Papers in Evolutionary Economic Geography

20.27

Mapping General Purpose Technologies with Patent Data

Sergio Petralia

Utrecht University Human Geography and Planning

Mapping General Purpose Technologies with Patent Data

Sergio Petralia*

July 17, 2020

Abstract

This article develops a three-dimension indicator to capture the main features of General Purpose Technologies (GPTs) in patent data. Technologies are evaluated based on their scope for improvement and elaboration, the variety of products and processes that use them, and their complementarity with existing and new technologies. Technologies' scope for improvement is measured using patenting growth rates. The range of its uses is mapped by implementing a text-mining algorithm that traces technology-specific vocabulary in the universe of all available patent documents. Finally, complementarity with other technologies is measured using the co-occurrence of technological claims in patents. These indicators are discussed and evaluated using widely studied examples of GPTs such as Electric & Electronic (at the beginning of the 20^{th} century) and Computer & Communications. These measures are then used to propose a simple way of identifying GPTs with patent data. It is shown there exist

*Department of Economic Geography, Utrecht University, Utrecht, The Netherlands. Email: S.G.Petralia@uu.nl. I would like to thank all the comments received at the CID Growth Lab at the Harvard Kennedy School, the Macro Connections Lab at MIT, the SPRU seminar series at Sussex University, the seminars at QMUL and LSE. Additionally, I am grateful to Ricardo Hausmann and Petra Moser for the comments received.

a positive association between the rate of adoption of GPTs in sectors, measured in terms of the number of GPT patents, and their growth.

Keywords: Technological Change, General Purpose Technologies, Disruptive Technologies.

JEL Classification: O33, O34.

1 Introduction

Scholars have long emphasised the importance of technological change as a key factor behind economic growth and the rise in modern living standards. In particular, considerable attention has been devoted to better understand how major inventions, those having far reaching and prolonged implications, have transformed the modes of production throughout history. Widely discussed examples of disruptive technologies are the steam engine, the electricity, and more recently Computer & Communication (C&C) technologies.

A significant amount of effort has focused on understanding what make these technologies so revolutionary, what are the common features, if any, that distinguish them above others and that forge their disruptive and pervasive nature. Such a characterization can be used to create measurable indicators that improve our understanding of these technologies, their impact, and help informing and guiding policy making.

The previous literature has often referred to these disruptive technologies as General Purpose Technologies (GPTs). It has been argued that GPTs differ from others by possessing a wider scope for continuous improvement and elaboration, on the one hand, and higher complementarity, on the other. The latter means that a GPT should be able to diffuse across a wide range of sectors, not only because it is used as an input in many different products and processes (wide pervasiveness), but also because it is a technological complement of existing and new technologies (high innovation complementarity). These characteristics are what make GPTs "engines of growth" (Bresnahan and Trajtenberg, 1995).

While theoretical models have advanced greatly, providing a precise and coherent characterization of what defines a GPT and the economic implications of its diffusion (Bresnahan and Trajtenberg, 1995; Helpman and Trajtenberg, 1998b,a; Aghion and Howitt, 2000), a lack of convincing and comprehensive empirical evidence has called into question the relevance and usefulness of the notion of GPTs (Field, 2008). This is because the empirical literature on the subject has struggled to provide a measurable way of characterizing GPTs. As Lipsey, Carlaw, and Bekar (2005) note, "if the concept of a GPT is to be useful, then GPTs must be identifiable".

For instance, Hall, Trajtenberg, et al. (2006) discuss how to characterize GPTs with patent data studying the case of Information & Communication Technologies (ICTs). They propose a series of indicators based on a group of selected patents granted by the United States Patent and Trademark Office (USPTO) and argue that these indicators are not able to fully portray ICTs as a GPT. In addition, Moser and Nicholas (2004) use a similar set of measures to evaluate whether electricity matches the GPT criteria based on a sample of historical patents assigned to publicly traded companies in the 1920s, ultimately concluding that chemical technologies evidenced more of the characteristics of a GPT than electrical technologies. Also using patent data, Feldman and Yoon (2012) argue that technologies of genetic material recombination exhibit some of the characteristics of a GPT.

As a consequence, the empirical evidence remains inconclusive: either there is no particular technology that is capable of fulfilling the criteria established by theory, or current measures are not suitable to identify GPTs.

This article develops a set of patent-based indicators to capture the main characteristics of GPTs in data. Technologies are evaluated based on their scope for improvement and elaboration, the variety of products and processes that use them, and their complementarity with existing and new technologies. While technologies' scope for improvement is measured using patenting growth rates, the range of its uses is mapped by implementing a text-mining algorithm that traces technology-specific vocabulary in the universe of all available patent documents. Finally, complementarity with other technologies is measured using the co-occurrence of technological claims in patents.

These indicators have several advantages. First, they can be calculated at different levels of aggregation and do not rely on broadly defined technological categories like C&C or E&E. In addition, they do not use current patent citations to evaluate the past behavior of technologies, instead, they are constructed based on information that is available for all patent documents at the moment of issue. This means that these indicator can be used to study the behaviour of technologies since 1836 in the US, where digitized versions of historical patent documents are available. Finally, they treat the GP-ness of technologies as a matter of degree. They contemplate the possibility that technologies fulfill some of the criteria of the GPT definition but not all of them and evaluate the intensity at which they do.

The usefulness of these indicators is discussed and evaluated from two different perspectives. On the one hand, a top down approach is followed, where indicators are evaluated in reference to the anecdotal and historical evidence surrounding widely studied examples of GPTs such as E&E at the beginning of the 20^{th} century and C&C more recently. This first approach follows closely what has been done in the literature (Moser and Nicholas, 2004; Hall, Trajtenberg, et al., 2006; Feldman and Yoon, 2012), thus facilitating comparisons. It centers the discussion around a pre-selected set of technologies that are often cited as canonical examples of GPTs. It is shown that indicators behave as expected in light of the anecdotal and historical evidence on E&E and C&C technologies.

On the other hand a bottom up approach is followed, in which the same indicators are calculated at a finer level of aggregation (technological class level in the USPTO classification system). I then propose a simple way of identifying GPTs in patent data without relying on any imposed categorization of technologies but based on how technologies rank according to the indicators described before. I classify as GPTs those that rank above the average in terms of all three patent indicators, thus defining a technological "frontier" after which technologies are consider to be GPTs. It is shown that results are in line with anecdotal and historical evidence, and that there exist a positive association between the rate of adoption of GPTs in sectors (measured in terms of the number of GPT patents) and their growth, as predicted by theoretical models (Helpman and Trajtenberg, 1998b,a; Aghion and Howitt, 2000).

These results challenge some of the notions and practices that have been applied to understand and measure GPTs in the past. First, it is shown there exists a high level of heterogeneity in the GP-ness of technologies that compose commonly used technological categories like E&E and C&C, such that very dynamic and complementary technologies coexist with stagnant and mature ones. In addition, there is also a considerable degree of heterogeneity in the type of technologies that compose this newly defined set of GPTs, as they form an interconnected cluster of related technologies that span multiple categories. This suggests that our understanding of what delimits the boundaries of a GPT is more elusive and diffuse than previously thought. It is perhaps because of this fact that scholars have struggled to find a connection between the emergence of GPTs and their economic impact (Field, 2008). If we allow technologies to organize themselves based on their potential for growth and complementarity, results suggest GPTs can be better understood as a cluster or a network of related technologies that are connected to one another by underlying principles and mutual dependencies.

This paper is organized as follows: Sections 2 establishes the foundations, explains the rationale, and implements this three-dimensional index using the two most iconic technologies of the last century as examples, E&E and C&C technologies. Section 3 calculates the same set of indicators at the level of technological class in the USPTO classification system (there are more than 400 classes) and proposes a simple way of identifying GPTs in patent data. It is also shown that results are in line with historical and anecdotal evidence, and that there is a positive association between the rate of adoption of GPTs in sectors, measured in terms of the number of GPT patents, and their growth. Section 4 concludes.

2 Measuring GP-ness of Technologies

The aim of this section is to propose and discuss a set of indicators to identify the main characteristics of a GPT using patent data. These indicators are evaluated in reference to the anecdotal and historical evidence surrounding widely studied examples of GPTs such as E&E at the beginning of the 20^{th} century and C&C more recently. Therefore, this section follows closely the methodological approach that has been implemented in the literature (Moser and Nicholas, 2004; Hall, Trajtenberg, et al., 2006; Feldman and Yoon, 2012), thus facilitating comparisons. It centers the discussion around a pre-selected set of technologies that are often cited as canonical examples of GPTs, showing that the proposed indicators behave as expected in light of the anecdotal and historical evidence on E&E and C&C technologies.

Given the amount of information and the level of detail contained in patent documents it is natural to start looking for ways of characterizing GPTs using patent data. Every patent provides information about the technological nature of the invention, the geographical location of the inventor, and the prior art, among other things. This means that one could identify whether a patent has claimed, for instance, rights on the invention of a new electrical device, a new function for a chemical compound, or both. Therefore, a patent can claim rights to different types of components or technologies that have been created and combined to come up with a new product.

Information about US patents' technological class(es) is made available by the USPTO.[1] The USPTO classifies patents into classes according to the type of invention to which they claim rights. There are currently more than 400 different technological classes in use, and whenever a new class is created, or an existing one re-defined, all available patents are re-classified to maintain temporal consistency. Furthermore, patents can be grouped into broad economically relevant categories (Chemical, C&C, Drugs & Medical (D&M), E&E,

[1] http://www.uspto.gov/learning-and-resources/electronic-bulk-data-products

Mechanical, and Others).[2]

Even though there is no agreement on how to measure GPTs, there is a clear understanding on what defines them. According to Bresnahan and Trajtenberg (1995); Helpman and Trajtenberg (1998b,a); Aghion and Howitt (2000); David and Wright (1999); Moser and Nicholas (2004); Jovanovic and Rousseau (2005); Hall, Trajtenberg, et al. (2006); Lipsey, Carlaw, and Bekar (2005), a GPT must have the following characteristics:

1. **Wide scope for improvement and elaboration.** The technology should be able to go through a process of technical advance after it is first introduced, a continuous process of technological improvement. This idea is build on the notion that a GPT would generally grant the possibility to cumulatively build knowledge around it.

2. **Potential for use in a wide variety of products and processes.** The technology should spread across and be used in most sectors. This feature has been extensively discussed by Moser and Nicholas (2004), Hall, Trajtenberg, et al. (2006), and Feldman and Yoon (2012), which use the diversity of patent citations to proxy for the generality of an innovation.

3. **Strong complementarity with existing or potential new technologies.** The technology should have an impact on other existing and new technologies, not only by creating the need to alter and combine many of the existing technologies but also by increasing the opportunities to develop new technologies in combination with it.[3]

The GPT indicators discussed in this section are constructed and explained using E&E and C&C technologies as a benchmark. This is because they constitute two widely accepted examples of a GPT that, given the vast anecdotal and historical evidence surrounding them,

[2] See Hall, Jaffe, and Trajtenberg (2001) for details. The concordance is available at `http://www.nber.org/patents/`.

[3] This is also related to the literature of windows of opportunities (Perez and Soete, 1988; Lee and Lim, 2001).

allow for a preliminary assessment about the fitness of the indicators. Later, this rigid administrative categorization is relaxed and technological classes are used as the basic unit of analysis.

2.1 Wide Scope for Improvement and Elaboration

One of the aspects that distinguish GPTs from the rest is their capacity to go through a continuous process of technological improvement. This notion is based on the idea that most technologies are originally introduced as unrefined versions of their best self. It is argued that there exist a potential in GPTs that is based on the distance to the most efficient/mature version of themselves, which requires developing and perfecting them for many different uses and adapting them to a wide arrange of complementary and yet potentially unrelated technologies

This is probably the least challenging characteristic to relate to data. Previous empirical approaches have often used patenting growth to measure the extent to, and pace at which, technologies have been advancing. For instance, Jovanovic and Rousseau (2005) examine the growth rate of total patenting activity in the US and relate changes in its pace to the electric and ICT era. Moser and Nicholas (2004) demonstrates that patent activity in the category of E&E technologies grew the fastest in the 1920s. Similarly, Hall, Trajtenberg, et al. (2006) finds that classes related to C&C technologies grew faster than others after the 1980s.

Let's consider how E&E technologies have grown in terms of patenting activity since 1880. Table I shows the share occupied by different technological categories from 1880 to 1950, using as a starting point the invention of the first electric light-bulb. This period covers the emergence, development, and diffusion of E&E technologies (David, 1990; Goldfarb, 2005). Since C&C and D&M technologies represented approximately 1% of total patenting activity at the time, in this table they are not treated as separate categories but placed within the category Others. During this period E&E technologies grew faster than any other, from representing a 3.7% of all patenting activity to almost 15%. They showed more than a four-fold increase in participation.

Table I: Share Occupied by Technological Categories 1880-1950

	Chemical	E&E	Mechanical	Others
1880	9	3.7	34.1	53.2
1890	7	6.8	38.1	48.2
1900	9.6	5.7	37.3	47.4
1910	8.1	6.9	38.6	46.4
1920	8.8	8.6	39.3	43.3
1930	12.5	10.6	34.9	42.1
1940	15.8	12.8	31.3	40.1
1950	17.6	15	27.8	39.6

Notes: Values are in percentages.

Similarly, C&C technologies evidenced a remarkable growth since the 1960s, one year after the first integrated circuit was patented in the US. In the 50-years period considered, C&C technologies grew from representing 4.4% to 33.4% of total patenting activity, which makes it the biggest category today.

These results provide a comprehensive overview of the emergence and evolution of these two technologies that enlarges but mainly agrees with what Moser and Nicholas (2004) and Hall, Trajtenberg, et al. (2006) have found using a selected subset of patent documents.

Table II: Distribution of Patenting Activity 1950-2010

	Chemical	C&C	D&M	E&E	Mechanical	Others
1960	19	5.1	1.9	17.3	29.3	27.5
1970	22.9	6	2.7	18.7	25.8	23.8
1980	23.7	6.8	6.4	15.1	24.1	23.9
1990	19.5	10.5	8.5	17.3	22.4	21.8
2000	14.7	19.5	11.5	19.5	17.8	17
2010	10.8	33.5	9.7	22.1	13.3	10.6

Notes: Values are in percentages.

2.2 Potential for Use in a Wide Variety of Products and Processes

It is argued that as GPTs evolve and develop they should spread throughout the economy, given their potential to be used as an input in many different applications. For example, electricity is used as a power source in a wide range of sectors, to power household appliances, transportation services, and a varied number of industries. Additionally, the ability of electricity to drive chemical reactions, or to be used to transport information, drastically expands its range of uses and its pervasiveness.

Several approaches have been used to evaluate the pace at which GPT candidates diffused throughout the economy. For the case of electricity, one possibility is to consider the overall nationwide electrification of factories and households. David (1990) documents that the electric power used for mechanical drive capacity in the U.S. reached more than 50% by 1920, while Goldfarb (2005) and DuBoff (1979) find that by 1929 the ratio of electric motor power to total motor power reached 82%, on average. Jovanovic and Rousseau (2005) show that by 1929, nearly 70% of households had electrical connections.

Another approach is to rely on patent data. Moser and Nicholas (2004); Hall, Trajtenberg, et al. (2006); Feldman and Yoon (2012) measure the range of applicability of E&E and C&C technologies through citations. They use the technological diversity of citing patents to evaluate the generality of any cited patent. Therefore, the generality of a technology depends on how heterogeneous its citing patents are.

One of the main concerns with the use of patent citations to trace back knowledge is that they may be actually reflecting the technological structure of citing rather than cited patents. For instance, it is well known that patents tend to cite disproportionately other patents within the same domain to delimit the scope of their claims (Hall, Jaffe, and Trajtenberg, 2001). The possibility of tracing knowledge embodied in patents back in time through citations relies on two assumptions. First, that direct citations provide a comprehensive picture of the type of knowledge contained in a patent and second, that the dynamics of patent citations are invariant enough so that prior and distant knowledge is still cited in a meaningful way. If patent citations are more concerned with delimiting the scope of the invention rather than comprehensively accounting for its knowledge composition, then the ability of a citation today to trace knowledge back in time is severely undermined.

Consider the problem of measuring the extent of use of E&E technologies in the 1920s bearing in mind that digitized citations are not available before 1975. One possibility is to assume that we can use citations made by patents after 1975 to obtain information about the state of the technological landscape in the 1920s. The pioneering studies of Moser and Nicholas (2004) and Hall, Trajtenberg, et al. (2006) are good examples of this, since they use citations from 1975 to 1999 to trace the evolution of key technologies that were introduced long before 1975. However, in the Appendix A it is shown that commonly used generality measures based on patent citations produce a ranking that remains almost invariant since the 1890s for all technological categories, suggesting that this approach may be actually reflecting the current state of the technological landscape rather than the intended one.

Here a different approach is considered, which exploits the wealth of information contained in patent documents and provides a characterization of technologies based on information that is available at the time of issue and for all patents. Note that all patents contain a detailed description of the invention, which can be scrutinized to identify keywords related the use of components, notions, or principles of any technology regardless of whether the patent produces that particular type of technology as an output. For instance, it could be possible to trace patents that contain specific wording related to E&E(C&C) technologies even if they don't belong to the category of E&E(C&C). This set of patents will typically include inventions that use E&E(C&C)-related terms because they rely on E&E(C&C)-related components, notions, or principles, but do not produce any particular technological improvement in that area.

Figure I can be used as an example. It shows the first page of the patent number 2,956,114 assigned to Ampex Corporation in 1960 for a broad band magnetic tape system (tape recorder). This particular patent falls under the category of C&C and cites patents only in C&C and Mechanical. This implies that using a citation based method, or its main technological classification, we are not able to identify that the invention contains electrical components.

Examining the patent description, however, provides enough clues for a text mining algorithm to detect the electrical nature of it. Figure I highlights electricity-related words that could be used to identify that this invention uses E&E components, notions, or principles.

It is clear then that the set of words used to identify technologies is a key element in this process. One could, in principle, choose to use a predetermined set of words that is unequivocally associated with a technology. In the case of E&E this seems relatively easy to do since most of the words containing "electro" or "electri" are likely to refer to E&E-related components, notions, or principles. However, for C&C technologies the task is not so straightforward because relevant words such as "port" or "screen" may be describing completely different technologies.[4]

[4]See for instance: https://www.google.com/patents/US2386950. This patent contains both words but describes a new mean for protecting ships at sea.

Figure I: Ampex Patent

United States Patent Office

2,956,114

Patented Oct. 11, 1960

1

2,956,114

BROAD BAND MAGNETIC TAPE SYSTEM AND METHOD

Charles P. Ginsburg, Los Altos, Shelby F. Henderson, Jr., Woodside, Ray M. Dolby, Cupertino, and Charles E. Anderson, Belmont, Calif., assignors to Ampex Corporation, Redwood City, Calif., a corporation of California

Filed July 25, 1955, Ser. No. 524,004

8 Claims. (Cl. 178—6.6)

This invention relates generally to electromagnetic tape systems, methods and apparatus, particularly to systems and methods of this character capable of recording and reproducing signal intelligence over a wide frequency spectrum, including for example, video frequencies.

Various problems are involved when it is attempted to record and reproduce frequencies over a wide spectrum, as for example frequencies ranging higher than one megacycle, on magnetic tape. Assuming the use of reasonable tape speeds, conventional equipment is limited with respect to its usable frequency range. The recordable range can be increased by increasing the speed of the tape, but the speeds required for the recording of such high frequencies are such that the system becomes impractical because of the large amount of tape employed for a given recording period. It is possible to reduce the linear tape speed by recording successive tracks extending laterally across the tape. Equipment with this purpose involves the use of magnetic record units which are mounted to sweep successively across the coated surface of the tape while the tape is being advanced in the direction of its length. While this arrangement makes it theoretically possible to provide relative speeds such that frequencies up to four megacycles or higher can be recorded, its application necessarily involves a number of problems. For example the outputs of the several heads are subject to amplitude variations, due to various causes such as lack of exact registration on the recorded track, amplitude variations in the record because of slight variations in pressure between the several heads, and slight variations in the electrical characteristics of the heads. The conventional magnetic tape recording system, using currents varying in amplitude for application to the recording head, is particularly susceptible to undesired amplitude variations. The undesired signal variations cause distortion of the reproduced signal, and make it difficult if not impossible to reproduce the original frequency spectrum with reasonable fidelity, and particularly with sufficient fidelity to permit the recording and reproduction of television or like visual images.

The present invention is predicated upon certain discoveries which we have made, and which we utilize to advantage in the present invention. Particularly we have found that a wide frequency spectrum can be successfully recorded and reproduced by the use of a frequency modulation system in which the deviation of the carrier is small relative to the highest frequency components to be transmitted. In other words we have found that it is practical to use what can be referred to as narrow band F.-M. Narrow band F.-M. means that the ratio of $\Delta f/f_m$ is relatively small, and in actual practice can be of the order of 0.2, where Δf represents deviation corresponding to maximum signal amplitude and f_m represents the highest modulating frequency. Likewise we have found that the limit of f_m can be made reasonably close to the carrier frequency. We have also discovered that the center or carrier frequency can be so selected that it is near the upper recordable frequency limit of the apparatus, which

2

as previously explained is generally determined by the relative speed between the heads and the tape and the characteristics of the head. When the carrier frequency is so selected the recording system depends upon single sideband or vestigial sideband transmission. In other words the upper band of frequencies containing the frequency modulation components is not recorded or reproduced to any substantial extent. We have found that such a magnetic record can be reproduced to provide, after demodulation, the original modulating frequencies with a good degree of fidelity.

In addition to the foregoing, a practical system for the recording and reproduction of frequencies over a wide spectrum requires highly accurate speed control means for both recording and reproduction.

It is an object of the present invention to provide a system and method for the recording and reproduction of a wide frequency band, which will be relatively immune to spurious variations in signal strength.

Another object of the invention is to provide a system and method of the above character which, when used for the recording and reproduction of video frequencies, makes possible the reproduction of visual images with good fidelity.

Another object of the invention is to provide a system and method of making use of narrow band frequency modulation for recording over a wide frequency band.

Another object of the invention is to provide improved means for controlling the speed of operation of various parts during recording and reproduction.

Another object of the invention is to provide a system and apparatus for the recording of frequency components over a wide spectrum, such as video frequencies, which utilizes a plurality of record heads sweeping laterally across a magnetic tape, but without causing troublesome distortion or disturbances of the reproduced signal due to amplitude variations.

Additional objects and features of the invention will appear from the following description in detail in conjunction with the accompanying drawings.

Referring to the drawings:

Figure 1 is a circuit diagram illustrating a complete recording and reproducing system incorporating the present invention.

Figure 2 is a circuit diagram illustrating a modification of Figure 1.

Figure 3 is a plan view schematically illustrating mechanism for mounting the magnetic heads and for transporting the tape.

Figure 4 is a cross sectional view taken along the line 4—4 of Figure 3.

Figure 5 is a cross sectional detail taken along the line 5—5 of Figure 3.

Figure 6 is an enlarged cross sectional detail illustrating the guide means for the tape and the manner in which the tape is contacted by the magnetic heads.

Figure 7 is an enlarged detail illustrating means for engaging the lower edge of the tape, while it is passing through the guide means.

Figure 8 is a cross sectional detail taken along the line 8—8 of Figure 3, and showing suitable pulse generating means.

Figure 9 is a schematic view illustrating the pulse generating means and the cathode follower which may connect to the same.

Figure 10 is a circuit diagram schematically illustrating the commutating means for making connections with the various magnetic heads.

Figure 11 is a diagram like Figure 10, but showing simplified connections.

Figure 12 is a plan view schematically illustrating a

This subsection develops a data-driven algorithm to identify relevant keywords (2-grams) in any category or group of patents. The entire procedure for identifying these keywords is explained and documented in the Appendix B. It requires creating a vocabulary of technically relevant keywords used by the USPTO to describe technologies, computing their frequency of use within and outside the category (in this section E&E and C&C technologies), and finally selecting those that were at least 5 times more likely to appear within them than in any other category. As a result, this procedure produces a list of approximately 2000 E&E(C&C)-related keywords of length two to be searched, from which Table III displays the most relevant in each category.

Table III: Top 10 of Most Characteristic Keywords

E&E	C&C
deflection current	coherency unit
fuse tube	register sender
vertical charge	trunk circuit
focusing electrode	cord circuit
overload current	idle trunk
focus electrode	calling station
lightning arrester	telegraph system
shallow junction	reservation station
deflection coil	branch history
accelerating electrode	dial pulse

After identifying the set of E&E and C&C-related keywords it is therefore possible to evaluate the variety of technologies using E&E or C&C components, notions, and principles. The most straightforward way to do this consists of counting the number of different technological categories (technological classes in the USPTO classification scheme) that use E&E

or C&C related vocabulary at any point in time. A technological class is considered to have "used" E&E or C&C technologies if more than 3 different keywords were found in at least one patent within that class.[5]

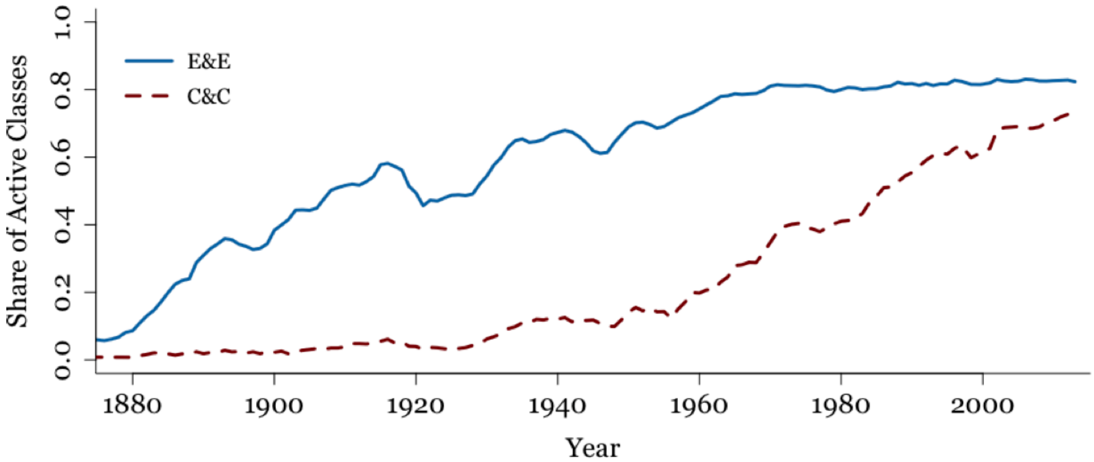
Therefore, the pervasiveness of use of a technology is measured at the extensive margin, in line with how it has been defined and modelled (Bresnahan and Trajtenberg, 1995; Helpman and Trajtenberg, 1998b). Note that GPT models usually measure diffusion based on the variety of sectors a technology has pervaded or is able to complement with.[6]

Figure II shows the share of technological classes in which E&E(C&C)-related vocabulary has been found. This share is calculated with respect to the total number of different non-E&E(C&C) technological classes available at any point in time. For instance, Figure II shows that by 1880, fewer than 5% of all non-E&E technological classes were using E&E-related vocabulary to describe inventions. By the end of the 1930s, this share had increased to approximately 70%. Similarly, C&C-related vocabulary appeared in less than 10% of non-C&C classes by 1950, increasing to more than 60% by 2010. These technologies pervaded the whole inventive structure, affecting the entire nature of technological production since their introduction.

[5]Results are not sensitive to the choice of the number of keywords or to the number of patents considered. For a detailed analysis on the intensive margin of diffusion please refer to the Appendix C.

[6]For instance, Helpman and Trajtenberg (1998a) and Aghion and Howitt (2000) characterizes each sector with a set of parameters that determine the order of adoption of the technology, to later evaluate the trajectory of the economy as one sector after the other adopts it.

Figure II: Share of Tech Classes Using E&E and C&C Vocabulary



2.3 Strong Complementarity with Existing and New Technologies

One of the most valuable features of a GPT is its capacity to act as an "enabling technology". This means that its introduction provides a vast number of opportunities to adapt and modify existing products and processes, to expand the space of possible inventions and innovations, and to create opportunities to develop new products, processes and technologies in combination with it. For instance, Bresnahan and Trajtenberg (1995) note that *"...the productivity gains associated with the introduction of electric motors in manufacturing were not limited to a reduction in energy costs. The new energy source fostered the more efficient design of factories, taking advantage of the new found flexibility of electric power."*

The far-reaching extent of its "innovation complementarities" (IC) is one of the most salient aspects of a GPT, as it is considered to be responsible for the creation and reinforcement of rapid technical advance and economic growth. Even though there is a vast literature collecting case-specific historical evidence (DuBoff, 1979; David, 1990; Helpman and Trajtenberg, 1998a; Rosenberg, 1998; Lipsey, Carlaw, and Bekar, 2005; Goldfarb, 2005; Bresnahan, 2010; Nuvolari, Verspagen, and von Tunzelmann, 2011), there has not been any systematic and comprehensive empirical study on this subject.

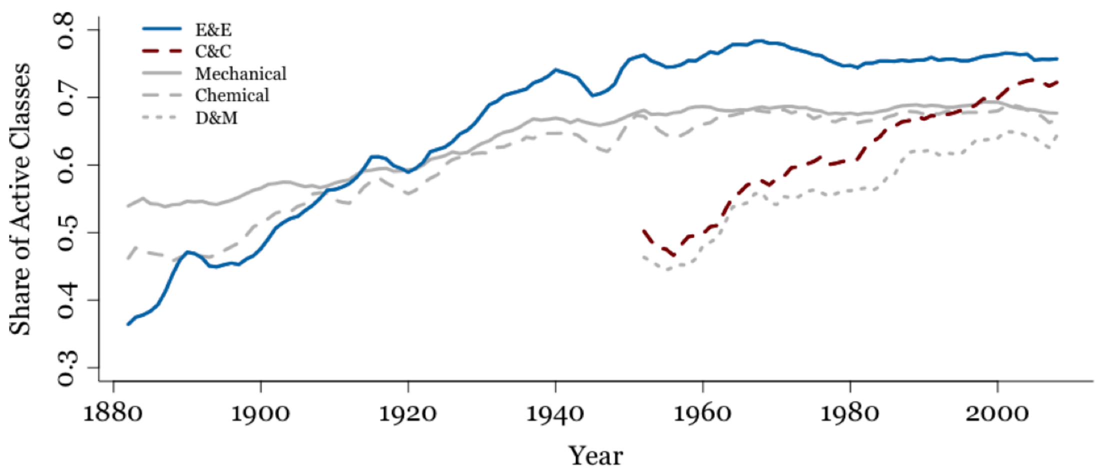
This subsection proposes a way of measuring the "innovation complementarity" (IC) of technologies using the co-occurrence structure of the technological classes within patent documents. Whenever a patent is issued, several claims are made regarding the inventiveness and scope of the patent. These claims specify all inventions contained within a particular patent for a product or process, and are classified based on their technological characteristics into different technological classes. Therefore, a patent can be classified into different technological categories, implying that a given product or process required multiple inventions in different fields to be realized. The extent and diversity of the IC of a technology can be measured by examining the diversity of its class co-occurrence profile.

For instance, consider the Ampex broad band magnetic tape system (1960) described in Section 2.2. This patent has claims in two different technological classes, class 360 (Dynamic Magnetic Information Storage or Retrieval) in C&C and class 386 (Motion Video Signal Processing for Recording or Reproducing) in E&E. This is because the patent introduces two main complementary improvements. The first concerns a more efficient way of comprising and recording frequencies (class 360), which allowed the size of the magnetic tapes to be reduced considerably. For this improvement to be properly used, higher precision in the speed of motion of the recording system needed to be achieved. This is when improvements in E&E technologies related to components regulating motion for recording devices had to be developed (class 386).

Therefore, the co-occurrence of technological classes within a patent can be used to measure the IC of technologies. A GPT should co-occur with a wide variety of different technologies (classes); this is because its IC allows it to be re-combined with existing technologies to improve existing products (such as tape systems), as well as to develop new-to-the-world and yet complementary technologies. Note that this measure varies considerably with respect to the one developed in Section 2.2. This is because the IC of a technology is evaluated using the the co-occurrence of different technological classes with itself, while the indicator devised in Section 2.2 is calculated scrutinizing patent documents outside the domain of the technology.

Consider again E&E and C&C technologies as examples, in this case their IC would be measured by counting the number of technological classes they co-occur with. Figure III shows the IC of all main technological categories,[7] providing a set of clear and straightforward messages: First, E&E and C&C technologies increased the variety and scope of their IC since their introduction over other types of technologies and in line with what is expected. They began as very narrow technological fields to later become the most complementary technologies. In the case of E&E technologies, the timing and dynamics of these results are in line with previous historical evidence suggesting that the transformative power of electricity did not acquire momentum until after the 1914s (David, 1990; Greenwood, 1997; Lipsey, Carlaw, and Bekar, 2005; Field, 2008). In addition, results show that C&C technologies surpassed others in terms of IC only after the 2000s.

Figure III: Innovation Complementarity of Technologies



[7] To avoid taking into account irrelevant, proximate combinations, co-occurrences of different classes within the same category are not counted. Thus, for instance, if a class within chemicals co-occurs with another class in the same category it is not considered an IC.

These subsections have developed a set of indicators aimed at characterizing the GP-ness of technologies with patent data. Even though grouping patents into broad technological categories may be useful to exemplify the construction of the indicators and to provide a first reality check, it is clear this approach imposes rigid administrative boundaries on technologies which are not desirable nor needed. The next section applies the same logic to create the same indicators at the level of technological class in the USPTO classification scheme and discusses a data-driven approach to identify GPTs.

3 The Concept of a GPT Frontier

3.1 Measuring GP-ness of Technologies in Detail

If the indicators developed before constitute a suitable empirical counterpart of what theory describes as the main characteristics of a GPT, how can we operationalize these principles in a way that is useful, flexible and informative? What would happen if we let technologies sort themselves according to this definition? Would technologies within E&E and C&C remain at the top of the classification and in line with the historical evidence? Can we learn something from this self-organizing exercise?

The most straightforward way to start answering these questions is to calculate all three metrics at a finer level of aggregation than the broad NBER categorization used before. Note that the indicators developed in the previous section can be constructed at technological class level in the USPTO classification scheme (there are more than 400 active classes).[8]

Consider Figure V, it shows a scatter-plot of all existing technologies (classes) according to their Growth, IC, and UC in the period 2005-2010. Growth rates are calculated as $\Delta P_t = \frac{P_t - P_{t-5}}{P_{t-5}} - 1$, where P_t represents the number of patents in a given year. Values for the period are then averaged. UC and IC values are expressed as the average number of technological classes they complement with in this period. Technologies with higher values

[8]In principle they could also be computed at finer levels of aggregation such as the subclass level.

of Growth, IC, and UC are located in the upper-right quadrant of the figure (in red). In this particular period, the type of technologies that appear at the top are mostly in C&C and E&E (Figure IV). Table IV shows the correlation of the different measures while Table V details the top and bottom ranked technologies.

Table IV: Correlation Table

	IC	UC	Growth
IC	1	0.251	0.180
UC	0.251	1	0.217
Growth	0.180	0.217	1

Figure V: Map of Technologies (2005-2010)

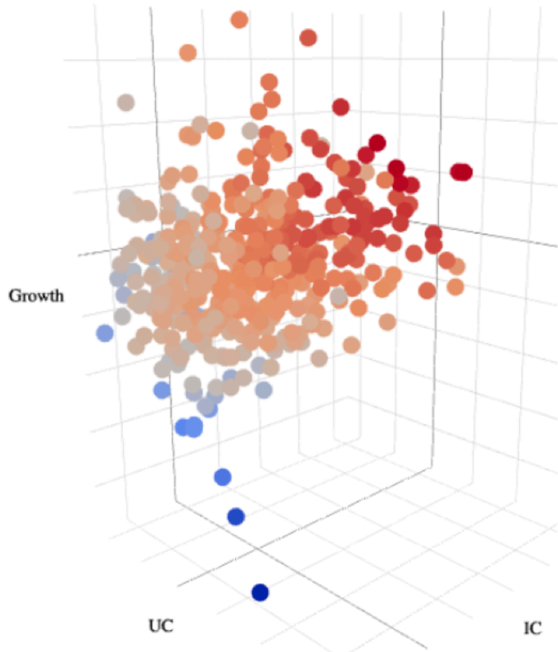
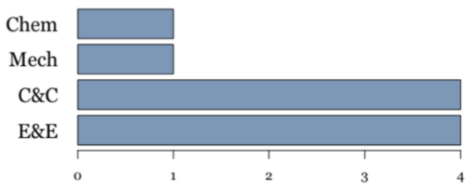


Figure IV: Distribution (Top 10)



Note: Growth, UC, and IC values were averaged over the period 2005-2010. Technologies are ranked based on the average normalized value (de-meaned and divided by the standard deviation) of the indicators (Growth, UC, IC). In Figure V warmer colors (red) represent higher values. Table IV shows the correlation between these measures. Figure IV shows the number of technological classes per category of the top ten ranked technologies.

Table V: Ranking of Technological Classes (2005-2010)

Rank	Category	Class	Growth	IC	UC
1	E&E	Television	0.4	147.3	163.2
2	C&C	Telecommunications	0.6	128.8	162.5
3	E&E	Radiant energy	0.4	191.7	182.5
4	E&E	Illumination	0.3	155.3	160.3
5	C&C	Communications: electrical	0.3	229.5	155.3
6	C&C	Image analysis	0.5	136.8	148.2
7	E&E	Active solid-state devices	0.2	188	146.2
.	.	.	.	.	.
413	Mechanical	Advancing material of indeterminate length	-1	33.2	59.3
414	Others	Heating systems	-1.1	21.8	103.5
415	Others	Industrial electric heating furnaces	-1.1	18	107.5
416	E&E	Recorders	-1.1	23.2	91.3
417	Chemical	Explosive and thermic compositions	-1.3	13.2	137.3
418	Chemical	Combinatorial chemistry technology	-1.6	27	133.5
419	E&E	Scanning-probe techniques or apparatus	-1.9	18.3	161.3

Notes: Growth rates are calculated as $\Delta\ P_t = \frac{P_t - P_{t-5}}{P_{t-5}}$, where P_t represents the number of patents in a given year. Values for the period are then averaged. UC and IC values are expressed as the average number of technological classes they complement with in the period. Technologies are ranked based on the average normalized value (de-meaned and divided by the standard deviation) of the indicators (Growth, UC, IC)

Several points are worth mentioning about the resulting ranking in table V. Note that there exist a considerable degree of heterogeneity in the type of technologies at the top. This group is formed by an interconnected cluster of related technologies that span outside the borders of commonly used categories. Most of the E&E technologies that appear at the top of Table V, such as those related to the production of solid state devices, are closely connected to C&C technologies. On the other hand, the bottom of Table V is occupied by mature technologies, those still having a broad impact in terms of UC but that do not function as a platform for other technologies to innovate with (low IC) and that have exhausted their potential for improvement (low Growth). In addition, note that certain domains such as E&E contain dynamic technologies (active solid state devices) that coexist with already stagnant or mature ones (recorders).

A remarkably similar pattern can be found if we inspect the ranking of technologies at the peak of the electrical revolution in the 1930s. Table VI shows top and bottom technologies for the period 1930-1940. As expected, the top of the ranking is occupied by technologies in the area of E&E and related, like refrigeration technologies. This detail is not minor since the development of the sector of electrical refrigerators in the US was one of the fastest growing industries after the electrification of households. The production of refrigerators jumped from 5,000 units in 1920 to 1,000,000 units in 1930, reaching 6,000,000 units by 1936 (Nebeker, 2009).[9]

Table VI: Ranking of Technological Classes (1930-1940)

Rank	Category	Class	Growth	IC	UC
1	E&E	Electricity: circuit makers and breakers	-0.100	167.400	299.300
2	Others	Refrigeration	0.400	126	232.100
3	E&E	Electric lamp and discharge devices	0.400	121.200	217.100
4	Mechanical	Clutches and power-stop control	0	136.400	260.400
5	E&E	Electric lamp and discharge devices: systems	0.600	101	186.600
.	.	.	.	.	.
389	Others	Printed matter	-0.600	29.200	14.100
390	Mechanical	Vehicle fenders	-0.700	22.100	35.700
391	C&C	Computer graphics processing & visual display systems	-0.500	13.700	4.700
392	Others	Merchandising	-0.600	8.500	17.200
393	Mechanical	Compound tools	-0.800	24.100	28.500

Notes: Growth rates are calculated as $\Delta\ P_t = \frac{P_t - P_{t-5}}{P_{t-5}}$, where P_t represents the number of patents in a given year. Values for the period are then averaged. UC and IC values are expressed as the average number of technological classes they complement with in the period. Technologies are ranked based on the average normalized value (de-meaned and divided by the standard deviation) of the indicators (Growth, UC, IC)

Therefore, these indicators are able to organize technologies in a way that is consistent with the historical and anecdotal evidence on the evolution of GPTs. In addition, results

[9]The highest share of this market was occupied by the Kelvinator Company of Detroit, Michigan, which introduced the first refrigerator with automated control based on the inventions of engineer Nathaniel B. Wales.

show that the top of this ranking is occupied by a cluster or network of related technologies that are connected to one another by underlying principles and mutual dependencies. Note, however, that this cluster of technologies span multiple technological categories, challenging some of the notions that have been used to identify GPTs in the literature.

If the indicators outlined before constitute suitable empirical counterparts of what theory describes as the main characteristics of a GPT, which technologies would qualify as GPTs? If as Lipsey, Carlaw, and Bekar (2005) point out, what distinguishes GPTs from others is a matter of degree, where do we draw the line to separate GPTs from the rest? Are GPTs to be identified in relative terms with respect to contemporaneous technologies or in absolute terms with respect to all technologies ever invented?

3.2 Introducing the Concept of a GPT Frontier

It seems clear, at least in theory, that GPTs should posses all of the characteristics mentioned before and not just some of them. One crucial factor in the definition of what constitutes a GPT is that the characteristics technologies are evaluated upon are not substitutes of one another. For instance, a technology that possesses low capacity to recombine with other technologies would not be able to compensate that by improving faster than the rest, since the propagation mechanisms described by Helpman and Trajtenberg (1998b,a); Aghion and Howitt (2000) will not be present; which means that positive feedbacks and externalities will not materialize. Similarly, it is to be expected that mature and already exhausted technologies have high pervasiveness of use across sectors. Regardless of how high the use of a mature technology is, this does not compensate for its lack of potential for improvement.

A simple way of identifying GPTs using the indicators developed before is to consider as such those that rank above the average in terms of IC, UC, and Growth. This implies that only those technologies located at the "frontier" (in the upper right quadrant of Figure V or at the top of Tables V & VI) would qualify as GPTs, as they fulfill all three criteria simultaneously.

In what follows I provide a first attempt to test this idea by evaluating whether there is a correlation between the rate of adoption of GPTs in a sector and its growth, in line with what theory predicts (Helpman and Trajtenberg, 1998b,a; Aghion and Howitt, 2000). GPT adoption is measured in terms of the number of patents within a sector that fulfill the criteria mentioned before (i.e. the number of patents in technological classes that rank above the average in terms of IC, UC, and Growth). Hence, the aim of this last section is not to provide causal evidence to prove or disprove the existence of GPTs and their impact, but rather to find support about the claim that the indicators proposed here can be considered as useful instruments in the discussion of what is and what is not a GPT.

One the one hand, Growth, UC, and IC values are calculated for all technological classes in the USPTO classification scheme since 1920 (the oldest year for which digitized patent documents are available).[10] Then, for any given year, technologies (technological classes in the USPTO scheme) are classified as GPTs if they rank above the average in terms of their Growth, UC, and IC (after normalizing their values).[11] Then patents are identified as GPT or non-GPT based on their technological classification.

On the other hand, sector level data is matched with patenting information using firms as intermediaries. There are few firm-level databases that can be used to link patenting information of firms within sectors that also contain information about their economic performance. The ORBIS database (compiled by the Bureau van Dijk Electronic Publishing, BvD) is a commercial dataset that provides economic and administrative data for more than 130 million firms worldwide, covering more than 100 countries since 2005.[12] In addition, each company possesses an unique identifier (usually at the corporation level) that links firms to their patent activity at the USPTO, making it possible to connect economic indicators of firm performance with their patenting profiles.[13] This later aspect is crucial because it allows

[10]Data can be downloaded here: `https://dataverse.harvard.edu/dataverse/GPT-Indicators`

[11]In every year and for each of the variables (Growth, UC, and IC) values are normalized by subtracting the mean and dividing by the standard deviation.

[12]Compustat Global and Compustat North America contain detailed information of listed firms, however they cover mostly large firms.

[13]Patenting activity in the US has been used extensively in economics and innovation studies to address

one to evaluate how changes in technological profile of companies within sectors relate to their overall performance.

There are 302,052 unique manufacturing companies in the 2017 version of the this dataset, which covers the period 2000-2017.[14] However, there are several aspects to be considered when using ORBIS data. First, it should be taken into account that before 2008 and after 2013 the number of firms in the sample represent only a fraction of those appearing between 2008 and 2013. In fact, Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, and Yesiltas (2015) show that firms in this database represent only a fraction of the total output for European countries prior 2005.[15] In addition, economic or financial indicators are not always available even if firms appear in the sample. For instance, crucial information such as the number of employees or the operating revenue is available for 57% and 70% of the companies, respectively, in the period 2008-2013. In the Appendix D there is a detailed description of the yearly, sector, and country composition of the ORBIS database.

In what follows the analysis is conducted based on a set of manufacturing firms for which there is information about their operating revenue in 2008 and 2013 and have at least one employee.[16] Therefore, the economic performance of sectors is evaluated at the beginning and end of the period for which the ORBIS database contains a stable number and composition of firms, while GPT adoption is calculated as the number of GPT patents granted to firms in it. Because the number of patents a company files varies greatly from year to year (Hall,

issues of global scope, this responds not only to the importance and size of the US technological market but also to the consistent and systematic way patents applications have been evaluated over the years; making data collected at the USPTO very suitable for comparisons, both across countries and over time.

[14] The focus on manufacturing companies responds to the fact that most of what is patented comes from these industries.

[15] Please refer to the Appendix D for more details.

[16] The operating revenue of a firm, as opposed to total revenue, constitutes a better instrument to evaluate the productivity and profitability of a firm. This is because the latter may include revenues from sources that are unrelated to the day-to-day activity of the firm, such as asset sales, interests earned from deposits, etc.

Jaffe, and Trajtenberg, 2001), the patenting activity of firms is evaluated using a 5-years windows previous to the date the economic data is available.

The final sample contains 106,739 firms from 80 countries that operate in 297 different NACE 4-Digit sectors. 10,668 companies (10%) have at least a patent granted by the USPTO in this period. For more details about the country and sector distribution of firms in the database please refer to the Appendix D[17] (See also Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, and Yesiltas (2015) for a more general discussion of the ORBIS database).

Table VII show the result of regressing the operating revenue per employee (in logs) in sectors using the 3-digits and 2-digits NACE sector classification available in ORBIS against the number and share of GPT patents in that sector. Since the coverage of this database varies greatly across countries and sectors, two different strategies are considered to avoid including countries for which sectors of the economy are not well represented. On the one hand, columns (1) and (4) in Table VII include countries that have data for at least 50% of the sectors in both 2008 and 2013 and sectors that produced at least 1 patent in the entire period.[18] On the other hand, columns (2) and (5) only consider countries for which the firms in the ORBIS database represent more than 50% of the official gross output and

[17] Linking patenting activity to economic sectors is not a a straightforward task. The USPTO provides a concordance between its technological classification scheme and the NAICS and SIC sectors each patent is most related to, however, this procedure usually assigns patents to multiple sectors, making the links between technologies and sectors very diffuse (The concordances can be found here: `https://www.uspto.gov/web/offices/ac/ido/oeip/taf/data/`). In addition, it is not clear that the proper way to link technologies with sectors is by their degree of relatedness. After all, the firm the technology is assigned to is the entity that appropriates the benefits of that invention despite of the fact that the invention may be related to a different sector the company operates under. The ORBIS database can be used to address this issue, since information about the main sector of activity is provided for all companies. However, one have to keep in mind that any aggregation of company data at sector level using ORBIS may not actually be representative of the true sector distribution in a given country.

[18] This is to avoid including sectors that do not rely on patenting, such as primary sectors.

employment data in Eurostat.[19] Columns (3) and (6) replicate the procedure in columns (1) and (4) but at a higher level of aggregation. All regressions include country, sector, and year fixed effects. Standard errors are clustered at the same level according to Cameron, Gelbach, and Miller (2012).

Results in columns 1-3 in Table VII show that the number of GPT patents in sectors is associated with higher operating revenues, after controlling for sector-specific effects and the total number of other patents (non-GPT). These results suggest there exists a positive association between sectors' growth and the number of GPT patents produced, an association that does not exist for other type of patents. Columns 3-6 replace the number of GPT patents by its share, in order to use a variable that is less influenced by sector size. In this case, the total number of patents is included as a control. In sum, results show a positive and strong association between the share of GPT patents in a sector and its growth. They suggest that a 10% increase in the number of GPT patents in a sector is associated with an increase in the operating revenue that ranges from 0.28% to 0.33%.

In the Appendix E it is shown that the number of GPT patents is also associated with higher operating revenues at firm level. Firm-level results suggest that a 10% increase in the number of GPT patents in firms is associated with an increase in their operating revenue per employee of approximately 0.3%, remarkably similar to those obtained at sector level.

[19] See Table 6.1 and 6.2 in Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, and Yesiltas (2015)

Table VII: GPT Patents & Sector Performance

	Operating Revenue per Employee			*Operating Revenue per Employee*		
	(1)	(2)	(3)	(4)	(5)	(6)
GPT Patents (in logs)	0.028***	0.033***	0.033*			
	(0.009)	(0.012)	(0.017)			
Non-GPT Patents (in logs)	−0.020***	−0.032***	−0.033***			
	(0.006)	(0.011)	(0.010)			
Share of GPT Patents				0.062*	0.102***	0.183**
				(0.034)	(0.032)	(0.079)
Total Patents (in logs)				0.006	−0.002	−0.008
				(0.009)	(0.009)	(0.013)
Years Included	2008 & 2013	2008 & 2013	2008 & 2013	2008 & 2013	2008 & 2013	2008 & 2013
Sector	3-Digits NACE	3-Digits NACE	2-Digits NACE	3-Digits NACE	3-Digits NACE	2-Digits NACE
Number of Countries	24	25	38	24	25	38
Sector, Year & Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,020	1,812	962	2,020	1,812	962
Adjusted R^2	0.612	0.599	0.614	0.612	0.599	0.617

*p<0.1; **p<0.05; ***p<0.01

To sum up, this section proposes a data-driven solution to identify the GPTs using patent data. It is shown that sectors (and firms) producing a higher number (or share) of GPT patents outperformed others in terms of their operating revenue. This evidence can be seen as a first reality check on the potential of the proposed indicators to identify GPTs. The next section discusses the implications and limitations of these findings.

4 Concluding Remarks

This article develops a three-dimension indicator to capture the main features of General Purpose Technologies (GPTs) in patent data. Technologies are evaluated based on their scope for improvement and elaboration, the variety of products and processes that use them, and their complementarity with existing and new technologies. Technologies' scope for improvement is measured using patenting growth rates. The range of its uses is mapped by implementing a text-mining algorithm that traces technology-specific vocabulary in the universe of all available patent documents. Finally, complementarity with other technologies is measured using the co-occurrence of technological claims in patents.

These indicators have several advantages. First, they can be calculated at different levels of aggregation and do not rely on broadly defined technological categories like C&C or E&E. In addition, they do not use current patent citations to evaluate the past behavior of technologies, instead, they are constructed based on information that is available for all patent documents at the moment of issue. This means that these indicator can be used to study the behaviour of technologies since 1836 to present in the US, where digitized versions of historical patent documents are available. Finally, they treat the GP-ness of technologies as a matter of degree. They contemplate the possibility that technologies fulfill some of the criteria of the GPT definition but not all of them and evaluate the intensity at which they do.

The proposed ranking of GP-ness of technologies is evaluated in reference to the anecdotal and historical evidence surrounding widely studied examples of GPTs such as E&E at the beginning of the 20^{th} century and C&C more recently. In line with what is expected, and when broad technological categories are used, E&E and C&C technologies occupy the top of the GP-ness ranking in their reference periods. However, when these indicators are constructed at a finer aggregation level, results challenge some of the notions and practices that have been applied to understand and measure GPTs in the past.

First, it is shown there exists a high level of heterogeneity in the GP-ness of technologies that compose commonly used technological categories like E&E and C&C, such that very dynamic and complementary technologies coexist with stagnant and mature ones. In addition, results show that GPTs can be better understood as an interconnected cluster of related technologies that span multiple categories.

One of the main goals of this article is to contribute to the discussion of what defines a GPT and how to measure its most salient features. Results suggests that our understanding of what delimits the boundaries of a GPT is more elusive and diffuse than what we think. It is perhaps because of this fact that scholars have struggled to find a connection between the emergence of GPTs and their economic impact.

An interesting yet unsolved discussion is whether we should understand the existence of GPTs in absolute or relative terms. If as Lipsey, Carlaw, and Bekar (2005) points out, the notion of a GPT is a matter of degree, where should we draw the line? Should technologies be evaluated contemporaneously or with respect to all other technologies ever invented? Although this article favours the former idea, it remains inconclusive about the potential implications of such a choice.

How can we evaluate GPT candidates in different periods of time if their potential for growth and complementarity are highly contextual? It is well documented that the emergence E&E technologies displaced steam engine related technologies at the beginning of the 20^{th} century, eroding their potential for growth and complementarity. What does this tell us about the transformative impact of steam technologies and their GPT potential? Would we even consider steam technologies as a GPT candidate if E&E technologies would have emerged 50 years earlier?

One of the main limitations of this study is that measures are inherently backward looking. They rely on current knowledge about the structure of technologies to evaluate present and past behaviour. A promising line of research could focus on improving our capabilities to predict the emergence of new GPTs given early technological trends.

5 Appendix

A The Use of Citations to Identify Historical Technological Trends

Given the lack of detailed data on historical inventive and innovative outputs, researchers interested in tracing the evolution of technologies prior 1975 have usually relied on historical accounts or have found clever ways to trace information of particular events in the available data.

The pioneering studies of Moser and Nicholas (2004) and Hall, Trajtenberg, et al. (2006) are a good example of the later, since they use patent citation data from 1975 to 1999 to trace the evolution of key technologies that were introduced long before data was made available.

Moser and Nicholas (2004) use the Herfindahl-Hirschman concentration index to measure the degree of generality of the citations received today by a sample of historical patents assigned to publicly traded companies in the 1920s. In this study they find that chemical technologies evidenced more of the characteristics of a GPT than electrical technologies. In a similar fashion, Hall, Trajtenberg, et al. (2006) propose a series of indicators to evaluate the generality of ICTs based on a group of selected patents granted by the USPTO. Similarly, these indicators are not able to fully portray ICTs as a GPT.

This subsection provides empirical evidence showing that a citation-based measures may not be an appropriate vehicle to identify historical trends in the evolution of technologies. It is argued that citations seem to be reflecting the technological structure of citing rather than cited patents.

One possible way of evaluating this hypothesis is to apply the procedure used in Moser and Nicholas (2004) and Hall, Trajtenberg, et al. (2006) from 1890 to 1960 (before the citation data starts). Ideally, we would like to see that this procedure is able to capture some of the dynamics that occurred in this period, such as the emergence and development of E&E technologies, the appearance of C&C technologies or the decline of Mechanical technologies. Table VIII replicates the generality measures used by Trajtenberg, Henderson, and Jaffe (1997); Moser and Nicholas (2004); Hall, Trajtenberg, et al. (2006) but considering the entire set of patent granted since E&E technologies had a considerable size (1890) and until the decade before patents started to be digitized (1960).

Therefore the decade of the 1920s in Table VIII can be directly compared to the results of Moser and Nicholas (2004), Table 1. The only difference is that Moser and Nicholas (2004) used a subset of patents assigned to publicly traded companies while Table VIII considers all patents granted in that period. Two things are worth mentioning about this exercise, first, that the generality attributed to different technological categories using the entire sample of patents is very similar to that reported by Moser and Nicholas (2004). Meaning that the subset of patents used in their analysis accurately represented the whole. In both cases, Chemical technologies has the highest generality index across technologies (0.12 in both cases) while E&E the lowest (0.09 in Table VIII and 0.08 in Moser and Nicholas (2004)).

Second, the generality index remains practically invariant in relative terms since the 1890s. Chemical technologies has the highest generality index across technologies and E&E the lowest regardless of the period considered.

Not surprisingly, Hall, Trajtenberg, et al. (2006) also finds chemical technologies to be the most general, especially when they consider the industry of use. As they mention, one possible explanation behind these results is that there are a number of chemical classes that are essentially the same class (532-570), which may bias the generality index up.

Table VIII: Generality Measures Applied Accross Decades

Decade	Chemical	Mechanical	Others	E&E
1890	0.09	0.07	0.08	0.07
1900	0.10	0.08	0.09	0.08
1910	0.10	0.09	0.10	0.08
1920	0.12	0.10	0.11	0.09
1930	0.15	0.13	0.14	0.11
1940	0.18	0.16	0.17	0.14
1950	0.23	0.20	0.20	0.17
1960	0.29	0.26	0.26	0.24

B Procedure to Identify E&E and C&C Keywords

Section 2.2 exploits the wealth of information contained in patent documents to provide a characterization of the pervasiveness of use of E&E and C&C technologies in their historical context. To do so, patents not belonging to E&E or C&C classes are scrutinized to identify keywords related to the use of E&E or C&C components, notions, or principles. Consequently, patents that contain specific wording related to E&E technologies but do not belong to the category of E&E can be considered "users of E&E", and the same holds for C&C.

This subsection describes how this set of words was chosen. The first step requires creating a vocabulary of technically relevant words, which was obtained from the manual used by the USPTO to describe technological classes. This manual contains a detailed description of all technological classes in the the U.S. Patent Classification System (USPC) and is used by examiners to classify patents.[20] This definition schedule not only contains a description of the class but also of all their subclasses. After parsing all definitions for the 54 classes contained in E&E and the 44 classes in C&C, a vocabulary of keywords of length two (two consecutive words) containing 208,689 keywords for E&E technologies and 122,255 keywords for C&C was created.

The second step is then to calculate how frequently these keywords appear within and outside E&E and C&C patents. If some of these keywords appear much more often in E&E or C&C patents than in any other then we can consider them to be highly associated to E&E or C&C technologies. For instance, Table IX shows the frequency of occurrence of the E&E keywords shown in Table III of Section 2.2, outside an inside E&E classes. Column 2 shows how many times a keyword was found in the entire set of patents granted since 1920, while column 3 shows only the number of occurrences in E&E classes.

Table IX: Use of top Keywords in E&E

Keyword	Ocurrences$_{(All)}$	Ocurrences$_{(E\&E)}$	$f_{E\&E}/f_{-E\&E}$
deflection current	$16,032$	$15,295$	109.4
fuse tube	$22,347$	$21,292$	106.4
vertical charge	$11,956$	$11,221$	80.5
focusing electrode	$28,343$	$26,363$	70.2
overload current	$11,639$	$10,819$	69.5
focus electrode	$13,343$	$12,283$	61.1
lightning arrester	$12,316$	$11,289$	57.9
shallow junction	$11,955$	$10,953$	57.6
lamp claimed	$13,408$	$12,257$	56.1
deflection coil	$37,931$	$34,656$	55.8
accelerating electrode	$19,625$	$17,903$	54.8

[20]All class definitions can be downloaded from: http://patents.reedtech.com/classdata.php

Finally, the last step in the procedure consist of selecting a suitable set of keywords that may be informative of the type of technology the patent is describing. This set of keywords is selected based on the ratio shown in column 4 of Table IX. This is the ratio of the share keyword k represents in E&E technologies ($f_{k,E\&E} = O_{k\in K,E\&E} \,/\, \sum_{i\in K} O_{i,E\&E}$) over the share keyword k represents in other technologies ($f_{k,-E\&E} = O_{k\in K,-E\&E} \,/\, \sum_{i\in K} O_{i,-E\&E}$). Where O represent the number of occurrences and K the set of all keywords. Therefore the set of keywords used to track E&E and C&C-related vocabulary in patents is selected based on this ratio being higher than 5, which results in approximately 2000 E&E-related and C&C-related keywords.

When the procedure is implemented at the class level there is only a small change. Instead of keeping all the keywords with a ratio higher than five, only the first 250 most representative keywords are kept. This adjustment is to keep the number of keywords to be searched under a certain limit. Note that this arrangement implies searching for approximately 100,000 keywords in 10,000,000 patent documents. Results are insensitive to choosing either 100 or 500 keywords instead.

C Intensive vs Extensive Margins in Use and ICs

In sections 2.2 and 2.3 adoption is measured at the extensive margin. This is to keep the characterization of GPTs in data as closely related as possible to how they have been defined and modelled. For instance, Helpman and Trajtenberg (1998a) and Aghion and Howitt (2000) characterize each sector with a set of parameters that determine the order of technological adoption of a GPT, to later evaluate the trajectory of the economy as one sector after the other adopt it.

This subsection evaluates, on the one hand, the robustness of the results showed in sections 2.2 and 2.3 to alternative measures of the pervasiveness of use and innovation complementarities of E&E and C&C technologies. This is done by considering the share of total patent activity occupied by E&E and C&C technologies instead of the share of the total number of available classes. On the other hand, a more detailed analysis of the intensive margin of diffusion is provided, by considering how use and innovation complementarities of E&E and C&C technologies were distributed across classes.

Figure II in Section 2.2 shows that by 1880 fewer than 5% of all non-E&E technological classes were using E&E-related vocabulary to describe inventions, while by the end of the 1930s this share had increased to approximately 70%. Similarly, the second column in Table X reports the share of non-E&E patents that were using E&E vocabulary from 1880 to 1950. The remarkable increase in the variety of technologies that were using E&E-related vocabulary was also followed by a sharp increase in the number of patents using this vocabulary, which multiplied by 7 since 1880, going from representing 1% of all non-E&E patents to 7% of them. The same trend can be found with respect to the expansion of the innovation complementarities of E&E technologies (third column). By 1950, the share of patents where E&E claims were made in combination with other types of technologies multiplied by 14 (going from 2% to 28%)[21].

Table X: Total Patents Share for E&E

Year	Use	IC
1880	0.01	0.02
1890	0.02	0.04
1900	0.02	0.04
1910	0.03	0.05
1920	0.02	0.08
1930	0.04	0.12
1940	0.07	0.22
1950	0.07	0.28

[21] This share is calculated with respect to the number of patents that did not complement with E&E in a given year.

Analogous results can be found when considering the diffusion of related vocabulary and innovation complementarities of C&C technologies. The number of patents using C&C vocabulary went from barely non existent in 1960 to represent more than 9% of patenting activity in 2010. Similarly, the number of complementary patents rose from 9% to 17%.

Section 2 characterized the diffusion of E&E and C&C technologies exclusively at the extensive margin of adoption. However, it is of interest to provide a more detailed analysis on how diffusion was distributed across classes, since it may be informative about the uneven nature of the diffusion of these technologies.

Figures VI and VII show the concentration across different classes using the Gini coefficient (the higher the value of the Gini coefficient the more concentrated the distribution of patents across classes). Figure VI considers the concentration in the use of E&E(C&C)-related vocabulary. As in described in Section 2.2, only non-E&E(C&C) classes are considered when evaluating the diffusion of E&E(C&C)-related vocabulary, respectively.

Table XI: Total Patents Share for C&C

Year	Use	IC
1960	0	0.10
1970	0.01	0.09
1980	0.02	0.09
1990	0.03	0.12
2001	0.05	0.12
2010	0.09	0.17

Note that the concentration of the use of E&E vocabulary has been decreasing since their introduction and up to the 1960s, with sharp and regular increases before the 1930s. These boosts in concentration correspond to the period of development of E&E technologies, where new products and processes were introduced, probably affecting few surrounding technologies at the beginning to later spread more widely. After the 1930s, when the technology reached a maturity phase, the concentration became less pronounced. Since the introduction of C&C technologies, which are highly complementary with E&E technologies, the concentration of E&E related vocabulary increased. This is explained by the amount of E&E technologies, notions, and principles that are used in C&C, the fastest growing technology since then. A remarkably similar trend can be found with respect to the concentration of the innovation complementarities of E&E technologies (Figure VII).

On the other hand, C&C technologies haven't experienced yet a period of de-concentration. This contrast between the pattern of diffusion experienced by E&E and C&C technologies can be explained by the degree of complementarity these technologies had with their most notable predecessor. While C&C technologies complemented and built upon advances in E&E technologies, the later basically replaced most of the uses of the technologies that preceded it, like the steam engine.

Figure VI: Concentration of the Use of E&E(C&C)-Related Vocabulary

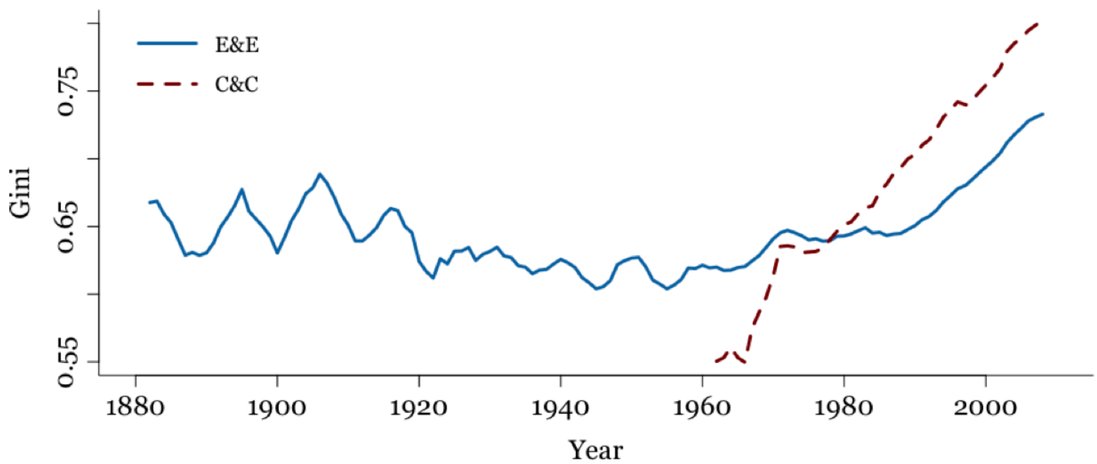
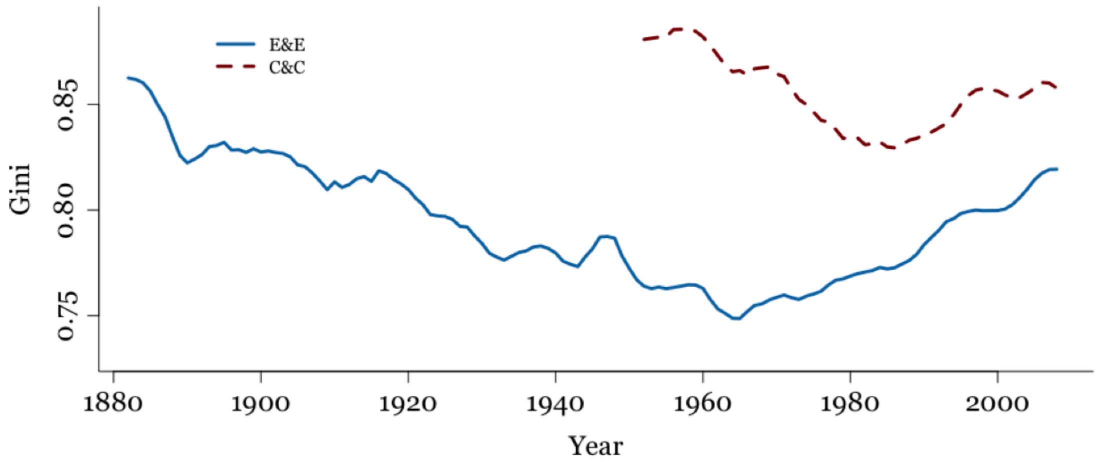


Figure VII: IC Gini



D Description of the ORBIS Database

Table XII shows the number of firms present in the 2017 version of the ORBIS database. The number of firms included in the sample before 2008 represent only a small fraction of the number of firms included between 2008 and 2013. In addition, note that after 2013 the number of firms decreases, which may represent a non-trivial attrition of the sampled firms. In the analysis carried out in Section 3 only the years 2008-2013 are considered.

Table XII: Number of Firms in ORBIS

Year	Number of Firms
2000	22, 951
2001	24, 088
2002	23, 897
2003	24, 403
2004	23, 702
2005	21, 951
2006	24, 055
2007	22, 396
2008	168, 549
2009	133, 529
2010	138, 443
2011	145, 932
2012	150, 727
2013	193, 019
2014	137, 820
2015	115, 643
2016	86, 405
2017	36, 921

Table XIII shows the share of firms with non-missing values for the main economic indicators this database provides, the number of employees ($Employees$) and the operating revenue (OR). Information about the number of employees and the operating revenue is available for at least 50% of the firms between 2008-2013.

Table XIII: Share of Non-Missing Observations (2008-2013)

Employees	OR
0.526	0.643

Approximately 10% of all firms that were included in the sample (those with more than one employee that have information about their OR and $Employees$) have at least one patent granted at the USPTO. Table XIV shows the number of firms per country with and without patenting activity in this period.[22] Column (1) of Table XIV shows the number of firms while Column (2) the share they represent of the total number of firms in the sample. Column (3) shows how many of these firms have at least one patent granted by the USPTO and Column(4) the share they represent among all patenting firms in the sample. Note that 98% of the firms in the sample come from the countries described in this table. Analogously, Table XV shows the same information but across the 25 most populated sectors in the sample. The distribution of firms across sectors is less concentrated than across countries, the 25th most populated sectors account for 50% of all firms.

[22]Countries are ordered by the number of firms in the sample. Only the first 25 entries are shown.

Table XIV: Country of Procedence of the Firms Included in the Sample

Country Code	Number of Firms	Share	Patenting Firms	Share
CN	25,809	0.242	356	0.033
KR	14,554	0.136	670	0.063
IT	13,404	0.126	1,244	0.117
DE	12,717	0.119	1,870	0.175
ES	6,456	0.060	254	0.024
JP	6,005	0.056	1,902	0.178
FR	4,138	0.039	636	0.060
GB	3,802	0.036	784	0.073
CH	2,899	0.027	432	0.040
SE	2,014	0.019	390	0.037
CZ	1,687	0.016	35	0.003
FI	1,654	0.015	221	0.021
PL	1,567	0.015	10	0.001
RU	1,350	0.013	20	0.002
US	1,108	0.010	562	0.053
TW	1,079	0.010	465	0.044
AT	946	0.009	145	0.014
BE	643	0.006	124	0.012
PT	535	0.005	14	0.001
AU	505	0.005	71	0.007
HU	485	0.005	21	0.002
NL	429	0.004	98	0.009
RO	335	0.003	1	0
SI	321	0.003	18	0.002
BG	303	0.003	4	0

Table XV: Number of Firms with Employee and OR Information (2008-2013)

NACE4	Description	Firms	Share	Patenting Firms	Share
2899	Manufacture of other special-purpose machinery n.e.c.	5,341	0.050	465	0.044
2611	Manufacture of electronic components	4,638	0.043	817	0.077
2651	Manufacture of instruments and appliances for measuring, testing and navigation	3,528	0.033	463	0.043
2229	Manufacture of other plastic products	3,119	0.029	215	0.020
2932	Manufacture of other parts and accessories for motor vehicles	3,014	0.028	340	0.032
2599	Manufacture of other fabricated metal products n.e.c.	2,703	0.025	170	0.016
2829	Manufacture of other general-purpose machinery n.e.c.	2,516	0.024	286	0.027
2120	Manufacture of pharmaceutical preparations	2,346	0.022	458	0.043
2511	Manufacture of metal structures and parts of structures	2,271	0.021	79	0.007
2790	Manufacture of other electrical equipment	2,257	0.021	239	0.022
2630	Manufacture of communication equipment	2,003	0.019	271	0.025
3250	Manufacture of medical and dental instruments and supplies	1,995	0.019	411	0.039
2059	Manufacture of other chemical products n.e.c.	1,917	0.018	222	0.021
2712	Manufacture of electricity distribution and control apparatus	1,887	0.018	124	0.012
3299	Other manufacturing n.e.c.	1,625	0.015	180	0.017
2562	Machining	1,576	0.015	58	0.005
2740	Manufacture of electric lighting equipment	1,274	0.012	89	0.008
2822	Manufacture of lifting and handling equipment	1,228	0.012	108	0.010
2849	Manufacture of other machine tools	1,205	0.011	130	0.012
2573	Manufacture of tools	1,185	0.011	115	0.011
3109	Manufacture of other furniture	1,161	0.011	46	0.004
2512	Manufacture of doors and windows of metal	1,146	0.011	35	0.003
2711	Manufacture of electric motors, generators and transformers	1,142	0.011	93	0.009
2620	Manufacture of computers and peripheral equipment	1,040	0.010	227	0.021
2751	Manufacture of electric domestic appliances	1,024	0.010	120	0.011

E Replication of Results in Table VII at Firm-Level

In this section the analysis is conducted based on a panel of manufacturing firms for which there is information about their operating revenue in 2008 and 2013 and have at least 20 employees. Therefore, the economic performance of firms is evaluated at the beginning and end of the period for which the ORBIS database contains a stable number and composition of firms for which a panel could be created. Because the number of patents a company files varies greatly from year to year (Hall, Jaffe, and Trajtenberg, 2001), the patenting activity of firms is evaluated using a 5-years windows previous to the date the economic data is available.

As before, the GPT profile of firms is evaluated by identifying patents with above-average values of Growth, UC, and IC in spans of 5 years (from 2004-2008 and 2009-2013).

Table XVI shows the result of regressing the operating revenue per employee of firms (in logs) against the number and share of GPT patents.[23] Results in columns (1) and (2) in Table XVI show that the number of GPT patents in firms is associated with higher operating revenues, after controlling for firm specific effects and the total number of other patents (non-GPT). The difference between Column (1) and (2) is that the latter only includes firms that have patenting activity in the period. These results suggest there exists a positive association between firms' growth and the number of GPT patents produced, an association that does not exist for other type of patents. Columns (3) and (4) replace the number of GPT patents by its share, in order to use a variable that is less influenced by firm size. In this case, the total number of patents is included as a control. Results show, as before, a positive and strong association between the share of GPT patents in a firm and its growth. Results suggest that a 10% increase in the number of firms' GPT patents is associated with an increase in their operating revenue per employee of approximately 0.3%.

[23]Standard errors are clustered by firm and year according to Cameron, Gelbach, and Miller (2012).

Table XVI: GPT Patents and Firm Performance

	Operating Revenue per Employee			
	(1)	(2)	(3)	(4)
GPT Patents (in logs)	0.034***	0.034***		
	(0.010)	(0.010)		
Non-GPT Patents (in logs)	0.006	0.011		
	(0.009)	(0.009)		
Share of GPT Patents			0.054***	0.051***
			(0.018)	(0.018)
Total Patents (in logs)			0.020**	0.024**
			(0.009)	(0.009)
Years Included	2008 & 2013	2008 & 2013	2008 & 2013	2008 & 2013
Number of Firms	26,854	5,340	26,854	5,340
Type of Firms	All	Patenting	All	Patenting
Firm & Year FE	Yes	Yes	Yes	Yes
Observations	53,708	10,680	53,708	10,680
Adjusted R^2	0.775	0.711	0.775	0.712

*p<0.1; **p<0.05; ***p<0.01

References

Aghion, P., and P. Howitt (2000): "On the macroeconomic effects of major technological change," in *The economics and econometrics of innovation*, pp. 31–53. Springer.

Bresnahan, T. (2010): "General purpose technologies," *Handbook of the Economics of Innovation*, 2, 761–791.

Bresnahan, T. F., and M. Trajtenberg (1995): "General purpose technologies 'Engines of growth'?," *Journal of econometrics*, 65(1), 83–108.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2012): "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*.

David, P., and G. Wright (1999): "General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution," Discussion paper, Economics Group, Nuffield College, University of Oxford.

David, P. A. (1990): "The dynamo and the computer: an historical perspective on the modern productivity paradox," *The American Economic Review*, 80(2), 355–361.

DuBoff, R. (1979): *Electrical Power in American Manufacturing 1889-1958*. New York: Arno Press.

Feldman, M. P., and J. W. Yoon (2012): "An empirical test for general purpose technology: an examination of the Cohen–Boyer rDNA technology," *Industrial and Corporate Change*, 21(2), 249–275.

Field, A. J. (2008): "Does Economic History Need GPTs?," *Available at SSRN 1275023.*

Goldfarb, B. (2005): "Diffusion of general-purpose technologies: understanding patterns in the electrification of US Manufacturing 1880–1930," *Industrial and Corporate Change*, 14(5), 745–773.

Greenwood, J. (1997): *The third industrial revolution: technology, productivity, and income inequality*, no. 435. American Enterprise Institute.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): "The NBER patent citation data file: Lessons, insights and methodological tools," Discussion paper, National Bureau of Economic Research.

Hall, B. H., M. Trajtenberg, et al. (2006): "Uncovering general purpose technologies with patent data," *Chapters*.

Helpman, E., and M. Trajtenberg (1998a): "Diffusion of General Purpose Technologies," in *General purpose technologies and economic growth*, ed. by E. Helpman, p. 85. MIT Press.

——— (1998b): "A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies," in *General purpose technologies and economic growth*, ed. by E. Helpman, p. 55. MIT Press.

Jovanovic, B., and P. L. Rousseau (2005): "General purpose technologies," *Handbook of economic growth*, 1, 1181–1224.

Kalemli-Ozcan, S., B. Sorensen, C. Villegas-Sanchez, V. Volosovych, and S. Yesiltas (2015): "How to construct nationally representative firm level data from the ORBIS global database," Discussion paper, National Bureau of Economic Research.

Lee, K., and C. Lim (2001): "Technological regimes, catching-up and leapfrogging: findings from the Korean industries," *Research policy*, 30(3), 459–483.

Lipsey, R. G., K. I. Carlaw, and C. T. Bekar (2005): *Economic transformations: general purpose technologies and long-term economic growth*. OUP Oxford.

Moser, P., and T. Nicholas (2004): "Was Electricity a General Purpose Technology?," in *The American Economic Review, Papers and Proceedings*, vol. 94, pp. 388–394.

Nebeker, F. (2009): *Dawn of the electronic age: Electrical technologies in the shaping of the modern world, 1914 to 1945*. John Wiley & Sons.

Nuvolari, A., B. Verspagen, and N. von Tunzelmann (2011): "The early diffusion of the steam engine in Britain, 1700–1800: a reappraisal," *Cliometrica*, 5(3), 291–321.

Perez, C., and l. Soete (1988): "Catching up in Technology Entry Barriers and Windows of Opportunity," in *Technical Change and Economic Theory*. Pinter, London.

Rosenberg, N. (1998): "Chemical engineering as a general purpose technology," *General purpose technologies and economic growth*, (7), 167–192.

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(1), 19–50.