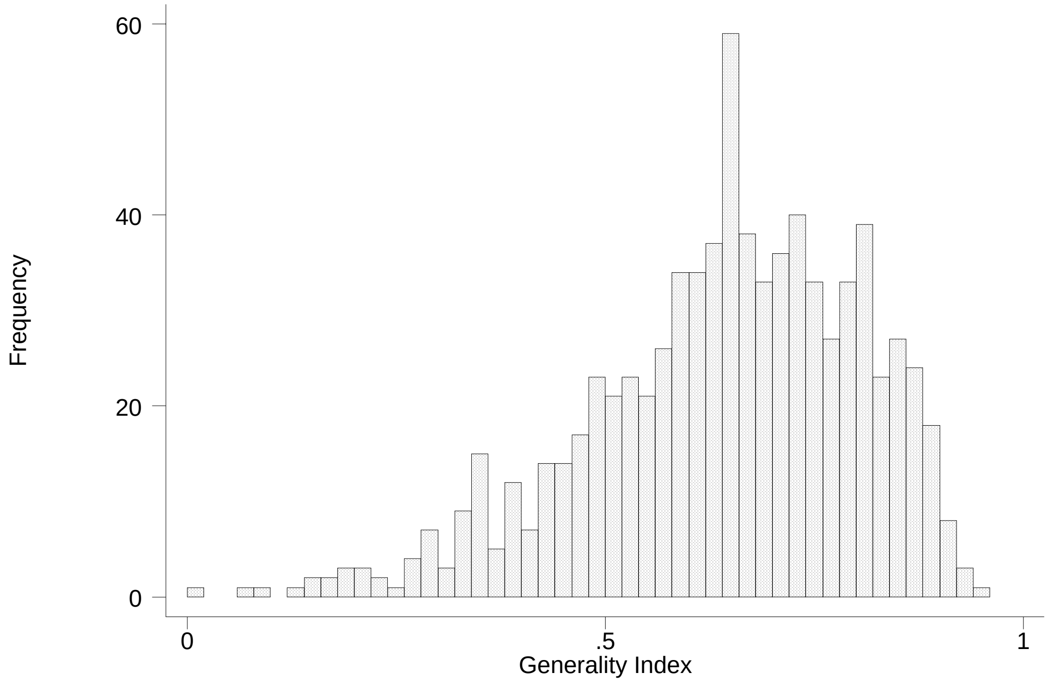


Figure 3

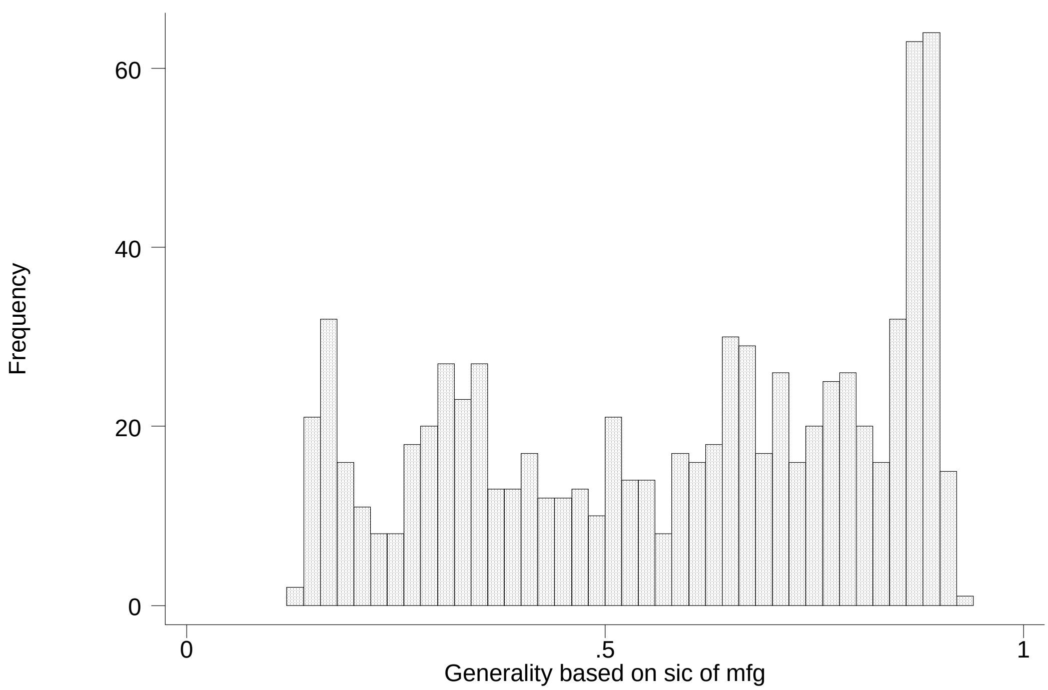


Figure 4

Figure 5
The Freeny Patent (4528643)

United States Patent ***4,528,643***

Freeny, Jr. **July 9, 1985**

System for reproducing information in material objects at a point of sale location

Abstract

The present invention contemplates a system for reproducing information in material objects at a point of sale location wherein the information to be reproduced is provided at the point of sale location from a location remote with respect to the point of sale location, an owner authorization code is provided to the point of sale location in response to receiving a request code from the point of sale location requesting to reproducing predetermined information in a material object, and the predetermined information is reproduced in a material object at the point of sale location in response to receiving the owner authorization code.

Inventors: **Freeny, Jr.; Charles C.** (Fort Worth, TX)

Assignee: **FPDC, Inc.** (Oklahoma City, OK)

Appl. No.: **456730**

Filed: **January 10, 1983**

http://www.e-data.com/e-freeny.htm

Table 1

Selecting the Sample of Highly Cited Patents

Grant Year	Cutoff No. Citations	Highly Cited Patents Number	Highly Cited Patents Median Cites	All Patents Number	All Patents Median Cites	Highly cited Share
1967	78	20	103	65,652	2	0.030%
1968	84	15	131	59,104	2	0.025%
1969	87	28	115	67,559	3	0.041%
1970	93	19	105	64,429	3	0.029%
1971	99	22	165	78,317	3	0.028%
1972	108	30	136	74,810	4	0.040%
1973	111	25	132	74,143	4	0.034%
1974	117	24	136.5	76,278	5	0.031%
1975	126	19	152	73,690	5	0.026%
1976	132	22	156.5	72,015	5	0.031%
1977	135	17	182	66,883	5	0.025%
1978	138	16	179	67,862	5	0.024%
1979	141	7	190	50,177	5	0.014%
1980	144	26	204.5	63,371	5	0.041%
1981	147	17	177	67,373	5	0.025%
1982	150	19	232	59,462	5	0.032%
1983	156	20	224.5	58,435	5	0.034%
1984	159	22	188.5	69,338	5	0.032%
1985	159	25	199	73,824	5	0.034%
1986	168	21	204	72,977	6	0.029%
1987	180	29	213	85,522	6	0.034%
1988	177	29	252	80,345	6	0.036%
1989	177	27	258	98,567	5	0.027%
1990	177	27	215	93,290	5	0.029%
1991	171	38	204.5	99,789	5	0.038%
1992	174	41	218	100,760	5	0.041%
1993	174	33	231	100,980	4	0.033%
1994	171	26	214.5	104,317	4	0.025%
1995	156	21	178	104,091	4	0.020%
1996	141	13	171	112832	3	0.012%
1997	114	28	136	115337	3	0.024%
1998	90	33	103	151745	2	0.022%
1999	63	21	69	153486	1	0.014%
All years		780	183	2,756,760	3	0.028%

*Patents with zero cites 1975-2002 are excluded

Table 2

Number of Top 20 Highly Cited Patents

HJT Sub-category	US patent class	Class description	US class	IPC	Generality measure: US subcategory	Industry of manufacture	Industry of use
11	442	Textiles: Web or Sheet Containing Structurally De	0	1	4	1	3
12		Coatings	0	0	1	1	0
	106	Compositions: Coating or Plastic	0	0	0	1	0
	118	Coating Apparatus	0	0	1	0	0
14	540,556,568	Organic Compounds -- Part of the Class 532-570 Series	3	3	0	1	0
15	521,523,524,528	Synthetic Resins or Natural Rubbers -- Part of the Class 520 Series	3	5	1	1	1
19		Miscellaneous chemicals	0	0	0	2	6
	156	Adhesive Bonding and Miscellaneous Chemicals	0	0	0	0	2
	366	Agitating	0	0	0	0	1
	430	Radiation Imagery Chemistry: Process, Composition, or Product Thereof	0	0	0	1	2
	510	Cleaning Compositions for Solid Surface	0	0	0	1	1
		Total chemicals	**6**	**9**	**6**	**6**	**10**
21	21	Communications	0	2	1	0	0
	340	Communications: Electrical	0	0	1	0	0
	342	Communications: Directive Radio Wave Systems & Devices (e.g., Radar, Radio Na	0	1	0	0	0
	455	Telecommunications	0	1	0	0	0
23	345	Selective Visual Display Systems	1	0	0	0	0
		Total computing	**1**	**2**	**1**	**0**	**0**
32		Surgery & Med Inst.	3	2	4	2	3
	128	Surgery	1	2	2	2	1
	604	Surgery	2	0	2	0	2
		Total drugs & medical instruments	**3**	**2**	**4**	**2**	**3**
41	174	Electricity: Conductors and Insulators	1	1	1	1	1
46	257	Active Solid-State Devices (e.g., Transistors)	1	1	1	0	0
49		Miscellaneous electrical	4	0	0	0	0
	348	Television	3	0	0	0	0
	386	Television Signal Processing for Dynamic Recording or Reproducing	1	0	0	0	0
		Total electrical	**6**	**2**	**2**	**1**	**1**
51	264	Plastic and Nonmetallic Article Shaping	3	3	4	2	1
54	359	Optics: Systems (including Communications)	0	0	0	0	2
59	49	Movable or Removable Closures	0	1	0	0	0
		Total mechanical	**3**	**4**	**4**	**2**	**3**
67	138	Pipes and Tubular Conduits	0	0	1	1	0
68	53	Package Making	0	0	0	1	0
69	248	Supports	1	1	2	6	3

Total other	**1**	**1**	**3**	**8**	**3**

Table 3

U. S. Patent Classes with Rapid Growth

Patents in 1999	Highly cited patents 1993-99		Annual growth 1992-99		
Number	**Number**	**Share (%)**	**(%)**	**Class description**	**Class**
335	0	0.00%	22.2%	Data Processing: Design and Analysis of Circuit or Semiconductor Mask	716
90	0	0.00%	22.2%	Interactive Video Distribution Systems	725
311	0	0.00%	22.1%	Data Processing: Software Development, Installation, or Management	717
152	12	7.89%	21.6%	Data Processing: Structural Design, Modeling, Simulation, and Emulation	703
523	7	1.34%	15.6%	Multicellular Living Organisms and Unmodified Parts Thereof and Related Processes	800
1061	0	0.00%	13.1%	Data Processing: Database and File Management, Data Structures, or Document Processing	707
4742	0	0.00%	13.0%	Semiconductor Device Manufacturing: Process	438
255	0	0.00%	12.1%	Amusement Devices: Games	463
37	0	0.00%	12.0%	Foundation Garments	450
42	0	0.00%	11.7%	Chemistry: Fischer-Torpsch Processes; or Purification or Recovery of Products Thereof	518
7548	19	0.25%			
146045	**148**	**0.10%**		**All classes**	
Patents in 1992	**Highly cited patents 1984-92**		**Annual growth 1983-92 (%)**	**Class description**	**Class**
240	0	0.00%	22.2%	Superconductor Technology: Apparatus, Material, Process	505
204	0	0.00%	21.8%	Data Processing: Artificial Intelligence	706
91	0	0.00%	18.2%	Multicellular Living Organisms and Unmodified Parts Thereof and Related Processes	800
300	4	1.33%	17.6%	Electrical Computers and Digital Processing Systems: Multiple Computer or Process	709
273	4	1.47%	17.2%	Data Processing: Database and File Management, Data Structures, or Document Processing	707
264	4	1.52%	16.1%	Information Processing System Organization	395
18	0	0.00%	15.9%	Textiles: Cloth Finishing	26
221	4	1.81%	15.7%	Electrical Computers and Digital Processing Systems: Support	713
32	0	0.00%	14.2%	Roll or Roller	492
27	0	0.00%	14.1%	Railway Switches and Signals	246
1670	16	0.96%		Total for selected classes	
106626	**259**	**0.24%**		**All classes**	
Patents in 1983	**Highly cited patents 1975-83**		**Annual growth 1975-83 (%)**	**Class description**	**Class**
42	0	0.00%	18.8%	Information Processing System Organization	395
96	0	0.00%	13.7%	Data Processing: Speech Signal Processing, Linguistics, Language Translation, and	704
38	1	2.63%	13.0%	Electrical Computers and Digital Processing Systems: Support	713
75	1	1.33%	12.1%	Electrical Computers and Digital Processing Systems: Processing Architectures and	712
126	1	0.79%	11.6%	Data Processing: Vehicles, Navigation, and Relative Location	701
223	0	0.00%	11.3%	Data Processing: Generic Control Systems or Specific Applications	700
54	3	5.56%	11.0%	Data Processing: Financial, Business Practice, Management, or Cost/Price Determination	705
135	0	0.00%	10.9%	Data Processing: Measuring, Calibrating, or Testing	702
35	0	0.00%	10.7%	Electrical Computers and Digital Processing Systems: Multiple Computer or Process	709
257	0	0.00%	10.5%	Error Detection/Correction and Fault Detection/Recovery	714
1081	6	0.56%		Total for selected classes	
63383	**322**	**0.51%**		**All classes**	

Patent classes with fewer than 10 patents at the end of each period have been omitted from the table.

Table 4

Patent classes with whose cited classes have high growth rates*

HJT Sub-category	US patent class	Class description	Highly cited patents	All patents excl. highly cited
		Total chemicals	**0**	**0**
21		Communications	4	
	370	Multiplex Communications	3	
	379	Telephonic Communications	1	
22		Computer Hardware & Software	11	18
	380	Cryptography	1	
	395	Information Processing System Organization		16
	704	Data Processing: Speech Signal Processing, Linguistics, Language	1	
	705	Data Processing: Financial, Business Practice, Management, or	2	
	707	Data Processing: Database and File Management, Data Structure	1	
	709	Electrical Computers and Digital Processing Systems: Multiple	3	
	712	Electrical Computers and Digital Processing Systems: Processimg		1
	713	Electrical Computers and Digital Processing Systems: Support	3	
	717	Data Processing: Software Development, Installation, or Management		1
23	345	Selective Visual Display Systems	3	1
		Total computing	**18**	**19**
33	435	Chemistry: Molecular Biology and Microbiology	1	
		Total drugs & medical instruments	**1**	**0**
49		Miscellaneous electrical		
	348	Television	2	
	505	Semiconductor technology: apparatus, etc		1
		Total electrical	**2**	**1**
		Total mechanical	**0**	**0**
		Total miscellaneous	**0**	**0**

*Classes for the top 20 patents in each category are shown. The average growth rate of the classes of the highly cited patents is above 28 per cent per annum; those for all patents above 52 per cent per annum.

Table 5

Patent classes with highly cited patents that have long cite lags

HJT Sub-category	US patent class	Class description	Number of patents
		Total chemicals	**0**
24	365	Static Information Storage and Retrieval	1
		Total computing	**1**
32	604,606	Surgery & Med Instruments	3
39	623	Prosthesis (i.e., Artificial Body Members), Parts Thereof, or A	1
		Total drugs & medical instruments	**4**
		Total electrical	**0**
51		Mat. Proc & Handling	2
	264	Plastic and Nonmetallic Article Shaping or Treating: Processes	1
	425	Plastic Article or Earthenware Shaping or Treating: Apparatus	1
54	359	Optics	2
		Total mechanical	**4**
61	47	Agriculture,Husbandry,Food	3
68		Receptacles	7
	53	Package Making	1
	206	Special Receptacle or Package	3
	383	Flexible Bags	3
69	428	Stock Material or Miscellaneous Articles	1
		Total miscellaneous	**11**

Table 6

U. S. Patents Granted 1967-1999

Statistic	All Patents	4% sample of patents (>1 cite)	Highly Cited Patents
Number of patents	2,768,011	100,634	780
Year applied for	1983.1	1982.8	1981.6
Year granted	1984.6	1984.8	1983.9
Average grant lag (years)	1.5	2.0	2.3
Number of claims	12.1	12.5	23.6
Number of forward citations (to 2002)	6.72	8.85	204.71
Average citation lag	8.73	9.98	13.48
Average class growth (5 years)	NA	2.98%	7.28%
Share US origin	59.9%	61.2%	88.3%
Share assigned to US corporations	46.5%	47.7%	75.9%
Share multiple assignees	0.5%	0.6%	1.0%
Generality 1 (US class)	0.3417	0.5261	0.6416
Generality 2 (IPC)	0.3548	0.5484	0.5716
Generality 3 (US subcategory)	0.2711	0.4167	0.4569
Generality 4 (SIC of mfg-IPC)	NA	NA	0.5856
Generality 5 (SIC of use-IPC)	NA	NA	0.6444
Average cites to citing patents	4.02	4.70	12.92
Total cites to citing patents	46.5	55.7	2663.9
Average growth of citing patent classes*	NA	3.57%	7.69%
Average generality of citing patents	0.3094	0.3487	0.3887
Broad technology classes			
Chemicals	20.8%	19.2%	18.0%
Computing	10.2%	11.4%	23.9%
Drugs & medical	7.3%	7.0%	32.6%
Electrical	17.1%	17.8%	9.7%
Mechanical	23.0%	22.9%	6.1%
Other	21.6%	21.8%	9.7%
Type of Assignee			
US corporation	1247030	47975	596
non-US corporation	885533	31798	75
US individual	378394	14347	98
non-US individual	135756	4645	9
US government	43048	1499	6
non-US government	9845	370	1

Table 7

Probit Regression for Highly Cited Patents (101,414 observations; 780 highly cited)*

Variable	**Cited patent characteristics**		**Cited & citing patent char.**		**Cited & citing patent char.**		**Including year dummies**	
	dp/dx	Std. err	dp/dx	Std. err	dp/dx	Std. err	dp/dx	Std. err
Number of claims/10	0.124%	0.011%	0.062%	0.006%	0.076%	0.007%	0.055%	0.006%
D (claims missing)+	0.400%	0.063%	*0.033%*	*0.025%*	*0.039%*	*0.029%*	*0.508%*	*0.452%*
Average grant lag (years)	0.062%	0.015%	0.057%	0.008%	0.066%	0.009%	0.054%	0.008%
Dummy for US origin+	0.397%	0.047%	0.146%	0.025%	0.163%	0.028%	0.137%	0.023%
Dummy for US corporation+	0.210%	0.046%	0.147%	0.026%	0.166%	0.029%	0.132%	0.024%
Generality 1 (US class)			0.226%	0.032%			0.202%	0.030%
Generality 5 (SIC of use)					*-0.096%*	*0.063%*		
Average citation lag (relative to year average)			0.068%	0.005%	0.078%	0.005%	0.066%	0.005%
Average generality of citing patents			0.331%	0.044%	0.470%	0.050%	0.334%	0.044%
Dummies for technology classes**								
Chemicals+	0.308%	0.085%	0.201%	0.052%	0.239%	0.060%	0.191%	0.049%
Computing+	1.085%	0.162%	1.007%	0.147%	1.005%	0.158%	0.904%	0.138%
Drugs & medical+	2.882%	0.318%	2.432%	0.291%	2.454%	0.292%	2.222%	0.275%
Electrical+	*0.079%*	*0.072%*	0.118%	0.048%	0.119%	0.052%	0.106%	0.044%
Mechanical+	*-0.157%*	*0.057%*	*-0.077%*	*0.029%*	*-0.085%*	*0.033%*	*-0.069%*	*0.027%*
Year dummies	no		no		no		yes	
Scaled R-squared	0.130		0.222		0.217		0.229	
Log likelihood	-3980.69		-3556.96		-3583.02		-3528.15	

Coefficient estimates in italics are not significant at the one per cent level.

*The sample of non-highly-cited patents is a 10 per cent sample of all patents that have 2 or more citations. The average probability of being highly cited in the sample is 0.77%.

**The excluded class is other technologies.

***Estimated derivative of probability with respect to independent variable. For dummy variables (+), the discrete change in probability from 0 to 1.

Table 8

Highly Cited Patents with High Generality, Class Growth, and Citing Patent Generality

Patent number	Number of Cites	Application year	Inventor state, country	Assignee	Patent description	Generality: By US class	Generality: By IPC	Generality: By tech sub-category	Generality: By ind of manu-facture	Generality: By industry of use
3624019	129	1970	IL, US	Nalco Chemical Company	Process for Rapidly Dissolving Water-soluble Polymers	0.846	0.907	0.659	0.856	0.798
3636956	181	1970	DE, US	Ethicon, Inc.	Polyactide sutures (absorbable)		0.841	0.696		0.825
3842194	125	1971	NJ, US	RCA Corporation	Information records and recording playback system therefore (video disc)		0.843	0.730		0.830
3956615	178	1974	CA, US	IBM	Transaction execution system with secure data storage and communications	0.801				
4528643	186	1983	TX, US	FPDC (Freeny patent)	System for reproducing information in material objects in a point of sale location	0.880	0.797	0.696		
4558413	377	1983	CA, US	Xerox	Software version management system	0.826				
4575621	186	1984	PA, US	Corpra Research Inc	Portable electronic transaction device and system therefor		0.804	0.714		
4672658	200	1986	NJ, US	AT&T	Spread spectrum wireless PBX		0.844			
4783695	195	1986	NY, US	General Electric Co	Multichip integrated circuit packaging configuration and method	0.824		0.674		
4821220	180	1986	WA, US	Tektronix	System for animating program operation and displaying time-based relationships	0.796				
4885717	183	1986	OR, US	Tektronix	System for graphically representing operation of object-oriented programs	0.816				
4916441	286	1988	CO, US	Clinicom Inc	Portable handheld terminal	0.912	0.824	0.827		
4953080	181	1988	CA, US	Hewlett-Packard Co	Object management facility for maintaining data in a computer system	0.794				
5133075	210	1988	CA, US	Hewlett-Packard Co	Method of monitoring changes in attribute values of object in an object-oriented database	0.812				
5155847	224	1988	ON, CA	Minicom Data Corp	Method and apparatus for updating software at remote locations	0.870				
5093914	217	1989	IL, US	AT&T	Method of controlling the execution of object-oriented programs	0.832				
5347632	255	1989	NJ, US	Prodigy Services Co	Reception system for an interactive computer network and method of operation	0.796				
5119475	200	1991	TX, US	Schlumberger Technology Corp	Object-oriented framework for menu definition	0.810				
5132992	178	1991	NY, US	*unassigned*	Audio and video transmission and receiving system (compression)	0.856		0.764		
5307456	173	1992	CA, US	Sony Electronics Inc	Integrated multi-media production and authoring system			0.682		

Table 8 (cont.)

Highly Cited Patents with High Generality, Class Growth, and Citing Patent Generality

Citing patents	Patent number	Mean cite lag	Mean cite lag relative to average	Sub-category	US class	Growth of class	Class description
0.566	3624019	17.3	1.9	15	523	15.2%	Synthetic Resins or Natural Rubbers -- Part of the Class 520 Series
0.573	3636956	21.5	6.1	32	606	13.0%	Surgery
0.546	3842194	8.6	-6.0	24	369	12.9%	Dynamic Information Storage or Retrieval
0.658	3956615	18.2	4.4	22	705	20.7%	Data Processing: Financial, Business Practice, Management, or Cost
0.616	4528643	14.6	5.0	22	705	14.7%	Data Processing: Financial, Business Practice, Management, or Cost
0.547	4558413	13.1	3.5	22	707	18.9%	Data Processing: Database and File Management, Data Structures
0.615	4575621	10.4	1.2	59	235	13.3%	Registers
0.521	4672658	9.9	1.4	21	455	13.2%	Telecommunications
0.609	4783695	9.9	1.4	46	257	20.1%	Active Solid-State Devices (e.g., Transistors, Solid-State Diodes)
0.570	4821220	8.8	0.3	22	395	19.9%	Information Processing System Organization
0.581	4885717	9.1	0.6	22	395	19.9%	Information Processing System Organization
0.539	4916441	9.2	1.4	23	345	15.1%	Selective Visual Display Systems
0.593	4953080	8.0	0.2	22	707	18.6%	Data Processing: Database and File Management, Data Structures
0.557	5133075	6.5	-1.3	22	707	18.6%	Data Processing: Database and File Management, Data Structures
0.541	5155847	8.3	0.5	22	709	30.2%	Electrical Computers and Digital Processing Systems: Multiple Computers
0.593	5093914	6.6	-0.7	22	395	18.6%	Information Processing System Organization
0.542	5347632	6.5	-0.9	22	709	24.1%	Electrical Computers and Digital Processing Systems: Multiple Computers
0.566	5119475	6.5	0.0	23	345	26.5%	Selective Visual Display Systems
0.707	5132992	6.6	0.1	21	375	18.5%	Pulse or Digital Communications
0.539	5307456	5.9	-0.1	23	345	23.7%	Selective Visual Display Systems

Table A1
SIC-Industry Correspondence for Generality Indices

Hall-Vopel Quasi 2-digit Industry	SIC Codes (1987)
01 Food & tobacco	20xx, 21xx
02 Textiles, apparel & footwear	22xx, 23xx, 31xx, 3021, 3961, 3965
03 Lumber & wood products	24xx
04 Furniture	25xx
05 Paper & paper products	26xx
06 Printing & publishing	27xx
07 Chemical products	28xx, excl 283x, 284x
08 Petroleum refining & prods	13xx, 29xx
09 Plastics & rubber prods	30xx, excl 3021
10 Stone, clay & glass	32xx
11 Primary metal products	33xx
12 Fabricated metal products	34xx
13 Machinery & engines	35xx, excl 357x, 358x, 3524
14 Computers & comp. equip.	357x
15 Electrical machinery	358x, 3596, 360x, 361x, 362x, 363x, 364x, 3677, 369x excl. 3690, 3695
16 Electronic inst. & comm. eq.	3651, 3652, 366x, 367x excl 3677,3678; 3690, 3695, 381x, 382x excl 3827
17 Transportation equipment	372x, 373x, 374x, 376x-379x, excl 3790, 3792, 3799
18 Motor vehicles	371x, excl 3714; 375x, 3790, 3792, 3799
19 Optical & medical instruments	3827, 384x, 386x
20 Pharmaceuticals	283x, 3851
21 Misc. manufacturing	387x, 39xx, excl 3961, 3965
22 Soap & toiletries	284x
23 Auto parts	3714
24 Computing software	737x
25 Telecommunications	48xx
26 Wholesale trade	50xx
27 Business services	73xx, excl 737x
28 Agriculture	01xx-09xx
29 Mining	10xx, 11xx, 12xx, 14xx
30 Construction	15xx-19xx
31 Transportation services	40xx-47xx
32 Utilities	49xx
33 Trade	51xx-59xx
34 Fire, Insurance, Real Estate	60xx-69xx
35 Health services	80xx
36 Engineering services	87xx
37 Other services	70xx-99xx and not 73xx, 80xx, 87xx

Table A2

Breakdown by Technology Subcategory

Subcategory	All Patents Number	All Patents Share	Highly Cited Patents Number	Highly Cited Patents Share	Ratio
Agriculture,Food,Textiles	24,134	0.9%	8	1.0%	1.17
Coating	42,235	1.5%	7	0.9%	0.58
Gas	13,614	0.5%	4	0.5%	1.04
Organic Compounds	116,334	4.2%	13	1.7%	0.39
Resins	96,948	3.5%	40	5.1%	1.45
Miscellaneous	282,717	10.2%	69	8.8%	0.86
Chemical technologies	**575,982**	**20.8%**	**141**	**18.0%**	**0.86**
Communications	118,316	4.3%	51	6.5%	1.52
Computer Hardware & Software	90,326	3.3%	93	11.8%	3.63
Computer Peripherials	24,147	0.9%	28	3.6%	4.09
Information Storage	49,963	1.8%	16	2.0%	1.13
Computer hardware & software	**282,752**	**10.2%**	**188**	**23.9%**	**2.34**
Drugs	83,410	3.0%	35	4.5%	1.48
Surgery & Med Instruments	69,344	2.5%	164	20.9%	8.34
Genetics	31,794	1.1%	24	3.1%	2.66
Miscellaneous	16,312	0.6%	33	4.2%	7.13
Drugs & med. instruments	**200,860**	**7.3%**	**256**	**32.6%**	**4.49**
Electrical Devices	92,508	3.3%	1	0.1%	0.04
Electrical Lighting	44,738	1.6%	6	0.8%	0.47
Measuring & Testing	80,315	2.9%	2	0.3%	0.09
Nuclear & X-rays	40,746	1.5%	4	0.5%	0.35
Power Systems	97,739	3.5%	4	0.5%	0.14
Semiconductor Devices	51,950	1.9%	38	4.8%	2.58
Miscellaneous	66,440	2.4%	21	2.7%	1.11
Electrical technologies	**474,436**	**17.1%**	**76**	**9.7%**	**0.56**
Mat. Proc & Handling	155,200	5.6%	16	2.0%	0.36
Metal Working	88,661	3.2%	11	1.4%	0.44
Motors & Engines + Parts	102,504	3.7%	1	0.1%	0.03
Optics	62,832	2.3%	4	0.5%	0.22
Transportation	82,854	3.0%	0	0.0%	0.00
Miscellaneous	143,849	5.2%	16	2.0%	0.39
Mechanical technologies	**635,900**	**23.0%**	**48**	**6.1%**	**0.27**
Agriculture,Husbandry,Food	59,793	2.2%	19	2.4%	1.12
Amusement Devices	28,095	1.0%	0	0.0%	0.00
Apparel & Textile	50,477	1.8%	0	0.0%	0.00
Earth Working & Wells	40,857	1.5%	0	0.0%	0.00
Furniture,House Fixtures	57,362	2.1%	0	0.0%	0.00
Heating	38,146	1.4%	0	0.0%	0.00
Pipes & Joints	25,198	0.9%	3	0.4%	0.42
Receptacles	58,616	2.1%	32	4.1%	1.93
Miscellaneous	239,537	8.7%	22	2.8%	0.32
Other technologies	**598,081**	**21.6%**	**76**	**9.7%**	**0.45**
All technologies	**2,768,011**	**100.0%**	**785**	**100.0%**	**1.00**

Table A3

Correlation matrix for Generality Indices (N=780)

	1 US class	2 IPC	3 US subcategory	4 Industry of manufacture	5 Industry of use
Generality 1 (US class)	1.000				
Generality 2 (IPC)	0.555	1.000			
Generality 3 (US subcategory)	0.621	0.523	1.000		
Generality 4 (industry of manufacture)	0.238	0.590	0.599	1.000	
Generality 5 (industry of use)	0.143	0.627	0.389	0.632	1.000